ALEA, Lat. Am. J. Probab. Math. Stat. **16**, 487–530 (2019)
DOI: 10.30757/ALEA.v16-19

An optimal Berry–Esseen type theorem for integrals of smooth functions

Lutz Mattner and Irina Shevtsova

Universität Trier, Fachbereich IV – Mathematik
54286 Trier, Germany.
E-mail address: mattner@uni-trier.de
URL: https://www.math.uni-trier.de/∼mattner/

Hangzhou Dianzi University, School of Science
310018, Hangzhou, Zhejiang Province, China;
Institute of Informatics Problems of FRC CSC RAS
119333, Vavilov str., 44/2, Moscow, Russia;
Lomonosov Moscow State University
119991, Leninskie Gory, 1/52, Moscow, Russia.
E-mail address: ishevtsova@cs.msu.ru
URL: https://cs.msu.ru/persons/168

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.

Received by the editors March 22th, 2018; accepted January 22th, 2019.

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.

Research was partially supported by the Russian Foundation for Basic Research (project 19-07-01220-a) and by the Ministry for Education and Science of Russia (grant No. MD-189.2019.1).

Contents

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 the case of approximating distributions of 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 the random summands or of the metric under consideration, as in Osipov's theorem presented e.g. in Petrov (1995, pp. 172–173), Senatov (2013, 2015) and further references therein for smooth distributions and "nonsmooth" (weighted) Kolmogorov metrics, and in Butzer et al. (1975), Barbour (1986, Theorem on p. 294), Goldstein (2010), Tyurin (2011, 2012) for possibly nonsmooth distributions but "smooth" Zolotarev type metrics; see also the references given on pp. 504 near the end of this section 1.1.

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 and for the Kolmogorov metric, of the order $O(n^{-2/3})$, see e.g. Arak (1981a,b, 1982), Arak and Zaitsev (1988), 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. Let us 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 by Fisher and Tippett (1928). 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 Boutsikas (2015), 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 Ledoux and Talagrand (2011, 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 (1.18) 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 (2009a,b, 2011) for some constellations (see Theorems 1.2, 1.15 below). This improvement is possible due to a more favourable dependence of our estimate on the moments of the convolved distributions.

First attempts at 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 (1959) and Zahl (1963), followed by Prawitz (1975) and Bentkus (1994) (for a detailed review see Shevtsova (2016, Sections 2.1.1 and 2.4)). The problem of optimal use of moment-type information in the estimates of the accuracy of the normal approximation was posed in Shevtsova (2012a,b,c, 2013) 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.

To be more precise, we should introduce some 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 $\mathrm{B}_{n,\frac12} := (\frac12(\delta_0+\delta_1))^{*n}$ 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 $X \sim P$. For $P, P_1,\ldots,P_n \in \mathrm{Prob}(\mathbb{R})$ let further $\ast_{i=1}^n P_i$ denote the convolution of the laws $P_1,\ldots,P_n$, and P^{*n} the nth convolution power of P.

With the above notation, the problem of optimization of the structure of *asymptotic* convergence rate estimates stated in Shevtsova (2012a,c,b, 2013) may be formulated as follows: Find the pointwise greatest lower bound to all functions $g : [1,\infty[\, \to \mathbb{R}_+$ such that we have

$$\Delta_n(P) \;:=\; \sup_{x\in\mathbb{R}}\left|\widetilde{P^{*n}}(]-\infty,x]) - \Phi(x)\right| \;\leq\; \frac{g(\varrho(P))}{\sqrt{n}} + \varepsilon_n(P) \quad \text{for } P \in \mathcal{P}_3 \text{ and } n \in \mathbb{N} \tag{1.1}$$

for some remainder term $\varepsilon_n(P) \geq 0$, possibly depending on g, satisfying

$$\lim_{\ell\to 0}\frac{\varepsilon(\ell)}{\ell} \;=\; 0 \quad \text{with} \quad \varepsilon(\ell) \;:=\; \sup_{P\in\mathcal{P}_3, n\in\mathbb{N}\colon\, \varrho(P)=\ell\sqrt{n}} \varepsilon_n(P) \quad \text{for} \quad \ell > 0, \tag{1.2}$$

that is, $\varepsilon_n(P) = o(\varrho(P)/\sqrt{n})$ for $\varrho(P)/\sqrt{n} \to 0$ with not only n but also $\varrho(P)$ allowed to vary. It is easy to see that for any g satisfying (1.1) and (1.2) with some $\varepsilon_n(P)$ we have $g \geq g_*$, where

$$g_*(\varrho) \;:=\; \lim_{\ell\to 0}\sup\left\{\sqrt{n}\,\Delta_n(P)\colon n\in\mathbb{N},\ P\in\mathcal{P}_3,\ \varrho(P) = \varrho \leq \ell\sqrt{n}\right\} \quad \text{for } \varrho \in [1,\infty[\,;$$

moreover, for $g = g_*$ and, say, $\varepsilon_n(P) := \max\left\{0, \Delta_n(P) - g(\varrho(P))/\sqrt{n}\right\}$, we have (1.1) and

$$\limsup_{\ell\to 0}\; \sup_{n\in\mathbb{N}, P\in\mathcal{P}_3\colon\, \varrho(P)=\varrho=\ell\sqrt{n}} \sqrt{n}\,\varepsilon_n(P) \;=\; 0 \quad \text{for } \varrho \in [1,\infty[, \tag{1.3}$$

which is weaker than (1.2) since $\varrho(P)$ is fixed in the supremum in (1.3). However, in some cases, it is possible to construct $\varepsilon_n(P)$ satisfying the stronger condition (1.2) such that inequality (1.1) holds with $g = g_*$. In what follows, we call g_* the *optimal* function.

The problem of explicitly determining the optimal function g_* is very complicated. Historically the first investigations were done for analogous problems with either the functions $\varepsilon_n \geq 0$ in (1.1) only required to satisfy a version of (1.2) pointwise rather than uniformly in P, namely

$$\sup_{P\in\mathcal{P}_3}\lim_{n\to\infty}\sqrt{n}\,\varepsilon_n(P) \;=\; 0 \tag{1.4}$$

(which is even weaker than (1.3)), solved by Esseen (1945, 1956) and thus yielding a lower bound for g_*, or the functions g satisfying (1.1) restricted to be linear (without constant term), where Chistyakov (2002a,b, 2003), significantly sharpening Esseen's (1956) result, eventually found the optimal one. More precisely, restricting now attention to $P \in \widetilde{\mathcal{P}_3}$ rather than $P \in \mathcal{P}_3$ for notational convenience and without loss of generality, from Esseen's (1945) short Edgeworth expansion

$$\widetilde{P^{*n}}\big(]-\infty, x]\big) \;=\; \Phi(x) + (1-x^2)\mathrm{e}^{-\frac{x^2}{2}} \cdot \frac{\mu_3(P)}{6\sqrt{2\pi n}} + \psi_n(x)\mathrm{e}^{-\frac{x^2}{2}} \cdot \frac{h(P)}{\sqrt{2\pi n}} + o\Big(\frac{1}{\sqrt{n}}\Big)$$

valid as $n \to \infty$ uniformly in $x \in \mathbb{R}$ for every fixed $P \in \widetilde{\mathcal{P}_3}$, where $h(P)$ is the span in case of a lattice distribution P and $h(P) = 0$ otherwise, and ψ_n is a certain $(h(P)/\sqrt{n})$-periodic $[-\frac{1}{2}, \frac{1}{2}]$-valued function, Esseen (1956) first deduced that

$$\lim_{n\to\infty} \sqrt{n}\,\Delta_n(P) \;=\; \frac{|\mu_3(P)| + 3h(P)}{6\sqrt{2\pi}} \quad \text{for} \quad P \in \widetilde{\mathcal{P}_3}. \tag{1.5}$$

Second, he considered and solved an extremal problem yielding an exact upper bound of the R.H.S. of (1.5) in terms of $\varrho(P)$ only, namely, he proved that

$$\sup_{P\in\widetilde{\mathcal{P}_3}} \frac{\mu_3(P) + 3h(P)}{\varrho(P)} \;=\; \sqrt{10} + 3, \tag{1.6}$$

with equality attained iff $P = P_{\varrho_\mathrm{E}}$, where for $\varrho \in [1, \infty[$ here and below, $P_\varrho \in \widetilde{\mathcal{P}_3}$ denotes the two-point distribution uniquely defined by the conditions $\mu_3(P_\varrho) \geq 0$ and $\nu_3(P_\varrho) = \varrho$, namely

$$\begin{aligned} P_\varrho\left(\left\{-\sqrt{\tfrac{p}{q}}\right\}\right) &= q := 1-p, \\ P_\varrho\left(\left\{\sqrt{\tfrac{q}{p}}\right\}\right) &= p = p_\varrho := \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\varrho}{2}\sqrt{\varrho^2+8} - \frac{\varrho^2}{2} - 1}\,, \end{aligned} \tag{1.7}$$

and having the span and the third moment

$$\begin{aligned} h_\varrho &:= h(P_\varrho) = 1/\sqrt{pq} = 2\sqrt{2}\Big/\sqrt{\varrho^2 - \varrho\sqrt{\varrho^2+8} + 4}\,, \\ B(\varrho) &:= \mu_3(P_\varrho) = (q-p)/\sqrt{pq} = \sqrt{\varrho^2/2 + \varrho\sqrt{\varrho^2+8}/2 - 2}\,, \end{aligned} \tag{1.8}$$

and where

$$\varrho_\mathrm{E} \;:=\; \sqrt{20(\sqrt{10}-3)/3} \;=\; 1.0401\ldots$$

corresponds to $p_E := p_{\varrho_\mathrm{E}} = (4-\sqrt{10})/2 = 0.4188\ldots$. Hence, (1.5) specialized to $P = P_\varrho$ yields the lower bound

$$g_*(\varrho) \;\geq\; \frac{B(\varrho) + 3h_\varrho}{6\sqrt{2\pi}} \;=\; \frac{1}{6\sqrt{2\pi}} \cdot \frac{2\sqrt{\varrho\sqrt{\varrho^2+8} - \varrho^2 - 2} + 6\sqrt{2}}{\sqrt{\varrho^2 - \varrho\sqrt{\varrho^2+8} + 4}} \;=:\; g_0(\varrho),\;\; \varrho \geq 1,$$

while combination of (1.5) and (1.6) allowed Esseen to find the so-called *asymptotically best* constant

$$C_\mathrm{E} \;:=\; \sup_{P\in\mathcal{P}_3} \lim_{n\to\infty} \frac{\sqrt{n}\,\Delta_n(P)}{\varrho(P)} \;=\; \frac{\sqrt{10}+3}{6\sqrt{2\pi}} \;=\; 0.4097\ldots \tag{1.9}$$

and consequently the necessary condition $c \geq C_\mathrm{E}$ for (1.1) to hold with the linear function $g(\varrho) = c\varrho$. Let us remark in passing that Dinev and Mattner (2013) present

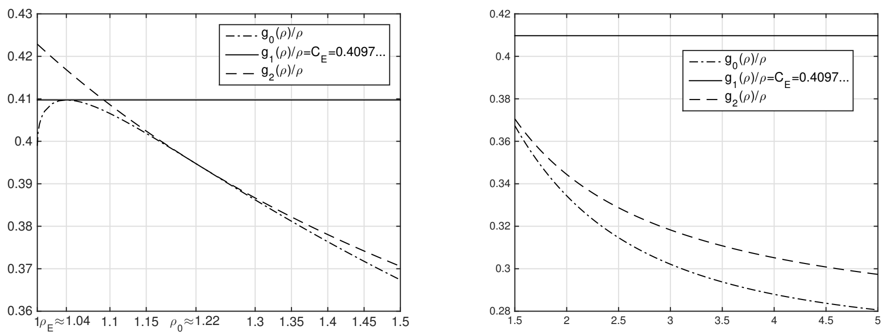

FIGURE 1.1. Plots of the functions $g_0(\varrho)/\varrho$ (dashdot line), $g_1(\varrho)/\varrho = C_{\mathrm{E}} = (\sqrt{10}+3)/(6\sqrt{2\pi}) = 0.4097\ldots$ (solid line), and $g_2(\varrho)/\varrho$ (dashed line) for $\varrho \in [1, 1.5]$ (left) and for $\varrho \in [1.5, 5]$ (right).

$\sqrt{2/\pi} = 0.7978\ldots$ as the asymptotically best constant analogous to C_{E} when arbitrary intervals replace the unbounded ones $]-\infty, x]$ in the definition of $\Delta_n(P)$ in (1.1).

We observe that (1.1) holds with $g = g_0$ and $\varepsilon_n(P)$ satisfying the weakest condition (1.4). A plot of the normalized function $g_0(\varrho)/\varrho$ is given on Fig. 1.1. About 40 years after Esseen's (1956) work, Chistyakov (1999, 2002a,b, 2003) finally managed to find in particular the value of the *asymptotically exact* constant

$$\lim_{\ell\to 0} \sup \left\{ \frac{\sqrt{n}\,\Delta_n(P)}{\varrho(P)} : n \in \mathbb{N},\ P \in \mathcal{P}_3,\ \varrho(P) \le \ell\sqrt{n} \right\} \quad = \quad C_{\mathrm{E}} \tag{1.10}$$

(we are not aware of any really convincing names for the "Esseen constant" in (1.9) and the "Chistyakov constant" in (1.10); the ones used above are at least compatible with some earlier literature such as Shevtsova (2011)) and to prove that (1.1) and (1.2) hold true with

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

and $\varepsilon(\ell) = O\left(\ell^{40/39} |\log \ell|^{7/6}\right)$ (see Chistyakov, 2002a,b, 2003). We also remark here that Chistyakov treats the more general non-i.i.d. case and that the results cited above represent the corresponding specializations to the i.i.d. case. In the more recent paper Shevtsova (2012a, Corollary 4.18 on p. 303), Chistyakov's upper bound for $\varepsilon(\ell)$ was improved to $\varepsilon(\ell) \le 4\ell^{4/3}$ in the general case and $\varepsilon(\ell) \le 3\ell^2$ in the i.i.d. case.

Discarding now the restriction to linear functions, we note that Chistyakov's result reported above yields $g_*(\varrho) \le g_1(\varrho)$ for every $\varrho \in [1, \infty[$, with equality in case of $\varrho = \varrho_{\mathrm{E}}$, and that a result of Hipp and Mattner (2008) yields $g_*(1) = 1/\sqrt{2\pi} = 0.3989\ldots < g_1(1)$. The recent papers Shevtsova (2012b) and Shevtsova (2012a, Theorem 4.13, Corollary 4.17) succeeded in particular in proving that in fact the equality $g_* = g_0$ holds on an interval containing the previously treated points 1 and ϱ_{E}, namely,

$$\begin{aligned} g_*(\varrho) \quad &= \quad g_0(\varrho) \quad \text{for } 1 \ \le\ \varrho \ \le\ \varrho_0 \ :=\ 3^{1/4}(4-\sqrt{3})/\sqrt{6} \ =\ 1.2185\ldots, \\ g_*(\varrho) \quad &\le \quad g_2(\varrho) \quad \text{for all } \varrho \ge 1 \end{aligned}$$

with

$$g_2(\varrho) \;\coloneqq\; \frac{2\varrho}{3\sqrt{2\pi}} + \sqrt{\frac{2\sqrt{3}-3}{6\pi}} \quad \text{for } \varrho \geq 1,$$

$g_2(\varrho_0) = g_0(\varrho_0)$, and g_2 is asymptotically optimal for $\varrho \to \infty$ in the sense of

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

Observe that $g_2(\varrho) < C_E \cdot \varrho$ for $\varrho > \frac{2}{3}\sqrt{2/\sqrt{3}-1}\,(\sqrt{10}+1) = 1.0914\ldots$, in particular, for $\varrho \geq \varrho_0$, and that each of the functions g_1 and g_2 is tangent to g_* at the points ϱ_{E} and ϱ_0, respectively. Plots of the normalized functions $g_k(\varrho)/\varrho$, $k = 0, 1, 2$, are given in Fig. 1.1. We also note that it follows from Schulz (2016, p. 16) that the equality $g_* = g_0$ cannot hold on the whole ray $[1, \infty[$ even in the binomial case, namely,

$$g_*(\varrho) \;>\; g_0(\varrho) \quad \text{for} \quad \varrho > 3.8021\ldots,$$

which corresponds to $p_\varrho < p_*$, where $p_* = 0.05822\ldots$ is the unique root of the equation $7 - 130p + 165p^2 + 50p^3 - 23p^4 = 0$ on the interval $0 < p < 1/3$.

In Shevtsova (2012b) and Shevtsova (2012a, Theorem 4.13 on p. 298 and Corollary 4.17 on p. 302) there have also been obtained explicit uniform upper bounds for the remainder term $\varepsilon_n(P)$ in (1.1) and (1.2) with the continuous function g defined by $g(\varrho) \coloneqq g_0(\varrho)$ for $\varrho \in [1, \varrho_0]$ and $g(\varrho) \coloneqq g_2(\varrho)$ for $\varrho > \varrho_0$, namely, $\varepsilon(\ell) \leq 2\ell^{3/2}$ for all $\ell > 0$.

Moreover, in the same papers an extension of (1.1) to the non-i.i.d. case was obtained in the form

$$\sup_{n\in\mathbb{N},\, P_1,\ldots,P_n\in\mathcal{P}_3}\ \sup_{x\in\mathbb{R}} \left| \widetilde{\mathop{\ast}_{i=1}^{n} P_i}\big(]-\infty, x]\big) - \Phi(x) \right| \;\leq\; \tau \cdot g(\ell/\tau) + 3\ell^{7/6}$$

with the same function g as defined in the preceding paragraph for the case of coinciding $P_1, \ldots, P_n$, where the supremum is taken over all n and all centred distributions $P_1, \ldots, P_n \in \mathcal{P}_3$ such that $\sum_{i=1}^n \nu_3(P_i)/(\sum_{i=1}^n \sigma_i^2)^{3/2} = \ell$, $\sum_{i=1}^n \sigma_i^3/(\sum_{i=1}^n \sigma_i^2)^{3/2} = \tau$, $\sigma_i^2 \coloneqq \sigma^2(P_i)$ for $i \in \{1, \ldots, n\}$. Moreover, there has also been proved a sharpened upper bound for $\varepsilon_n(P)$ in (1.1) and (1.2) with the linear function $g(\varrho) = C_{\mathrm{E}} \cdot \varrho$, namely, $\varepsilon(\ell) \leq 3\ell^2$ in the i.i.d. case and $\varepsilon(\ell) \leq 4\ell^{4/3}$ in the non-i.i.d. case. These bounds improve the earlier results of Prawitz (1975) and Bentkus (1994).

Recently Schulz (2016, Theorem 1 on p. 1) proved that the remainder term $\varepsilon_n(P)$ in (1.1) with $g = g_0$ can be omitted in case of two-point distribution $P = P_\varrho$ for $1 \leq \varrho \leq 5\sqrt{2}/6 = 1.1785\ldots$ (which corresponds to $p_\varrho \in [1/3, 1/2]$), generalizing the earlier result by Hipp and Mattner (2008) originally obtained for $\varrho = 1$. Also, in Schulz (2016, Theorem 1 on p. 1) it is proved that $\Delta_n(P_\varrho) \leq C_{\mathrm{E}} \cdot \varrho$ for every $\varrho \geq 1$ 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 (1928, p. 45, (41a)) in the case of characteristic functions, and to von Bahr (1965) in the case of moments and absolute moments, we may refer in chronological order to the surveys in Bhattacharya and Rao (2010, section 25, that section apparently unchanged from its earlier 1986 edition), Ghosh (1994, Chapter 2), and Petrov (1995, pp. 196–197), and to the more recent papers Borisov and Skilyagina (1996), Borisov et al. (1998), Jiao (2012). 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 mention only the recent paper Angst and Poly (2017).

This paper is organized as follows. Subsections 1.2, 1.3, and 1.4 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 (1955) and Tyurin's (2009a,b, 2011) 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 Shevtsova (2014) 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, see e.g. Denuit et al. (1998), as treated in section 4, and the properties of the Krawtchouk polynomials, see e.g. MacWilliams and Sloane (1977), associated to the symmetric binomial law used in section 6.

The main results of this paper have been announced without proofs in Mattner and Shevtsova (2017).

1.2. *Further notation, properties of the function B.* Terms like "positive", "increasing", and "convex" are understood 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. We use the de Finetti indicator notation, (statement) $:= 1$ if "statement" is true, (statement) $:= 0$ otherwise, for example in (2.20) below.

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

$$\|f^{(m)}\|_{\mathrm{L},\alpha} \quad := \quad \sup_{x,\,y\in I,\,x\neq y} \frac{|f^{(m)}(x) - f^{(m)}(y)|}{|x-y|^\alpha} \tag{1.11}$$

of order $\alpha \in \,]0,1]$. It is well known that for E finite-dimensional, and also more generally as discussed in Diestel and Uhl Jr. (1977), 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

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

Recalling the definition of $B(\varrho)$ given in (1.8) for $\varrho \in [1,\infty[$, let us also put

$$A(\varrho) \;:=\; \varrho^{-1}B(\varrho) \;=\; \sqrt{\tfrac{1}{2}\sqrt{1+8\varrho^{-2}} + \tfrac{1}{2} - 2\varrho^{-2}} \qquad \text{for } \varrho \in [1,\infty[. \tag{1.12}$$

The notation A here is as used in Shevtsova (2014, pp. 194, 208), so let us note that there is an inconsequential typo in the formula for $A'(\varrho)$ in Shevtsova (2014, p. 208), where $\varrho^{3/2}$ should be $\varrho^3/2$.

Lemma 1.1. *The functions A and B are continuous, strictly concave and increasing, with $A(1) = B(1) = 0$, $\lim_{\varrho\to 1} A(\varrho)/\sqrt{\varrho - 1} = \sqrt{8/3}$, 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 $\big(A^2(\varrho)\big)' = 4\varrho^{-3}(1-(1+8\varrho^{-2})^{-1/2})$ 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 $\lim_{\varrho\to 1} A^2(\varrho)/(\varrho-1) = \big(A^2\big)'(1) = 8/3$. B is obviously strictly increasing and, by Shevtsova (2014, 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. □

1.3. *The main result (Rademacher average approximation) and some consequences.* 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*

$$\left| \widetilde{\mathop{\ast}_{i=1}^{n} P_i}\, f - \widetilde{\mathop{\ast}_{i=1}^{n} Q_i}\, f \right| \quad\leq\quad \frac{\|f''\|_{\mathrm{L}}}{6} \sum_{i=1}^{n} \frac{\sigma_i^3}{\sigma^3} B(\varrho_i) \tag{1.13}$$

with $\sigma_i \coloneqq \sigma(P_i)$, $Q_i \coloneqq \frac{1}{2}(\delta_{-\sigma_i} + \delta_{\sigma_i})$, $\sigma \coloneqq \big(\sum_{i=1}^n \sigma_i^2\big)^{1/2}$, 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 (1.13).

The proof of Theorem 1.2 is given in section 5 on p. 529.

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 (1.13) further simplifies to

$$\left| \widetilde{P^{*n}} f - \widetilde{\mathrm{B}_{n,\frac{1}{2}}} f \right| \quad\leq\quad \frac{B(\varrho(P))}{6\sqrt{n}} \|f''\|_{\mathrm{L}}, \tag{1.14}$$

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 Shevtsova (2014, Theorem 6), namely

$$\max_{P\in\mathcal{P}_3 :\, \varrho(P)=\varrho} \left| \int x^3 \,\mathrm{d}\widetilde{P}(x) \right| \quad=\quad B(\varrho) \qquad \text{for } \varrho \in [1,\infty[\tag{1.15}$$

with equality attained for two-point laws, by applying (1.14) with $E = \mathbb{R}$, $n = 1$, and $f(x) \coloneqq x^3$, since for $P \in \mathcal{P}_3$, we have

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

and $\|f''\|_{\mathrm{L}} = 6$. However, (1.15) is used in Step 6 of our proof of Theorem 1.2.

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

$$\left|\varphi(t) - \prod_{i=1}^{n} \cos\left(\frac{\sigma_i t}{\sigma}\right)\right| \;\leq\; \frac{|t|^3}{6}\sum_{i=1}^{n} \frac{\sigma_i^3 B(\varrho_i)}{\sigma^3}, \tag{1.16}$$

since here $\|f''\|_{\mathrm{L}} = \sup_{x\in\mathbb{R}} |f'''(x)| = |t|^3$. In (1.16), 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 (1.13) for $f = (\cdot)^3$ and by a Taylor expansion inside the modulus on the left hand side of (1.16).

Moreover, using

$$0 \;\leq\; \prod_{i=1}^{n} \cos t_i - 1 + \frac{1}{2}\sum_{i=1}^{n} t_i^2 \;\leq\; \frac{1}{24}\sum_{i=1}^{n} t_i^4 + \frac{1}{4}\sum_{i<j} t_i^2 t_j^2 \quad \text{for } t \in \mathbb{R}^n, \tag{1.17}$$

which follows by rewriting the central term in (1.17) with the help of i.i.d. Rademacher variables $\xi_1, \ldots, \xi_n$ as

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

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 (1.16) the following estimate for the accuracy of the approximation of φ by the first terms of its Taylor expansion:

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

In particular, with $n = 1$ we have

$$\left|\mathbb{E}\mathrm{e}^{\mathrm{i}tX} - 1 + \frac{t^2}{2}\right| \;\leq\; A(\varrho)\frac{\varrho|t|^3}{6} + \frac{t^4}{24} \tag{1.18}$$

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. (more precisely, either $X \sim P_\varrho$ or $-X \sim P_\varrho$ with P_ϱ defined in (1.7)).

Inequality (1.18) 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 Shevtsova (2014, 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

$$\left|\mathbb{E}\left|\widetilde{\textstyle\sum_{i=1}^n X_i}\right|^3 - \widetilde{\mathrm{B}_{n,\frac12}}|\cdot|^3\right| \quad\leq\quad \frac{B(\varrho(P))}{\sqrt{n}}, \tag{1.19}$$

by $\|f''\|_{\mathrm{L}} = 6$, where by formula (6.4) stated and proved below, we have explicitly

$$\widetilde{\mathrm{B}_{n,\frac12}}|\cdot|^3 \quad=\quad \begin{cases} \left(2n^{\frac12} + n^{-\frac12} - n^{-\frac32}\right)\mathrm{b}_{n,\frac12}(\lfloor\frac{n}{2}\rfloor) & \text{if } n \text{ is odd,} \\ 2n^{\frac12}\mathrm{b}_{n,\frac12}(\frac{n}{2}) & \text{if } n \text{ is even.} \end{cases}$$

Let us note that $\widetilde{\mathrm{B}_{n,\frac12}}|\cdot|^3$ can not be replaced by any other function of n without invalidating (1.19), since the R.H.S. of (1.19) 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 1.10 below, to obtain the quite sharp normal approximation result in Theorem 1.11. Since in turn the proof of Theorem 1.10 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 Zolotarev (1997, Chapter 1), Rachev (1991), Senatov (1998, Chapter 2).

We will use the notation introduced around (1.11), 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

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

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

$$\zeta_s(P,Q) \quad:=\quad \sup_{f\in\mathcal{F}_s^\infty} |Pf - Qf| \tag{1.20}$$

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

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

which is also called the s-th *absolute pseudomoment* Zolotarev (1997, p. 67).

Let us note that in Zolotarev (1997, p. 44) and Senatov (1998, 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 (1.20) 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 Neininger and Sulzbach (2015, (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*

$$\begin{aligned}\zeta_s(P,Q) < \infty \quad &\Leftrightarrow \quad P \in \mathrm{Prob}_s(\mathbb{R}) \text{ and } \mu_j(P) = \mu_j(Q) \text{ for } j \in \{1,\ldots,m\} \\ &\Leftrightarrow \quad P \in \mathrm{Prob}_s(\mathbb{R}) \text{ and } \zeta_s(P,Q) \le \frac{\Gamma(1+\alpha)}{\Gamma(1+s)}\nu_s(P,Q). \end{aligned} \tag{1.21}$$

Hence, if $c_1, \ldots, c_m \in \mathbb{R}$ are given, then ζ_s is a metric on the (possibly empty) set $\{P \in \mathrm{Prob}_s(\mathbb{R}) : \mu_j(P) = c_j \text{ for } j \in \{1,\ldots,m\}\}$. 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 (1.20), that is, we have*

$$\zeta_s(P,Q) \quad = \quad \sup_{f\in\mathcal{F}_s} |Pf - Qf|, \tag{1.22}$$

and we further have

$$|Pf - Qf| \;\le\; \|f^{(m)}\|_{\mathrm{L},\alpha}\, \zeta_s(P,Q) \quad \text{if } f \in \mathcal{C}^{m,\alpha}(\mathbb{R},\mathbb{R}) \text{ and } \zeta_s(P,Q) < \infty. \tag{1.23}$$

References or proofs for Theorem 1.7 are given in section 4 on p. 523, 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 $P_i, \sigma_i, Q_i, \sigma, \varrho_i$ for $i \in \{1,\ldots,n\}$ be as in Theorem 1.2. Then we have*

$$\zeta_3\left(\widetilde{\mathop{\ast}_{i=1}^{n} P_i}\,,\ \widetilde{\mathop{\ast}_{i=1}^{n} Q_i}\right) \quad \le \quad \frac{1}{6\sigma^3}\sum_{i=1}^{n}\sigma_i^3 B(\varrho_i), \tag{1.24}$$

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 (1.24), and using also (1.22) from Theorem 1.7(d) yields the accompanying equality statement. Conversely, if (1.24) is proved, then, using (1.23), we get Theorem 1.2 in the case of $E = \mathbb{R}$ and except for the equality statement.

Remark 1.9. Under the assumptions of Theorem 1.8, we have the equivalence

$$\text{L.H.S.}(1.24) = 0 \quad \Leftrightarrow \quad \text{R.H.S.}(1.24) = 0. \tag{1.25}$$

Here the implication "$\Leftarrow$" of course follows trivially from (1.24). Conversely, if we have L.H.S.(1.24) $= 0$, then we get $\widetilde{\ast_{i=1}^{n} P_i} = \widetilde{\ast_{i=1}^{n} Q_i}$ and hence, assuming from now on without loss of generality the P_i to be centred, and recalling that $\sigma(P_i) = \sigma_i = \sigma(Q_i)$ for each i, we have

$$\mathop{\ast}_{i=1}^{n} P_i \quad = \quad \mathop{\ast}_{i=1}^{n} Q_i. \tag{1.26}$$

We now use some well-known elementary facts about cumulants, for which we may refer to Hald (2000) and Mattner (1999, 2004). Cumulants are certain functions $\kappa_\ell : \mathrm{Prob}_\ell(\mathbb{R}) \to \mathbb{R}$ for $\ell \in \mathbb{N}$, most importantly $\kappa_1 = \mu(\cdot)$, $\kappa_2 = \sigma^2(\cdot)$, $\kappa_3(P) = \int (x-\mu(P))^3 \,\mathrm{d}P(x)$ for $P \in \mathrm{Prob}_3(\mathbb{R})$, and $\kappa_4(P) = \int (x-\mu(P))^4 \,\mathrm{d}P(x) - 3\sigma^4(P)$ for $P \in \mathrm{Prob}_4(\mathbb{R})$, designed to enjoy the additivity

$$\kappa_\ell(P * Q) \quad = \quad \kappa_\ell(P) + \kappa_\ell(Q) \qquad \text{for } \ell \in \mathbb{N} \text{ and } P, Q \in \mathrm{Prob}_\ell(\mathbb{R}). \tag{1.27}$$

Observing now that, for a centred $P \in \mathrm{Prob}_4(\mathbb{R})$, we have

$$\kappa_4(P) \quad = \quad \int x^4 \,\mathrm{d}P(x) - 3\sigma^4(P) \quad \geq \quad \left(\int x^2 \,\mathrm{d}P(x) \right)^2 - 3\sigma^4(P) = -2\sigma^4(P)$$

with equality throughout iff $P = \frac{1}{2}\left(\delta_{-\sigma(P)} + \delta_{\sigma(P)}\right)$, by, say, Jensen's inequality with the strictly convex square function and by centredness of P, we get from (1.26), using (1.27) with $\ell = 4$ in the first step,

$$\sum_{i=1}^n \kappa_4(P_i) \quad = \quad \sum_{i=1}^n \kappa_4(Q_i) \quad = \quad \sum_{i=1}^n \left(-2\sigma^4(Q_i)\right) \quad = \quad \sum_{i=1}^n \left(-2\sigma^4(P_i)\right),$$

and thus $P_i = Q_i$ and hence $\varrho(P_i) = 1$ for each i, and hence R.H.S.(1.24) $= 0$ due to $B(1) = 0$.

Thus the error bound (1.24) in Theorem 1.8 enjoys the property (1.25) in analogy to classical refinements of the Berry–Esseen bound for normal approximations to convolution products first obtained in the i.i.d. case, after a preliminary result of Zolotarev (1965), by Paulauskas (1969), and then quickly generalized or sharpened in publications up to 1973 by Sazonov (1972), Nagaev and Rotar' (1973), and Zolotarev (1973); reviews by Sazonov (1981, pp. 9, 68), Rotar' (1982, §2), Petrov (1995, pp. 190–191, subsections 5.10.16–5.10.18), and Zolotarev (1997, section 6.5.1) point to further relevant works, including several ones by the authors already mentioned here and by Ulyanov, in particular Ul'yanov (1976, 1979), to which one can add, among others, the papers of Shiganov (1989), Paditz (1988), and, treating asymptotic expansions, Yaroslavtseva (2009).

In contrast to our bound in (1.24), those refinements have to use some so-called (absolute) pseudo– or difference–moments instead of ordinary absolute moments of the involved distributions.

1.4. *Normal approximation.* Coming now to the normal approximation results following from Theorem 1.8, let us first consider in Theorem 1.11 below the i.i.d. case. There

$$\varepsilon_n \coloneqq \zeta_3\left(\widetilde{\mathrm{B}_{n,\frac{1}{2}}}, \mathrm{N}\right) \qquad \text{for } n \in \mathbb{N} \tag{1.28}$$

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

Theorem 1.10. *For $n \in \mathbb{N}$, we have, with the first equality to be read from right to left due to the $O(n^{-2})$,*

$$\begin{aligned} \frac{1}{6\sqrt{2\pi}\,n} + O\left(\frac{1}{n^2}\right) &= \frac{1}{6}\left\{\begin{array}{ll} \left|\left(2n^{\frac12}+n^{-\frac12}-n^{-\frac32}\right)\mathrm{b}_{n,\frac12}(\lfloor\frac{n}{2}\rfloor)-\frac{4}{\sqrt{2\pi}}\right| & \text{if } n \text{ is odd,} \\ \left|2n^{\frac12}\mathrm{b}_{n,\frac12}(\frac{n}{2})-\frac{4}{\sqrt{2\pi}}\right| & \text{if } n \text{ is even} \end{array}\right\} \\ &= \left|\left(\widetilde{\mathrm{B}_{n,\frac12}}-\mathrm{N}\right)\frac{|\cdot|^3}{6}\right| \;\leq\; \varepsilon_n \\ &< \frac{1}{3\sqrt{2\pi}\,n}+\left(\frac{4+\zeta(\frac12)}{\sqrt{2\pi}}-1\right)\frac{1}{6n^{3/2}} \\ &< \frac{0.1330}{n}+\frac{0.0022}{n^{3/2}} \;\leq\; \frac{0.1352}{n}, \end{aligned}$$

where $\zeta(\cdot)$ is the Riemann zeta-function, in particular $\zeta(\frac12) = -1.4603\ldots$.

The proof of Theorem 1.10 is given in section 6 on p. 532.

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

$$\zeta_3\left(\widetilde{P^{*n}},\mathrm{N}\right) \;\leq\; \frac{B(\varrho(P))}{6\sqrt{n}}+\varepsilon_n, \tag{1.29}$$

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

$$\frac{B(\varrho)}{6} \;=\; \lim_{n\to\infty}\sqrt{n}\,\zeta_3\left(\widetilde{P_\varrho^{*n}},\mathrm{N}\right) \;=\; \sqrt{k}\left|\widetilde{P_\varrho^{*k}}f-\mathrm{N}f\right| \quad \text{for } \varrho \geq 1 \text{ and } k \in \mathbb{N}, \tag{1.30}$$

with $P_\varrho \in \mathcal{P}_3$ being the two-point law defined in (1.7) and satisfying $\varrho(P_\varrho) = \varrho$, and with $f \in \mathcal{F}_3$ given by $f(x) = x^3/6$ for $x \in \mathbb{R}$, and the leading term for $\varrho \to 1$ is asymptotically exact in the sense of

$$\varepsilon_n \;=\; \lim_{P\in\mathcal{P}_3:\ \varrho(P)\to 1}\zeta_3\left(\widetilde{P^{*n}},\mathrm{N}\right) \quad \text{for } n \in \mathbb{N}. \tag{1.31}$$

The proof of Theorem 1.11 is given in section 6 on p. 534.

Remark 1.12. In view of (1.29) and $\varepsilon_n = O(n^{-1})$, the first equation in (1.30) yields, as an alternative formulation of the large n optimality of (1.29):

$$\frac{B(\varrho)}{6} \;=\; \max_{P\in\mathcal{P}_3:\ \varrho(P)=\varrho}\ \underline{\lim}_{n\to\infty}\sqrt{n}\,\zeta_3\left(\widetilde{P^{*n}},\mathrm{N}\right) \quad \text{for } \varrho \in [1,\infty[, \tag{1.32}$$

with the maximum attained for $P = P_\varrho$. We suspect that in (1.32) one can replace "$\underline{\lim}$" by "lim", since if $P \in \mathcal{P}_3$ is given and if also $f \in \mathcal{F}_3$ is fixed, then we have

$$\lim_{n\to\infty}\sqrt{n}\left|\widetilde{P^{*n}}f-Nf\right| \;=\; |Ef| \tag{1.33}$$

with E denoting here the signed measure on $\mathbb{R}$ with the distribution function $x \mapsto (1-x^2)\mathrm{e}^{-x^2/2}\mu_3(P)/(6\sqrt{2\pi})$ occurring in the short Edgeworth expansion for $\widetilde{P^{*n}}$, by applying Götze and Hipp (1978, Theorem (3.6) in the i.i.d. case with $k = 1$, $s = s_0 = 3$, $p = 2$ for $|\alpha| = 1$). However, for an arbitrary $P \in \mathcal{P}_3$, we are not aware of a reference conveniently yielding the convergence in (1.33) uniformly in $f \in \mathcal{F}_3$, which would then yield the existence of $\lim_{n\to\infty}\sqrt{n}\,\zeta_3\left(\widetilde{P^{*n}},\mathrm{N}\right) = \sup_{f\in\mathcal{F}_3}|Ef|$.

In the special case of $P = P_\varrho$ this limit exists, as claimed in (1.30), by the proof of Theorem 1.11.

Remark 1.13. Inequality (1.29) often improves estimates in Tyurin (2009a, Theorem 4), Tyurin (2009b), Tyurin (2011, Theorem 4) (with Tyurin, 2009a actually being the final one among the three papers)

$$\zeta_3\Big(\widetilde{\mathop{\ast}_{i=1}^{n} P_i}, \mathrm{N}\Big) \quad\leq\quad \frac{1}{6\sigma^3}\sum_{i=1}^{n}\sigma_i^3\varrho_i \quad \text{for } P_1\ldots,P_n\in\mathcal{P}_3 \tag{1.34}$$

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

$$\zeta_3(\widetilde{P^{*n}}, \mathrm{N}) \quad\leq\quad \frac{\varrho(P)}{6\sqrt{n}} \qquad \text{for } P\in\mathcal{P}_3 \text{ and } n\in\mathbb{N} \tag{1.35}$$

and is optimal in the sense that the constant 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 (1.29) improves (1.35) for *every* value of $\varrho \geq 1$ and every sufficiently large $n \in \mathbb{N}$, namely iff

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

which, by Theorem 1.10, is surely true for

$$n \quad\geq\quad \Big(\frac{6\cdot 0.1352}{\varrho - B(\varrho)}\Big)^2 \quad=\quad \frac{0.65804\ldots}{(\varrho - B(\varrho))^2}. \tag{1.36}$$

Here is a table of the values of ϱ and n satisfying condition (1.36), where, for convenience, we also provide values of $B(\varrho)$ rounded up:

$\varrho\leq$	1.01	1.10	1.18	1.24	1.30	1.52	1.66	1.77	1.94	2.17	2.33	2.519
$B(\varrho)$	0.17	0.53	0.72	0.83	0.94	1.27	1.45	1.59	1.80	2.06	2.24	2.438
$n\geq$	1	2	3	4	5	10	15	20	30	50	70	100

Example 1.14. Let P be an exponential distribution. Then $\varrho = 12\mathrm{e}^{-1} - 2 = 2.4145\ldots$, $B(\varrho) = 2.3248\ldots$, and condition (1.36) holds for $n \geq 82$.

If P is a uniform distribution on an interval, then $\varrho = 3\sqrt{3}/4 = 1.2990\ldots$, $B(\varrho) = 0.9302\ldots$, and condition (1.36) holds for $n \geq 5$.

If P is the Bernoulli distribution with parameter $p \in]0, \frac{1}{2}]$, then, denoting $q \coloneqq 1 - p$, 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 (1.36) 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 $p \geq 0.2$, $n \geq 149$ if $p \geq 0.1$.

In particular, in the symmetric case ($p = 1/2$) our bound (1.29) is of course sharper than (1.35) for every $n \in \mathbb{N}$.

If P is the Poisson distribution with parameter $\lambda > 0$, then:
if $\lambda = 1$ we have $\varrho = 1.7357\ldots$, $B(\varrho) = 1.5448\ldots$, and (1.36) holds for $n \geq 19$;
if $\lambda = 2$ we have $\varrho = 1.6640\ldots$, $B(\varrho) = 1.4543\ldots$, and (1.36) holds for $n \geq 15$;
if $\lambda = 4$ we have $\varrho = 1.6294\ldots$, $B(\varrho) = 1.4096\ldots$, and (1.36) holds for $n \geq 14$;
if $\lambda = 8$ we have $\varrho = 1.6125\ldots$, $B(\varrho) = 1.3874\ldots$, and (1.36) 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 (1.36) holds for $n \geq 83$.

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

Theorem 1.15. *For $P_i, Q_i, \varrho_i, \sigma_i, \sigma$ as in Theorem 1.2 we have*

$$\zeta_3\Big(\widetilde{\mathop{\ast}_{i=1}^{n} P_i}, \mathrm{N}\Big) \quad\le\quad \frac{1}{6\sigma^3}\sum_{i=1}^{n}\sigma_i^3 B(\varrho_i) + \zeta_3\Big(\widetilde{\mathop{\ast}_{i=1}^{n} Q_i}, \mathrm{N}\Big). \tag{1.37}$$

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

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

The proof of Theorem 1.15 is given in section 6 on p. 531.

Remark 1.16. Inequality (1.37) improves Tyurin's already optimal bound (1.34) iff

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

Thus, as already indicated at the end of subsection 1.1, Theorems 1.2 and 1.15 can be regarded as extensions of the results previously obtained in Shevtsova (2012b,c), Shevtsova (2012a, Corollary 4.7 on p. 284, Theorem 4.13 on p. 298, Theorem 4.14 on p. 300, Corollary 4.17 on p. 302), Shevtsova (2013, Theorems 2.3, 2.4) for the uniform metric to ζ_3-metric, so that the inequalities (1.29) and (1.37) can be called *estimates with an asymptotically optimal structure.*

2. Auxiliary analytic results

2.1. *Two-point Hermite interpolation, and approximation in $\mathcal{F}_s$.* The purpose of the presumably well-known 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 $y = (y_0, y_1) = \big((y_{0,j})_{j=0}^{m_0}, (y_{1,j})_{j=0}^{m_1}\big) \in V$, 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*

$$p^{(j)}(x_i) \quad=\quad y_{i,j} \qquad \textit{for } i \in \{0,1\} \textit{ and } j \in \{0,\ldots,m_i\}. \tag{2.1}$$

(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*

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

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*

$$y_{0,j} \;\ge\; 0 \textit{ for } j \in \{0,\ldots,m_0\}, \qquad (-1)^j y_{1,j} \;\ge\; 0 \textit{ for } j \in \{0,\ldots,m_1\}. \tag{2.2}$$

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*

$$\sup_{x\in[x_0,x_1]} \left|p^{(k)}_{x_0,x_1,y}(x)\right| \quad\leq\quad c\,\|y\|\,\frac{1 \vee |x_1-x_0|^{m_0\vee m_1}}{|x_1-x_0|^k} \qquad \text{for } k \in \mathbb{N}_0. \tag{2.3}$$

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

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

We then have

$$p^{(j)}(x) \quad>\quad 0 \qquad \text{for } x > 0 \text{ sufficiently close to } 0 \tag{2.4}$$

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 (2.4) 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 (2.4) 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 (2.3) 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. Schwartz (1991, 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

$$\begin{aligned}\sup_{x\in[x_0,x_1]}\left|p^{(k)}_{x_0,x_1,y}(x)\right| \quad&=\quad \sup_{x\in[x_0,x_1]}\left|\tfrac{1}{(x_1-x_0)^k}\,p^{(k)}_{0,1,z}\big(\tfrac{x-x_0}{x_1-x_0}\big)\right| \\ &\leq\quad \tfrac{c}{(x_1-x_0)^k}\|z\|_\infty \quad\leq\quad \text{R.H.S.(2.3)},\end{aligned}$$

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. Schwartz (1991, p. 279). □

We recall the definitions of $\mathcal{F}^\infty_s$ 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 (2.3), so that $p_n := p_{n,b_n,y}$ satisfies

$$\left|p_n^{(m+1)}(x)\right| \quad\le\quad \tfrac{1}{2n}(b_n - n)^{\alpha-1} \qquad \text{for } x \in [n, b_n]. \tag{2.5}$$

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

$$\left|q_n^{(m+1)}(x)\right| \quad\le\quad \tfrac{1}{2n}(-n - a_n)^{\alpha-1} \qquad \text{for } x \in [a_n, -n]. \tag{2.6}$$

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

$$f_n(x) \quad:=\quad (a_n \le x \le -n)q_n(x) + (|x| < n)\tfrac{n-1}{n} f(x) + (n \le x \le b_n)p_n(x)$$

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

$$\sup_{u,\,v\in\mathbb{R},\,u<v} \frac{|f_n^{(m)}(v) - f_n^{(m)}(u)|}{|v-u|^\alpha} \quad\le\quad 1. \tag{2.7}$$

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 (2.5) and (2.6), and also (2.7) with f in place of f_n, to get

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

The remaining cases needed to prove (2.7) 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

$$g(x) \quad=\quad \sum_{j=0}^{m-1} \frac{g^{(j)}(0)}{j!}x^j + \int_0^1 \frac{(1-\lambda)^{m-1}}{(m-1)!} g^{(m)}(\lambda x)x^m \,\mathrm{d}\lambda \tag{2.8}$$

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 Pinkus and Wulbert (2005), 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

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

where

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

$V(x_0, \ldots, x_s) \coloneqq U(x_0, \ldots, x_s; (\cdot)^s) = \prod_{i<j}(x_j - x_i)$ is the Vandermonde determinant. Alternatively one can set (see e.g. Kuczma, 2009, Chapter 15)

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

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 (1926) and was further extensively developed by Popoviciu (1933).

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, by Popoviciu (1933) or Kuczma (2009, 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 $-\infty =: x_0 < x_1 < \ldots < x_s < x_{s+1} \coloneqq +\infty$ 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, see Hopf (1926), Popoviciu (1933), or Kuczma (2009). If f is s times differentiable on I, then f is s-convex iff $f^{(s)} \geq 0$ on I, see Popoviciu (1933) or Kuczma (2009).

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*

$$f(s) \;=\; f'(s) \;=\; f(t) \;=\; f'(t) \;=\; 0 \tag{2.9}$$

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 DeVore and Lorentz (1993, p. 119, Theorem 6.3).

To prove now the lemma, we may assume $s < t$. Assumption (2.9) 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*

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

satisfies

$$g(s) = f(s),\;\; g'(s) = f'(s),\;\; g(t) = f(t),\;\; g'(t) = f'(t). \tag{2.10}$$

(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*

$$f \leq g \text{ on } \mathbb{R} \quad \Leftrightarrow \quad d \geq 0 \quad \Leftrightarrow \quad D \cdot (s+t) \geq 0, \tag{2.11}$$

$$f \geq g \text{ on } \mathbb{R} \quad \Leftrightarrow \quad d \leq 0 \quad \Leftrightarrow \quad D \cdot (s+t) \leq 0, \tag{2.12}$$

and the inequality between f and g in (2.11) or (2.12) 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,*

$$|x - r|^3 \quad \leq \quad a + bx + cx^2 + d|x|^3, \tag{2.13}$$

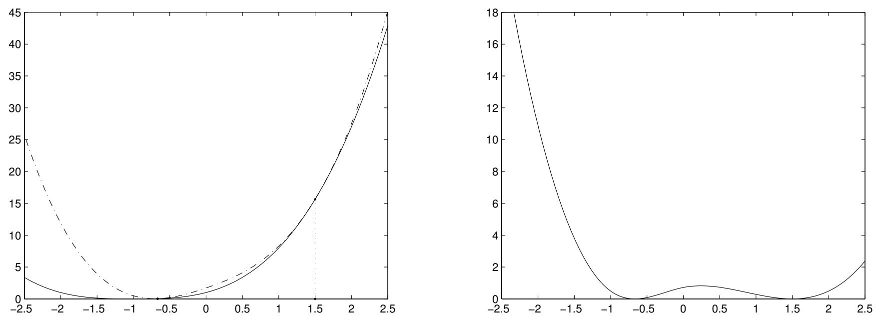

FIGURE 2.2. Left: plots of the functions $f(x) = |x+1|^3$ (solid line) and $g(x) = a + bx + cx^2 + d|x|^3$ (dashdot line) from Lemma 2.4(c) with $u = 3/2$, $v = 2/3$. Right: plot of the difference $g(x) - f(x)$.

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

$$a_r(u,v) = |r|^3 + \frac{4u^3v^3}{(u-v)(u^2+4uv+v^2)}, \tag{2.14}$$

$$b_r(u,v) = -\mathrm{sgn}(r)\left(3r^2 + \frac{6u^2v^2}{u^2+4uv+v^2}\right), \tag{2.15}$$

$$c_r(u,v) = 3|r| - \frac{12u^2v^2}{(u-v)(u^2+4uv+v^2)}, \tag{2.16}$$

$$d_r(u,v) = \frac{(u+v)^3}{(u-v)(u^2+4uv+v^2)} \tag{2.17}$$

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

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

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

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

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

for $v > |r|$. *Equality in* (2.13) *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 Shevtsova (2018, Lemma 2), where the stated inequality was proved only in the case of $v > |r|$.

Proof of Lemma 2.4: (a) Condition (2.10) is a system of linear equations for a, b, c, d with the determinant

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

(b) Lemma 2.3 applied to $g-f$ or to $f-g$ yields the first equivalences in (2.11) and (2.12), 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 (2.10). 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 (2.11) and (2.12).

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 Shevtsova (2018, 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

$$-1 \;\leq\; s \;=\; -v \;\leq\; 0 \;\leq\; v \;<\; t \;=\; u, \quad v \;\leq\; 1. \tag{2.18}$$

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 (2.11) 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

$$h(x) \;\coloneqq\; g(x) - f(x) \;=\; a + 1 + (b+3)x + (c+3)x^2 + (1-d)x^3$$

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 (2.18) and also (2.17), we get

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

and hence, for $x \in\,]-\infty,-1[$, using (2.16) with $r=-1$ in the central step, and $v \in [0,1]$ from (2.18) in the last,

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

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

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

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

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

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 Denuit et al. (1998, 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 Karlin (1968, p. 20), one calls

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

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

$$S^-(f|_{A\cup B}) \quad \leq \quad S^-(f|_A)+S^-(f|_B)+1 \qquad \text{for } A, B \subseteq D \text{ with } A \leq B. \tag{2.19}$$

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*

$$S^-(f) \quad \le \quad S^-(f') + 1 - \left(\lim_{x\to a+} f(x) = 0\right) - \left(\lim_{x\to b-} f(x) = 0\right) \qquad (2.20)$$

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 $z \in I^n_\le$ denoting any sign change tuples of f and f' respectively, and with

$$J \;:=\; \Big\{ j \in \{0,\ldots,m\} : 1 \le j \le m-1, \text{ or } j = 0 \text{ and } \lim_{x\to a+} f(x) = 0, \qquad (2.21)$$
$$\text{or } j = m \text{ and } \lim_{x\to b-} f(x) = 0 \Big\}$$

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 $\#J \le n$ and thus (2.20), 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 $z \in I^n_\le$ 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 $f' \le 0$ on I_j or $f' \ge 0$ on I_j, and hence $S^-(f|_{I_j}) \le 1$, and hence $m := S^-(f) \le 2n+1 < \infty$, by applying (2.19) n times. So let $y \in I^m_<$ be a sign change tuple of f, let J be defined by (2.21), $y_0 := a$, $y_{m+1} := b$, 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 (2.22) below, see `oeis.org/A059750` in Sloane (2010).

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

$$\sum_{k=1}^{n-1} \frac{1}{\sqrt{k}} - 2\sqrt{n} \quad < \quad \zeta(\tfrac{1}{2}) \quad = \quad -1.46035\ldots, \qquad (2.22)$$

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

Proof: Let a_n denote the left hand side of the inequality in (2.22). 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 $(a_n)_{n\ge1}$ is strictly increasing. Since we have $\lim_{n\to\infty} a_n =$

$\zeta(1/2)$ by Hardy (1949, p. 333, (13.10.7) with $s = \sigma = \frac{1}{2}$), or see Wirths (2015, p. 192, (4.1)) for a more elementary proof, the inequality in (2.22) 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)**.** *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 part (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 Ferguson (1967, p. 74, Lemma 3) or Dudley (2002, 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 as in Hiriart-Urruty and Lemaréchal (2001, 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 in Hiriart-Urruty and Lemaréchal (2001, 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 Richter (1957, 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 Richter (1957, 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 Kemperman (1968), 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 Hoeffding's (1955, p. 269, Theorem 2.1 with $n = 1$) result, combined with Theorem 3.1(a) and with the concavity of the function B from (1.8), 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. So in the end, of the theory of this section, we later use in the present paper only Theorems 3.1(a) and 3.3, and in the former "finite" in place of "$k+1$ or fewer" would actually be enough.

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)**.** *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

$$K \quad := \quad \{Q \in \mathrm{Prob}(\mathcal{X}) : Q f_i = P f_i \text{ for } i \in \{1, \ldots, k\}\}$$

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 in Hiriart-Urruty and Lemaréchal (2001, 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

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

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

Theorem 3.3 (Tyurin, 2009a,b, 2011)**.** *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*

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

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

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

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 Winkler (1988) and Pinelis (2016) 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 (1.22) follows from Lemma 2.2 using dominated convergence. Inequality (1.23) follows from (1.21) in case of $\|f\|_{\mathrm{L},\alpha} = 0$, and otherwise from (1.22) applied to $f/\|f\|_{\mathrm{L},\alpha}$.

(c) If $\zeta_s(P,Q) < \infty$, then we apply Lemma 2.2 to $f \coloneqq \left(\prod_{j=0}^{m-1}(s-j)\right)^{-1} |\cdot|^s \in \mathcal{F}_s$ 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 (1.22) 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 (1.21) holds, then the third follows easily using (2.8), compare Senatov (1998, pp. 102–103). Finally, the third condition in (1.21) 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*

$$\zeta_s(P*R, Q*R) \quad \le \quad \zeta_s(P,Q) \qquad \text{for } P,Q,R \in \mathrm{Prob}(\mathbb{R}) \tag{4.1}$$

proved e.g. in Senatov (1998, p. 101), which, given Theorem 1.7(b), is equivalent to its *semiadditivity*

$$\zeta_s\big(\mathop{\ast}_{i=1}^{n} P_i, \mathop{\ast}_{i=1}^{n} Q_i\big) \quad \le \quad \sum_{i=1}^{n} \zeta_s(P_i,Q_i) \qquad \text{for } n \in \mathbb{N} \text{ and } P_i, Q_i \in \mathrm{Prob}(\mathbb{R}), \tag{4.2}$$

compare Senatov (1998, p. 48). To formulate the second, we use here, as well as later in some proofs, the obvious random variable notation $\zeta_s(X,Y) \coloneqq \zeta_s(P,Q)$ if X, Y are $\mathbb{R}$-valued r.v.'s with $X \sim P$ and $Y \sim Q$. Then we have the *homogeneity*

$$\zeta_s(aX, aY) \quad = \quad a^s \zeta_s(X,Y) \quad \text{for } a \in [0,\infty[\text{ and } \mathbb{R}\text{-valued r.v.'s } X \text{ and } Y, \tag{4.3}$$

the obvious proof of which being given in Senatov (1998, p. 102).

The following Lemma, which is presented in Senatov (1998, pp. 108-112) without explicit constants, allows us in the proof of Theorem 1.15, in a case where $aX \sim P$ and $aY \sim Q$ with small a, to use the homogeneity (4.3) with a better exponent than possible by just using (4.1). 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*

$$\zeta_s(P*\mathrm{N}_\sigma, Q*\mathrm{N}_\sigma) \quad \le \quad C_{s,t}\frac{\zeta_{s+t}(P,Q)}{\sigma^t} \tag{4.4}$$

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

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

and letting φ denote the standard normal density, we put

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

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

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 Senatov (1998, 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

$$g(x) \;:=\; \int f(x+z)\varphi_\sigma(z)\,\mathrm{d}z \quad \text{and} \quad h(x) \;:=\; \frac{\sigma^t g(x)}{C_{s,t}} \quad \text{for } x\in\mathbb{R}, \tag{4.5}$$

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

$$|(P*\mathrm{N}_\sigma)f-(Q*\mathrm{N}_\sigma)f| \;=\; |Pg-Qg| \;=\; \frac{C_{s,t}}{\sigma^t}|Ph-Qh| \;\le\; \text{R.H.S.(4.4)}$$

as desired. So let $f\in\mathcal{F}_s^\infty$ and let g and h be defined through (4.5). 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

$$\left|g^{(n)}(x)-g^{(n)}(y)\right| \;\le\; \frac{C_{s,t}}{\sigma^t}|x-y|^\gamma \quad \text{for } x,y\in\mathbb{R}. \tag{4.6}$$

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

$$g^{(\ell)}(x) \;=\; \int f^{(\ell)}(x+z)\varphi_\sigma(z)\,\mathrm{d}z \;=\; \int f^{(\ell)}(z)\varphi_\sigma(x-z)\,\mathrm{d}z, \tag{4.7}$$

$$g^{(k)}(x) \;=\; \int f^{(\ell)}(z)\varphi_\sigma^{(k-\ell)}(x-z)\,\mathrm{d}z \;=\; \int f^{(\ell)}(x-z)\varphi_\sigma^{(k-\ell)}(z)\,\mathrm{d}z, \tag{4.8}$$

$$\begin{aligned}|g^{(k)}(x)-g^{(k)}(y)| \;&\le\; \int\left|f^{(\ell)}(x-z)-f^{(\ell)}(y-z)\right|\left|\varphi_\sigma^{(k-\ell)}(z)\right|\,\mathrm{d}z\\ &\le\; |x-y|^\alpha\frac{D_{k-\ell}}{\sigma^{k-\ell}}\end{aligned} \tag{4.9}$$

where, to justify differentiation under the integral, we may in (4.7) apply the dominated convergence theorem successively using polynomial bounds on the derivatives $f',\ldots,f^{(\ell)}$, compare (2.8) and the ensuing line, and we may treat (4.8) similarly, or remember it as a well-known special case of the differentiability of Laplace transforms, see for example Mattner (2001, Example); in the last step in (4.9) we used

$f \in \mathcal{F}_s$ and the change of variables $z \mapsto \sigma z$. Specializing (4.8) to $k := \ell + m + 1$ and using in the first step below $\int \varphi_\sigma^{(m+1)}(z)\,\mathrm{d}z = 0$ yields

$$\begin{aligned} |g^{(\ell+m+1)}(x)| &= \left| \int \Big(f^{(\ell)}(x-z) - f^{(\ell)}(x) \Big) \varphi_\sigma^{(m+1)}(z)\,\mathrm{d}z \right| \\ &\le \int |z|^\alpha \left| \varphi_\sigma^{(m+1)}(z) \right| \mathrm{d}z \;=\; \frac{D_{m+1,\alpha}}{\sigma^{m+1-\alpha}} \quad \text{for } x \in \mathbb{R}. \end{aligned} \tag{4.10}$$

Let us now first assume that we have $\alpha + \beta \le 1$, and hence $n = \ell + m$ and $\gamma = \alpha + \beta$. Then, using (4.10) in the second step below, we get

$$\text{L.H.S.(4.6)} \;\le\; \|g^{(n+1)}\|_\infty \cdot |x-y| \;\le\; \frac{D_{m+1,\alpha}}{\sigma^{m+1-\alpha}} |x-y| \quad \text{for } x, y \in \mathbb{R},$$

and taking a geometric mean of this bound and the one from (4.9) with $k := n$, with the exponents $u := \beta/(1-\alpha) \in \,]0,1]$ and $1-u$, yields (4.6) 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 (4.10) to x and to y, we get

$$\text{L.H.S.(4.6)} \;\le\; \frac{2D_{m+1,\alpha}}{\sigma^{m+1-\alpha}} \quad \text{for } x, y \in \mathbb{R},$$

and taking a geometric mean of this bound and the one from (4.9) with $k := n$, with the exponents $v := (1-\beta)/\alpha \in [0,1[$ and $1-v$, yields (4.6) 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 Zolotarev (1997), Denuit et al. (1998), and Boutsikas and Vaggelatou (2002). In particular, Theorem 4.2(b) contains Denuit et al. (1998, Theorems 3.3 and 4.3) and Boutsikas and Vaggelatou (2002, 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 Denuit et al. (1998, p. 590), Boutsikas and Vaggelatou (2002, p. 351), Müller and Stoyan (2002, p. 39, Definition 1.6.2 a)), and Shaked and Shanthikumar (2007, p. 139), but being here somewhat more explicit with respect to the appropriate integrability assumptions: If $s \in \mathbb{N}$, then

$$P \quad \le_{s\text{-cx}} \quad Q \tag{4.11}$$

is defined to mean that $P, Q \in \mathrm{Prob}_{s-1}(\mathbb{R})$ and that $Pf \le Qf$ holds for every s-convex function $f : \mathbb{R} \to \mathbb{R}$ such that Pf and Qf are well-defined (possibly infinite). Thus $\le_{1\text{-cx}}$ is just the usual stochastic order $\le_{\mathrm{st}}$ on $\mathrm{Prob}(\mathbb{R})$, $\le_{2\text{-cx}}$ is the usual convex order $\le_{\mathrm{cx}}$ on $\mathrm{Prob}_1(\mathbb{R})$, and $\le_{3\text{-cx}}$ 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 (4.11) necessitates

$$\mu_j(P) \;=\; \mu_j(Q) \;\in\; \mathbb{R} \quad \text{for } j \in \{1, \ldots, s-1\}. \tag{4.12}$$

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 $x_-^0 := (x \le 0)$ and $x_+^0 := (x \ge 0)$, 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 $\overline{F}(x) \coloneqq P([x,\infty[)$ 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 $F_1 \coloneqq F$, $\overline{F}_1 \coloneqq \overline{F}$,

$$F_{k+1}(t) \;\coloneqq\; \int_{-\infty}^{t} F_k(x)\,\mathrm{d}x, \qquad \overline{F}_{k+1}(t) \;\coloneqq\; \int_{t}^{\infty} \overline{F}_k(x)\,\mathrm{d}x, \tag{4.13}$$

and hence get, as follows by inserting the right hand sides from (4.14) into the integrals in (4.13) and using Fubini,

$$F_k(t) \;=\; \int \frac{(x-t)_-^{k-1}}{(k-1)!}\,\mathrm{d}P(x), \qquad \overline{F}_k(t) \;=\; \int \frac{(x-t)_+^{k-1}}{(k-1)!}\,\mathrm{d}P(x). \tag{4.14}$$

By (4.14), the functions F_k and $\overline{F}_k$ are finite-valued in particular if $P \in \mathrm{Prob}_{k-1}(\mathbb{R})$, and then (4.13) with $k-1$ in place of k yields

$$\lim_{t\to-\infty} F_k(t) = 0, \qquad \lim_{t\to\infty} \overline{F}_k(t) = 0. \tag{4.15}$$

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 (4.12). 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 (4.13) and (4.14), let $H_k \coloneqq G_k - F_k$ and $\overline{H}_k \coloneqq \overline{G}_k - \overline{F}_k$ for $k \in \{1,\ldots,s\}$.*

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

$$(-1)^{k-1} H_k(t) + \overline{H}_k(t+) \;=\; 0, \tag{4.16}$$

$$(-1)^{k-1} H_k(t-) + \overline{H}_k(t) \;=\; 0, \tag{4.17}$$

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

$$\overline{H}_k(-t) \;=\; (-1)^k \overline{H}_k(t+)\,; \tag{4.18}$$

here the one-sided limit signs, namely "+" in the argument of $\overline{H}_k$ in (4.16) and (4.18), and "−" in the argument of H_k in (4.17), 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

$$\lim_{t\to-\infty} \overline{H}_k(t) \;=\; \lim_{t\to\infty} \overline{H}_k(t) \;=\; 0. \tag{4.19}$$

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

$$\zeta_s(P,Q) \;=\; \int |\overline{H}_s(x)|\,\mathrm{d}x, \tag{4.20}$$

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

$$\zeta_s(P,Q) \;=\; Qf - Pf \tag{4.21}$$

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

$$f^{(s)}(x) \;=\; \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} \quad \textit{if} \quad \overline{H}_s(x) \begin{Bmatrix} < \\ > \end{Bmatrix} 0, \qquad \textit{for Lebesgue-a.e. } x \in I. \tag{4.22}$$

(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*

$P \le_{s-\mathrm{cx}} Q$ is further equivalent to $\zeta_s(P,Q) = \frac{1}{s!}(\mu_s(Q) - \mu_s(P))$, that is, to (4.21) holding for the function $f \in \mathcal{F}_s$ given by

$$f(x) \quad := \quad \tfrac{1}{s!}x^s \qquad \text{for } x \in \mathbb{R}. \tag{4.23}$$

(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 (4.21) holding, with some sign change point x_0 of $\overline{H}_s$, for the function $f \in \mathcal{F}_s$ given by*

$$f(x) \quad := \quad \tfrac{1}{s!}|x - x_0|^s \qquad \text{for } x \in \mathbb{R}, \tag{4.24}$$

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 (4.21) holds with $f(x) := |x|^s/s!$ for $x \in \mathbb{R}$.*

Proof: (a) For every $t \in \mathbb{R}$, (4.14) 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 (4.12) yields (4.16) and (4.17). If now $P - Q$ is assumed to be symmetric, then, using this in the second step below, and using (4.14) applied to Q and to P in the first and fourth steps, and (4.16) in the fifth, we get (4.18) 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 $\le k-1$ 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 (4.19) follows using (4.15) and (4.16).

Assume now $P, Q \in \mathrm{Prob}_s(\mathbb{R})$. If $f \in \mathcal{F}_s$, then the representation $f(x) = \sum_{j=0}^{s-1}\frac{f^{(j)}(0)}{j!}x^j + \int_0^x \frac{(x-y)^{s-1}}{(s-1)!}f^{(s)}(y)\,\mathrm{d}y = \sum_{j=0}^{s-1}\frac{f^{(j)}(0)}{j!}x^j + \int_{\mathbb{R}}\big((0 \le y < x) - (x \le y < 0)\big)\frac{(x-y)^{s-1}}{(s-1)!}f^{(s)}(y)\,\mathrm{d}y$ and a Fubini calculation, valid due to $\|f^{(s)}\|_\infty \le 1$ and the moment assumption just introduced, and using (4.16) with $k = s$, yield the formula

$$Qf - Pf \quad = \quad \int f^{(s)}(x)\overline{H}_s(x)\,\mathrm{d}x. \tag{4.25}$$

By applying (4.25) to $f \in \mathcal{F}_s^\infty$ and using $\|f^{(s)}\|_\infty \le 1$ we get "$\le$" in (4.20). By applying (4.25) to a function $f \in \mathcal{F}_s$ with $f^{(s)}(x) = \mathrm{sgn}(\overline{H}_s(x))$ for Lebesgue-a.e. x, and using Theorem 1.7(d), we get "$\ge$" in (4.20). Finally, (4.20) and (4.25) yield the claim involving (4.22).

(b) Using (4.19), the implications $(B_1) \Rightarrow (B_2) \Rightarrow \ldots \Rightarrow (B_s)$ follow from Lemma 2.8 up to the statement involving (2.20), since (4.13) 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 Denuit et al. (1998, Theorem 3.2), using (4.14). Since (4.22) holds for f from (4.23) 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 (4.16), 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 (4.17) and the recursion (4.13), a contradiction. Thus indeed (C_{k+1}) holds.

Let $x_0 \in \mathbb{R}$ and f be as in (4.24). Then $f^{(s)}(x) = \mathrm{sgn}(x - x_0)$ for $x \in \mathbb{R} \setminus \{x_0\}$, and hence (4.22) 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 (4.18) 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 (4.21) holds with f from (4.24) with $x_0 = 0$.

Hence, if s were even, then (4.21) would hold with f from (4.23), 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 (1.28), the results (4.26) and (4.27) are used in the proofs of Theorems 1.15 and 1.10 in section 6 below.

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

$$\varepsilon_1 = \zeta_3(Q, \mathrm{N}) = \frac{1}{6}\left(\frac{4}{\sqrt{2\pi}} - 1\right) < 0.0993, \tag{4.26}$$

$$\zeta_4(Q, \mathrm{N}) = \frac{1}{12} < 0.0834, \tag{4.27}$$

$$\zeta_s(Q, \mathrm{N}) = \infty \quad \textit{for } s \in]4, \infty[. \tag{4.28}$$

Proof: Claim (4.28) follows from Theorem 1.7(c) with $m \geq 4$, since $\mu_4(Q) = 1 \neq 3 = \mu_4(\mathrm{N})$.

For proving (4.26) and (4.27) using Theorem 4.2, let us change here the notation and put for the rest of this proof

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

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 (4.12), 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 $f(x) \coloneqq x^4/4!$ from (4.23), we accordingly get

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

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

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

□

5. Proof of the main result

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 (1.13) occurs under the stated conditions. Indeed, we then have $\widetilde{\ast_{i=1}^n Q_i}\, f = 0$ by symmetry, and thus

$$\begin{aligned}\text{L.H.S.}(1.13) \quad &= \quad \left|\mathbb{E}\, c\left(\frac{1}{\sigma}\sum_{i=1}^n X_i\right)^3\right| \quad = \quad \frac{1}{6\sigma^3}\left|\sum_{i=1}^n \mathbb{E}X_i^3\right| 6|c| \\ &= \quad \frac{1}{6\sigma^3}\sum_{i=1}^n \sigma_i^3\left|\mathbb{E}\left(\frac{X_i}{\sigma_i}\right)^3\right| \|f''\|_{\mathrm{L}} \quad = \quad \text{R.H.S.}(1.13)\end{aligned}$$

by using in the third step above the additivity of the third centred moment for independent random variables, that is, (1.27) with $\ell = 3$, and in the last step the equality statement in Example 1.3, that is, a rather easy part of Shevtsova (2014, 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 as in Rudin (1987, Theorem 5.20) yields an $\mathbb{R}$-linear functional $\ell : E \to \mathbb{R}$ of norm 1 satisfying the first of the following equalities

$$\text{L.H.S.}(1.13) \quad = \quad \ell\left(\widetilde{\mathop{\ast}_{i=1}^n P_i}\, f - \widetilde{\mathop{\ast}_{i=1}^n Q_i}\, f\right) \quad = \quad \widetilde{\mathop{\ast}_{i=1}^n P_i}\, \ell\circ f - \widetilde{\mathop{\ast}_{i=1}^n Q_i}\, \ell\circ f,$$

and thus an application of inequality (1.13) 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 (1.13) 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 $c \cdot \big(\widetilde{\ast_{i=1}^n P_i}\, f - \widetilde{\ast_{i=1}^n Q_i}\, f\big)$ is real and ≥ 0).

Step 3. It is enough to prove inequality (1.24), 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 (1.23) with $s := 3$, and in view of Steps 1 and 2.

Step 4. It is enough to prove inequality (1.24) in case of $n = 1$, since assuming this special case to be true yields the penultimate step below in

$$\begin{aligned}
\text{L.H.S.(1.24)} \quad &= \quad \zeta_3\left(\frac{1}{\sigma}\sum_{i=1}^n X_i\,,\, \frac{1}{\sigma}\sum_{i=1}^n Y_i\right) \quad = \quad \frac{1}{\sigma^3}\zeta_3\left(\sum_{i=1}^n X_i\,,\, \sum_{i=1}^n Y_i\right)\\
&\le \quad \frac{1}{\sigma^3}\sum_{i=1}^n \zeta_3(X_i, Y_i) \quad = \quad \frac{1}{\sigma^3}\sum_{i=1}^n \sigma_i^3 \zeta_3(\widetilde{X_i}, \widetilde{Y_i})\\
&\le \quad \frac{1}{\sigma^3}\sum_{i=1}^n \sigma_i^3 \frac{B(\varrho_i)}{6} \quad = \quad \text{R.H.S.(1.24)},
\end{aligned}$$

where we have used the homogeneity (4.3) in the second and fourth steps, and the semiadditivity (4.2) in the third.

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

$$Q \quad := \quad \tfrac{1}{2}(\delta_{-1} + \delta_1). \tag{5.1}$$

By Step 4, it remains to prove that we have

$$\zeta_3(P,Q) - \frac{B(\varrho(P))}{6} \quad \le \quad 0 \tag{5.2}$$

for $P \in \widetilde{\mathcal{P}_3}$ or, equivalently in view of the alternative representation (1.22) of ζ_3, that

$$Pf - Qf - \frac{B(\varrho(P))}{6} \quad \le \quad 0 \tag{5.3}$$

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 (5.3) is a function of Pf_3 and Pf_4, that it is enough to prove (5.3) 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 (5.2) for $P \in \mathcal{P}$ where

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

Let $F(P)$ be the left hand side of (5.2) 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 (5.2) for P standardized and having at most three support points. So, for the remaining two steps, let

$$P \quad = \quad p\delta_\alpha + q\delta_\beta + (1-p-q)\delta_\gamma \tag{5.4}$$

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 (5.4) and Q from (5.1). 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 Shevtsova (2014, Theorem 6), that is, from (1.15) 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

$$\zeta_3(P,Q) \quad = \quad \frac{1}{6}\left(\int |x-r|^3\,\mathrm{d}P(x) - \int |x-r|^3\,\mathrm{d}Q(x)\right). \tag{5.5}$$

If $r=0$, then R.H.S.(5.5) $= \frac{1}{6}(\varrho(P)-1) \le \frac{1}{6}B(\varrho(P))$, 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 $v\cdot\operatorname{sgn}(r)$ and $-u\cdot\operatorname{sgn}(r)$ 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 (5.5) 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 involving ζ_3-distances between normal and convolutions of symmetric two-point laws

Proof of Theorem 1.15: Inequality (1.37) follows from (1.24) in Theorem 1.8 by using the triangle inequality for ζ_3 recalled in Theorem 1.7(b). For the remaining claim, we assume without loss of generality that

$$\sum_{i=1}^n \sigma_i^2 \quad = \quad 1. \tag{6.1}$$

Let $Y, Y_1, \ldots, Y_n, Z, Z_1, \ldots, Z_n$ be independent r.v.'s with $Y \sim \frac{1}{2}(\delta_{-1}+\delta_1)$, $Y_i \sim Q_i = \frac{1}{2}(\delta_{-\sigma_i}+\delta_{\sigma_i})$ and hence $Y_i \sim \sigma_i Y$, $Z \sim \mathrm{N}$, and $Z_i \sim \mathrm{N}_{\sigma_i}$, so that $Z_i \sim \sigma_i Z$, 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 (6.1), we get $T_0 \sim \ast_{i=1}^n Q_i = \widetilde{\ast_{i=1}^n Q_i}$ and $T_n \sim \mathrm{N}$ and hence, writing in this proof ε_n for a quantity more general than the one introduced in (1.28), 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 in the last step.

The regularity (4.1) and the homogeneity (4.3) 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

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

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 (6.1), 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

$$\varepsilon_n \;\le\; \zeta_3(Y, Z)\sigma_1^3 + \sqrt{\frac{2}{\pi}}\, \zeta_4(Y, Z) \sum_{k=1}^{n-1} \frac{\sigma_{k+1}^3 \min\{1, \sqrt{n}\,\sigma_{k+1}\}}{\sqrt{k}}. \tag{6.2}$$

Inserting now the values for $\zeta_3(Y, Z)$ and $\zeta_4(Y, Z)$ from (4.26) and (4.27) in Example 4.3 yields the claim. □

Proof of Theorem 1.10: For the upper bound we observe that formula (6.2), specialized to the homoscedastic case $\sigma_1 = \ldots = \sigma_n = 1/\sqrt{n}$, yields

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

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

$$\varepsilon_n \;<\; 2\sqrt{\frac{2}{\pi}} \cdot \frac{\zeta_4(Y, Z)}{n} + \frac{\zeta_3(Y, Z) + \zeta(\frac{1}{2})\sqrt{\frac{2}{\pi}}\zeta_4(Y, Z)}{n^{3/2}},$$

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 $\mathrm{B}_{n,\frac{1}{2}}$ as defined in MacWilliams and Sloane (1977, pp. 130, 151–154, the case of $q = 2$ and hence $\gamma = 1$) and also, with the unnecessary restriction $k \le n$, in Diaconis and Zabell (1991, section 6.2 on p. 298, the special case of $p = \frac{1}{2}$ and hence $\gamma = 1$), that is,

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

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

$$(k+1)P_{k+1}^n(x) \quad = \quad (n-2x)P_k^n(x) - (n-k+1)P_{k-1}^n(x) \qquad \text{for } k \in \{1, \ldots, n-1\}$$

and hence further

$$\begin{aligned} P_2^n(x) &= 2\left(\left(x - \tfrac{n}{2}\right)^2 - \tfrac{n}{4}\right), \\ P_3^n(x) &= -\tfrac{4}{3}\left(x - \tfrac{n}{2}\right)^3 + \left(n - \tfrac{2}{3}\right)\left(x - \tfrac{n}{2}\right). \end{aligned}$$

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

$$\sum_{x=0}^{a} P_k^n(x) \mathrm{b}_{n,\frac{1}{2}}(x) \quad = \quad \tfrac{n-a}{k} P_{k-1}^{n-1}(a) \mathrm{b}_{n,\frac{1}{2}}(a), \tag{6.3}$$

and hence in particular

$$\begin{aligned} \sum_{x=0}^{a} P_1^n(x) \mathrm{b}_{n,\frac{1}{2}}(x) &= (n-a)\mathrm{b}_{n,\frac{1}{2}}(a), \\ \sum_{x=0}^{a} P_3^n(x) \mathrm{b}_{n,\frac{1}{2}}(x) &= \tfrac{2}{3}(n-a)\left(\left(a - \tfrac{n-1}{2}\right)^2 - \tfrac{n-1}{4}\right) \mathrm{b}_{n,\frac{1}{2}}(a) \end{aligned}$$

and thus

$$\begin{aligned} \sum_{x=0}^{a} \left(x - \tfrac{n}{2}\right) \mathrm{b}_{n,\frac{1}{2}}(x) &= \tfrac{a-n}{2} \mathrm{b}_{n,\frac{1}{2}}(a), \\ \sum_{x=0}^{a} \left(x - \tfrac{n}{2}\right)^3 \mathrm{b}_{n,\frac{1}{2}}(x) &= \sum_{x=0}^{a} \left(-\tfrac{3}{4} P_3^n(x) - \left(\tfrac{3}{8}n - \tfrac{1}{4}\right) P_1^n(x)\right) \mathrm{b}_{n,\frac{1}{2}}(x) \\ &= \tfrac{a-n}{2} \left(\left(a - \tfrac{n-1}{2}\right)^2 + \tfrac{1}{2}n - \tfrac{1}{4}\right) \mathrm{b}_{n,\frac{1}{2}}(a), \end{aligned}$$

and finally

$$\begin{aligned} \sum_{x=0}^{n} \left|x - \tfrac{n}{2}\right|^3 \mathrm{b}_{n,\frac{1}{2}}(x) &= -2 \sum_{x=0}^{\lfloor \frac{n}{2} \rfloor} \left(x - \tfrac{n}{2}\right)^3 \mathrm{b}_{n,\frac{1}{2}}(x) \\ &= \begin{cases} \tfrac{1}{4}n^2 \mathrm{b}_{n,\frac{1}{2}}\left(\tfrac{n}{2}\right) & \text{if } n \text{ is even,} \\ \left(\tfrac{1}{4}n^2 + \tfrac{1}{8}n - \tfrac{1}{8}\right) \mathrm{b}_{n,\frac{1}{2}}\left(\lfloor \tfrac{n}{2} \rfloor\right) & \text{if } n \text{ is odd.} \end{cases} \end{aligned} \tag{6.4}$$

Recalling the local Edgeworth expansion for binomial laws (see e.g. Gnedenko and Kolmogorov, 1954, § 51, Theorem 1)

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

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 in the last step below $2\alpha_n^2 = \alpha_n$,

$$\begin{aligned}
\varepsilon_n &\geq \left| \int \frac{|\cdot|^3}{6} \,\mathrm{d}\left(\mathrm{N} - \widetilde{\mathrm{B}_{n,\frac{1}{2}}}\right) \right| \\
&= \frac{1}{6} \left| \frac{4}{\sqrt{2\pi}} - \frac{2^3}{n^{3/2}} \sum_{x=0}^{n} \left| x - \tfrac{n}{2} \right|^3 \mathrm{b}_{n,\frac{1}{2}}(x) \right| \\
&= \frac{1}{6} \left| \frac{4}{\sqrt{2\pi}} - \frac{2^3}{n^{3/2}} \text{R.H.S.(6.4)} \right| \\
&= \frac{1}{6} \left| \frac{4}{\sqrt{2\pi}} - \frac{2^3}{n^{3/2}} \left(\tfrac{n^2}{4} + \alpha_n \tfrac{n}{4} + O(1) \right) \sqrt{\tfrac{4}{n}} \right. \\
&\qquad \left. \times \left(\Phi'\left(-\alpha_n / \sqrt{\tfrac{n}{4}} \right) \left(1 - \tfrac{3 + O(n^{-1})}{12n} \right) + O(n^{-2}) \right) \right| \\
&= \frac{4}{6\sqrt{2\pi}} \left| 1 - \left(1 + \tfrac{\alpha_n}{n} \right) \left(1 - \tfrac{2\alpha_n^2}{n} \right) \left(1 - \tfrac{1}{4n} \right) + O(n^{-2}) \right| \\
&= \frac{1}{6\sqrt{2\pi}\, n} + O\left(\tfrac{1}{n^2} \right).
\end{aligned}$$

□

Proof of Theorem 1.11: Inequality (1.29) results from (1.37) in Theorem 1.15, already proved above, when specialized to the i.i.d. case. Alternatively, we may first specialize Theorem 1.8 to the i.i.d. case and then apply the triangle inequality similarly to (6.5) below.

Let now $\varrho \in [1, \infty[$ and $f(x) = x^3/6$ for $x \in \mathbb{R}$. Then we have

$$\sqrt{k} \left| \widetilde{P_\varrho^{*k}} f - \mathrm{N} f \right| = \sqrt{k} \left| \widetilde{P_\varrho^{*k}} f \right| = \frac{1}{k} \left| P_\varrho^{*k} f \right| = \frac{B(\varrho)}{6} \quad \text{for } k \in \mathbb{N}$$

by using in the last step above (1.27) with $\ell = 3$, as we did in Step 1 of the proof of Theorem 1.2, and hence we get

$$\frac{B(\varrho)}{6} \leq \varliminf_{n \to \infty} \sqrt{n}\, \zeta_3 \left(\widetilde{P_\varrho^{*n}}, \mathrm{N} \right) \leq \varlimsup_{n \to \infty} \sqrt{n}\, \zeta_3 \left(\widetilde{P_\varrho^{*n}}, \mathrm{N} \right) \leq \frac{B(\varrho)}{6}$$

using in the last step (1.29) with $P = P_\varrho$ and $\varepsilon_n = O(n^{-1})$. This proves (1.30).

Let finally $n \in \mathbb{N}$. For $P \in \mathcal{P}_3$ using the triangle inequality for ζ_3 in the first step below and the i.i.d. case of Theorem 1.8 in the second we then have

$$\left| \zeta_3 \left(\widetilde{P^{*n}}, \mathrm{N} \right) - \varepsilon_n \right| \leq \zeta_3 \left(\widetilde{P^{*n}}, \widetilde{\mathrm{B}_{n,\frac{1}{2}}} \right) \leq \tfrac{1}{6} B(\varrho(P)), \qquad (6.5)$$

and (1.31) follows using $\lim_{\varrho \to 1} B(\varrho) = 0$. □

Acknowledgements

We thank Jürgen Müller for showing us the limit result in Hardy (1949) and Wirths (2015) used in the proof of Lemma 2.9, Bero Roos for helpful remarks on an earlier version of the present paper, and a very careful anonymous referee for suggestions leading to a dozen improvements.

References

J. Angst and G. Poly. A weak Cramér condition and application to Edgeworth expansions. *Electron. J. Probab.* **22** (59), 1–24 (2017). MR3683368.

T. V. Arak. On the convergence rate in Kolmogorov's uniform limit theorem. I. *Theory Probab. Appl.* **26** (2), 219–239 (1981a). DOI: 10.1137/1126026. Original Russian text: *Teor. Veroyatn. Primen.* **26**, 225–245.

T. V. Arak. On the convergence rate in Kolmogorov's uniform limit theorem. II. *Theory Probab. Appl.* **26** (3), 437–451 (1981b). DOI: 10.1137/1126052. Original Russian text: *Teor. Veroyatn. Primen.* **26**, 449–463.

T. V. Arak. An improvement of the lower bound for the rate of convergence in Kolmogorov's uniform limit theorem. *Theory Probab. Appl.* **27** (4), 826–832 (1982). DOI: 10.1137/1127090. Original Russian text: *Teor. Veroyatn. Primen.* **27**, 767–772.

T. V. Arak and A. Yu. Zaitsev. Uniform limit theorems for sums of independent random variables. *Proc. Steklov Instit. Math.* **174**, 1–222 (1988). Original Russian text: *Tr. MIAN USSR.* **174**, 3–214 (1986).

B. von Bahr. On the convergence of moments in the central limit theorem. *Ann. Math. Statist.* **36**, 808–818 (1965). MR0179827.

A. D. Barbour. Asymptotic expansions based on smooth functions in the central limit theorem. *Probab. Theory Relat. Fields* **72**, 289–303 (1986). MR836279.

V. Bentkus. On the asymptotical behavior of the constant in the Berry-Esseen inequality. *J. Theoret. Probab.* **7** (2), 211–224 (1994). MR1270601.

R. N. Bhattacharya and R. R. Rao. *Normal approximation and asymptotic expansions*, volume 64 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2010). ISBN 978-0-898718-97-3. Updated reprint of the 1986 edition, corrected edition of the 1976 original. MR3396213.

I. S. Borisov, D. A. Panchenko and G. I. Skilyagina. On minimal smoothness conditions for asymptotic expansions of moments in the CLT. *Siberian Adv. Math.* **8** (1), 80–95 (1998). MR1651906.

I. S. Borisov and G. I. Skilyagina. On asymptotic expansion of the moments of smooth functions in the central limit theorem. *Sib. Math. J.* **37**, 447–453 (1996). Original Russian text: *Sibirsk. Mat. Zh.* **37**, 519–525.

M. Boutsikas. Penultimate gamma approximation in the CLT for skewed distributions. *ESAIM: Probability and Statistics* **19**, 590–604 (2015). DOI: 10.1051/ps/2015010.

M. Boutsikas and E. Vaggelatou. On the distance between convex-ordered random variables, with applications. *Adv. in Appl. Probab.* **34** (2), 349–374 (2002). MR1909919.

P. L. Butzer, L. Hahn and U. Westphal. On the rate of approximation in the central limit theorem. *J. Approximation Theory* **13**, 327–340 (1975). MR0394809.

G. P. Chistyakov. Asymptotically proper constants in the Lyapunov theorem. *J. Math. Sci.* **93** (3), 480–483 (1999). DOI: 10.1007/BF02364834.

G. P. Chistyakov. A new asymptotic expansion and asymptotically best constants in Lyapunov's theorem. I. *Theory Probab. Appl.* **46** (2), 226–242 (2002a). DOI: 10.1137/S0040585X97978932. Original Russian text: *Teor. Veroyatn. Primen.* **46**, 326–344 (2001).

G. P. Chistyakov. A new asymptotic expansion and asymptotically best constants in Lyapunov's theorem. II. *Theory Probab. Appl.* **46** (3), 516–522 (2002b). DOI: 10.1137/S0040585X97979172. Original Russian text: *Teor. Veroyatn. Primen.* **46**, 573–579 (2001).

G. P. Chistyakov. A new asymptotic expansion and asymptotically best constants in Lyapunov's theorem. III. *Theory Probab. Appl.* **47** (3), 395–414 (2003). DOI: 10.1137/S0040585X97979822. Original Russian text: *Teor. Veroyatn. Primen.* **46**, 475–497 (2002).

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

M. Denuit, C. Lefevre and M. Shaked. The s-convex orders among real random variables, with applications. *Math. Inequal. Appl.* **1** (4), 585–613 (1998). MR1646686.

R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin (1993). ISBN 3-540-50627-6. MR1261635.

P. Diaconis and S. Zabell. Closed form summation for classical distributions: variations on a theme of de Moivre. *Statist. Sci.* **6** (3), 284–302 (1991). MR1144242.

J. Diestel and J. J. Uhl Jr. *Vector measures*. American Mathematical Society, Providence, R.I. (1977). MR0453964.

T. Dinev and L. Mattner. The asymptotic Berry-Esseen constant for intervals. *Theory Probab. Appl.* **57** (2), 323–325 (2013). MR3201658. Original text: *Teor. Veroyatn. Primen.* **57** (2), 381–325 (2012).

R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge (2002). ISBN 0-521-00754-2. Revised reprint of the 1989 original. MR1932358.

C.-G. Esseen. Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.* **77**, 1–125 (1945). MR0014626.

C.-G. Esseen. A moment inequality with an application to the central limit theorem. *Skand. Aktuarietidskr.* **39**, 160–170 (1956). MR0090166.

T. S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach*. Probability and Mathematical Statistics, Vol. 1. Academic Press, New York-London (1967). MR0215390.

R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philosphical Soc.* **24**, 180–190 (1928).

J. K. Ghosh. *Higher Order Asymptotics*, volume 4 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Mathematical Statistics (1994).

B. V. Gnedenko and A. N. Kolmogorov. *Limit distributions for sums of independent random variables*. Addison-Wesley Publishing Company, Inc., Cambridge, Mass. (1954). MR0062975.

L. Goldstein. Bounds on the constant in the mean central limit theorem. *Ann. Probab.* **38** (4), 1672–1689 (2010). MR2663641.

F. Götze and C. Hipp. Asymptotic expansions in the central limit theorem under moment conditions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **42** (1), 67–87 (1978). MR0467882.

A. Hald. The early history of cumulants and the Gram–Charlier series. *International Statistical Review* **68**, 137–153 (2000). DOI: 10.1093/acprof:oso/9780198509721.003.0007.

G. H. Hardy. *Divergent Series.* Oxford, at the Clarendon Press (1949). MR0030620.

C. Hipp and L. Mattner. On the normal approximation to symmetric binomial distributions. *Theory Probab. Appl.* **52** (3), 516–523 (2008). DOI: 10.1137/S0040585X97983213. Original text: *Teor. Veroyatn. Primen.* **52**, 610–617 (2007).

J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis.* Grundlehren Text Editions. Springer-Verlag, Berlin (2001). ISBN 3-540-42205-6. MR1865628.

W. Hoeffding. The extrema of the expected value of a function of independent random variables. *Ann. Math. Statist.* **26**, 268–275 (1955). MR0070087.

E. Hopf. *Über die Zusammenhänge zwischen gewissen höheren Differenzenquotienten reeller Funktionen einer reellen Variablen und deren Differenzierbarkeitseigenschaften.* Norddeutsche Buchdruckerei und Verlagsanstalt A.-G., Berlin (1926).

S. Ikeda. A note on the normal approximation to the sum of independent random variables. *Ann. Inst. Statist. Math. Tokyo* **11**, 121–130 (1959). MR0130710.

Y. Jiao. Zero bias transformation and asymptotic expansions. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** (1), 258–281 (2012). MR2919206.

S. Karlin. *Total Positivity, Vol. I.* Stanford University Press, Stanford, Calif (1968). MR0230102.

J. H. B. Kemperman. The general moment problem, a geometric approach. *Ann. Math. Statist* **39**, 93–122 (1968). MR0247645.

M. Kuczma. *An introduction to the theory of functional equations and inequalities. Cauchy's equation and Jensen's inequality.* Birkhäuser Verlag, Basel, second edition (2009). ISBN 978-3-7643-8748-8. MR2467621.

M. Ledoux and M. Talagrand. *Probability in Banach spaces. Isoperimetry and processes.* Classics in Mathematics. Springer-Verlag, Berlin (2011). ISBN 978-3-642-20211-7. Reprint of the 1991 edition. MR2814399.

F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes.* North-Holland Publishing Co., third printing of the 1981 edition (1977). ISBN 0-444-85010-4. North-Holland Mathematical Library, Vol. 16. MR0465510.

L. Mattner. What are cumulants? *Doc. Math.* **4**, 601–622 (1999). MR1727425.

L. Mattner. Complex differentiation under the integral. *Nieuw Arch. Wiskd. (5)* **2** (1), 32–35 (2001). MR1823156.

L. Mattner. Cumulants are universal homomorphisms into Hausdorff groups. *Probab. Theory Related Fields* **130** (2), 151–166 (2004). MR2093760.

L. Mattner and I. Shevtsova. An optimal Berry-Esseen type inequality for expectations of smooth functions. *Dokl. Math.* **95** (3), 250–253 (2017). DOI: 10.1134/S1064562417030188. Original Russian text: *Dokl. Akad. Nauk* **474**, 535–539.

A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risk.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester (2002). ISBN 0-471-49446-1. MR1889865.

S. V. Nagaev and V. I. Rotar'. On strenghtening Lyapunov type estimates (the case when the distribution of the summands is close to the normal distribution). *Theory Probab. Appl.* **18**, 107–119 (1973). DOI: 10.1137/1118008. Original Russian text: *Teor. Veroyatn. Primen.* **18**, 109–121.

R. Neininger and H. Sulzbach. On a functional contraction method. *Ann. Probab.* **43** (4), 1777–1822 (2015). MR3353815.

L. Paditz. A nonclassical error-estimate in the central limit theorem. *Math. Nachr.* **136**, 59–68 (1988). MR952462.

V. Paulauskas. A certain strengthening of Ljapunov's theorem. *Litovsk. Mat. Sb.* **9**, 323–328 (1969). MR0266280.

V. V. Petrov. *Limit Theorems of Probability Theory. Sequences of Independent Random Variables*, volume 4 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York (1995). ISBN 0-19-853499-X. MR1353441.

I. Pinelis. On the extreme points of moments sets. *Math. Methods Oper. Res.* **83** (3), 325–349 (2016). MR3513195.

A. Pinkus and D. Wulbert. Extending n-convex functions. *Studia Math.* **171** (2), 125–152 (2005). MR2183481.

T. Popoviciu. *Sur quelques propriétés des fonctions d'une ou de deux variables réelles*. Ph.D. Thesis, Faculté des Sciences de Paris., Paris (1933). See also Mathematica (Cluj) **8**, 1–85 (1934). MR3533006.

H. Prawitz. 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 (1975). MR0397839.

S. T. Rachev. *Probability metrics and the stability of stochastic models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons (1991). ISBN 0-471-92877-1. MR1105086.

H. Richter. Parameterfreie Abschätzung und Realisierung von Erwartungswerten. *Bl. Deutsch. Ges. Versicherungsmath.* **3**, 147–162 (1957). MR0088130.

V. I. Rotar'. On summation of independent variables in a non-classical situation. *Russian Mathematical Surveys* **37**, 151–175 (1982). DOI: 10.1070/rm1982v037n06abeh004026.

W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., third edition (1987). ISBN 0-07-054234-1. MR924157.

V. V. Sazonov. On a bound for the rate of convergence in the multidimensional central limit theorem. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume II, pages 563–581. Univ. California Press (1972). MR0400351.

V. V. Sazonov. *Normal approximation—some recent advances*, volume 879 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin-New York (1981). ISBN 3-540-10863-7. MR643968.

J. Schulz. *The Optimal Berry–Esseen Constant in the Binomial Cases*. Ph.D. Thesis, Trier University, Trier, Germany (2016). http://ubt.opus.hbz-nrw.de/volltexte/2016/1007.

L. Schwartz. *Analyse I. Théorie des ensembles et topologie*, volume 42 of *Collection Enseignement des Sciences*. Hermann, Paris (1991). ISBN 2-7056-6161-8. MR1143061.

V. V. Senatov. *Normal Approximation: New Results, Methods and Problems*. Modern Probability and Statistics. VSP, Utrecht (1998). ISBN 90-6764-292-4.

MR1686374.

V. V. Senatov. Nonimprovable estimates for asymptotic expansions in the central limit theorem. *Theory Probab. Appl.* **57** (4), 633–648 (2013). MR3201669. Original Russian text: *Teor. Veroyatn. Primen.* **57**, 649–666 (2012).

V. V. Senatov. On Nonuniform Estimates of Approximation Exactness in the Central Limit Theorem. *Theory Probab. Appl.* **59** (2), 279–310 (2015). MR3416051. Original Russian text: *Teor. Veroyatn. Primen.* **59**, 276–312 (2014).

M. Shaked and J. G. Shanthikumar. *Stochastic Orders*. Springer Series in Statistics. Springer, New York (2007). ISBN 978-0-387-32915-4; 0-387-32915-3. MR2265633.

I. Shevtsova. On the asymptotically exact constants in the Berry–Esseen–Katz inequality. *Theory Probab. Appl.* **55** (2), 225–252 (2011). MR2768905. Original Russian text: *Teor. Veroyatn. Primen.* **55**, 271–304 (2010).

I. Shevtsova. Moment-type estimates with asymptotically optimal structure for the accuracy of the normal approximation. *Ann. Math. Inform.* **39**, 241–307 (2012a). MR2959891.

I. Shevtsova. On the accuracy of the normal approximation for sums of independent random variables. *Dokl. Math.* **85** (2), 274–278 (2012b). MR2984662.

I. Shevtsova. On the accuracy of the normal approximation for sums of independent symmetric random variables. *Dokl. Math.* **85** (2), 292–296 (2012c). DOI: 10.1134/S1064562412020354.

I. Shevtsova. Moment–type estimates with an improved structure for the accuracy of the normal approximation to distributions of sums of independent symmetric random variables. *Theory Probab. Appl.* **57** (3), 468–496 (2013). MR3196782. Original Russian text: *Teor. Veroyatn. Primen.* **57**, 499–532 (2012).

I. Shevtsova. 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** (1), 185–210 (2014). MR3198873.

I. Shevtsova. *Accuracy of the Normal Approximation: Methods of Estimation and New Results* (in Russian). Argamak–Media, Moscow (2016).

I. Shevtsova. A moment inequality with application to convergence rate estimates in the global CLT for Poisson–binomial random sums. *Theory Probab. Appl.* **62** (2), 278–294 (2018). DOI: 10.1137/S0040585X97T988605. Original Russian text: *Teor. Veroyatn. Primen.* **62**, 345–364 (2017).

I. S. Shiganov. A note on numerical rate of convergence estimates in central limit theorem. *J. Soviet Math.* **47** (5), 2810–2816 (1989). MR1040115. Original Russian text in: *Problemy Ustoichivosti Stokhasticheskikh Modelei, Trudy Seminara*, pp. 142–149 (1987).

N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences (2010). https://oeis.org.

I. S. Tyurin. New estimates of the convergence rate in the Lyapunov theorem. *ArXiv Mathematics e-prints* (2009a). arXiv: 0912.0726.

I. S. Tyurin. On the accuracy of the Gaussian approximation. *Dokl. Math.* **80** (3), 840–843 (2009b). DOI: 10.1134/S1064562409060155.

I. S. Tyurin. On the convergence rate in Lyapunov's theorem. *Theory Probab. Appl.* **55** (2), 253–270 (2011). DOI: 10.1137/S0040585X97984760. Original Russian text: *Teor. Veroyatn. Primen.* **55**, 250–270 (2010).

I. S. Tyurin. Some optimal bounds in CLT using zero biasing. *Stat. Prob. Letters* **82** (3), 514–518 (2012). DOI: 10.1016/j.spl.2011.11.010.

V. V. Ul'yanov. A non-uniform estimate for the speed of convergence in the central limit theorem in *R*. *Theory Probab. Appl.* **21**, 270–282 (1976). DOI: 10.1214/10.1137/1121033. Original Russian text: *Teor. Veroyatn. Primen.* **21**, 280–292.

V. V. Ul'yanov. On more precise convergence rate estimates in the central limit theorem. *Theory Probab. Appl.* **23** (3), 660–663 (1979). DOI: 10.1137/1123080. Correction note **24**, 237. DOI: 10.1137/1124029. Original Russian text: *Teor. Veroyatn. Primen.* **23**, 684–688 (1978); correction note **24**, 236.

G. Winkler. Extreme points of moment sets. *Math. Oper. Res.* **13** (4), 581–587 (1988). MR971911.

K. J. Wirths. Telescoping estimates for smooth series. *Canad. Math. Bull.* **58** (1), 188–195 (2015). MR3303223.

L. S. Yaroslavtseva. Nonclassical error bounds for asymptotic expansions in the central limit theorems. *Theory Probab. Appl.* **53** (2), 365–367 (2009). DOI:10.1137/S0040585X97983699. Original Russian text: *Teor. Veroyatn. Primen.* **53**, 390–393 (2008).

S. Zahl. *Bounds for the Central Limit Theorem Error.* Ph.D. Thesis, Harward University, Cambridge, MA (1963). MR2939434.

V. M. Zolotarev. On the closeness of the distributions of two sums of independent random variables. *Theory Probab. Appl.* **10** (3), 472–479 (1965). DOI: 10.1137/1110055. Original Russian text: *Teor. Veroyatn. Primen.* **10**, 519–526 (1965).

V. M. Zolotarev. Exactness of an approximation in the central limit theorem. In *Proceedings of the Second Japan-USSR Symposium on Probability Theory (Kyoto, 1972)*, pages 531–543. Lecture Notes in Math., Vol. 330. Springer, Berlin (1973). MR0443048.

V. M. Zolotarev. *Modern Theory of Summation of Random Variables.* Modern Probability and Statistics. VSP, Utrecht (1997). ISBN 90-6764-270-3. MR1640024.