

Fisher information and the central limit theorem

Sergey G. Bobkov · Gennadiy P. Chistyakov ·
Friedrich Götze

Received: 21 July 2012 / Revised: 2 April 2013 / Published online: 23 April 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract An Edgeworth-type expansion is established for the relative Fisher information distance to the class of normal distributions of sums of i.i.d. random variables, satisfying moment conditions. The validity of the central limit theorem is studied via properties of the Fisher information along convolutions.

Keywords Entropy · Entropic distance · Central limit theorem · Edgeworth-type expansions

Mathematics Subject Classification (1991) Primary 60E

Contents

1 Introduction	2
2 General properties of densities with finite Fisher information	6
3 Fisher information as a functional	11
4 Convolution of three densities of bounded variation	16
5 Bounds in terms of characteristic functions	24

Research partially supported by NSF grant DMS-1106530 and SFB 701.

S. G. Bobkov
School of Mathematics, University of Minnesota, 127 Vincent Hall, 206 Church St. S.E.,
Minneapolis, MN 55455, USA
e-mail: bobkov@math.umn.edu

G. P. Chistyakov · F. Götze (✉)
Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany
e-mail: goetze@math.uni-bielefeld.de

G. P. Chistyakov
e-mail: chistyak@math.uni-bielefeld.de

6	Classes of densities representable as convolutions	27
7	Bounds under moment assumptions	30
8	Fisher information in terms of the second derivative	33
9	Normalized sums. Proof of Theorem 1.3	38
10	Edgeworth-type expansions	40
11	Behaviour of densities not far from the origin	44
12	Moderate deviations	49
13	Theorem 1.1 in the case $s = 2$ and Corollary 1.2	51
14	Extensions to non-integer s . Lower bounds	54

1 Introduction

Given a random variable X with an absolutely continuous density p , the Fisher information of X (or its distribution) is defined by

$$I(X) = I(p) = \int_{-\infty}^{\infty} \frac{p'(x)^2}{p(x)} dx,$$

where p' denotes a Radon–Nikodym derivative of p . In all other cases, let $I(X) = \infty$.

With the first two moments of X being fixed, this quantity is minimized for the normal distribution (which is a variant of Cramér–Rao’s inequality). That is, if $\mathbf{E}X = a$ $\text{Var}(X) = \sigma^2$, then we have $I(X) \geq I(Z)$ for $Z \sim N(a, \sigma^2)$ with density

$$\varphi_{a,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-(x-a)^2/2\sigma^2}.$$

Moreover, the equality $I(X) = I(Z)$ holds if and only if X is normal.

In many applications the relative Fisher information

$$I(X||Z) = I(X) - I(Z) = \int_{-\infty}^{\infty} \left(\frac{p'(x)}{p(x)} - \frac{\varphi'_{a,\sigma}(x)}{\varphi_{a,\sigma}(x)} \right)^2 p(x) dx$$

is used as a strong measure of non-Gaussianity of X . For example, it dominates the relative entropy, or Kullback–Leibler distance of the distribution of X to the standard normal distribution; more precisely (cf. Stam [S]),

$$\frac{\sigma^2}{2} I(X||Z) \geq D(X||Z) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{\varphi_{a,\sigma}(x)} dx. \quad (1.1)$$

We consider the scheme of independent identically distributed random variables $(X_n)_{n \geq 1}$. Assuming that $\mathbf{E}X_1 = 0$ $\text{Var}(X_1) = 1$, define the normalized sums

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

Since Z_n are weakly convergent in distribution to $Z \sim N(0, 1)$, one may wonder whether the convergence holds in a stronger sense. A remarkable observation in this respect is due to Barron and Johnson proving in [3] that

$$I(Z_n) \rightarrow I(Z), \quad \text{as } n \rightarrow \infty, \tag{1.2}$$

i.e., $I(Z_n|Z) \rightarrow 0$, if and only if $I(Z_{n_0})$ is finite for some n_0 . In particular, it suffices to require that $I(X_1) < \infty$, although choosing larger values of n_0 considerably enhances the range of applicability of this theorem.

Quantitative estimates on the relative Fisher information in the central limit theorem are partly developed, as well. In the i.i.d. case Barron and Johnson [3], and Artstein et al. [1] derived an asymptotic bound $I(Z_n|Z) = O(1/n)$ under the hypothesis that the distribution of X_1 admits an analytic inequality of Poincaré-type

$$c \operatorname{Var}(u(X_1)) \leq \mathbf{E} u'(X_1)^2.$$

Here, u is an arbitrary bounded smooth function on the real line, and c is a constant depending on the distribution of X_1 , only (the spectral gap). More precisely, they established the bound

$$I(Z_n|Z) \leq \frac{2}{2 + c(n - 1)} I(X_1|Z),$$

leading to the $1/n$ convergence in case $c > 0$ and $I(X_1) < \infty$. The work [1], which brought important ideas from [4], also provides a similar bound for weighted sums of X_k in terms of the Lyapunov coefficient of order 4. Note that Poincaré inequalities involve a large variety of “nice” probability distributions on the line all having finite exponential moments.

One of the aims of this paper is to study the exact asymptotics (or rates) of $I(Z_n|Z)$ under standard moment conditions. We prove:

Theorem 1.1 *Let $\mathbf{E} |X_1|^s < \infty$ for an integer $s \geq 2$, and assume that $I(Z_{n_0}) < \infty$, for some n_0 . Then for certain coefficients c_j we have, as $n \rightarrow \infty$,*

$$I(Z_n|Z) = \frac{c_1}{n} + \frac{c_2}{n^2} + \dots + \frac{c_{\lfloor (s-2)/2 \rfloor}}{n^{\lfloor (s-2)/2 \rfloor}} + o\left(n^{-\frac{s-2}{2}} (\log n)^{-\frac{(s-3)_+}{2}}\right). \tag{1.3}$$

As it turns out, a similar expansion holds as well for the entropic distance $D(Z_n|Z)$, cf. [11], showing a number of interesting analogies in the asymptotic behavior of these two distances. In particular, in both cases each coefficient c_j is given by a certain polynomial in the cumulants $\gamma_3, \dots, \gamma_{2j+1}$. In order to describe these polynomials, we first note that, by the moment assumption, the cumulants

$$\gamma_r = i^{-r} \frac{d^r}{dt^r} \log \mathbf{E} e^{itX_1} \Big|_{t=0}$$

are well-defined for all positive integers $r \leq s$, and one may introduce the well-known functions

$$q_k(x) = \varphi(x) \sum H_{k+2j}(x) \frac{1}{r_1! \dots r_k!} \left(\frac{\gamma_3}{3!}\right)^{r_1} \dots \left(\frac{\gamma_{k+2}}{(k+2)!}\right)^{r_k}$$

involving the Chebyshev-Hermite polynomials H_l . Here $\varphi = \varphi_{0,1}$ denotes the density of the standard normal law, and the summation runs over all non-negative integer solutions (r_1, \dots, r_k) to the equation $r_1 + 2r_2 + \dots + kr_k = k$ with $j = r_1 + \dots + r_k$.

The functions q_k are correctly defined for $k = 1, \dots, s - 2$. They appear in Edgeworth-type expansions approximating the density of Z_n . We shall employ them to derive an expansion in powers of $1/n$ for the distance $I(Z_n||Z)$, which leads us to the following description of the coefficients in (1.3),

$$c_j = \sum_{k=2}^{2j} (-1)^k \sum_{-\infty}^{+\infty} \int (q'_{r_1} + xq_{r_1})(q'_{r_2} + xq_{r_2}) q_{r_3} \dots q_{r_k} \frac{dx}{\varphi^{k-1}}. \quad (1.4)$$

Here, the inner summation is carried out over all positive integer tuples (r_1, \dots, r_k) such that $r_1 + \dots + r_k = 2j$.

For example, $c_1 = \frac{1}{2} \gamma_3^2$, and in the case $s = 4$ the relation (1.3) becomes

$$I(Z_n||Z) = \frac{1}{2n} \left(\mathbf{E}X_1^3\right)^2 + o\left(\frac{1}{n(\log n)^{1/2}}\right). \quad (1.5)$$

Hence, under the 4-th moment condition, we have $I(Z_n||Z) \leq \frac{C}{n}$ with some constant C (which can actually be chosen to depend on $\mathbf{E}X_1^4$ and $I(X_1)$, only).

For $s = 6$, the result involves the coefficient c_2 which depends on γ_3 , γ_4 , and γ_5 . If $\gamma_3 = 0$ (i.e. $\mathbf{E}X_1^3 = 0$), we have $c_1 = 0$ $c_2 = \frac{1}{6} \gamma_4^2$, and then

$$I(Z_n||Z) = \frac{1}{6n^2} \left(\mathbf{E}X_1^4 - 3\right)^2 + o\left(\frac{1}{n^2(\log n)^{3/2}}\right).$$

More generally, (1.3) is simplified, when the first $k - 1$ moments of X_1 coincide with the corresponding moments of $Z \sim N(0, 1)$.

Corollary 1.2 *Let $\mathbf{E}|X_1|^s < \infty$ ($s \geq 4$), and assume $I(Z_{n_0}) < \infty$, for some n_0 . Given $k = 3, 4, \dots, s$, assume that $\gamma_j = 0$ for all $3 \leq j < k$. Then*

$$I(Z_n||Z) = \frac{\gamma_k^2}{(k-1)!} \cdot \frac{1}{n^{k-2}} + O\left(\frac{1}{n^{k-1}}\right) + o\left(\frac{1}{n^{(s-2)/2}(\log n)^{(s-3)/2}}\right). \quad (1.6)$$

This relation is consistent with an observation of Johnson who noticed that, if $\gamma_k \neq 0$, $I(Z_n||Z)$ cannot be asymptotically better than $n^{-(k-2)}$ ([15], Lemma 2.12).

Let us note that if $k < \frac{s}{2}$, the O -term in (1.6) dominates the o -term. But when $k \geq \frac{s}{2}$, it can be removed, and if $k > \frac{s}{2} + 1$, (1.6) just says that

$$I(Z_n||Z) = o\left(n^{-(s-2)/2} (\log n)^{-(s-3)/2}\right). \tag{1.7}$$

As for the remaining values $s = 2, 3$, there are no coefficients c_j in the sum (1.3). In case $s = 2$ Theorem 1.1 reduces to Barron–Johnson’s theorem (1.2), while under a 3-rd moment assumption we only have

$$I(Z_n||Z) = o\left(\frac{1}{\sqrt{n}}\right).$$

In fact, a similar observation holds for the whole range of reals $2 < s < 4$. Here the expansion (1.3) should be replaced by the bound (1.7). Although this bound is worse than (1.5), it cannot be essentially improved. As shown in [11], it may happen that $\mathbf{E}|X_1|^s < \infty$ with $D(X_1) < \infty$ (and actually with $I(X_1) < \infty$), while

$$D(Z_n||Z) \geq \frac{c}{n^{(s-2)/2} (\log n)^\eta}, \quad n \geq n_1(X_1),$$

where the constant $c > 0$ depends on s and an arbitrary prescribed value $\eta > s/2$. In view of (1.1), a similar lower bound therefore holds for $I(Z_n||Z)$, as well.

Another interesting issue connected with the convergence theorem (1.2) and the expansion (1.3) is the characterization of distributions for which these results hold. Indeed, the condition $I(X_1) < \infty$ corresponding to $n_0 = 1$ in Theorem 1.1 seems to be way too strong. To this aim, we establish an explicit criterion such that $I(Z_{n_0}) < \infty$ holds for sufficiently large n_0 in terms of the characteristic function $f_1(t) = \mathbf{E} e^{itX_1}$ of X_1 .

Theorem 1.3 *Given independent identically distributed random variables $(X_n)_{n \geq 1}$ with finite second moment, the following assertions are equivalent:*

- (a) *For some n_0 , Z_{n_0} has finite Fisher information;*
- (b) *For some n_1 , Z_{n_1} has density of bounded total variation;*
- (c) *For some n_2 , Z_{n_2} has a continuously differentiable density p_{n_2} such that*

$$\int_{-\infty}^{\infty} |p'_{n_2}(x)| dx < \infty;$$

- (d) *For some $\varepsilon > 0$, $|f_1(t)| = O(t^{-\varepsilon})$, as $t \rightarrow \infty$;*
- (e) *For some $\nu > 0$,*

$$\int_{-\infty}^{\infty} |f_1(t)|^\nu |t| dt < \infty. \tag{1.8}$$

Property (c) is a formally strengthened variant of (b), although in general they are not equivalent with $n_1 = n_2$. (For example, the uniform distribution has density of bounded total variation, but its density is not everywhere differentiable.)

Properties (a)–(c) are equivalent to each other without any moment assumption, while (d)–(e) are always necessary for the finiteness of $I(Z_n)$ with large n . These two last conditions show that the range of applicability of Theorem 1.1 is indeed rather wide, since almost all reasonable absolutely continuous distributions satisfy (1.8). The latter should be compared to and viewed as a certain strengthening of the following condition (sometimes called a smoothness condition)

$$\int_{-\infty}^{\infty} |f_1(t)|^\nu dt < \infty, \quad \text{for some } \nu > 0.$$

It is equivalent to the property that, for some n , Z_n has a bounded continuous density p_n (cf. e.g. [5]). In this and only in this case, a uniform local limit theorem holds:

$$\Delta_n = \sup_x |p_n(x) - \varphi(x)| \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

That this assertion is weaker compared to the convergence in Fisher information distance such as (1.2) can be seen by Shimizu's inequality $\Delta_n^2 \leq cI(Z_n|Z)$, which holds with some absolute constant c ([3, 21], Lemma 1.5). Note in this connection that Shimizu's inequality may be strengthened in terms of the total variation distance as $\|p_n - \varphi\|_{\text{TV}}^2 \leq cI(Z_n|Z)$. Using Theorem 1.3, this shows in the i.i.d. case that (1.2) is equivalent to the convergence $\|p_n - \varphi\|_{\text{TV}} \rightarrow 0$.

The paper is organized in the following way. We start with the description of general properties of densities having finite Fisher information (Sect. 2) and properties of Fisher information as a functional on spaces of densities (showing lower semi-continuity and convexity, Sect. 3). Some of the properties and relations which we state for completeness may be known already. We apologize for being unable to find references for them.

In Sects. 4 and 5 we turn to upper bounds needed mainly in the proof of Theorem 1.3. Further properties of densities emerging after several convolutions, as well as, bounds under additional moment assumptions are discussed in Sects. 6–8. In Sect. 9 we complete the proof of Theorem 1.3, and in the next section we state basic lemmas on Edgeworth-type expansions which are needed in the proof of Theorem 1.1. Sections 11 and 12 are devoted to the proof itself. Some remarks leading to the particular case $s = 2$ in Theorem 1.1 (Barron–Johnson theorem) are given in Sect. 13. Finally, in the last section we briefly describe the modifications needed to obtain Theorem 1.1 under moment assumptions with arbitrary real values of s .

2 General properties of densities with finite Fisher information

Definition If a random variable X has an absolutely continuous density p with Radon–Nikodym derivative p' , put

$$I(X) = I(p) = \int_{\{p(x)>0\}} \frac{p'(x)^2}{p(x)} dx. \quad (2.1)$$

In this case, if $\tilde{p}(x) = p(x)$ for almost all x , i.e., if \tilde{p} is another representative of the density, put $I(\tilde{p}) = I(p)$. In all other cases, put $I(X) = \infty$. The quantity $I(X)$ is called the Fisher information of X .

With this definition, I is correctly defined as a functional on the space of all densities (and on the space of all probability distributions). However, when $I(X) < \infty$ and p is the density of X , we will always assume that p is chosen to be absolutely continuous. In particular, in this case the derivative $p'(x)$ exists and is finite on a set of full Lebesgue measure.

One may write an equivalent definition by involving the score function $\rho(x) = \frac{p'(x)}{p(x)}$. In general $\mathbf{P}\{p(X) > 0\} = 1$, so the random variable $\rho(X)$ is defined with probability 1, and thus

$$I(X) = \mathbf{E} \rho(X)^2. \tag{2.2}$$

For different purposes, it is useful to realize how the ratio $\frac{p'(x)^2}{p(x)}$ may behave when $p(x)$ is small and is even vanishing. The behavior cannot be arbitrary, when the Fisher information is finite. The following statement will allow us to make more rigorous the derivation of various Fisher information bounds on the density and its derivatives.

Proposition 2.1 *Assume X has density p with finite Fisher information. If p is differentiable at the point x_0 such that $p(x_0) = 0$, then $p'(x_0) = 0$.*

Proof If p is differentiable in some neighborhood of x_0 , and its derivative is continuous at this point, the statement is obvious. In the general case, for simplicity of notations let $x_0 = 0$, and assume $c = p'(0) > 0$. Since $p(\varepsilon) = c\varepsilon + o(\varepsilon)$, as $\varepsilon \rightarrow 0$, one may choose $\varepsilon_0 > 0$ such that

$$\frac{3c}{4} |x| \leq p(x) \leq \frac{5c}{4} |x|, \quad \text{for all } |x| \leq \varepsilon_0.$$

In particular, p is positive on $(0, \varepsilon_0]$. Hence, according to (2.1),

$$I(X) \geq \int_0^{\varepsilon_0} \frac{p'(x)^2}{p(x)} dx \geq \frac{4}{5c} \int_0^{\varepsilon_0} \frac{p'(x)^2}{x} dx.$$

We split the last integral into the intervals $\Delta_n = (2^{-(n+1)}\varepsilon_0, 2^{-n}\varepsilon_0)$, which leads to

$$\frac{5c}{4} I(X) \geq \sum_{n=0}^{\infty} \frac{2^n}{\varepsilon_0} \int_{\Delta_n} p'(x)^2 dx.$$

Now, applying Cauchy's inequality and using $p(x) - p(\frac{x}{2}) \geq \frac{c}{8}x$ for $0 \leq x \leq \varepsilon_0$, we obtain

$$\begin{aligned} \int_{\Delta_n} p'(x)^2 dx &\geq \frac{2^{n+1}}{\varepsilon_0} \left(\int_{\Delta_n} p'(x) dx \right)^2 \\ &= \frac{2^{n+1}}{\varepsilon_0} \left(p(2^{-n}\varepsilon_0) - p(2^{-(n+1)}\varepsilon_0) \right)^2 \geq 2^{-(n-1)} \frac{c^2\varepsilon_0}{64}. \end{aligned}$$

As a result,

$$\frac{5c}{4} I(X) \geq \sum_{n=0}^{\infty} 2^n \cdot 2^{-(n-1)} \cdot \frac{c^2}{64} = \infty,$$

a contradiction with finiteness of the Fisher information. \square

As an example illustrating a possible behavior as in Proposition 2.1, one may consider the beta distribution with parameters $\alpha = \beta = 3$, which has the density

$$p(x) = 30(x(1-x))^2, \quad 0 \leq x \leq 1.$$

Then X has finite Fisher information, although $p(x_0) = p'(x_0) = 0$ at $x_0 = 0$ and $x_0 = 1$.

More generally, if a density p is supported and twice differentiable on a finite interval $[a, b]$, and if p has finitely many zeros $x_0 \in [a, b]$, and $p'(x_0) = 0$ $p''(x_0) > 0$ at any such point, then X has finite Fisher information.

Now, let us return to the definitions (2.1)–(2.2). By Cauchy's inequality,

$$I(X)^{1/2} = \left(\mathbf{E} \rho(X)^2 \right)^{1/2} \geq \mathbf{E} |\rho(X)| = \int_{\{p(x)>0\}} |p'(x)| dx.$$

Here, by Proposition 2.1, the last integral may be extended to the whole real line without any change, and then it represents the total variation of the function p in the usual sense of the Theory of Functions:

$$\|p\|_{\text{TV}} = \sup \sum_{k=1}^n |p(x_k) - p(x_{k-1})|,$$

where the supremum runs over all finite collections $x_0 < x_1 < \dots < x_n$.

In the sequel, we consider this norm also for densities which are not necessarily continuous, and then it is natural to require that, for each x , the value $p(x)$ lies in the closed segment $\Delta(x)$ with endpoints $p(x-)$ and $p(x+)$. Note that if we change $p(x)$ at a point of discontinuity such that $p(x)$ goes out of $\Delta(x)$, then the probability measure $\mu(dx) = p(x)dx$ with density p is unchanged, while $\|p\|_{\text{TV}}$ will increase.

Let us note that the same notation $\|\nu\|_{\text{TV}}$ in the sense of the Measure Theory is commonly used to denote the total variation of a signed Borel measure ν on the real line. The connection with the Theory of Functions is simply $\|\nu\|_{\text{TV}} = \|F\|_{\text{TV}}$ in terms of the cumulative “distribution function” $F(x) = \nu((-\infty, x])$.

Returning to the Fisher information, we thus observed that, if $I(X)$ is finite, the density p of X is a function of bounded variation. Hence, the limits

$$p(-\infty) = \lim_{x \rightarrow -\infty} p(x), \quad p(\infty) = \lim_{x \rightarrow \infty} p(x)$$

exist and are finite. But, since p is a density (hence integrable), these limits must be zero. In addition, for any x ,

$$p(x) = \int_{-\infty}^x p'(y) dy \leq \int_{-\infty}^x |p'(y)| dy \leq \sqrt{I(X)}.$$

We can summarize these observations in the following:

Proposition 2.2 *If X has density p with finite Fisher information $I(X)$, then $p(-\infty) = p(\infty) = 0$, and the density has finite total variation satisfying*

$$\|p\|_{\text{TV}} = \int_{-\infty}^{\infty} |p'(x)| dx \leq \sqrt{I(X)}.$$

In particular, p is bounded: $\max_x p(x) \leq \sqrt{I(X)}$.

Let $f(t) = \mathbf{E} e^{itX}$ denote the characteristic function of a random variable X with density p . Since in general $|f(t)| \leq \frac{\|p\|_{\text{TV}}}{|t|}$, an immediate consequence of Proposition 2.2 is a similar bound

$$|f(t)| \leq \frac{1}{|t|} \sqrt{I(X)},$$

involving the Fisher information. Here, as noticed by Zhang [24], the behavior near the origin can be improved by using the Cramér–Rao inequality which yields:

Proposition 2.3 *If X has finite Fisher information, then its characteristic function $f(t)$ admits the bound*

$$|f(t)|^2 \leq \frac{I(X)}{I(X) + t^2}, \quad t \in \mathbf{R}. \tag{2.3}$$

Indeed, for any smooth function $u : \mathbf{R} \rightarrow \mathbf{C}$ such that $\mathbf{E} |u'(X)| < \infty$, one has, by integration by parts and applying Cauchy’s inequality,

$$|\mathbf{E} u'(X)|^2 \leq I(X) \mathbf{E} |u(X)|^2.$$

In case $u(x) = e^{itx} - f(t)$, this gives (2.3); cf. also [24] for a slightly different argument.

Another immediate consequence of Proposition 2.2 is that both p and p' are square integrable, that is, p belongs to the Sobolev space $W_1^2 = W_1^2(-\infty, \infty)$ of all real-valued absolutely continuous functions on the real line with finite (Hilbert) norm

$$\|u\|_{W_1^2}^2 = \int_{-\infty}^{\infty} u(x)^2 dx + \int_{-\infty}^{\infty} u'(x)^2 dx.$$

More precisely,

$$\int_{-\infty}^{\infty} p'(x)^2 dx = \int_{-\infty}^{\infty} \frac{p'(x)^2}{p(x)} p(x) dx \leq \max_x p(x) \int_{-\infty}^{\infty} \frac{p'(x)^2}{p(x)} dx \leq I(X)^{3/2}. \tag{2.4}$$

By the inverse Fourier formula, the resulting inequality in (2.4) is equivalent to the following integral analogue of the pointwise bound (2.3).

Proposition 2.4 *The characteristic function $f(t)$ of any random variable X satisfies*

$$\int_{-\infty}^{\infty} |tf(t)|^2 dt \leq 2\pi I(X)^{3/2}. \tag{2.5}$$

In particular, when the Fisher information of X is finite, so is the integral in (2.5)

Let us return to Proposition 2.2. Since the estimate on the total variation norm $\|p\|_{TV}$ can be given in terms of the Fisher information, it is natural to ask whether or not it is possible to bound the total variation distance from p to a normal density in terms of the relative Fisher information. This suggests the following bound.

Proposition 2.5 *If X has mean zero, variance one, and a density p with finite Fisher information, then*

$$\|p - \varphi\|_{TV} \leq 4\sqrt{I(X||Z)}, \tag{2.6}$$

where Z has the standard normal density φ .

Proof Using

$$p'(x) - \varphi'(x) = \left(\frac{p'(x)}{p(x)} - \frac{\varphi'(x)}{\varphi(x)} \right) p(x) - x(p(x) - \varphi(x)) \quad (p(x) > 0)$$

and applying Cauchy's inequality, we may write

$$\begin{aligned} \|p - \varphi\|_{TV} &= \int_{-\infty}^{\infty} |p'(x) - \varphi'(x)| dx \\ &\leq I(X||Z)^{1/2} + \int_{-\infty}^{\infty} |x| |p(x) - \varphi(x)| dx. \end{aligned} \tag{2.7}$$

The last integral represents a weighted total variation distance between the distributions of X and Z with weight function $w(x) = |x|$.

On this step we apply the following extention of Csiszár-Kullback-Pinsker’s inequality (CKP) to the scheme of weighted total variation distances, which is proposed by Bolley and Villani, cf. [12], Theorem 2.1 (ii). If X and Y are random variables with densities p and q , and $w(x) \geq 0$ is a measurable function, then

$$\left(\int_{-\infty}^{\infty} w(x) |p(x) - q(x)| dx \right)^2 \leq CD(X||Y) = C \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx,$$

where

$$C = 2 \left(1 + \log \int_{-\infty}^{\infty} e^{w(x)^2} q(x) dx \right).$$

The inequality also holds in the setting of abstract measurable spaces, and when $w = 1$ it yields the classical CKP inequality with an additional factor 2.

In our case, $Y = Z$, $q = \varphi$, and taking $w(x) = \sqrt{t/2} |x|$ ($0 < t < 1$), we get

$$\frac{t}{2} \left(\int_{-\infty}^{\infty} |x| |p(x) - \varphi(x)| dx \right)^2 \leq \left(2 + \log \frac{1}{1-t} \right) D(X||Z).$$

One may choose, for example, $t = 1 - \frac{1}{e}$, and recalling (1.1), we arrive at

$$\int_{-\infty}^{\infty} |x| |p(x) - \varphi(x)| dx \leq 3.1 D(X||Z)^{1/2} \leq \frac{3.1}{\sqrt{2}} I(X||Z)^{1/2}.$$

It remains to use this bound in (2.7), and (2.6) follows. □

3 Fisher information as a functional

It is worthwhile to discuss separately a few general properties of the Fisher information viewed as a functional on the space of densities. We start with topological properties.

Proposition 3.1 *Let $(X_n)_{n \geq 1}$ be a sequence of random variables, and X be a random variable such that $X_n \Rightarrow X$ weakly in distribution. Then*

$$I(X) \leq \liminf_{n \rightarrow \infty} I(X_n). \tag{3.1}$$

Denote by \mathfrak{P}_1 the collection of all (probability) densities on the real line with finite Fisher information, and let $\mathfrak{P}_1(I)$ denote the subset of all densities which have Fisher information of at most size $I > 0$. On the set \mathfrak{P}_1 the relation (3.1) may be written as

$$I(p) \leq \liminf_{n \rightarrow \infty} I(p_n), \quad (3.2)$$

which holds under the condition that the corresponding distributions are convergent weakly, i.e.,

$$\lim_{n \rightarrow \infty} \int_{-\infty}^a p_n(x) dx = \int_{-\infty}^a p(x) dx, \quad \text{for all } a \in \mathbf{R}. \quad (3.3)$$

Hence, every $\mathfrak{P}_1(I)$ is closed in the weak topology. In fact, inside such sets (3.3) can be strengthened to the convergence in the L^1 -metric,

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} |p_n(x) - p(x)| dx = 0. \quad (3.4)$$

Proposition 3.2 *On every set $\mathfrak{P}_1(I)$ the weak topology with convergence (3.3) and the topology generated by the L^1 -norm coincide, and the Fisher information is a lower semi-continuous functional on this set.*

Proof For the proof of Proposition 3.1, one may assume that $I(X_n) \rightarrow I$, for some (finite) constant I . Then, for sufficiently large n , the X_n have absolutely continuous densities p_n with Fisher information at most $I + 1$. By Proposition 2.2, such densities are uniformly bounded and have uniformly bounded variations. Hence, by the second Helly theorem (cf. e.g. [16]), there are a subsequence p_{n_k} and a function p of bounded variation, such that $p_{n_k}(x) \rightarrow p(x)$, as $k \rightarrow \infty$, for all points x . Necessarily, $p(x) \geq 0$ and $\int_{-\infty}^{\infty} p(x) dx \leq 1$. Since the sequence of distributions of X_n is tight (or weakly pre-compact), it also follows that $\int_{-\infty}^{\infty} p(x) dx = 1$. Hence, X has an absolutely continuous distribution with p as its density, and the weak convergence (3.3) holds.

For the proof of Proposition 3.2, a similar argument should be applied to an arbitrary prescribed subsequence p_{n_k} , where we obtain $p(x) = \lim_{l \rightarrow \infty} p_{n_{k_l}}(x)$ for some further subsequence. By Scheffe's lemma, this property implies the convergence in L^1 -norm, that is, (3.4) holds along n_{k_l} . This implies the convergence in L^1 for the whole sequence p_n , which is the assertion of Proposition 3.2 (the first part).

To continue the proof of Proposition 3.1, for simplicity of notations, assume that the subsequence constructed in the first step is actually the whole sequence. By (2.4),

$$\int_{-\infty}^{\infty} p_n'(x)^2 dx \leq (I + 1)^{3/2},$$

which implies that the derivatives are uniformly integrable on every finite interval. By the Dunford-Pettis compactness criterion for the space L^1 (over finite measures), there is a subsequence p'_{n_k} which is convergent to some locally integrable function u in the sense that

$$\int_A p'_{n_k}(x) dx \rightarrow \int_A u(x) dx, \tag{3.5}$$

for any bounded Borel set $A \subset \mathbf{R}$. (This is the weak $\sigma(L^1, L^\infty)$ convergence on finite intervals.) Note that, according to Proposition 2.1, $p'_{n_k}(x)$ may be replaced in (3.5) with the sequence $p'_{n_k}(x) 1_{\{p_{n_k}(x)>0\}}$ which is thus convergent to $u(x)$ as well.

Taking finite intervals $A = (a, b)$ in (3.5), we get

$$\int_a^b u(x) dx = p(b) - p(a),$$

which means that p is (locally) absolutely continuous. Furthermore, since

$$\|p\|_{TV} = \int_{-\infty}^{\infty} |u(x)| dx,$$

and since p has finite total variation, we conclude that $u \in L^1(\mathbf{R})$, thus representing a Radon–Nikodym derivative: $u(x) = p'(x)$. Again, for simplicity of notations, assume the subsequence of derivatives obtained is actually the whole sequence.

Next, consider the sequence of functions

$$\xi_n(x) = \frac{p'_n(x)}{\sqrt{p_n(x)}} 1_{\{p_n(x)>0\}}.$$

They have $L^2(\mathbf{R})$ -norm bounded by $\sqrt{I+1}$ (for large n). Since the unit ball of L^2 is weakly compact, there is a subsequence ξ_{n_k} which is weakly convergent to some function $\xi \in L^2$, i.e.,

$$\int_{-\infty}^{\infty} \xi_{n_k}(x) q(x) dx \rightarrow \int_{-\infty}^{\infty} \xi(x) q(x) dx,$$

for any $q \in L^2$. As a consequence,

$$\int_{-\infty}^{\infty} \xi_{n_k}(x) \sqrt{p_{n_k}(x)} q(x) dx \rightarrow \int_{-\infty}^{\infty} \xi(x) \sqrt{p(x)} q(x) dx,$$

due to the uniform boundedness and pointwise convergence of p_n . In other words, again omitting sub-indices, the functions $p'_n 1_{\{p_n>0\}}$ are weakly convergent in L^2 to the function $\xi \sqrt{p}$. In particular, for $q = 1_A$ with an arbitrary bounded Borel set $A \subset \mathbf{R}$,

$$\int_A p'_n 1_{\{p_n>0\}} dx \rightarrow \int_A \xi(x) \sqrt{p(x)} dx.$$

As a result, we have obtained two limits for $p'_n \cdot 1_{\{p_n > 0\}}$ which must coincide, i.e., we get $\xi \sqrt{p} = u = p'$ a.e. Hence, $p = 0 \Rightarrow p' = 0$ and $\xi = \frac{p'}{\sqrt{p}}$ a.e. on the set $\{p(x) > 0\}$. Finally, the weak convergence $\xi_{n_k} \rightarrow \xi$ in L^2 , as in any Banach space, yields

$$I(p) = \|\xi \cdot 1_{\{p>0\}}\|_{L^2}^2 \leq \|\xi\|_{L^2}^2 \leq \liminf_{k \rightarrow \infty} \|\xi_{n_k}\|_{L^2}^2 = \liminf_{n \rightarrow \infty} I(p_{n_k}) = I.$$

Thus, Proposition 3.1 is proved. □

Another general property of the Fisher information is its convexity, that is, we have the inequality

$$I(p) \leq \sum_{i=1}^n \alpha_i I(p_i), \tag{3.6}$$

where $p = \sum_{i=1}^n \alpha_i p_i$ with arbitrary densities p_i and weights $\alpha_i > 0$ $\sum_{i=1}^n \alpha_i = 1$. This readily follows from the fact that the homogeneous function $R(u, v) = u^2/v$ is convex on the upper half-plane $u \in \mathbf{R} \ v > 0$. Moreover, Cohen [14] showed that the inequality (3.6) is strict.

As a consequence, the collection $\mathfrak{P}_1(I)$ of all densities on the real line with Fisher information $\leq I$ represents a convex closed set in the space $L^1 = L^1(\mathbf{R})$ (for strong or weak topologies).

We need to extend Jensen’s inequality (3.6) to arbitrary “continuous” convex mixtures of densities. In order to formulate this more precisely, recall the definition of mixtures. Denote by \mathfrak{P} the collection of all densities which represents a closed subset of L^1 with the weak $\sigma(L^1, L^\infty)$ topology. For any Borel set $A \subset \mathbf{R}$, the functionals $q \rightarrow \int_A q(x) dx$ are bounded and continuous on \mathfrak{P} . So, given a Borel probability measure π on \mathfrak{P} , one may introduce the probability measure on the real line

$$\mu(A) = \int_{\mathfrak{P}} \left[\int_A q(x) dx \right] d\pi(q). \tag{3.7}$$

It is absolutely continuous with respect to Lebesgue measure and has some density $p(x) = \frac{d\mu(x)}{dx}$ called the (convex) mixture of densities with mixing measure π . For short,

$$p(x) = \int_{\mathfrak{P}} q(x) d\pi(q).$$

Proposition 3.3 *If p is a convex mixture of densities with mixing measure π , then*

$$I(p) \leq \int_{\mathfrak{P}} I(q) d\pi(q). \tag{3.8}$$

Proof Note that the integral in (3.8) makes sense, since the functional $q \rightarrow I(q)$ is lower semi-continuous and hence Borel measurable on \mathfrak{P} (Proposition 3.1). We may assume that this integral is finite, so that π is supported on the convex (Borel measurable) set $\mathfrak{P}_1 = \cup_I \mathfrak{P}_1(I)$.

Identifying densities with corresponding probability measures (having these densities), we consider \mathfrak{P}_1 as a subset of the locally convex space E of all finite Borel measures μ on the real line endowed with the weak topology.

Step 1. Suppose that the measure π is supported on some convex compact set K contained in $\mathfrak{P}_1(I)$. Since the functional $q \rightarrow I(q)$ is finite, convex and lower semi-continuous on K , it admits the representation

$$I(q) = \sup_{l \in \mathcal{L}} l(q), \quad q \in K,$$

where \mathcal{L} denotes the family of all continuous affine functionals l on E such that $l(q) < I(q)$, for all $q \in K$ (cf. e.g. Meyer [18], Chapter XI, Theorem T7). In our particular case, any such functional acts on probability measures as $l(\mu) = \int_{-\infty}^{\infty} \psi(x) d\mu(x)$ with some bounded continuous function ψ on the real line. Hence,

$$I(q) = \sup_{\psi \in \mathcal{C}} \int_{-\infty}^{\infty} q(x)\psi(x) dx,$$

for some family \mathcal{C} of bounded continuous functions ψ on \mathbf{R} . An explicit description of \mathcal{C} would be of interest, but this question will not be pursued here. As a consequence, by the definition (3.7) for the measure μ with density p ,

$$\begin{aligned} \int_{\mathfrak{P}} I(q) d\pi(q) &\geq \sup_{\psi \in \mathcal{C}} \int_{\mathfrak{P}} \left[\int_{-\infty}^{\infty} q(x)\psi(x) dx \right] d\pi(q) \\ &= \sup_{\psi \in \mathcal{C}} \int_{-\infty}^{\infty} p(x)\psi(x) dx = I(p), \end{aligned}$$

which is the desired inequality (3.8).

Step 2. Suppose that π is supported on $\mathfrak{P}_1(I)$, for some $I > 0$. Since any finite measure on E is Radon, and since the set $\mathfrak{P}_1(I)$ is closed and convex, there is an increasing sequence of compact subsets $K_n \subset \mathfrak{P}_1(I)$ such that $\pi(\cup_n K_n) = 1$. Moreover, K_n can be chosen to be convex (since the closure of the convex hull will be compact, as well). Let π_n denote the normalized restriction of π to K_n with sufficiently large n so that $c_n = \pi(K_n) > 0$, and define its baricenter

$$p_n(x) = \int_{K_n} q(x) d\pi_n(q). \tag{3.9}$$

From (3.7) it follows that the measures with densities p_n are weakly convergent to the measure μ with density p , hence the relation (3.2) holds: $I(p) \leq \liminf_{n \rightarrow \infty} I(p_n)$. On the other hand, by the previous step,

$$I(p_n) \leq \int_{K_n} I(q) d\pi_n(q) = \frac{1}{c_n} \int_{K_n} I(q) d\pi(q) \rightarrow \int_{\mathfrak{P}_1(I)} I(q) d\pi(q), \quad (3.10)$$

and we obtain (3.8).

Step 3. In the general case, we may apply Step 2 to the normalized restrictions π_n of π to the sets $K_n = \mathfrak{P}_1(n)$. Again, for the densities p_n defined as in (3.9), we obtain (3.10), where $\mathfrak{P}_1(I)$ should be replaced with \mathfrak{P}_1 . Another application of the lower semi-continuity of the Fisher information finishes the proof. \square

4 Convolution of three densities of bounded variation

Although densities with finite Fisher information must be functions of bounded variation, the converse is not always true. Nevertheless, starting from a density of bounded variation and taking several convolutions with itself, the resulting density will have finite Fisher information. Our nearest aim is to prove:

Proposition 4.1 *If independent random variables X_1, X_2, X_3 have densities p_1, p_2, p_3 with finite total variation, then $S = X_1 + X_2 + X_3$ has finite Fisher information, and moreover,*

$$I(S) \leq \frac{1}{2} \left[\|p_1\|_{\text{TV}} \|p_2\|_{\text{TV}} + \|p_1\|_{\text{TV}} \|p_3\|_{\text{TV}} + \|p_2\|_{\text{TV}} \|p_3\|_{\text{TV}} \right]. \quad (4.1)$$

One may further extend (4.1) to sums of more than 3 independent summands, but this will not be needed for our purposes (since the Fisher information may only decrease when adding an independent summand.)

In the i.i.d. case the above estimate can be simplified. By a direct application of the inverse Fourier formula, the right-hand side of (4.1) may be related furthermore to the characteristic functions of X_j . We will return to this in the next section.

First let us look at the particular case where X_j are uniformly distributed over intervals. This important example already shows that the Fisher information $I(X_1 + X_2)$ does not need to be finite, while it is finite for 3 summands. (This somewhat curious fact was pointed out to one of the authors by K. Ball.) In fact, there is a simple quantitative bound.

Lemma 4.2 *If independent random variables X_1, X_2, X_3 are uniformly distributed on intervals of lengths a_1, a_2, a_3 , then*

$$I(X_1 + X_2 + X_3) \leq 2 \left[\frac{1}{a_1 a_2} + \frac{1}{a_1 a_3} + \frac{1}{a_2 a_3} \right]. \quad (4.2)$$

The density of the sum $S = X_1 + X_2 + X_3$ may easily be evaluated and leads to a rather routine problem of estimation of $I(S)$ as a function of the parameters a_j . Alternatively, there is an elegant approach based on the Brunn–Minkowski inequality and the fact that the density p of S behaves like the beta density near the end points of the supporting interval.

To describe the argument, first let us recall the volume relation (Brunn’s theorem)

$$|tA + (1 - t)B|^{1/2} \geq t|A|^{1/2} + (1 - t)|B|^{1/2}, \quad 0 < t < 1, \tag{4.3}$$

which holds for arbitrary non-empty Borel sets A, B lying in parallel hyperplanes of the Euclidean space \mathbf{R}^3 . Here

$$tA + (1 - t)B = \{ta + (1 - t)b : a \in A, b \in B\}$$

stands for the Minkowski sum, and $|C|$ is used to denote the two-dimensional Lebesgue measure of a set C in the hyperplane where it lies (cf. e.g. [13]). But the random vector (X_1, X_2, X_3) is uniformly distributed in the cube $Q \subset \mathbf{R}^3$ with sides a_j , so, the density of S is given by

$$p(x) = \frac{1}{a_1 a_2 a_3} |\{(x_1, x_2, x_3) \in Q : x_1 + x_2 + x_3 = x\}|.$$

Hence, by (4.3), the function $p^{1/2}$ is concave on the supporting interval.

The latter property may also be formulated in terms of the transform

$$L(t) = p(F^{-1}(t)), \quad 0 < t < 1.$$

Here, $F^{-1} : (0, 1) \rightarrow (x_0, x_1)$ denotes the inverse of the distribution function $F(x) = \mu(x_0, x)$, associated to a given probability measure μ which is supported and has a positive continuous density p on some interval (x_0, x_1) , finite or not. Namely, $p^{1/2}$ is concave on (x_0, x_1) , if and only if the function $L^{3/2}$ is concave on $(0, 1)$. Indeed, assuming without loss of generality that p has a continuous derivative, we have $L'(F(x)) = \frac{p'(x)}{p(x)}$ and thus

$$\frac{1}{3} (L^{3/2})'(F(x)) = (p^{1/2})'(x), \quad x_0 < x < x_1.$$

Therefore, the derivative $(p^{1/2})'$ does not increase on (x_0, x_1) , if and only if $(L^{3/2})'$ does not increase on $(0, 1)$. We refer to [8] for related issues about the so-called κ -concave probability measures and more general characterizations.

Note also that the Fisher information of a random variable X with density p is expressed in terms of the associated function L as

$$I(X) = \int_0^1 L'(t)^2 dt. \tag{4.4}$$

This general formula holds whenever p is absolutely continuous and positive on the supporting interval (without any concavity assumption).

Proof of Lemma 4.2 Let X_j take values in $[0, a_j]$. As was just explained, the distribution of $S = X_1 + X_2 + X_3$ has density p such that $p^{1/2}$ is concave on the interval $[0, a_1 + a_2 + a_3]$, or equivalently, $L^{3/2}$ is concave on $(0, 1)$, where L is the associated function for S .

Note that S has an absolutely continuous density p , which is thus vanishing at the end points $x = 0$ and $x = a_1 + a_2 + a_3$. Hence, $L(0+) = L(1-) = 0$. By the concavity, there is a non-increasing Radon–Nikodym derivative $(L^{3/2})' = \frac{3}{2} L^{1/2} L'$. Since also L is symmetric about the point $\frac{1}{2}$, we get, for all $0 < t < 1$,

$$L'(t)^2 L(t) \leq c, \quad \text{where } c = \lim_{t \rightarrow 0} L'(t)^2 L(t).$$

Hence, by (4.4),

$$I(X) \leq \int_0^1 \frac{c}{L(t)} dt = c(a_1 + a_2 + a_3). \tag{4.5}$$

It remains to find the constant c . Putting $a = a_1 a_2 a_3$, it should be clear that, for all $x > 0$ and $t > 0$ small enough,

$$F(x) = \mathbf{P}\{S \leq x\} = \frac{x^3}{6a}, \quad p(x) = \frac{x^2}{2a}, \quad F^{-1}(t) = (6at)^{1/3}, \quad L(t) = \frac{1}{2a} (6at)^{2/3},$$

and finally $L'(t)^2 L(t) = \frac{2}{a}$. Hence, $c = \frac{2}{a}$. Thus, in (4.5) we arrive at $I(X) \leq \frac{2}{a}(a_1 + a_2 + a_3)$ which is exactly (4.2). □

Lemma 4.2 allows us to reduce Proposition 4.1 to the case of uniform distributions. Note that if a density p is written as a convex mixture

$$p(x) = \int_{\mathfrak{P}} q(x) d\pi(q), \tag{4.6}$$

then by the convexity of the total variation norm,

$$\|p\|_{\text{TV}} \leq \int_{\mathfrak{P}} \|q\|_{\text{TV}} d\pi(q). \tag{4.7}$$

Recall that we understand (4.6) as the equality (3.7) of the corresponding measures. So, (4.7) uses our original agreement that, for each x , the value $p(x)$ lies in the closed segment with endpoints $p(x-)$ and $p(x+)$.

In order to apply Lemma 4.2 together with Jensen’s inequality for Fisher information, we need however to require that π has to be supported on uniform densities (that

is, densities of the normalized Lebesgue measures on finite intervals) and secondly to reverse (4.7). Indeed this turns out to be possible, which may be a rather interesting observation.

Lemma 4.3 *Any density p of bounded variation can be represented as a convex mixture (4.6) of uniform densities with a mixing measure π such that*

$$\|p\|_{\text{TV}} = \int_{\mathfrak{P}} \|q\|_{\text{TV}} d\pi(q). \tag{4.8}$$

For example, if p is supported and non-increasing on $(0, +\infty)$, there is a canonical representation

$$p(x) = \int_0^\infty \frac{1}{x_1} 1_{\{0 < x < x_1\}} d\pi(x_1) \quad \text{a.e.}$$

with a unique mixing probability measure π on $(0, \infty)$. In this case $\|p\|_{\text{TV}} = 2p(0+)$, and (4.8) is obvious. One may write a similar representation for densities of unimodal distributions. In general, another way to write (4.6) and (4.8) is

$$p(x) = \int_{x_1 > x_0} \frac{1}{x_1 - x_0} 1_{\{x_0 < x < x_1\}} d\pi(x_0, x_1),$$

$$\|p\|_{\text{TV}} = 2 \int_{x_1 > x_0} \frac{1}{x_1 - x_0} d\pi(x_0, x_1),$$

where π is a Borel probability measure on the half-plane $x_1 > x_0$ (i.e., above the main diagonal). It was noticed by Maurey [17] that a mixing measure π satisfying (4.6) and (4.8) is not unique in general. This can be seen on the example of $p(x) = \frac{1}{4} \cdot 1_{\{0 < x < 3\}} + \frac{1}{4} \cdot 1_{\{1 < x < 2\}}$.

Let us also note that the sets $\text{BV}(c)$ of all densities p with $\|p\|_{\text{TV}} \leq c$ are closed under the weak convergence (3.3) of the corresponding probability distributions. Moreover, the weak convergence in $\text{BV}(c)$ coincides with convergence in L^1 -norm, which can be proved using the same arguments as in the proof of Proposition 3.2. In particular, the functional $q \rightarrow \|q\|_{\text{TV}}$ is lower semi-continuous and hence Borel measurable on \mathfrak{P} , so the integrals (4.7)–(4.8) make sense.

Denote by U the collection of all uniform densities which thus may be identified with the half-plane $\tilde{U} = \{(a, b) \in \mathbf{R}^2 : b > a\}$ via the map $(a, b) \rightarrow q_{a,b}(x) = \frac{1}{b-a} 1_{\{a < x < b\}}$. The usual convergence on \tilde{U} in the Euclidean metric coincides with the weak convergence (3.3) of $q_{a,b}$. The closure of U for the weak topology contains U and all delta-measures, hence U is a Borel measurable subset of \mathfrak{P} .

Proof We split the argument into two steps.

Step 1. First consider the discrete case, where p is piecewise constant. That is, assume that p is supported and constant on consecutive semiopen intervals $\Delta_k =$

$[x_{k-1}, x_k)$ $k = 1, \dots, n$, where $x_0 < \dots < x_n$. Putting $p(x) = c_k$ on Δ_k , we then have

$$\|p\|_{\text{TV}} = c_1 + |c_2 - c_1| + \dots + |c_n - c_{n-1}| + c_n.$$

In this case the existence of the representation (4.6), moreover—with a discrete mixing measure π , satisfying (4.8), can be proved by induction on n .

If $n = 1$, there is nothing to prove. For $n = 2$, if $c_1 = c_2$ or $\min(c_1, c_2) = 0$, we are reduced to the case $n = 1$. Otherwise, let for definiteness $c_2 > c_1 > 0$. Then one can write

$$p = c_1 1_{[x_0, x_2)} + (c_2 - c_1) 1_{[x_1, x_2)} = \alpha_1 q_1 + \alpha_2 q_2,$$

where q_1 is the uniform density on $\Delta_1 \cup \Delta_2$ and q_2 is the uniform density on Δ_2 (with certain $\alpha_1, \alpha_2 > 0$, $\alpha_1 + \alpha_2 = 1$). This representation corresponds to (4.6) with π having the atoms at q_1 and q_2 . In addition,

$$\alpha_1 \|q_1\|_{\text{TV}} + \alpha_2 \|q_2\|_{\text{TV}} = \|c_1 1_{[x_0, x_2)}\|_{\text{TV}} + \|(c_2 - c_1) 1_{[x_1, x_2)}\|_{\text{TV}} = 2c_2 = \|p\|_{\text{TV}},$$

so (4.8) is fulfilled.

If $n \geq 3$, first we distinguish between several cases. If $c_1 = 0$ or $c_n = 0$, we are reduced to the smaller number of supporting intervals. If $c_k = 0$ for some $1 < k < n$, one can write $p = f + g$ with $f(x) = p(x) 1_{\{x < x_{k-1}\}}$, $g(x) = p(x) 1_{\{x \geq x_k\}}$. These functions are supported on disjoint half-axes, so $\|p\|_{\text{TV}} = \|f\|_{\text{TV}} + \|g\|_{\text{TV}}$. Moreover, the induction hypothesis may be applied to both f and g (or one can first normalize these functions to work with densities, but this is less convenient). As a result,

$$f = f_1 + \dots + f_k, \quad g = g_1 + \dots + g_l \quad \text{a.e.}$$

where each f_i is supported and constant on some interval inside $[x_0, x_{k-1})$, each g_j is supported and constant on some interval inside $[x_k, x_n)$, and

$$\|f\|_{\text{TV}} = \|f_1\|_{\text{TV}} + \dots + \|f_k\|_{\text{TV}}, \quad \|g\|_{\text{TV}} = \|g_1\|_{\text{TV}} + \dots + \|g_l\|_{\text{TV}}.$$

Hence,

$$p = \sum_i f_i + \sum_j g_j \quad \text{with} \quad \|f\|_{\text{TV}} = \sum_i \|f_i\|_{\text{TV}} + \sum_j \|g_j\|_{\text{TV}}.$$

Finally, assume that $c_k > 0$ for all $k \leq n$. Putting $c_* = \min_k c_k$, write $p = f + g$, where $f = c_* 1_{[x_0, x_n)}$ and g thus takes the values $c_k - c_*$ on Δ_k . Clearly,

$$\|p\|_{\text{TV}} = 2c_* + \|g\|_{\text{TV}} = \|f\|_{\text{TV}} + \|g\|_{\text{TV}}.$$

By the definition, g takes the value zero on one of the intervals (where $c_k = c_*$), so we are reduced to the previous step. On that step, we obtained a representation

$g = g_1 + \dots + g_l$ such that $\|g\|_{TV} = \|g_1\|_{TV} + \dots + \|g_l\|_{TV}$, where each g_j is supported and constant on some interval inside $[x_0, x_n)$. Hence,

$$p = f + \sum_j g_j \quad \text{with} \quad \|p\|_{TV} = \|f\|_{TV} + \sum_j \|g_j\|_{TV}.$$

Although the measure π has not been constructed constructively, one may notice that it should be supported on the densities of the form

$$q_{ij}(x) = \frac{1}{x_j - x_i} 1_{\{x_i \leq x < x_j\}}, \quad 0 \leq i < j \leq n.$$

Step 2. (Approximation) In the general case, one may assume that p is right-continuous. Consider the collection of piecewise constant densities of the form

$$\tilde{p}(x) = dq(x), \quad q(x) = \sum_{k=1}^N c_k 1_{\{x_{k-1} \leq x < x_k\}}, \quad c_k = \min_{x_{k-1} \leq x \leq x_k} p(x), \quad (4.9)$$

with arbitrary points $x_0 < \dots < x_N$ of continuity of p , such that p is not vanishing on (x_0, x_N) , and where $d \geq 1$ is a normalizing constant so that $\int_{-\infty}^{\infty} \tilde{p}(x) dx = 1$. Denoting by y_k a point of minimum of p on $[x_{k-1}, x_k]$, we first note that

$$\frac{1}{d} \|\tilde{p}\|_{TV} = \|q\|_{TV} = p(y_1) + p(y_N) + \sum_{k=2}^N |p(y_k) - p(y_{k-1})| \leq \|p\|_{TV}.$$

If the endpoints x_0 and x_N are fixed, while the maximal step of partition $\max_k (x_k - x_{k-1})$ is getting small, the integral $\int_{x_0}^{x_N} q(x) dx$ will approximate $\int_{x_0}^{x_N} p(x) dx$ (since p has bounded total variation). Hence, it is possible to construct a sequence $p_n(x) = d_n q_n(x)$ of the form (4.9) which is convergent to p in L^1 -norm and with $d_n \rightarrow 1$. By the construction,

$$p_n(x) \leq d_n p(x) \quad \text{and} \quad \|p_n\|_{TV} \leq d_n \|p\|_{TV}. \quad (4.10)$$

Now, using the previous step, one can define discrete probability measures π_n supported on U and such that

$$p_n(x) = \int_U q(x) d\pi_n(q), \quad \|p_n\|_{TV} = \int_U \|q\|_{TV} d\pi_n(q). \quad (4.11)$$

Since U has been identified with the half-plane \tilde{U} , replacing $d\pi_n(q)$ with $d\pi_n(a, b)$ should not lead to confusion. In particular, the second equality in (4.11) may be written as

$$\|p_n\|_{\text{TV}} = 2 \int_{\tilde{U}} \frac{1}{b-a} d\pi_n(a, b). \tag{4.12}$$

Let n be large enough, say $n \geq n_0$ (when $d_n \leq 2$). From the first equality in (4.11) and by (4.10), it then follows that, for any $T > 0$,

$$\int_U \left[\int_{|x| \geq T} q(x) dx \right] d\pi_n(q) = \int_{|x| \geq T} p_n(x) \leq 2 \int_{|x| \geq T} p(x) dx.$$

Hence, by Chebyshev’s inequality, the sets $U(\varepsilon, T) = \{q \in U : \int_{|x| \geq T} q(x) dx > \varepsilon\}$ have π_n -measure

$$\pi_n(U(\varepsilon, T)) \leq \frac{2}{\varepsilon} \int_{|x| \geq T} p(x) dx \quad (\varepsilon, T > 0). \tag{4.13}$$

Next we choose two sequences $\varepsilon = \varepsilon_k \downarrow 0$ and $T = T_k \uparrow \infty$, for which the right-hand side of (4.13), say δ_k , will tend to zero sufficiently fast, as $k \rightarrow \infty$. Let $\delta_k < 2^{-k}$. Identifying q with corresponding probability distributions, by the Prokhorov compactness criterion (cf. e.g. [6]), the collection of densities

$$F_k = \bigcap_{l=k}^{\infty} \left\{ q \in \mathfrak{P} : \int_{|x| \geq T_l} q(x) dx \leq \varepsilon_l \right\}$$

is pre-compact in the space $M(\mathbf{R})$ of all probability distributions on the real line with the weak topology. Moreover, by (4.13),

$$1 - \pi_n(F_k) \leq \sum_{l=k}^{\infty} \pi_n(U(\varepsilon_l, T_l)) \leq \sum_{l=k}^{\infty} \delta_l < 2^{-(k-1)}.$$

Therefore, by the same criterion, but now applied to the Polish space $M(M(\mathbf{R}))$ of all probability distributions on $M(\mathbf{R})$ (with the weak topology), π_n contains a weakly convergent subsequence π_{n_k} with some limit π . This measure is supported on the weak closure of U , which is a larger set, since it contains delta-measures, or the main diagonal in \mathbf{R}^2 , if we identify U with \tilde{U} . However, using (4.12) together with Chebyshev’s inequality, and then applying (4.10), we see that, for any $\varepsilon > 0$ and all $n \geq n_0$,

$$\pi_n\{(a, b) : b - a < \varepsilon\} = \pi_n \left\{ (a, b) : \frac{1}{b-a} > \frac{1}{\varepsilon} \right\} \leq \frac{\varepsilon}{2} \|p_n\|_{\text{TV}} \leq \varepsilon \|p\|_{\text{TV}}.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that π is actually supported on U .

Moreover, taking the limit along n_k in the first equality in (4.11), we obtain the representation (4.6). Indeed, (4.11) implies that, for all $a < b$,

$$\int_a^b p_n(x) dx = \int_U \left[\int_a^b q(x) dx \right] d\pi_n(q).$$

The functional $q \rightarrow \int_a^b q(x) dx$ is bounded and continuous on the space \mathfrak{P} with the weak topology (3.3), so the limit yields a similar equality

$$\int_a^b p(x) dx = \int_U \left[\int_a^b q(x) dx \right] d\pi(q).$$

But the latter is equivalent to (4.6).

Finally, the sets $G(t) = \{q \in U : \|q\|_{TV} > t\}$ are open in the weak topology (by the lower semicontinuity of the total variation norm), hence, $\liminf_{k \rightarrow \infty} \pi_{n_k}(G(t)) \geq \pi(G(t))$. Applying Fatou’s lemma and then again (4.10) and the second equality in (4.11), we get

$$\begin{aligned} \int_U \|q\|_{TV} d\pi(q) &= \int_0^\infty \pi(G(t)) dt \leq \liminf_{k \rightarrow \infty} \int_0^\infty \pi_{n_k}(G(t)) dt \\ &= \liminf_{k \rightarrow \infty} \int_U \|q\|_{TV} d\pi_{n_k}(q) = \liminf_{k \rightarrow \infty} \|p_{n_k}\|_{TV} \leq \|p\|_{TV}. \end{aligned}$$

In view of Jensen’s inequality (4.7), we obtain (4.8) thus proving the lemma. □

Proof of Proposition 4.1 We may write down the representation (4.6) from Lemma 4.2 for each of the densities p_j ($j = 1, 2, 3$). That is,

$$p_j(x) = \int q(x) d\pi_j(q) \quad \text{a.e.}$$

with some mixing probability measures π_j , supported on U and satisfying

$$\|p_j\|_{TV} = \int \|q\|_{TV} d\pi_j(q). \tag{4.14}$$

Taking the convolution, we have a similar representation

$$(p_1 * p_2 * p_3)(x) = \iiint (q_1 * q_2 * q_3)(x) d\pi_1(q_1) d\pi_2(q_2) d\pi_3(q_3) \quad \text{a.e.}$$

One can now use Jensen’s inequality (3.8) for the Fisher information and apply (4.2) to bound $I(p_1 * p_2 * p_3)$ from above by

$$\frac{1}{2} \iiint [\|q_1\|_{TV} \|q_2\|_{TV} + \|q_1\|_{TV} \|q_3\|_{TV} + \|q_2\|_{TV} \|q_3\|_{TV}] d\pi_1(q_1) d\pi_2(q_2) d\pi_3(q_3).$$

In view of (4.14), the triple integral coincides with the right-hand side of (4.1). □

5 Bounds in terms of characteristic functions

In view of Proposition 4.1, let us describe how to bound the total variation norm of a given density p of a random variable X in terms of the characteristic function

$$f(t) = \mathbf{E} e^{itX} = \int_{-\infty}^{\infty} e^{itx} p(x) dx.$$

There are many different bounds depending on the integrability properties of f and its derivatives, which may also depend on assumptions on the finiteness of moments of X . We shall present two of them here.

Recall that, if p is absolutely continuous, then

$$\|p\|_{TV} = \int_{-\infty}^{\infty} |p'(x)| dx.$$

Proposition 5.1 *If X has finite second moment and*

$$\int_{-\infty}^{\infty} |t| (|f(t)| + |f'(t)| + |f''(t)|) dt < \infty, \tag{5.1}$$

then X has a continuously differentiable density p with finite total variation

$$\|p\|_{TV} \leq \frac{1}{2} \int_{-\infty}^{\infty} (|tf''(t)| + 2|f'(t)| + |tf(t)|) dt. \tag{5.2}$$

Proof The argument is standard, and we recall it here for completeness.

First, by the moment assumption, f is twice continuously differentiable. Using the inverse Fourier transform, the assumption (5.1) implies that X has a continuously differentiable density

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t) dt \tag{5.3}$$

with derivative

$$p'(x) = -\frac{i}{2\pi} \int_{-\infty}^{\infty} e^{-itx} t f(t) dt. \tag{5.4}$$

By the Riemann–Lebesgue theorem, $f(t) \rightarrow 0$, as $|t| \rightarrow \infty$, and the same is true for the derivatives $f'(t)$ and $f''(t)$ (since they are Fourier transforms of integrable functions). Therefore, one may integrate in (5.3) by parts to get, for all $x \in \mathbf{R}$,

$$xp(x) = -\frac{i}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f'(t) dt \tag{5.5}$$

and

$$x^2 p(x) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f''(t) dt.$$

By (5.1), we are allowed to differentiate the last equality by performing differentiation under the integral sign, which together with (5.4) and (5.5) gives

$$(1 + x^2)p'(x) = \frac{i}{2\pi} \int_{-\infty}^{\infty} e^{-itx} (tf''(t) + 2f'(t) - tf(t)) dt.$$

Hence, $|p'(x)| \leq \frac{C}{2\pi(1+x^2)}$ with a constant described as the integral in (5.2). After integration of this pointwise bound, the proposition follows. \square

One can get rid of the assumption of existing second derivative in the bound above and remove any moment assumption in Proposition 5.1. But we still need to insist on some integrability and differentiability properties for the characteristic function on the positive half-axis.

Proposition 5.2 *Assume that the characteristic function $f(t)$ of a random variable X has a continuous derivative for $t > 0$ with*

$$\int_{-\infty}^{\infty} t^2 (|f(t)|^2 + |f'(t)|^2) dt < \infty. \tag{5.6}$$

Then X has an absolutely continuous density p with finite total variation

$$\|p\|_{\text{TV}} \leq \left(\int_{-\infty}^{\infty} |tf(t)|^2 dt \int_{-\infty}^{\infty} |(tf(t))'|^2 dt \right)^{1/4}. \tag{5.7}$$

Proof First assume additionally that f decays at infinity sufficiently fast. Then $tf(t)$ is integrable, so that X has a smooth density p with derivative p' represented by (5.4). One may integrate therein by parts over the intervals $(-T, -\varepsilon)$ and (ε, T) with $\varepsilon \downarrow 0, T \uparrow \infty$, using the property that $(tf(t))'$ is integrable near zero. Then we get in the limit a similar representation

$$xp'(x) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} (tf(t))' dt,$$

where the integral is understood in the improper sense (at infinity), and with resulting function in $L^2(-\infty, \infty)$. Write $|p'(x)| = \frac{1}{|1+ix|} |(1+ix)p'(x)|$ and use Cauchy's inequality together with Plancherel's formula, to get

$$\begin{aligned} \left(\int_{-\infty}^{\infty} |p'(x)| dx \right)^2 &\leq \int_{-\infty}^{\infty} \frac{dx}{1+x^2} \int_{-\infty}^{\infty} (1+x^2) p'(x)^2 dx \\ &= \frac{1}{2} \int_{-\infty}^{\infty} [|tf(t)|^2 + |(tf(t))'|^2] dt. \end{aligned}$$

Applying the same inequality to λX and optimizing over $\lambda > 0$, we arrive at (5.7).

In the general case, one may apply (5.7) to the regularized random variables $X_\sigma = X + \sigma Z$ with small parameter $\sigma > 0$, where $Z \sim N(0, 1)$ is independent of X . They have smooth densities p_σ and characteristic functions $f_\sigma(t) = f(t) e^{-\sigma^2 t^2/2}$. Repeating the previous argument for the difference of densities, we obtain an analogue of (5.7),

$$\|p_{\sigma_1} - p_{\sigma_2}\|_{TV}^4 \leq \int_{-\infty}^{\infty} |t(f_{\sigma_1}(t) - f_{\sigma_2}(t))|^2 dt \int_{-\infty}^{\infty} |(t(f_{\sigma_1}(t) - f_{\sigma_2}(t)))'|^2 dt \quad (5.8)$$

with arbitrary $\sigma_1, \sigma_2 > 0$. Since the integrals in (5.7) are finite, by the Lebesgue dominated convergence theorem, the right-hand side of (5.8) tends to zero, as $\sigma_1, \sigma_2 \rightarrow 0$. Hence, the family $\{p_\sigma\}$ is fundamental (Cauchy) for $\sigma \rightarrow 0$ in the Banach space of all functions of bounded variation on the real line that are vanishing at infinity. As a result, there exists the limit $p = \lim_{\sigma \rightarrow 0} p_\sigma$ in this space in total variation norm.

Necessarily, $p(x) \geq 0$, for all x , and $\int_{-\infty}^{\infty} p(x) dx = 1$. Hence, X has an absolutely continuous distribution with density p . In addition, by (5.7) applied to p_σ ,

$$\|p\|_{TV} = \lim_{\sigma \rightarrow 0} \|p_\sigma\|_{TV} \leq \lim_{\sigma \rightarrow 0} \left(\int_{-\infty}^{\infty} |tf_\sigma(t)|^2 dt \int_{-\infty}^{\infty} |(tf_\sigma(t))'|^2 dt \right)^{1/4}.$$

The last limit exists and coincides with the right-hand side of (5.7).

Finally, using Plancherel’s formula in (5.4) for the regularized random variables, we have, for all $\sigma_1, \sigma_2 > 0$,

$$\int_{-\infty}^{\infty} |p'_{\sigma_1}(x) - p'_{\sigma_2}(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} t^2 |f_{\sigma_1}(t) - f_{\sigma_2}(t)|^2 dt.$$

This relation shows that $\{p_\sigma\}$ is a fundamental family also in the Sobolev space $W_1^2(-\infty, \infty)$, and necessarily $p = \lim_{\sigma \rightarrow 0} p_\sigma$ in the norm of W_1^2 . Thus, p belongs to W_1^2 and is therefore absolutely continuous. \square

Combining Proposition 4.1 with Propositions 5.1–5.2, one can bound the Fisher information of the sum of three independent random variables in terms of their characteristic functions. In particular, in the i.i.d. case, we have:

Corollary 5.3 *If the independent random variables X_1, X_2, X_3 have finite first absolute moment and a common characteristic function $f(t)$, then*

$$I(X_1 + X_2 + X_3) \leq \frac{3}{2} \left(\int_{-\infty}^{\infty} |tf(t)|^2 dt \int_{-\infty}^{\infty} |(tf(t))'|^2 dt \right)^{1/2}. \tag{5.9}$$

If X_1 has finite second moment, we also have

$$I(X_1 + X_2 + X_3) \leq \frac{3}{8} \left(\int_{-\infty}^{\infty} (|tf''(t)| + 2|f'(t)| + |tf(t)|) dt \right)^2.$$

It is interesting to note that, in turn, the first integral in (5.9) is bounded from above by $I(X_1)^{3/2}$ up to a constant (Proposition 2.4). The same can also be shown for the second integral under the 4th moment assumption (cf. Sect. 7).

6 Classes of densities representable as convolutions

General bounds like those in Proposition 2.2 may considerably be sharpened in the case where p is representable as convolution of several densities with finite Fisher information.

Definition 6.1 Given an integer $k \geq 1$ and a real number $I > 0$, denote by $\mathfrak{P}_k(I)$ the collection of all functions on the real line which can be represented as convolution of k probability densities with Fisher information at most I . Correspondingly, let

$$\mathfrak{P}_k = \cup_{I>0} \mathfrak{P}_k(I)$$

denote the collection of all functions representable as convolution of k probability densities with finite Fisher information.

The collection \mathfrak{P}_1 of all densities with finite Fisher information has been already discussed in connection with general properties of the functional I . For growing k , the classes $\mathfrak{P}_k(I)$ decrease, since the Fisher information may only decrease when adding an independent summand. This also follows from the following general inequality of Stam

$$\frac{1}{I(X_1 + \dots + X_k)} \geq \frac{1}{I(X_1)} + \dots + \frac{1}{I(X_k)}, \tag{6.1}$$

which holds for all independent random variables X_1, \dots, X_k (cf. [7, 15, 22]). Moreover, it implies that $p = p_1 * \dots * p_k \in \mathfrak{P}_1(I/k)$, as long as $p_i \in \mathfrak{P}_1(I)$, $i = 1, \dots, k$.

Any function p in \mathfrak{P}_k is $k - 1$ times differentiable, and its $(k - 1)$ -th derivative is absolutely continuous and has a Radon–Nikodym derivative, which we denote by $p^{(k)}$. Let us illustrate this property in the important case $k = 2$. Write

$$p(x) = \int_{-\infty}^{\infty} p_1(x - y)p_2(y) dx \tag{6.2}$$

in terms of absolutely continuous densities p_1 and p_2 of independent summands X_1 and X_2 of a random variable X with density p . Differentiating under the integral sign, we obtain a Radon–Nikodym derivative of the function p ,

$$p'(x) = \int_{-\infty}^{\infty} p'_1(x - y)p_2(y) dy = \int_{-\infty}^{\infty} p'_1(y)p_2(x - y) dy. \tag{6.3}$$

The latter expression shows that p' is absolutely continuous and has a Radon–Nikodym derivative

$$p''(x) = \int_{-\infty}^{\infty} p'_1(y)p'_2(x - y) dy, \tag{6.4}$$

which is well-defined for all x . In other words, p'' appears as the convolution of the functions p'_1 and p'_2 (which are integrable, according to Proposition 2.2).

These formulas may be used to derive a number of elementary relations within the class \mathfrak{P}_k , and here we shall describe some of them for the cases \mathfrak{P}_2 and \mathfrak{P}_3 .

Proposition 6.2 *Given a density $p \in \mathfrak{P}_2(I)$, for all $x \in \mathbf{R}$,*

$$|p'(x)| \leq I^{3/4} \sqrt{p(x)} \leq I. \tag{6.5}$$

Moreover, p' has finite total variation

$$\|p'\|_{\text{TV}} = \int_{-\infty}^{\infty} |p''(x)| dx \leq I.$$

The last bound immediately follows from (6.4) and Proposition 2.2. To obtain the pointwise bound on the derivative, we appeal to Proposition 2.1 and rewrite the first equality in (6.3) as

$$p'(x) = \int_{-\infty}^{\infty} \frac{p'_1(x - y)}{\sqrt{p_1(x - y)}} \mathbf{1}_{\{p_1(x - y) > 0\}} \sqrt{p_1(x - y)} p_2(y) dy.$$

By Cauchy’s inequality,

$$\begin{aligned} p'(x)^2 &\leq I(X_1) \int_{-\infty}^{\infty} p_1(x - y) p_2(y)^2 dy \\ &\leq I(X_1) \max_y p_2(y) \int_{-\infty}^{\infty} p_1(x - y) p_2(y) dy \leq I(X_1) I(X_2)^{1/2} p(x), \end{aligned}$$

where we applied Proposition 2.2 to the random variable X_2 on the last step. This gives the first inequality in (6.5), while the second follows from $p(x) \leq \sqrt{I}$.

Now, we state similar bounds for the second derivative.

Proposition 6.3 *For any density $p \in \mathfrak{F}_2(I)$ $p(x) = 0 \Rightarrow p''(x) = 0$, for all x . Moreover,*

$$\int_{\{p(x) > 0\}} \frac{p''(x)^2}{p(x)} dx \leq I^2.$$

Proof Let us start with the representation (6.4) for a fixed value $x \in \mathbf{R}$. By Proposition 2.1, the integral in (6.4) may be restricted to the set $\{y : p_2(y) > 0\}$. By the same reason, it may also be restricted to the set $\{y : p_1(x - y) > 0\}$. Hence,

$$p''(x) = \int_{-\infty}^{\infty} p'_1(y) p'_2(x - y) \mathbf{1}_A(y) dy, \tag{6.6}$$

where $A = \{y : p_1(x - y) p_2(y) > 0\}$. On the other hand, $p(x) = 0$ in the equality (6.2) implies that $p_1(y) p_2(x - y) = 0$ for almost all y . Therefore, $\mathbf{1}_A(y) = 0$ a.e., and thus the integral (6.6) is vanishing, that is, $p''(x) = 0$.

Next, introduce the functions $u_i(x) = \frac{p'_i(x)}{\sqrt{p_i(x)}} \mathbf{1}_{\{p_i(x) > 0\}}$ ($i = 1, 2$) and rewrite (6.4) as

$$p''(x) = \int_{-\infty}^{\infty} (u_1(x - y) u_2(y)) \sqrt{p_1(x - y) p_2(y)} dy.$$

By Cauchy's inequality,

$$p''(x)^2 \leq \int_{-\infty}^{\infty} u_1(x-y)^2 u_2(y)^2 dy \int_{-\infty}^{\infty} p_1(x-y) p_2(y) dy = u(x)^2 p(x), \quad (6.7)$$

where $u \geq 0$ is defined by

$$u(x)^2 = \int_{-\infty}^{\infty} u_1(x-y)^2 u_2(y)^2 dy. \quad (6.8)$$

Clearly,

$$\int_{-\infty}^{\infty} u(x)^2 dx = I(X_1)I(X_2) \leq I^2,$$

which is the inequality of the proposition. \square

Proposition 6.4 *Given a density $p \in \mathfrak{P}_3(I)$, we have, for all x ,*

$$|p''(x)| \leq I^{5/4} \sqrt{p(x)} \leq I^{3/2}. \quad (6.9)$$

Proof By the assumption, one may write $p = p_1 * p_2$ with $p_1 \in \mathfrak{P}_1(I)$ and $p_2 \in \mathfrak{P}_2(I)$. Returning to (6.7)–(6.8) and applying Proposition 6.2 to p_2 , we get $u_2(y) \leq I^{3/4}$, so

$$u(x)^2 \leq I^{3/2} \int_{-\infty}^{\infty} u_1(x-y)^2 dy \leq I^{5/2}.$$

This proves the first inequality in (6.9). The second bound follows from the uniform bound $p(x) \leq \sqrt{I}$, cf. Proposition 2.2. \square

7 Bounds under moment assumptions

Another way to sharpen the bounds obtained in Sect. 2 for general densities with finite Fisher information is to invoke conditions on the absolute moments

$$\beta_s = \beta_s(X) = \mathbf{E}|X|^s \quad (s > 0 \text{ real}).$$

By Proposition 2.1 and Cauchy’s inequality, if the Fisher information is finite,

$$\begin{aligned} \int_{-\infty}^{\infty} |x|^s |p'(x)| dx &= \int_{\{p(x)>0\}} |x|^s p(x)^{1/2} \frac{|p'(x)|}{p(x)^{1/2}} dx \\ &\leq \left(\int_{\{p(x)>0\}} |x|^{2s} p(x) dx \right)^{1/2} \left(\int_{\{p(x)>0\}} \frac{p'(x)^2}{p(x)} dx \right)^{1/2}. \end{aligned}$$

Hence, we arrive at:

Proposition 7.1 *If X has an absolutely continuous density p , then, for any $s > 0$,*

$$\int_{-\infty}^{\infty} |x|^s |p'(x)| dx \leq \sqrt{\beta_{2s} I(X)}.$$

This bound holds irrespectively of the Fisher information or the $2s$ -th absolute moment β_{2s} being finite or not. Below we describe several applications of this proposition.

First, let us note that, when $s \geq 1$, the function $u(x) = (1 + |x|^s)p(x)$ is (locally) absolutely continuous and has a Radon–Nikodym derivative satisfying

$$|u'(x)| \leq s|x|^{s-1} p(x) + (1 + |x|^s) |p'(x)|.$$

Integrating this inequality and assuming for a moment that both $I(X)$ and β_{2s} are finite, we see that u is a function of bounded variation. Since u is also integrable,

$$u(-\infty) = \lim_{x \rightarrow -\infty} u(x) = 0, \quad u(\infty) = \lim_{x \rightarrow \infty} u(x) = 0.$$

Therefore, applying Propositions 2.2 and 7.1, we get

$$\begin{aligned} u(x) &= \int_{-\infty}^x u'(y) dy \leq \int_{-\infty}^{\infty} |u'(y)| dy \\ &\leq s \int_{-\infty}^{\infty} |x|^{s-1} p(x) dx + \int_{-\infty}^{\infty} (1 + |x|^s) |p'(x)| dx \\ &\leq s\beta_{s-1} + \sqrt{I(X)} + \sqrt{\beta_{2s} I(X)}. \end{aligned}$$

One can summarize.

Corollary 7.2 *If X has density p , then, given $s \geq 1$, for any $x \in \mathbf{R}$,*

$$p(x) \leq \frac{C}{1 + |x|^s}$$

with $C = s\beta_{s-1} + \sqrt{2(1 + \beta_{2s})I(X)}$. If this constant is finite, we also have

$$\lim_{x \rightarrow \infty} (1 + |x|^s) p(x) = 0.$$

In the resulting inequality no requirements on the density are needed.

Under stronger moment assumptions, one can obtain better bounds for the decay of the density. For example, if for some $\lambda > 0$, the exponential moment

$$\beta = \mathbf{E} e^{2\lambda|X|} = \int_{-\infty}^{\infty} e^{2\lambda|x|} p(x) dx$$

is finite, then by similar arguments, $p(x) \leq C e^{-\lambda|x|}$, for any $x \in \mathbf{R}$, with some constant C depending on λ , β and $I(X)$.

Applying Proposition 7.1 and Corollary 7.2 (the last assertion) with $s = 1$, we obtain the following analogue of Proposition 2.3.

Corollary 7.3 *If X has finite second moment and finite Fisher information $I(X)$, then for its characteristic function $f(t) = \mathbf{E} e^{itX}$ we have*

$$|f'(t)| \leq \frac{C}{|t|}, \quad t \in \mathbf{R},$$

with constant $C = 1 + \sqrt{\beta_2 I(X)}$.

Indeed, if p is density of X and $t \neq 0$, one may integrate by parts

$$f'(t) = \frac{1}{t} \int_{-\infty}^{\infty} x p(x) de^{itx} = -\frac{1}{t} \int_{-\infty}^{\infty} (p(x) + x p'(x)) e^{itx} dx,$$

which yields $|t f'(x)| \leq 1 + \sqrt{\beta_2 I(X)}$.

One can also derive a similar integral bound with the help of Corollary 7.2 with $s = 2$, that is, assuming that β_4 is finite. Alternatively (so that to improve the resulting constant), let us repeat the argument used in the proof of Corollary 7.2 with the particular function $u(x) = x^2 p(x)$. Then we readily get

$$x^2 p(x) \leq 2\beta_1 + \sqrt{\beta_4 I(X)}.$$

But, by the Cramér–Rao inequality, $\beta_4 I(X) \geq \beta_2^2 I(X) \geq \beta_2 \geq \beta_1^2$, and the above estimate is simplified to $x^2 p(x) \leq 3\sqrt{\beta_4 I(X)}$. Hence,

$$\int_{-\infty}^{\infty} x^2 p'(x)^2 dx = \int_{\{p(x)>0\}} x^2 p(x) \frac{p'(x)^2}{p(x)} dx \leq 3\sqrt{\beta_4 I(X)} I(X).$$

Since $x p'(x)$ represents the inverse Fourier transform for $-(tf(t))'$, one may use the Plancherel formula which leads to:

Corollary 7.4 *If X has finite 4th moment and finite Fisher information $I(X)$, then*

$$\int_{-\infty}^{\infty} |(tf(t))'|^2 dt \leq 6\pi\sqrt{\beta_4} I(X)^{3/2}.$$

The left integral appeared in the bound (5.9) of Corollary 5.3. Combining Proposition 2.4 and Corollary 7.4, (5.9) may be thus complemented by a similar I -containing bound, namely,

$$\begin{aligned} I(X_1 + X_2 + X_3) &\leq \frac{3}{2} \left(\int_{-\infty}^{\infty} |tf(t)|^2 dt \int_{-\infty}^{\infty} |(tf(t))'|^2 dt \right)^{1/2} \\ &\leq 3\sqrt{3} \pi \beta_4^{1/4} I(X_1)^{3/2}, \end{aligned}$$

where random variables X_1, X_2, X_3 are independent and have a common characteristic function $f(t)$ with finite 4th moment $\beta_4 = \beta_4(X_1)$.

8 Fisher information in terms of the second derivative

It will be convenient to work with the formulas for the Fisher information and for the parts of corresponding integrals over half-axes, which involve the second derivative of the density. First we consider convolutions of two densities with finite Fisher information.

Proposition 8.1 *If a random variable X has density $p \in \mathfrak{F}_2$, then*

$$I(X) = - \int_{-\infty}^{\infty} p''(x) \log p(x) dx, \tag{8.1}$$

provided that

$$\int_{-\infty}^{\infty} |p''(x) \log p(x)| dx < +\infty. \tag{8.2}$$

The latter condition holds, if $\mathbf{E} |X|^s < \infty$ for some $s > 2$.

Strictly speaking, the integration in (8.1)–(8.2) should be performed over the open set $G = \{x : p(x) > 0\}$. One may extend this integration to the whole real line by using the convention $0 \log 0 = 0$. This is consistent with the property that $p''(x) = 0$, as soon as $p(x) = 0$ (according to Proposition 6.3).

Proof The assumption $p \in \mathfrak{P}_2$ ensures that p has an absolutely continuous derivative p' with Radon–Nikodym derivative p'' . By Proposition 6.2, p' has bounded total variation, which justifies the possibility of integration by parts.

More precisely, assuming that $p \in \mathfrak{P}_2$, let us decompose the set G into disjoint open intervals (a_n, b_n) , bounded or not. In particular, $p(a_n) = p(b_n) = 0$, and by the bound (6.5),

$$|p'(x) \log p(x)| \leq I^{3/4} \sqrt{p(x)} |\log p(x)| \rightarrow 0, \quad \text{as } x \downarrow a_n,$$

and similarly for b_n . Integrating by parts, we get for $a_n < T_1 < T_2 < b_n$,

$$\begin{aligned} \int_{T_1}^{T_2} \frac{p'(x)^2}{p(x)} dx &= \int_{T_1}^{T_2} p'(x) d \log p(x) \\ &= p'(x) \log p(x) \Big|_{x=T_1}^{T_2} - \int_{T_1}^{T_2} p''(x) \log p(x) dx. \end{aligned}$$

Letting $T_1 \rightarrow a_n$ and $T_2 \rightarrow b_n$, we get

$$\int_{a_n}^{b_n} \frac{p'(x)^2}{p(x)} dx = - \int_{a_n}^{b_n} p''(x) \log p(x) dx,$$

where the second integral is understood in the improper sense. It remains to perform summation over n on the basis of (8.2), and then we obtain (8.1).

To verify the integrability condition (8.2), one may apply an integral bound of Proposition 6.3. Namely, using Cauchy’s inequality, for the integral in (8.2) we have

$$\left(\int_{\{p(x)>0\}} \frac{|p''(x)|}{\sqrt{p(x)}} \sqrt{p(x)} |\log p(x)| dx \right)^2 \leq I^2 \int_{-\infty}^{\infty} p(x) \log^2 p(x) dx.$$

If the moment $\beta_s = \mathbf{E} |X|^s$ is finite, Corollary 7.2 yields

$$p(x) \log^2 p(x) \leq C \frac{\log^2(e + |x|)}{1 + |x|^{s/2}}$$

with a constant C depending on I and β_s . The latter function is integrable in case $s > 2$, so the integral in (8.2) is finite. □

As the above argument shows, without the requirement that $p \in \mathfrak{P}_2$ and the integrability condition (8.2), formula (8.1) still remains valid under the following assumptions:

- $p(x)$ is twice continuously differentiable on the real line;
- $p(x) > 0$, for all x ;
- $p'(x) \log p(x) \rightarrow 0$, as $|x| \rightarrow \infty$.

However, then the integral (8.1) should be understood in the improper sense, i.e., we have

$$I(X) = - \lim_{T_1 \rightarrow -\infty, T_2 \rightarrow \infty} \int_{T_1}^{T_2} p''(x) \log p(x) dx,$$

where the limit exists regardless of whether the Fisher information is finite or not.

In order to involve the standard moment assumption—the finiteness of the second moment, we consider densities representable as convolutions of more than two densities with finite Fisher information.

Proposition 8.2 *If a random variable X has finite second moment and density $p \in \mathfrak{P}_5$, then condition (8.2) holds, and for all $-\infty \leq a < b \leq \infty$,*

$$\int_a^b \frac{p'(x)^2}{p(x)} 1_{\{p(x)>0\}} dx = p'(b) \log p(b) - p'(a) \log p(a) - \int_a^b p''(x) \log p(x) dx. \tag{8.3}$$

In particular, X has finite Fisher information given by (8.1).

Here we use the convention $p'(\pm\infty) \log p(\pm\infty) = 0$ for the case where a and/or b are infinite, together with

$$p'(x) \log p(x) = p''(x) \log p(x) = 0 \quad \text{in the case } p(x) = 0,$$

as before in (8.1)–(8.2). To show that (8.2) is indeed fulfilled, it will be sufficient to prove the following pointwise bounds which are of independent interest.

Proposition 8.3 *If $EX^2 \leq 1$ and X has density $p \in \mathfrak{P}_5(I)$, then with some absolute constant C , for all x ,*

$$|p''(x)| \leq CI^3 \frac{1}{1+x^2} \tag{8.4}$$

and

$$|p''(x) \log p(x)| \leq CI^3 \frac{\log(e + |x|)}{1+x^2}. \tag{8.5}$$

Proof The assumption $\mathbf{E}X^2 \leq 1$ implies $I \geq 1$ (by Cramer–Rao’s inequality). Also, the characteristic function $f(t) = \mathbf{E}e^{itX}$ is twice differentiable, and by Proposition 2.3, it satisfies

$$|f(t)| \leq \frac{I^{5/2}}{|t|^5}.$$

Hence, p may be described as the inverse Fourier transform

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t) dt,$$

and a similar representation is also valid for the second derivative,

$$p''(x) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} t^2 f(t) dt. \tag{8.6}$$

Write $X = X_1 + \dots + X_5$ with independent summands such that $I(X_j) \leq I$ and assume (without loss of generality) that they have equal means. Then $\mathbf{E}X_j^2 \leq 1$, hence the characteristic functions $f_j(t)$ of X_j have second derivatives $|f_j''(t)| \leq 1$. Moreover, by Proposition 2.3 and Corollary 7.3,

$$|f_j(t)| \leq \frac{I^{1/2}}{|t|}, \quad |f_j'(t)| \leq \frac{1 + I^{1/2}}{|t|}.$$

Now, differentiation of the equality $f(t) = f_1(t) \dots f_5(t)$ leads to

$$f'(t) = f_1'(t) f_2(t) \dots f_5(t) + \dots + f_1(t) \dots f_4(t) f_5'(t),$$

hence $|f'(t)| \leq \frac{5I^2(1+I^{1/2})}{|t|^5}$. Differentiating once more, it should be clear that

$$|f''(t)| \leq \frac{5I^2}{t^4} + \frac{20I^{3/2}(1 + I^{1/2})^2}{|t|^5}.$$

These estimates imply that

$$|(t^2 f(t))'| \leq \frac{CI^{5/2}}{|t|^3}, \quad |(t^2 f(t))''| \leq \frac{CI^{5/2}}{t^2} \quad (|t| \geq 1)$$

with some absolute constant C . As a consequence, one may integrate in (8.6) by parts with $x \neq 0$ to get

$$p''(x) = \frac{1}{2\pi x^2} \int_{-\infty}^{\infty} (t^2 f(t))'' e^{-itx} dx.$$

Hence, for all $x \in \mathbf{R}$,

$$|p''(x)| \leq \frac{CI^{5/2}}{1+x^2} \tag{8.7}$$

with some absolute constant C , implying the required pointwise bound (8.4).

Now, to derive the second pointwise bound, first we recall that $p(x) \leq I^{1/2}$. Hence,

$$|\log p(x)| \leq \log(I^{1/2}) + \log \frac{I^{1/2}}{p(x)}, \tag{8.8}$$

where the last term is thus non-negative. Next, we partition the real line into the sets $A = \{x : p(x) \leq \frac{I^{1/2}}{2(1+x^4)}\}$ and its complement B . On the set A , by Proposition 6.4,

$$|p''(x)| \log \frac{I^{1/2}}{p(x)} \leq I^{5/4} \sqrt{p(x)} \log \frac{I^{1/2}}{p(x)} \leq C_1 I^{3/2} \frac{\log(e + |x|)}{1+x^2},$$

and similarly, by (8.7), on the set B we have an analogous inequality

$$|p''(x)| \log \frac{I^{1/2}}{p(x)} \leq |p''(x)| \log (2(1+x^4)) \leq C_2 I^{5/2} \frac{\log(e + |x|)}{1+x^2}.$$

Thus, for all x , applying (8.8) and again (8.7),

$$\begin{aligned} |p''(x) \log p(x)| &\leq |p''(x)| \log(I^{1/2}) + |p''(x)| \log \frac{I^{1/2}}{p(x)} \\ &\leq CI^{5/2} (1 + \log I) \frac{\log(e + |x|)}{1+x^2}. \end{aligned}$$

Proposition 8.3 is proved. □

Proof of Proposition 8.2 Like in the proof of Proposition 8.1, first one should decompose the open set $G = \{x \in (a, b) : p(x) > 0\}$ into disjoint open intervals (a_n, b_n) . If $G = (a, b)$, then for $a < T_1 < T_2 < b$, we have

$$\int_{T_1}^{T_2} \frac{p'(x)^2}{p(x)} dx = p'(x) \log p(x) \Big|_{x=T_1}^{T_2} - \int_{T_1}^{T_2} p''(x) \log p(x) dx. \tag{8.9}$$

In case $b = \infty$ $p(x) \rightarrow 0$, as $x \rightarrow \infty$, by Corollary 7.2, so that $p'(x) \log p(x) \rightarrow 0$, due to Proposition 6.2. If $b < \infty$, then $p'(x) \log p(x) \rightarrow p'(b) \log p(b)$, as $x \rightarrow b$, with the limit being zero in case $p(b) = 0$. A similar conclusion is also true about the

point a . Hence, letting $T_1 \rightarrow a$ and $T_2 \rightarrow b$ in (8.9), we arrive at the desired equality (8.3). Moreover, the pointwise bound (8.5) confirms that the right integral in (8.3) is absolutely convergent.

If the decomposition of G contains more than one interval, similar arguments should be applied in every interval (a_n, b_n) with the following remark. If $b_n < b$, then necessarily $p(b_n) = 0$, so $p'(x) \log p(x) \rightarrow 0$, as $x \rightarrow b_n$ (and likewise for the end points $a_n > a$). Then it will remain to perform summation of the obtained equalities over all n . □

9 Normalized sums. Proof of Theorem 1.3

By the definition of classes \mathfrak{P}_k ($k = 1, 2, \dots$), the normalized sum

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

of independent random variables X_1, \dots, X_n with finite Fisher information has density p_n belonging to \mathfrak{P}_k , as long as $n \geq k$.

Moreover, if all $I(X_j) \leq I$ for all j , then $p_n \in \mathfrak{P}_k(2kI)$. Indeed, one can partition the collection X_1, \dots, X_n into k groups and write $Z_n = U_1 + \dots + U_k$ with

$$U_i = \frac{1}{\sqrt{n}} \sum_{j=i}^m X_{(i-1)m+j} \quad (1 \leq i \leq k-1), \quad U_k = \frac{1}{\sqrt{n}} \sum_{j=(k-1)m+1}^n X_j,$$

where $m = \lfloor \frac{n}{k} \rfloor$. By Stam's inequality (6.1), for $1 \leq i \leq k-1$

$$\frac{1}{I(U_i)} \geq \frac{1}{n} \sum_{j=i}^m \frac{1}{I(X_{(i-1)m+j})} \geq \frac{m}{nI} \geq \frac{1}{2kI},$$

and similarly $\frac{1}{I(U_k)} \geq \frac{1}{2kI}$.

Therefore, the previous observations about densities from \mathfrak{P}_k are applicable to Z_n with sufficiently large n , as soon as the X_j have finite Fisher information with a common bound on $I(X_j)$.

In the i.i.d. case, a similar application of (6.1) also yields $I(Z_n) \leq 2I(Z_{n_0})$. Here, the factor 2 may actually be removed as a consequence of one generalization of Stam's inequality obtained by Artstein, Ball, Barthe and Naor. It is formulated below as a separate proposition (although for our purposes the weaker inequality is sufficient).

Proposition 9.1 [2] *If $(X_n)_{n \geq 1}$ are independent and identically distributed, then $I(Z_n) \leq I(Z_{n_0})$, for all $n \geq n_0$.*

We are now ready to return to Theorem 1.3 and complete its proof.

Proof of Theorem 1.3 Let $(X_n)_{n \geq 1}$ have finite second moment and a common characteristic function f_1 . The characteristic function of Z_n is thus

$$f_n(t) = \mathbf{E} e^{itZ_n} = f_1 \left(\frac{t}{\sqrt{n}} \right)^n. \tag{9.1}$$

(a) \Rightarrow (b), according to Proposition 2.2 applied to $X = Z_n$.

(b) \Rightarrow (a) and (c). If Z_{n_1} has density p_{n_1} of bounded total variation, Proposition 4.1 yields

$$I(Z_{3n_1}) = I(p_{3n_1}) \leq \frac{3}{2} \|p_{3n_1}\|_{\text{TV}}^2 < \infty.$$

In particular, p_{3n_1} has a continuous derivative and finite total variation.

(c) \Rightarrow (a), by the same reason, and thus the conditions (a)–(c) are equivalent.

(a) \Rightarrow (d). Assume that $I(Z_{n_0}) < \infty$, for some fixed $n_0 \geq 1$. Applying Proposition 2.3 with $X = Z_{n_0}$, it follows that

$$|f_{n_0}(t)| \leq \frac{1}{t} \sqrt{I(Z_{n_0})}, \quad t > 0.$$

Hence, $|f_1(t)| \leq Ct^{-\varepsilon}$ with constants $\varepsilon = \frac{1}{n_0}$ and $C = (I(Z_{n_0})/n_0)^{1/2n_0}$.

(d) \Rightarrow (e) is obvious.

(e) \Rightarrow (c). Differentiating the formula (9.1) and using the integrability assumption (1.8) on f_1 , we see that, for all $n \geq \nu + 2$, the characteristic function f_n and its first two derivatives are integrable with weight $|t|$. This implies that Z_n has a continuously differentiable density

$$p_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f_n(t) dt, \tag{9.2}$$

which, by Proposition 5.1, has finite total variation

$$\|p_n\|_{\text{TV}} = \int_{-\infty}^{\infty} |p'_n(x)| dx \leq \frac{1}{2} \int_{-\infty}^{\infty} (|tf''_n(t)| + 2|f'_n(t)| + |tf_n(t)|) dt.$$

Thus, Theorem 1.3 is proved.

Remark 9.2 If we assume in Theorem 1.3 finiteness of the first absolute moment of X_1 (rather than the finiteness of the second moment), the statement will remain valid, provided that the integrability condition (e) is replaced with a stronger condition like

$$\int_{-\infty}^{\infty} |f_1(t)|^\nu t^2 dt < \infty, \quad \text{for some } \nu > 0. \tag{9.3}$$

In this case, it follows from (9.1) that, for all $n \geq \nu + 1$, the characteristic function f_n and its derivative are integrable with weight t^2 . Therefore, according to Proposition 5.2, the normalized sum Z_n has density p_n with finite total variation

$$\|p_n\|_{TV} \leq \left(\int_{-\infty}^{\infty} |tf_n(t)|^2 dt \int_{-\infty}^{\infty} |(tf_n(t))'|^2 dt \right)^{1/4}.$$

As a result, we obtain the chain of implications (9.3) \Rightarrow (b) \Rightarrow (a) \Rightarrow (d). The latter condition ensures that p_n admits the representation (9.2) and has a continuous derivative for sufficiently large n . That is, we obtain (c).

10 Edgeworth-type expansions

In the sequel, let $(X_n)_{n \geq 1}$ be independent identically distributed random variables with mean $\mathbf{E}X_1 = 0$ and variance $\text{Var}(X_1) = 1$. Here we collect some auxiliary results about Edgeworth-type expansions for the distribution functions $F_n(x) = \mathbf{P}\{Z_n \leq x\}$ and the densities p_n of the normalized sums $Z_n = (X_1 + \dots + X_n)/\sqrt{n}$.

We recall that

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbf{R},$$

stands the density of the standard normal law. If the absolute moment $\beta_s = \mathbf{E}|X_1|^s$ is finite for a given integer $s \geq 2$, define

$$\varphi_s(x) = \varphi(x) + \sum_{k=1}^{s-2} q_k(x) n^{-k/2}$$

with the functions q_k described in the introductory section, i.e.,

$$q_k(x) = \varphi(x) \sum H_{k+2j}(x) \frac{1}{r_1! \dots r_k!} \left(\frac{\gamma_3}{3!}\right)^{r_1} \dots \left(\frac{\gamma_{k+2}}{(k+2)!}\right)^{r_k}. \tag{10.1}$$

Here, H_l denotes the Chebyshev-Hermite polynomial of degree $l \geq 0$ with leading coefficient 1, and the summation is running over all non-negative solutions (r_1, \dots, r_k) to the equation $r_1 + 2r_2 + \dots + kr_k = k$ with notation $j = r_1 + \dots + r_k$. Put also

$$\Phi_s(x) = \int_{-\infty}^x \varphi_s(y) dy = \Phi(x) + \sum_{k=1}^{s-2} Q_k(x) n^{-k/2}.$$

Similarly to q_k , the functions Q_k have an explicit description involving the cumulants $\gamma_3, \dots, \gamma_{k+2}$ of X_1 , namely,

$$Q_k(x) = -\varphi(x) \sum H_{k+2j-1}(x) \frac{1}{r_1! \dots r_k!} \left(\frac{\gamma_3}{3!}\right)^{r_1} \dots \left(\frac{\gamma_{k+2}}{(k+2)!}\right)^{r_k}, \tag{10.2}$$

where the summation is the same as in (10.1), cf. [5] or [20].

The functions φ_s and Φ_s are used to approximate the density and the distribution function of Z_n with error of order smaller than $n^{-(s-2)/2}$. The following lemma is classical.

Lemma 10.1 *Assume that $\limsup_{|t| \rightarrow \infty} |f_1(t)| < 1$. If $\mathbf{E}|X_1|^s < \infty$ ($s \geq 2$), then as $n \rightarrow \infty$, uniformly over all x*

$$(1 + |x|^s) (F_n(x) - \Phi_s(x)) = o\left(n^{-(s-2)/2}\right). \tag{10.3}$$

Actually, the relation (10.3) remains valid for real values $s \geq 2$, in which case Φ_s should be replaced with $\Phi_{[s]}$. For the range $2 \leq s < 3$ the Cramer condition for the characteristic function is not used, cf. [19]; the range $s \geq 3$ is treated in [20] (Theorem 2, Ch.VI, p. 168).

We also need to describe the approximation of densities. Recall that Z_n have the characteristic functions

$$f_n(t) = f_1\left(\frac{t}{\sqrt{n}}\right)^n,$$

where f_1 is for the characteristic function of X_1 . If the Fisher information $I = I(Z_{n_0})$ is finite, then, by Proposition 2.3,

$$|f(t)|^{2n_0} \leq \frac{I}{I + n_0 t^2}, \quad t \in \mathbf{R}. \tag{10.4}$$

Hence, given $m \geq 1$, we have a polynomial bound $|f_n(t)| \leq c|t|^{-m}$ for $n \geq mn_0$ and with c which does not depend on t . So, for all sufficiently large n Z_n have continuous bounded densities

$$p_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f_n(t) dt,$$

which have continuous derivatives

$$p_n^{(l)}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (-it)^l e^{-itx} f_n(t) dt$$

of any prescribed order.

Lemma 10.2 Assume that $I(Z_{n_0}) < \infty$, for some n_0 , and let $\mathbf{E} |X_1|^s < \infty$ ($s \geq 2$). Fix $l = 0, 1, \dots$. Then, for all sufficiently large n ,

$$(1 + |x|^s) |p_n^{(l)}(x) - \varphi_s^{(l)}(x)| \leq \psi_{l,n}(x) \frac{\varepsilon_n}{n^{(s-2)/2}}, \quad x \in \mathbf{R}, \tag{10.5}$$

where $\varepsilon_n \rightarrow 0$, as $n \rightarrow \infty$, and

$$\sup_x |\psi_{l,n}(x)| \leq 1, \quad \int_{-\infty}^{\infty} \psi_{l,n}(x)^2 dx \leq 1.$$

For the proof of Theorem 1.1, the lemma will be used with the values $l = 0, 1, 2$, only. In case $l = 0$, this lemma with the first bound $\sup_x |\psi_{l,n}(x)| \leq 1$ is a well-known result. It does not need to require the finiteness of Fisher information, but only uses the assumption of the boundedness of p_n for large n . We can refer to [20], p. 211 in case $s \geq 3$ and to [20], pp. 198–201 for the case $s = 2$ when $\varphi_s = \varphi$.

Proof The result follows from the corresponding approximation of f_n by the Fourier transforms of φ_s on growing intervals, and here we remind a standard argument. Introduce the “corrected normal characteristic” function

$$g_s(t) = e^{-t^2/2} + e^{-t^2/2} \sum_{k=1}^{s-2} P_k(it) n^{-k/2}, \quad t \in \mathbf{R},$$

where

$$P_k(it) = \sum_{r_1+2r_2+\dots+kr_k=k} \frac{1}{r_1! \dots r_k!} \left(\frac{\gamma_3}{3!}\right)^{p_1} \dots \left(\frac{\gamma_{k+2}}{(k+2)!}\right)^{p_k} (it)^{k+2(r_1+\dots+r_k)}.$$

This function may also be defined as the Fourier transform of φ_s , i.e.,

$$g_s(t) = \int_{-\infty}^{\infty} e^{itx} \varphi_s(x) dx.$$

Note that $g_2(t) = e^{-t^2/2}$ in the case $s = 2$.

If $s \geq 3$, by Lemma 3 in [20], p. 209, in the interval $|t| \leq n^{-1/7}$, or even for $|t| \leq n^{-1/6}$ (cf. e.g. [9, 10], Proposition 9.1), we have

$$\left| f_n^{(m)}(t) - g_s^{(m)}(t) \right| \leq \frac{\varepsilon_n}{n^{(s-2)/2}} \left(|t|^{s-m} + |t|^{2s^2} \right) e^{-t^2/2}, \quad m = 0, 1, \dots, s, \tag{10.6}$$

where $\varepsilon_n \rightarrow 0$, as $n \rightarrow \infty$ (not depending on t). In case $s = 2$, one only has

$$|f_n^{(m)}(t) - g^{(m)}(t)| \leq \varepsilon_n e^{-t^2/2}, \quad |t| \leq T_n, \quad m = 0, 1, 2, \tag{10.7}$$

with some $\varepsilon_n \rightarrow 0$ and $T_n \rightarrow \infty$, as $n \rightarrow \infty$ (cf. e.g. [9, 10], Proposition 5.1). On the other hand, on larger intervals $|t| \leq \sqrt{n}$, with some positive constants C and c , there is a simple subgaussian bound

$$\left| f_n^{(m)}(t) \right| \leq C e^{-ct^2} \quad (0 \leq m \leq s, n \geq 2s), \tag{10.8}$$

which easily follows from

$$|f_1(u)| \leq e^{-c_1 u^2}, \quad |f_1'(u)| \leq |u|, \quad \left| f_1^{(s)}(u) \right| \leq \beta_s \quad (|u| \leq 1).$$

Combining (10.8) with (10.6)–(10.7), we get a unified estimate

$$\left| f_n^{(m)}(t) - g_s^{(m)}(t) \right| \leq \frac{\varepsilon_n}{n^{(s-2)/2}} e^{-ct^2}, \quad |t| \leq \sqrt{n}, \quad m = 0, 1, \dots, s, \tag{10.9}$$

with some sequence $\varepsilon_n \rightarrow 0$ and some $c > 0$ depending on the distribution of X_1 , only.

Now, since f_n is integrable for large n , one may write

$$p_n(x) - \varphi_s(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} (f_n(t) - g_s(t)) dt.$$

Moreover, with our assumptions on f_1 , one can differentiate this equality l times and then integrate by parts $m \leq s$ times to get

$$(ix)^m \left(p_n^{(l)}(x) - \varphi_s^{(l)}(x) \right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{d^m}{dt^m} \left[(-it)^l (f_n(t) - g_s(t)) \right] dt. \tag{10.10}$$

More precisely, by the polynomial differentiation formula, for any $r = 0, 1, \dots, s$,

$$\left| \frac{d^r}{dt^r} f_1(t)^n \right| \leq \beta_r n^r |f_1(t)|^{n-r},$$

and then, by the Newton binomial formula,

$$\begin{aligned} \left| \frac{d^m}{dt^m} [t^l f_1(t)^n] \right| &\leq \sum_{r=0}^m \frac{m!}{r! (m-r)!} |(t^l)^{(r)}| \cdot \beta_{m-r} n^{m-r} |f_1(t)|^{n-(m-r)} \\ &\leq \beta_m n^m l! \sum_{r=0}^{\min(l,m)} \frac{m!}{r! (m-r)!} |t|^{l-r} \cdot |f_1(t)|^{n-(m-r)}. \end{aligned}$$

But $\sup_{|t| \geq 1} |f_1(t)| = \alpha < 1$, so that, by (10.4), for $n \geq n_1 = s + (l + 2)n_0$, one can write

$$\begin{aligned} |f_1(t)|^{n-(m-r)} &= |f_1(t)|^{(l-r+2)n_0} \cdot |f_1(t)|^{n-(m-r)-(l-r+2)n_0} \\ &\leq \left(\frac{I}{I+t^2}\right)^{(l-r+2)/2} \alpha^{n-n_1}. \end{aligned}$$

Hence, just using $\frac{t^2}{I+t^2} \leq 1$, we have

$$\begin{aligned} \left| \frac{d^m}{dt^m} [t^l f_1(t)^n] \right| &\leq \beta_m n^m l! \alpha^{n-n_1} \sum_{r=0}^{\min(m,l)} \frac{m!}{r!(m-r)!} \left(\frac{It^2}{I+t^2}\right)^{(l-r)/2} \frac{I}{I+t^2} \\ &\leq \beta_s (2nI)^m l! \alpha^{n-n_1} \frac{I}{I+t^2}. \end{aligned}$$

This estimate easily implies

$$\left| \frac{d^m}{dt^m} [t^l f_n(t)] \right| \leq C \alpha_1^n \frac{1}{1+t^2}, \quad \text{for } |t| \geq \sqrt{n}, \quad n \geq n_1, \quad (10.11)$$

where the positive constants C and $\alpha_1 < 1$ may depend on m, l , and the distribution of X_1 , but not on t . In particular, the representation (10.10) is quite justified.

The estimate (10.11) also shows that the part of the integral in (10.10) over the region $|t| \geq \sqrt{n}$ decays exponentially fast uniformly over all x . As for the interval $|t| \leq \sqrt{n}$, one may use the bound (10.9) in (10.10), so that eventually

$$\sup_x |x|^m \left| p_n^{(l)}(x) - \varphi_s^{(l)}(x) \right| \leq \frac{\varepsilon_n}{n^{(s-2)/2}}, \quad \varepsilon_n \rightarrow 0.$$

By the same reasons, we obtain a similar bound for the L^2 norm of the right-hand side of (10.10) as a function of x , by applying Plancherel’s formula. \square

11 Behaviour of densities not far from the origin

To study the asymptotic behavior of the Fisher information distance

$$I(Z_n || Z) = \int_{-\infty}^{\infty} \frac{(p_n'(x) + xp_n(x))^2}{p_n(x)} dx,$$

we split the domain of integration into the interval $|x| \leq T_n$ and its complement. Thus, define

$$J_0 = \int_{|x| \leq T_n} \frac{(p_n'(x) + xp_n(x))^2}{p_n(x)} dx$$

and similarly J_1 for the region $|x| > T_n$. If T_n is not too large, the first integral can be treated with the help of Lemma 10.2. Namely, we take

$$T_n = \sqrt{(s - 2) \log n + s \log \log n + \rho_n} \quad (s > 2), \tag{11.1}$$

where $\rho_n \rightarrow \infty$ is a sufficiently slowly growing sequence whose growth is restricted by the decay of the sequence ε_n in (10.5). In other words, $[-T_n, T_n]$ represents an asymptotically largest interval, where we can guarantee that the densities p_n of Z_n are separated from zero, and moreover, $\sup_{|x| \leq T_n} |\frac{p_n(x)}{\varphi(x)} - 1| \rightarrow 0$. To cover the case $s = 2$, one may put $T_n = \sqrt{\rho_n}$, where $T_n \rightarrow \infty$ is a sufficiently slowly growing sequence. With this choice of T_n , an estimation of the integral J_1 can be performed via moderate inequalities.

In this section we focus on J_0 and provide an asymptotic expansion for it with a remainder term which turns out to be slightly better in comparison with the resulting expansion (1.3) of Theorem 1.1.

Lemma 11.1 *Let $s \geq 3$ be an integer. If $I(Z_{n_0}) < \infty$, for some n_0 , then*

$$J_0 = \frac{c_1}{n} + \frac{c_2}{n^2} + \dots + \frac{c_{[(s-2)/2]}}{n^{[(s-2)/2]}} + o\left(\frac{1}{n^{(s-2)/2} (\log n)^{(s-1)/2}}\right),$$

where the coefficients c_j are defined in (1.4).

Proof Let us adopt the convention to write δ_n for any sequence of functions satisfying $|\delta_n(x)| \leq \varepsilon_n n^{-(s-2)/2}$ with $\varepsilon_n \rightarrow 0$, as $n \rightarrow \infty$, at least on the intervals $|x| \leq T_n$. For example, the statement of Lemma 10.2 with $l = 0$ may be written as

$$p_n(x) = (1 + u_s(x))\varphi(x) + \frac{\delta_n}{1 + |x|^s}, \tag{11.2}$$

where

$$u_s(x) = \frac{\varphi_s(x) - \varphi(x)}{\varphi(x)} = \sum_{k=1}^{s-2} \frac{q_k(x)}{\varphi(x)} \frac{1}{n^{k/2}}.$$

Combining the lemma with $l = 0$ and $l = 1$, we obtain another representation

$$p'_n(x) + xp_n(x) = w_s(x) + \frac{\delta_n}{1 + |x|^{s-1}}, \tag{11.3}$$

where

$$w_s(x) = \sum_{k=1}^{s-2} \frac{q'_k(x) + xq_k(x)}{n^{k/2}}.$$

Note that the functions u_s and w_s depend on n as parameter and are getting small for growing n . More precisely, it follows from the definition of q_k that, for all $x \in \mathbf{R}$,

$$\frac{|w_s(x)|}{\varphi(x)} \leq C_s \frac{1 + |x|^{3(s-1)}}{\sqrt{n}} \quad \text{and} \quad |u_s(x)| \leq C_s \frac{1 + |x|^{3(s-2)}}{\sqrt{n}} \tag{11.4}$$

with some constants depending on s and the cumulants of X_1 , only. In particular, for $|x| \leq T_n$ and any prescribed $0 < \varepsilon < \frac{1}{2}$,

$$\frac{|w_s(x)|}{\varphi(x)} < \frac{1}{n^{\frac{1}{2}-\varepsilon}} \quad \text{and} \quad |u_s(x)| < \frac{1}{4} \tag{11.5}$$

with sufficiently large n . In addition, with a properly chosen sequence ρ_n , we have

$$\frac{\delta_n}{T_n^s \varphi(T_n)} < \frac{1}{4}. \tag{11.6}$$

Hence, by Lemma 10.2, $|\frac{p_n(x)}{\varphi(x)} - 1| < \frac{1}{2}$ on the interval $|x| \leq T_n$.

Now, for $|x| \leq T_n$

$$(1 + u_s(x))^{-1} - \left(1 + u_s(x) + \frac{\delta_n}{(1 + |x|^s)\varphi(x)}\right)^{-1} = \frac{\delta_n}{(1 + |x|^s)\varphi(x)},$$

and we obtain from (11.2)

$$\frac{1}{p_n(x)} = \frac{1}{(1 + u_s(x))\varphi(x)} + \frac{\delta_n}{(1 + |x|^s)\varphi(x)^2}.$$

Combining this with (11.3) and using (11.5), we will be lead to

$$\frac{(p'_n(x) + xp_n(x))^2}{p_n(x)} = \frac{w_s(x)^2}{(1 + u_s(x))\varphi(x)} + \sum_{j=1}^5 r_{nj}(x), \quad |x| \leq T_n,$$

where

$$\begin{aligned} r_{n1} &= \frac{w_s(x)}{(1 + |x|^{s-1})\varphi(x)} \delta_n, & r_{n2} &= \frac{w_s(x)^2}{(1 + |x|^s)\varphi(x)^2} \delta_n, \\ r_{n3} &= \frac{w_s(x)}{(1 + |x|^{2s-1})\varphi(x)^2} \delta_n^2, & r_{n4} &= \frac{1}{(1 + |x|^{2s-2})\varphi(x)} \delta_n^2, \\ r_{n5} &= \frac{1}{(1 + |x|^{3s-2})\varphi(x)^2} \delta_n^3. \end{aligned}$$

Here, according to the left inequality in (11.5), the remainder terms $r_{n1}(x)$ and $r_{n2}(x)$ are uniformly bounded on $[-T_n, T_n]$ by $|\delta_n| n^{-1/3}$. A similar bound also holds for

$r_{n3}(x)$, by taking into account (11.6). In addition, integrating by parts, for large n and with some constants (independent of n), we have

$$\begin{aligned} \int_{|x| \leq T_n} |r_{n4}(x)| dx &\leq \frac{C\varepsilon_n}{n^{s-2}} \int_1^{T_n} \frac{1}{x^{2s-2}} e^{x^2/2} dx \\ &\leq \frac{C'\varepsilon_n}{n^{s-2}} \frac{1}{T_n^{2s-1}} e^{T_n^2/2} = o\left(\frac{1}{T_n^{s-1} n^{(s-2)/2}}\right). \end{aligned}$$

With a similar argument, the same o -relation also holds for the integral of $|r_{n5}(x)|$. Thus,

$$\int_{|x| \leq T_n} \frac{(p'_n + xp_n)^2}{p_n} dx = \int_{|x| \leq T_n} \frac{w_s^2}{(1 + u_s)\varphi} dx + o\left(\frac{1}{T_n^{s-1} n^{(s-2)/2}}\right). \tag{11.7}$$

Now, by Taylor’s expansion around zero, in the interval $|u| \leq \frac{1}{4}$ we have

$$\frac{1}{1 + u} = \sum_{k=0}^{s-4} (-1)^k u^k + \theta u^{s-3}, \quad |\theta| < 2$$

(there are no terms in the sum for $s = 3$). Hence, with some $-2 < \theta_n < 2$

$$\int_{|x| \leq T_n} \frac{w_s^2}{(1 + u_s)\varphi} dx = \sum_{k=0}^{s-4} (-1)^k \int_{|x| \leq T_n} w_s^2 u_s^k \frac{dx}{\varphi} + \theta_n \int_{|x| \leq T_n} w_s^2 u_s^{s-3} \frac{dx}{\varphi}.$$

At the expense of a small error, these integrals may be extended to the whole real line. Indeed, for large enough n , by (11.4), we have, for $k = 0, 1, \dots, s - 4$ with some common constant C_s

$$\int_{|x| > T_n} w_s^2 |u_s|^k \frac{dx}{\varphi} \leq \frac{C_s}{n^{(k+2)/2}} \int_{|x| > T_n} (1 + |x|^{(3k+6)(s-1)}) \varphi(x) dx = o\left(\frac{1}{n^{(s-1)/2}}\right).$$

Moreover,

$$\int_{-\infty}^{\infty} w_s^2 |u_s|^{s-3} \frac{dx}{\varphi} = O\left(\frac{1}{n^{(s-1)/2}}\right).$$

Therefore,

$$\int_{|x| \leq T_n} \frac{w_s^2}{(1 + u_s)\varphi} dx = \sum_{k=0}^{s-4} (-1)^k \int_{-\infty}^{\infty} w_s^2 u_s^k \frac{dx}{\varphi} + O\left(\frac{1}{n^{(s-1)/2}}\right).$$

Inserting this in (11.7), we thus arrive at

$$J_0 = \sum_{k=0}^{s-4} (-1)^k \int_{-\infty}^{\infty} w_s^2 u_s^k \frac{dx}{\varphi} + o\left(\frac{1}{T_n^{s-1} n^{(s-2)/2}}\right). \tag{11.8}$$

In the next step, we develop this representation by expressing u_s and w_s in terms of q_k while expanding the sum in (11.8) in powers of $1/\sqrt{n}$ as

$$\sum_{j=2}^{s-2} \frac{a_j}{n^{j/2}} + O\left(\frac{1}{n^{(s-1)/2}}\right).$$

More precisely, here the coefficients are given by

$$a_j = \sum_{k=2}^j (-1)^k \int_{-\infty}^{\infty} (q'_{r_1} + xq_{r_1})(q'_{r_2} + xq_{r_2})q_{r_3} \dots q_{r_k} \frac{dx}{\varphi^{k-1}} \tag{11.9}$$

with summation over all positive solutions (r_1, \dots, r_k) to $r_1 + \dots + r_k = j$. Moreover, when j are odd, the above integrals are vanishing. Indeed, differentiating the equality (10.1) which defines the functions q_k and using the property $H'_n(x) = nH_{n-1}(x)$ ($n \geq 1$), we obtain a similar equality

$$q'_k(x) + xq_k(x) = \varphi(x) \sum (k + 2l) H_{k+2l-1}(x) \frac{1}{r_1! \dots r_k!} \left(\frac{\gamma_3}{3!}\right)^{r_1} \dots \left(\frac{\gamma_{k+2}}{(k+2)!}\right)^{r_k} \tag{11.10}$$

with summation over all non-negative solutions (r_1, \dots, r_k) to $r_1 + 2r_2 + \dots + kr_k = k$, and where $l = r_1 + \dots + r_k$. Hence, the integrand in (11.9) represents a linear combination of the functions of the form

$$H_{r_1+2l_1-1} H_{r_2+2l_2-1} H_{r_3+2l_3} \dots H_{r_k+2l_k} \varphi.$$

Note that here the sum of indices is mod 2 the same as j . We can now apply the following property of the Chebyshev-Hermite polynomials (see [23]). If the sum of indices d_1, \dots, d_k is odd, then necessarily

$$\int_{-\infty}^{\infty} H_{d_1}(x) \dots H_{d_k}(x) \varphi(x) dx = 0.$$

Hence, $a_j = 0$, when j is odd, and putting $c_j = a_{2j}$, we arrive at the assertion of the lemma. □

Remark In formula (11.9) with $c_j = a_{2j}$ we perform summation over all integers $r_l \geq 1$ such that $r_1 + \dots + r_k = 2j$. Hence, all $r_l \leq 2j - 1$, and thus the functions q_{r_l}

are determined by the cumulants up to order $2j + 1$. Hence, c_j represents a polynomial in $\gamma_3, \dots, \gamma_{2j+1}$.

12 Moderate deviations

We now consider the second integral

$$J_1 = \int_{|x|>T_n} \frac{(p'_n(x) + xp_n(x))^2}{p_n(x)} dx$$

participating in the Fisher information distance $I(Z_n||Z)$.

Lemma 12.1 *Let $s \geq 3$ be an integer. If $I(Z_{n_0}) < \infty$, for some n_0 , then*

$$J_1 = o\left(\frac{1}{n^{(s-2)/2}(\log n)^{(s-3)/2}}\right).$$

Proof Write

$$J_1 \leq 2J_{1,1} + 2J_{1,2} = 2 \int_{|x|>T_n} \frac{p'_n(x)^2}{p_n(x)} dx + 2 \int_{|x|>T_n} x^2 p_n(x) dx. \tag{12.1}$$

Using Lemma 10.1, we conclude that, for $s = 3, \dots$,

$$J_{1,2} = o\left(\frac{1}{(n \log n)^{(s-2)/2}}\right). \tag{12.2}$$

Indeed, integrating by parts we have

$$\int_{T_n}^{\infty} x^2 p_n(x) dx = T_n^2 (1 - F_n(T_n)) + 2 \int_{T_n}^{\infty} x(1 - F_n(x)) dx.$$

Recalling the definition of the approximating functions Φ_s , cf. (10.2), and applying an elementary inequality $1 - \Phi(x) < \frac{1}{x} \varphi(x)$ ($x > 0$), we get from (10.3) that

$$\begin{aligned} T_n^2 (1 - F_n(T_n)) &= T_n^2 (1 - \Phi_s(T_n)) + T_n^2 (\Phi_s(T_n) - F_n(T_n)) \\ &\leq T_n \varphi(T_n) + C \varphi(T_n) \sum_{k=1}^{s-2} T_n^{3k} n^{-k/2} + o\left(\frac{1}{T_n^{s-2} n^{(s-2)/2}}\right) \\ &= o\left(\frac{1}{(n \log n)^{(s-2)/2}}\right) \end{aligned}$$

with some constant C . In addition,

$$\int_{T_n}^{\infty} x(1 - F_n(x)) dx \leq 1 - \Phi(T_n) + C \sum_{k=1}^{s-2} \frac{1}{n^{k/2}} \int_{T_n}^{\infty} x^{3k} \varphi(x) dx + o\left(\frac{1}{T_n^{s-2} n^{(s-2)/2}}\right) = o\left(\frac{1}{(n \log n)^{(s-2)/2}}\right).$$

With similar estimates for the half-axis $x < -T_n$, we arrive at the relation (12.2).

Let us now estimate $J_{1,1}$. Denote by $J_{1,1}^+$ the part of this integral corresponding to the interval $x > T_n$. By Propositions 8.2 with $a = T_n$ $b = \infty$, for sufficiently large n we have the formula

$$J_{1,1}^+ = -p'_n(T_n) \log p_n(T_n) - \int_{T_n}^{\infty} p''_n(x) \log p_n(x) dx. \tag{12.3}$$

Since $p_n(x) \leq \sqrt{I(Z_{n_0})}$, for all x (cf. Propositions 2.2 and 9.1) and since, by Lemma 10.2, $p_n(T_n) \geq \frac{1}{2} \varphi(T_n)$, we see that, for all sufficiently large n $|\log p_n(T_n)| \leq cT_n^2$ with some constant c . Therefore, by Lemma 10.2 for the derivative of the density p_n , we get

$$\begin{aligned} |p'_n(T_n) \log p_n(T_n)| &\leq cT_n^2 |p'_n(T_n)| \\ &\leq cT_n^2 |\varphi'(T_n)| + o\left(\frac{1}{T_n^{s-2} n^{(s-2)/2}}\right) = o\left(\frac{1}{T_n^{s-3} n^{(s-2)/2}}\right). \end{aligned} \tag{12.4}$$

A similar relation holds at the point $-T_n$, as well.

It remains to evaluate the integral in (12.3). First we integrate over the set $A = \{x > T_n : p_n(x) \leq \varphi(x)^4\}$. By the upper bound of Proposition 6.4 and applying Proposition 9.1 once more, we have, for all x and all sufficiently large n ,

$$|p''_n(x)| \leq I(p_n)^{5/4} \sqrt{p_n(x)} \leq I(Z_{n_0})^{5/4} \sqrt{p_n(x)}.$$

Hence, with some constants c, c'

$$\begin{aligned} \int_A |p''_n(x) \log p_n(x)| dx &\leq c \int_A \sqrt{p_n(x)} |\log p_n(x)| dx \\ &\leq c' \int_{T_n}^{\infty} x^2 \varphi(x)^2 dx = o\left(\frac{1}{n^{s-2}}\right). \end{aligned}$$

On the other hand, for the complementary set $B = (T_n, \infty) \setminus A$, we have

$$\int_B |p_n''(x) \log p_n(x)| dx \leq c \int_B x^2 |p_n''(x)| dx. \tag{12.5}$$

We now apply Lemma 10.2 to approximate the second derivative. It yields

$$\int_{T_n}^{+\infty} x^2 |p_n''(x)| dx \leq \int_{T_n}^{+\infty} x^2 |\varphi_s''(x)| dx + \int_{T_n}^{\infty} \frac{|\psi_{2,n}(x)|}{1 + |x|^{s-2}} dx \cdot o\left(\frac{1}{n^{(s-2)/2}}\right).$$

Here, the first integral on the right-hand side is bounded by

$$\int_{T_n}^{\infty} x^2 |\varphi_s''(x) - \varphi''(x)| dx + \int_{T_n}^{\infty} x^2 |x^2 - 1| \varphi(x) dx = o\left(\frac{1}{T_n^{s-3} n^{(s-2)/2}}\right).$$

To estimate the second integral, we use Cauchy’s inequality, which gives

$$\int_{T_n}^{\infty} \frac{1}{1 + |x|^{s-2}} |\psi_{2,n}(x)| dx \leq \frac{1}{T_n^{s-5/2}} \left(\int_{-\infty}^{\infty} \psi_{2,n}(x)^2 dx \right)^{1/2} \leq \frac{1}{T_n^{s-5/2}}.$$

Therefore, returning to (12.5), we get

$$\int_B |p_n''(x) \log p_n(x)| dx = o\left(\frac{1}{n^{(s-2)/2} (\log n)^{(s-3)/2}}\right).$$

Together with the bound for the integral over the set A , we thus have

$$J_{1,1}^+ = o\left(\frac{1}{n^{(s-2)/2} (\log n)^{(s-3)/2}}\right).$$

The part of the integral $J_{1,1}$ taken over the axis $x < -T_n$ admits a similar bound, hence the lemma is proved. □

The statement of Theorem 1.1 in case $s \geq 3$ thus follows from Lemmas 11.1 and 12.1.

13 Theorem 1.1 in the case $s = 2$ and Corollary 1.2

In the most general case $s = 2$ the proof of Theorem 1.1 does not need Edgeworth-type expansions. With tools developed in the previous sections the argument is straightforward and may be viewed as an alternative approach to Barron–Johnson’s theorem.

Proof of Theorem 1.1 (case $s = 2$) Once the Fisher information $I(Z_{n_0})$ is finite, the normalized sums Z_n with $n \geq 2n_0$ have uniformly bounded densities p_n with bounded continuous derivatives p'_n (Proposition 6.2). Moreover, we have a well-known local limit theorem for densities; we described one of its variants in Lemma 10.2. In particular,

$$\sup_x (1 + x^2) |p_n(x) - \varphi(x)| = o(1), \quad (13.1)$$

$$\sup_x (1 + x^2) |p'_n(x) - \varphi'(x)| = o(1), \quad (13.2)$$

as $n \rightarrow \infty$, where the convergence of the derivatives relies upon the finiteness of the Fisher information.

Splitting the integration in

$$I(Z_n|Z) = \int_{-\infty}^{\infty} \frac{(p'_n(x) + xp_n(x))^2}{p_n(x)} dx$$

into the two regions, we have therefore, for every fixed $T > 1$,

$$J_0 = \int_{|x| \leq T} \frac{(p'_n(x) + xp_n(x))^2}{p_n(x)} dx = o(1), \quad n \rightarrow \infty. \quad (13.3)$$

On the other hand, write as we did before

$$\begin{aligned} J_1 &= \int_{|x| > T} \frac{(p'_n(x) + xp_n(x))^2}{p_n(x)} dx \leq 2J_{1,1} + 2J_{1,2} \\ &= 2 \int_{|x| > T} \frac{p'_n(x)^2}{p_n(x)} dx + 2 \int_{|x| > T} x^2 p_n(x) dx. \end{aligned}$$

As we saw in (12.3),

$$J_{1,1} = -p'_n(T) \log p_n(T) + p'_n(-T) \log p_n(-T) - \int_{|x| > T} p''_n(x) \log p_n(x) dx.$$

By (13.1)–(13.2), $|p'_n(\pm T) \log p_n(\pm T)| \leq 2T^3 e^{-T^2/2}$ for all sufficiently large $n \geq n_T$. By Proposition 8.3, with some constant c , for all x ,

$$u |p''_n(x) \log p_n(x)| \leq c \frac{\log(e + |x|)}{1 + x^2},$$

implying

$$\int_{|x|>T} |p_n''(x) \log p_n(x)| dx \leq c'T^{-1/2}$$

with some other constant c' . In addition, by (13.1),

$$\begin{aligned} \int_{|x|>T} x^2 p_n(x) dx &= \int_{|x|>T} x^2(p_n(x) - \varphi(x)) dx + \int_{|x|>T} x^2 \varphi(x) dx \\ &= - \int_{|x|\leq T} x^2(p_n(x) - \varphi(x)) dx + \int_{|x|>T} x^2 \varphi(x) dx \\ &\leq \int_{|x|\leq T} x^2 |p_n(x) - \varphi(x)| dx + \int_{|x|>T} x^2 \varphi(x) dx \leq 2T^3 o(1) + 4T\varphi(T). \end{aligned}$$

Hence, given $\varepsilon > 0$, one can choose T such that $J_1 < \varepsilon$, for all n large enough. This means that $J_1 = o(1)$, and recalling (13.3), we get $I(Z_n||Z) = o(1)$. \square

Let us now return to the case $s \geq 3$.

Proof of Corollary 1.2 According to the expansion (11.8) which appeared in the proof of Lemma 11.1, Theorem 1.1 may equivalently be formulated as

$$I(Z_n||Z) = \sum_{l=0}^{s-4} (-1)^l \int_{-\infty}^{\infty} w_s(x)^2 u_s(x)^l \frac{dx}{\varphi(x)} + o\left(\frac{1}{n^{(s-2)/2} (\log n)^{(s-3)/2}}\right), \tag{13.4}$$

where as before

$$w_s(x) = \sum_{j=1}^{s-2} (q'_j(x) + xq_j(x)) n^{-j/2}, \quad u_s(x) = \sum_{j=1}^{s-2} \frac{q_j(x)}{\varphi(x)} n^{-j/2}.$$

This representation for the Fisher information distance is more convenient for applications such as Corollary 1.2 in comparison with (1.3). Assume that $s \geq 4$ and $\gamma_3 = \dots = \gamma_{k-1} = 0$ for a given integer $3 \leq k \leq s$ (with no restriction when $k = 3$). Then, by the definition (10.2), $q_1 = \dots = q_{k-3} = 0$, so

$$w_s(x) = \sum_{j=k-2}^{s-2} (q'_j(x) + xq_j(x)) n^{-j/2}, \quad u_s(x) = \sum_{j=k-2}^{s-2} \frac{q_j(x)}{\varphi(x)} n^{-j/2}. \tag{13.5}$$

Hence, in order to isolate the leading term in (1.3) with the smallest power of $1/n$, one should take $l = 0$ in (13.4) and $j = k - 2$ in the first sum of (13.5). This gives

$$I(Z_n||Z) = n^{-(k-2)} \int_{-\infty}^{\infty} (q'_{k-2}(x) + xq_{k-2}(x))^2 \frac{dx}{\varphi(x)} + O\left(n^{-(k-1)}\right) + o\left(\frac{1}{n^{(s-2)/2} (\log n)^{(s-3)/2}}\right).$$

Now, again according to (10.2), or as found in (11.10),

$$q'_{k-2}(x) + xq_{k-2}(x) = \frac{\gamma_k}{(k-1)!} H_{k-1}(x) \varphi(x).$$

Therefore, the sum in (1.3) will contain powers of $1/n$ starting from $1/n^{k-2}$ with leading coefficient

$$c_{k-2} = \frac{\gamma_k^2}{(k-1)!^2} \int_{-\infty}^{\infty} H_{k-1}(x)^2 \varphi(x) dx = \frac{\gamma_k^2}{(k-1)!}.$$

Thus, $c_1 = \dots = c_{k-3} = 0$ and we get

$$I(Z_n||Z) = \frac{\gamma_k^2}{(k-1)!} \frac{1}{n^{k-2}} + O\left(n^{-(k-1)}\right) + o\left(\frac{1}{n^{(s-2)/2} (\log n)^{(s-3)/2}}\right).$$

□

14 Extensions to non-integer s . Lower bounds

If $s \geq 2$ is not necessary integer, put $m = [s]$ (integer part). Theorem 1.1 admits the following generalization. As before, let the normalized sums

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

be defined for independent identically distributed random variables with mean $\mathbf{E}X_1 = 0$ and variance $\text{Var}(X_1) = 1$.

Theorem 14.1 *If $I(Z_{n_0}) < \infty$ for some n_0 , and $\mathbf{E}|X_1|^s < \infty$ ($s > 2$), then*

$$I(Z_n||Z) = \frac{c_1}{n} + \frac{c_2}{n^2} + \dots + \frac{c_{\lfloor (s-2)/2 \rfloor}}{n^{\lfloor (s-2)/2 \rfloor}} + o\left(\frac{1}{n^{(s-2)/2} (\log n)^{(s-3)/2}}\right), \tag{14.1}$$

where the coefficients c_j are the same as in (1.4).

The proof is based on a certain extension and refinement of the local limit theorem described in Lemma 10.2.

Lemma 14.2 *Assume that $I(Z_{n_0}) < \infty$ for some n_0 , and let $\mathbf{E} |X_1|^s < \infty$ ($s \geq 2$). Fix $l = 0, 1, \dots$. Then for all n large enough, Z_n have densities p_n of class C^l satisfying, as $n \rightarrow \infty$,*

$$(1 + |x|^m) \left(p_n^{(l)}(x) - \varphi_m^{(l)}(x) \right) = \psi_{l,n}(x) o(n^{-(s-2)/2}), \quad m = [s], \quad (14.2)$$

uniformly for all x , with $\sup_x |\psi_{l,n}(x)| \leq 1$ and $\int_{-\infty}^{\infty} \psi_{l,n}(x)^2 dx \leq 1$. Moreover,

$$(1 + |x|^s) \left(p_n^{(l)}(x) - \varphi_m^{(l)}(x) \right) = \psi_{l,n,1}(x) o(n^{-(s-2)/2}) + (1 + |x|^{s-m}) \psi_{l,n,2}(x) \left(O(n^{-(m-1)/2}) + o(n^{-(s-2)}) \right), \quad (14.3)$$

uniformly for all x , where $\sup_x |\psi_{l,n,j}(x)| \leq 1$ and $\int_{-\infty}^{\infty} \psi_{l,n,j}(x)^2 dx \leq 1$ ($j = 1, 2$).

Here we use the approximating functions $\varphi_m = \varphi + \sum_{k=1}^{m-2} q_k n^{-k/2}$ as before.

When $l = 0$ and in a simpler form, namely, with $\psi_{l,s,j}(x, n) = 1$, this result has recently been obtained in [9, 10]. In this case, the finiteness of the Fisher information may be relaxed to the boundedness of the densities. The more general case involving derivatives can be carried out by a similar analysis as that developed in [9, 10], so we omit details.

If $s = m$ is integer, the Edgeworth-type expansions (14.2) and (14.3) coincide, and we are reduced to the statement of Lemma 10.2. However, if $s > m$, (14.3) gives an improvement over (14.2) on relatively large intervals such as $|x| \leq T_n$ defined in (11.1).

Proof of Theorem 14.1 With a few modifications one can argue in the same way as we did in the proof of Theorem 1.1. First, in case $l = 0$ (14.3) yields, uniformly in $|x| \leq T_n$

$$p_n(x) = \varphi_m(x) + \frac{1}{1 + |x|^s} o\left(n^{-(s-2)/2}\right),$$

which being combined with a similar relation for the derivative ($l = 1$) yields

$$p_n'(x) + xp_n(x) = w_m(x) + \frac{1}{1 + |x|^{s-1}} o\left(n^{-(s-2)/2}\right),$$

where $w_m(x) = \sum_{k=1}^{m-2} (q_k'(x) + xq_k(x)) n^{-k/2}$. These two relations thus extend (11.2) and (11.3) which were only needed in the proof of Lemma 11.1. Repeating the same arguments using the functions $u_m(x) = \frac{\varphi_m(x) - \varphi(x)}{\varphi(x)}$, we can extend the expansion of Lemma 11.1 with the same remainder term to general values $s > 2$.

In order to prove Lemma 12.1 with real $s > 2$, let us return to (12.1). The fact that the relation (12.2) extends to non-integer s follows from the extended variant of Lemma 10.1, which was already mentioned before. Thus our main concern has to be the integral $J_{1,1}$ which is responsible for the most essential contribution in the resulting remainder term. Thus, consider the part of this integral on the positive half-axis

$$J_{1,1}^+ = \int_{T_n}^{\infty} \frac{p'_n(x)^2}{p_n(x)} dx = -p'_n(T_n) \log p_n(T_n) - \int_{T_n}^{\infty} p''_n(x) \log p_n(x) dx. \tag{14.4}$$

Applying (14.3) at $x = T_n$, we obtain (12.4) for real $s > 2$, that is,

$$|p'_n(T_n) \log p_n(T_n)| = o\left(\frac{1}{n^{(s-2)/2} (\log n)^{(s-3)/2}}\right).$$

To prove (14.1), it remains to estimate the last integral in (14.4) which has to be treated with an extra care. The argument uses both (14.2) and (14.3) which are applied on different parts of the half-axis $x > T_n$. For the set $A = \{x \geq T_n : p_n(x) \leq \varphi(x)^4\}$ we have already obtained a general relation

$$\int_A |p''_n(x) \log p_n(x)| dx = o\left(\frac{1}{n^{s-2}}\right),$$

which holds for all sufficiently large n (without any moment assumption). Hence, with some constant c

$$\int_{T_n}^{4T_n^4} |p''_n(x) \log p_n(x)| dx \leq c \int_{T_n}^{4T_n^4} x^2 |p''_n(x)| dx + o\left(\frac{1}{n^{s-2}}\right). \tag{14.5}$$

Now, on the interval $[T_n, 4T_n^4]$ we apply Lemma 14.2 with $l = 2$ to approximate the second derivative. It yields

$$\begin{aligned} \int_{T_n}^{4T_n^4} x^2 |p''_n(x)| dx &\leq \int_{T_n}^{\infty} x^2 |\varphi''_m(x)| dx + \int_{T_n}^{4T_n^2} \frac{|\psi_{2,n,1}(x)|}{1+|x|^{s-2}} dx \cdot o\left(\frac{1}{n^{(s-2)/2}}\right) \\ &+ \int_{T_n}^{4T_n^4} \frac{1}{1+|x|^{m-2}} |\psi_{2,n,2}(x)| dx \cdot \left(O(n^{-(m-1)/2}) + o(n^{-(s-2)})\right). \end{aligned}$$

Here, as in the proof of Lemma 12.1, the first integral on the right-hand side is bounded, up to a constant, by

$$\int_{T_n}^{+\infty} x^4 \varphi(x) dx = o\left(\frac{1}{T_n^{s-3} n^{(s-2)/2}}\right),$$

and for the second one, we use Cauchy’s inequality to estimate it by $T_n^{-(s-5/2)}$. Similarly, the last integral is bounded by

$$2T_n^2 \left(\int_{-\infty}^{\infty} \psi_{2,n,2}(x)^2 dx \right)^{1/2} \leq 2T_n^2.$$

Since T_n^2 has a logarithmic growth, we conclude that

$$\int_{T_n}^{4T_n^4} x^2 |p_n''(x)| dx = o\left(\frac{1}{n^{(s-2)/2} (\log n)^{(s-3)/2}}\right),$$

so a similar bound also holds for the left integral in (14.5).

To deal with the remaining values of x , we will consider the set $S_1 = \{x > 4T_n^4 : p_n(x) \leq \frac{1}{2} e^{-4\sqrt{x}}\}$ and its complement $S_2 = (4T_n^4, \infty) \setminus S_1$. By Proposition 6.3, for all sufficiently large n , and with some constants c, c' we have

$$\begin{aligned} \int_{S_1} |p_n''(x) \log p_n(x)| dx &\leq c \int_{S_1} \sqrt{p_n(x)} |\log p_n(x)| dx \\ &\leq c' \int_{4T_n^4}^{\infty} \sqrt{x} e^{-2\sqrt{x}} dx = o\left(\frac{1}{n^{s-2}}\right). \end{aligned}$$

On the other hand, applying (14.2) on the set S_2 , we get

$$\begin{aligned} \int_{S_2} |p_n''(x) \log p_n(x)| dx &\leq c \int_{S_2} |p_n''(x)| \sqrt{x} dx \\ &\leq c' \int_{4T_n^4}^{\infty} x^{5/2} \varphi(x) dx + c' \int_{4T_n^4}^{\infty} \frac{dx}{x^{m-1/2}} \cdot o\left(\frac{1}{n^{(s-2)/2}}\right) \\ &= o\left(\frac{1}{T_n^{2(2m-3)} n^{(s-2)/2}}\right). \end{aligned}$$

Combining the two estimates, the theorem is proved. □

Remark If $2 < s < 4$, the expansion (14.1) becomes

$$I(Z_n||Z) = o\left(\frac{1}{n^{(s-2)/2}(\log n)^{(s-3)/2}}\right). \quad (14.6)$$

This formulation does not include the case $s = 2$. In case $s > 2$, we expect that the bound (14.6) may be improved further. However, a possible improvement may concern the power of the logarithmic term, only. This can be illustrated by means of the example of densities of the form

$$p(x) = \int_{\sigma_0}^{\infty} \varphi_{\sigma}(x) dP(\sigma) \quad (x \in \mathbf{R}),$$

that is, mixtures of densities of normal distributions on the line with mean zero, where P is a (mixing) probability measure supported on the half-axis (σ_0, ∞) with $\sigma_0 > 0$. The variance constraint on P is that

$$\int_{-\infty}^{\infty} x^2 p(x) dx = \int_{\sigma_0}^{\infty} \sigma^2 dP(\sigma) = 1, \quad (14.7)$$

so we should assume that $0 < \sigma_0 < 1$.

First, let us note that, by the convexity of the Fisher information,

$$I(p) \leq \int_{\sigma_0}^{\infty} I(\varphi_{\sigma}) dP(\sigma) = \int_{\sigma_0}^{\infty} \frac{1}{\sigma^2} dP(\sigma) \leq \frac{1}{\sigma_0^2},$$

hence, $I(p)$ is finite. On the other hand, given $\eta > s/2$, it is possible to construct the measure P to satisfy (14.7) and with

$$D(Z_n||Z) \geq \frac{c}{n^{(s-2)/2}(\log n)^{\eta}},$$

for all n large enough, and with a constant c depending on s and η , only (cf. [11]). For example, one may define P on the half-axis $[2, \infty)$ by its density

$$\frac{dP(\sigma)}{d\sigma} = \frac{c}{\sigma^{s+1}(\log \sigma)^{\eta}}, \quad \sigma > 2,$$

and then extend it to any interval $[\sigma_0, 2]$ in an arbitrary way so that to obtain a probability measure satisfying the requirement (14.7). Hence, (14.6) is sharp up to a logarithmic factor.

Finally, let us mention that in case $s = 2$ $D(Z_n||Z)$ and therefore $I(Z_n||Z)$ may decay at an arbitrary slow rate.

Acknowledgments The authors would like to thank the referee for careful reading of the manuscript and useful comments.

References

1. Artstein, S., Ball, K.M., Barthe, F., Naor, A.: On the rate of convergence in the entropic central limit theorem. *Probab. Theory Relat. Fields* **129**(3), 381–390 (2004)
2. Artstein, S., Ball, K.M., Barthe, F., Naor, A.: Solution of Shannon’s problem on the monotonicity of entropy. *J. Am. Math. Soc.* **17**(4), 975–982 (2004)
3. Barron, A.R., Johnson, O.: Fisher information inequalities and the central limit theorem. *Probab. Theory Relat. Fields* **129**(3), 391–409 (2004)
4. Ball, K.M., Barthe, F., Naor, A.: Entropy jumps in the presence of a spectral gap. *Duke Math. J.* **119**(1), 41–63 (2003)
5. Bhattacharya, R.N., Ranga Rao, R.: *Normal Approximation and Asymptotic Expansions*. Wiley, London (1976) (Soc. for Industrial and Appl. Math., Philadelphia 2010)
6. Billingsley, P.: *Convergence of Probability Measures*. Wiley, New York, p. xii+253 (1968)
7. Blachman, N.M.: The convolution inequality for entropy powers. *IEEE Trans. Inform. Theory* **11**, 267–271 (1965)
8. Bobkov, S.G.: Large deviations and isoperimetry over convex probability measures with heavy tails. *Electr. J. Probab.* **12**, 1072–1100 (2007)
9. Bobkov, S.G., Chistyakov, G.P., Götze, F.: Non-uniform bounds in local limit theorems in case of fractional moments. I. *Math. Methods Stat.* **20**(3), 171–191 (2011)
10. Bobkov, S.G., Chistyakov, G.P., Götze, F.: Non-uniform bounds in local limit theorems in case of fractional moments. II. *Math. Methods Stat.* **20**(4), 269–287 (2011)
11. Bobkov, S.G., Chistyakov, G.P., Götze, F.: Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. *Ann. Probab.* (to appear)
12. Bolley, F., Villani, C.: Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Ann. Fac. Sci. Toulouse Math.* (6) **14**(3), 331–352 (2005)
13. Burago, Yu.D., Zalgaller V.A.: *Geometric inequalities*. Springer Series in Soviet Mathematics (trans: Sosinskii A.B.). Springer, Berlin, p. xiv+331 (1988)
14. Cohen, M.: The Fisher information and convexity. *IEEE Trans. Inform. Theory* **14**, 591–592 (1968)
15. Johnson, O.: *Information Theory and the Central Limit Theorem*. Imperial College Press, London, 9 p. xiv+20 (2004)
16. Kolmogorov, A.N., Fomin S.V.: *Elements of the Theory of Functions and Functional Analysis*. With a Supplement “Banach algebras”, by V. M. Tikhomirov. (Russian), 6th edition. “Nauka”, Moscow, 6 p. 24 (1989)
17. Maurey, B.: Personal communication
18. Meyer, P.-A.: *Probability and Potentials*. Blaisdell Publishing Co/Ginn and Co., Waltham/London. p. xiii+266 (1966)
19. Osipov, L.V., Petrov, V.V.: On the estimation of the remainder term in the central limit theorem (Russian). *Theory Probab. Appl.* **12**, 322–329 (1967)
20. Petrov, V.V.: *Sums of Independent Random Variables*. Springer, New York, p. x+345 (1975)
21. Shimizu, R.: On Fisher’s amount of information for location family. In: Patil, G.P. et al. (eds), *Statistical Distributions in Scientific Work*, Vol. 3, Reidel, Dordrecht, pp. 305–312 (1975)
22. Stam, A.J.: Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Control* **2**, 101–112 (1959)
23. Szegő, G.: *Orthogonal polynomials*, 3rd edn. Amer. Math. Soc. Colloquium Publications, vol. 23. Amer. Math. Soc., Providence, RI, p. xiii+423 (1967)
24. Zhang, Z.: Inequalities for characteristic functions involving Fisher information. *C. R. Math. Acad. Sci. Paris* **344**(5), 327–330 (2007)