FISHER-TYPE INFORMATION INVOLVING HIGHER ORDER DERIVATIVES

SERGEY G. BOBKOV[1]

ABSTRACT. Basic general properties are considered for the Fisher-type information involving higher order derivatives. They are used to explore various properties of probability densities and derive Stam-type inequalities.

1. **Introduction**

Given a random variable X with an absolutely continuous density f, the Fisher information hidden in the distribution of X is defined by

$$I(X) = \mathbb{E}\,\rho(X)^2 = \int_{-\infty}^{\infty} \frac{f'(x)^2}{f(x)}\,dx, \tag{1.1}$$

where the integration may be restricted to the set of points where $f(x) > 0$. Here, $\rho = f'/f$ represents the logarithmic derivative of f, which is also called the score function (often being taken with the minus sign). Since $f(X) > 0$ almost surely, the random variable $\rho(X)$, called the score of X, is well-defined and finite with probability one.

The functional (1.1) has two natural generalizations motivated by various problems in different fields. In particular, one is interested in the behaviour of absolute moments of the scores

$$I_p(X) = \mathbb{E}\,|\rho(X)|^p = \int_{-\infty}^{\infty} \frac{|f'(x)|^p}{f(x)^{p-1}}\,dx, \quad p \geq 1. \tag{1.2}$$

As a partial case, the first absolute moment $I_1(X) = \|f\|_{\rm TV}$ describes the total variation norm of the density function f. Another closely related functional defined for positive integers p is

$$I^{(p)}(X) = \mathbb{E}\,\rho_p(X)^2 = \int_{f(x)>0} \frac{f^{(p)}(x)^2}{f(x)}\,dx. \tag{1.3}$$

Here $\rho_p = f^{(p)}/f$ may be viewed as the "p-th order" score function.

These functionals were introduced by Lions and Toscani [10] in their study of convergence of densities (and of their powers) in Sobolev spaces towards the central limit theorem. Previously, the functional I_4 was also considered by Gabetta [7] in the context of the kinetic theory of gases to study the convergence to equilibrium in Kac's model. In paper [2], the moments of the scores together with exponential and Gaussian moments of $\rho(X)$ appear with the aim to control the translates of product probability measures. See also [3] and [4] for various upper bounds on the Fisher information and moments of the scores.

2010 *Mathematics Subject Classification.* Primary 60E, 60F.
Key words and phrases. Fisher information, Stam-type inequalities.
1) School of Mathematics, University of Minnesota, Minneapolis, MN, USA .

The quantity $I^{(p)}(X)$ may be called the Fisher(-type) information of order p. Denote by $\mathfrak{C}^p$ the collection of all continuous functions f on the real line which have continuous derivatives up to order $p-1$, such that $f^{(p-1)}$ is (locally) absolutely continuous. We denote by $f^{(p)}$ a Radon-Nikodym derivative of $f^{(p-1)}$ which is defined and finite almost everywhere.

Definition 1.1. If the random variable X has a density f from the class $\mathfrak{C}^p$ for an integer $p \geq 1$, the Fisher information $I^{(p)}(X) = I^{(p)}(f)$ of order p is defined by (1.3). In all other cases, put $I^{(p)}(X) = \infty$.

Since $f^{(0)} = f$, it is natural to put $I^{(0)}(X) = 1$.

In this paper we explore general properties of the functional $I^{(p)}(X)$ and its relationship to various properties of densities f. Many of them extend and sharpen corresponding properties obtained under the hypothesis that the classical Fisher information $I(X)$ is finite. These properties include the integrability of the first p derivatives of f and assertions about their decay at infinity under moment assumptions posed on X. This will allow us to consider the relative Fisher-type information with respect to the standard normal distribution and to prove, for example, the following comparison. In the sequel, we use the notation $Z \sim N(a,\sigma^2)$ for the case where the random variable Z is normal with mean a and variance σ^2.

Theorem 1.2. *Let $I^{(p)}(X)$ be finite for an integer $p \geq 1$. Then, for $Z \sim N(0,1)$,*

$$\mathbb{E}\, H_p(X)^2 = \mathbb{E}\, H_p(Z)^2 \ \Rightarrow \ I^{(p)}(X) \geq I^{(p)}(Z). \tag{1.4}$$

Here and below H_p denotes the Chebyshev-Hermite polynomial of degree p with a leading coefficient 1 (let us note that the moment $\mathbb{E}X^{2p}$ should be finite as well). In the case $p=1$, (1.4) recovers a well-known statement that the Fisher information $I(X)$ is minimized for the normal distribution when the variance is fixed.

One interesting question which we partly address is: How can one compare $I^{(p)}(X)$ for different p? For example, in the case of moments of the scores defined as in (1.2), the L^p-norms $p \rightarrow I_p(X)^{1/p}$ are non-decreasing. However, it may occur that the Fisher-type information is finite for a given odd order $p \geq 3$, while $I^{(q)}(X)$ are infinite for all even $q < p$ (cf. Example 2.5 below). Nevertheless, using the so-called isoperimetric profiles, one can derive the following relations for the case $p = 2$.

Theorem 1.3. *For any random variable X,*

$$I^{(2)}(X) \geq \frac{1}{3}\, I_4(X) \geq \frac{1}{3}\, I(X)^2. \tag{1.5}$$

Thus, the finiteness of $I^{(2)}(X)$ guarantees the finiteness of the usual Fisher information.

Part of the proof of Theorem 1.3 is based on the lower semi-continuity of the Fisher-type information with respect to the weak convergence, as well as on the convexity of this functional in the space of probability distributions on the real line. These two important properties reduce many relations such as (1.5) to the case where X has a C^∞-smooth positive density, by means of the following continuity property.

Theorem 1.4. *For all independent random variables X and Z,*

$$\lim_{\varepsilon\to 0} I^{(p)}(X+\varepsilon Z) = I^{(p)}(X). \tag{1.6}$$

In particular, if the distribution of X is not absolutely continuous, then $I^{(p)}(X+\varepsilon Z)\to\infty$ regardless of whether or not Z has a smooth density.

If $Z \sim N(0,1)$, then $I^{(p)}(X+\varepsilon Z)$ is finite for any $\varepsilon > 0$, and the convergence in (1.6) is monotone in ε. Hence, this equality may be taken as an equivalent definition of $I^{(p)}(X)$, which was actually proposed in [10].

The property (1.6) can be also used to study in full generality various generalizations of the classical Stam inequality ([11], [6], [8])

$$\frac{1}{I(X+Y)} \geq \frac{1}{I(X)} + \frac{1}{I(Y)}. \tag{1.7}$$

In particular, we have:

Theorem 1.5. *Given independent random variables X and Y, for all $k = 1,\dots,p-1$, $p\geq 2$,*

$$\frac{1}{I^{(p)}(X+Y)} \geq \frac{1}{I^{(p)}(X)} + \frac{1}{I^{(p)}(Y)} + \frac{1}{I^{(k)}(X)I^{(p-k)}(Y)}. \tag{1.8}$$

In the case $p = 2$, the family (1.8) contains only one inequality, in which an equality is attained for the class of normal distributions similarly to (1.7).

Thus, (1.7) is satisfied for all $I^{(p)}$ in place of I. Another immediate consequence of (1.8) is that the finiteness of $I^{(k)}(X)$ and $I^{(p-k)}(Y)$ with $1\leq k\leq p-1$ guarantees the finiteness of $I^{(p)}(X+Y)$ in view of the following immediate consequence from (1.8)

$$I^{(p)}(X+Y) \leq I^{(k)}(X)I^{(p-k)}(Y).$$

By induction, it also follows that

$$I^{(p)}(X_1+\dots+X_p) \leq I(X_1)\dots I(X_p)$$

whenever the random variables $X_1,\dots,X_p$ are independent. In this connection, let us recall that the convolution of 3 probability densities with a finite total variation norm has a finite Fisher information ([4, 5]). Hence, the sum of $3p$ independent random variables whose densities are functions of bounded total variation has a finite Fisher-type information of order p.

In the proof of (1.8), we follow the argument by Lions and Toscani [10]. However, in Lemma 2.3 they mistakenly state a Stam-type inequality for the functional $I^{(p)}$ as a sharper relation

$$I^{(p)}(X+Y) \leq \sum_{k=0}^{p} \alpha_k^2\, I^{(k)}(X)I^{(p-k)}(Y)$$

with arbitrary $\alpha_i\geq 0$ such that $\alpha_0+\dots+\alpha_p = 1$. Optimizing over the coefficients α_i, it is equivalent to

$$\frac{1}{I^{(p)}(X+Y)} \geq \sum_{k=0}^{p} \frac{1}{I^{(k)}(X)I^{(p-k)}(Y)}. \tag{1.9}$$

Testing this inequality in the class of the Gamma distributions with $p = 3$, we had come to the conclusion that it was not correct in general. Nevertheless, one can give simple sufficient conditions for the validity of (1.9), including the case where one of the summands is normal.

Theorem 1.6. *Let X and Y be independent random variables, and let X have a normal distribution. Then* (1.9) *holds true.*

A more general sufficient condition for (1.9) to hold is that $I^{(k)}(X)$ is finite for any $k \leq p$, and that the density f of X satisfies

$$\int_{-\infty}^{\infty} \frac{f^{(k)}(x) f^{(l)}(x)}{f(x)}\, dx = 0,$$

whenever $k \neq l$ $(1 \leq k, l \leq p-1)$. In the standard Gaussian case, this property means the orthogonality of the Chebyshev-Hermite polynomials in L^2 over the Gaussian measure.

We start with several examples illustrating the Fisher-type information and then discuss basic properties of densities assuming that $I^{(p)}(X)$ is finite (Sections 2-5). A more general form of Theorem 1.2 is presented in Section 6. Sections 7-9 contain detailed arguments towards the lower semi-continuity and convexity of this functional, with proof of Theorem 1.4. Sections 10-11 are aimed at proving Theorem 1.3, and the remaining Sections 12-14 deal with the Stam-type inequalities. We use the following plan.

2. **Examples**

It is useful to keep in mind that the functional $I^{(p)}$ is shift invariant and homogeneous of order $-2p$ with respect to X, that is,

$$I^{(p)}(a + bX) = b^{-2p}\, I^{(p)}(X), \quad a \in \mathbb{R},\ b \neq 0.$$

Example 2.1. If $Z \sim N(0,1)$, then $I(Z) = 1$. The standard normal density

$$f(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$$

of Z has derivatives $f^{(p)}(x) = (-1)^p\, H_p(x)\varphi(x)$. Hence $\rho_p(x) = (-1)^p\, H_p(x)$ and

$$I^{(p)}(Z) = \mathbb{E}\, H_p(Z)^2 = p!$$

More generally, if $X \sim N(a, \sigma^2)$ with parameters $a \in \mathbb{R}$ and $\sigma > 0$, then $I^{(p)}(X) = p!\, \sigma^{-2p}$.

Example 2.2. Let X have a beta distribution with parameters $\alpha, \beta > 0$, i.e. with density

$$f(x) = \frac{1}{B(\alpha,\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1.$$

Near zero $\frac{f^{(p)}(x)^2}{f(x)} \sim \mathrm{const} \cdot x^{\alpha-2p-1}$ which is integrable in a neighborhood of zero, if and only if $\alpha > 2p$. In this case, the derivatives are continuous at zero for all $k = 0, 1, \dots, p-1$. A similar conclusion is true about the point $x = 1$, and we conclude that

$$I^{(p)}(X) < \infty \iff \min(\alpha, \beta) > 2p.$$

Example 2.3. Suppose that the random variable X has an even positive density f on the real line, which is C^∞-smooth and such that

$$f(x) = cx^{-q}, \quad x \geq 1,$$

with parameter $q > 1$ for some constant $c > 0$. In this case $f^{(p)}(x) = c_1\, x^{-q-p}$ for $x \geq 1$, where $c_1 \neq 0$ does not depend on x. Hence $I^{(p)}(X) < \infty$ for all integers $p \geq 1$.

Example 2.4. If X has density $f(x) = xe^{-x^2/2}$ supported on the half-axis $x > 0$, then $f'(x) = (1-x^2)\, e^{-x^2/2}$ and $f''(x) = (x^3 - 3x)\, e^{-x^2/2}$. Hence $I(X) = \infty$, while

$$\int_0^\infty \frac{f''(x)^2}{f(x)}\, dx < \infty.$$

Nevertheless, $I^{(2)}(X) = \infty$, since f' is not continuous: $f'(0-) = 0$, $f'(0+) = 1$.

Example 2.5. Consider the C^∞-smooth density

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

Put $\varphi_p = H_p\varphi$ and note that $\varphi_p' = -\varphi_{p+1}$, so that $f = \varphi + \varphi_2$ and $f^{(p)} = (-1)^p\, (\varphi_p + \varphi_{p+2})$. Since $H_{2p}(0) + H_{2p+2}(0) = c_p$ and $H_{2p-1}(x) + H_{2p+1}(x) \sim -c_p x$ as $x \to 0$ with constants $c_p = (-1)^{p-1}\, \frac{(2p)!}{(p-1)!2^{p-1}}$, we conclude that

$$I^{(2p-1)}(X) < \infty, \quad I^{(2p)}(X) = \infty \quad (p \geq 1).$$

Example 2.6. Let X have a Gamma distribution with n degrees of freedom, that is, with density

$$f(x) = \frac{x^{n-1}}{\Gamma(n)}\, e^{-x}, \quad x > 0$$

(where n may be a real positive number). Similarly to the beta distributions, $I^{(p)}(X)$ is finite if and only if $n > 2p$. For the first three values of p, direct computations show that

$$I(X) = \frac{1}{n-2}, \tag{2.1}$$

$$I^{(2)}(X) = \frac{2}{(n-3)(n-4)}, \tag{2.2}$$

$$I^{(3)}(X) = \frac{6\,(n^2+13n+6)}{(n-2)(n-3)(n-4)(n-5)(n-6)} \tag{2.3}$$

for the parameters $n > 2$, $n > 4$, and $n > 6$, respectively (the formula (2.1) was already mentioned in [8]). We postpone the derivation of these formulas to Section 14.

3. **First elementary properties**

It is well-known that, if $I(X)$ is finite, then the density f of X represents a function of bounded variation on the real line with a total variation norm satisfying

$$\|f\|_{\rm TV} = \int_{-\infty}^{\infty} |f'(x)|\,dx = \mathbb{E}\,|\rho(X)| \leq \sqrt{I(X)}.$$

In particular, $f(-\infty) = f(\infty) = 0$, and f is bounded by $\sqrt{I(X)}$. The latter implies

$$\int_{-\infty}^{\infty} f'(x)^2\,dx \leq I(X)^{3/2}.$$

We now extend these relations to the Fisher-type information functionals of orders $p \geq 1$. Here and in the sequel, the following elementary observation will be needed.

Proposition 3.1. *Let $I^{(p)}(X)$ be finite. If $f(x) = 0$ at the point $x \in \mathbb{R}$ and $f^{(p-1)}$ has a finite derivative $f^{(p)}(x)$, then necessarily $f^{(p)}(x) = 0$. We also have $f'(x) = 0$.*

Proof. Since f is non-negative, necessarily $f'(x) = 0$, and we are done in the case $p = 1$. If $p \geq 2$, then, by Taylor's formula in the Peano form,

$$f(x+h) = \frac{a_2}{2!}\,h^2 + \dots + \frac{a_p}{p!}\,h^p + o(|h|^p), \qquad a_k = f^{(k)}(x),\ 2 \leq k \leq p,$$

and $f^{(p)}(x) = a_p + o(|h|)$ as $h \to 0$. Assuming that $f^{(p)}(x) \neq 0$, let k be the smallest integer in the interval $2 \leq k \leq p$ such that $f^{(k)}(x) \neq 0$. Then $a_k \neq 0$, $a_p \neq 0$, so that

$$\frac{f^{(p)}(x+h)^2}{f(x+h)} = \frac{a_p^2 + o(|h)}{\frac{a_k}{k!}\,h^k + o(|h|^k)}.$$

But this function is not integrable over $h \in (-\varepsilon, \varepsilon)$ with $\varepsilon > 0$ small enough. $\square$

Proposition 3.2. *If $I^{(p)}(X)$ is finite, the derivative $f^{(p-1)}$ represents a function of bounded variation with a total variation norm*

$$\|f^{(p-1)}\|_{\rm TV} = \int_{-\infty}^{\infty} |f^{(p)}(x)|\,dx \leq \sqrt{I_p(X)}.$$

In particular, $f^{(p-1)}(-\infty) = f^{(p-1)}(\infty) = 0$, *and*

$$\max_x |f^{(p-1)}(x)| \le \sqrt{I^{(p)}(X)}.$$

Proof. By the assumption, the derivative $f^{(p-1)}$ is differentiable on a set $E \subset \mathbb{R}$ of full Lebesgue measure. By Proposition 3.1, $f^{(p)}(x) \neq 0 \Rightarrow f(x) > 0$ for all $x \in E$. Hence, applying the Cauchy inequality, we have

$$\begin{aligned}\int_{-\infty}^{\infty} |f^{(p)}(x)|\,dx &= \int_{f(x)>0} |f^{(p)}(x)|\,dx \\ &= \int_{f(x)>0} \frac{|f^{(p)}(x)|}{\sqrt{f(x)}}\,\sqrt{f(x)}\,dx \ \le \ \sqrt{I^{(p)}(X)},\end{aligned}$$

proving the first assertion. As a consequence, the limits

$$f^{(p-1)}(-\infty) = \lim_{x\to-\infty} f^{(p-1)}(x), \qquad f^{(p-1)}(\infty) = \lim_{x\to\infty} f^{(p-1)}(x)$$

exist and are finite. Necessarily, these limits must be zero, since otherwise $f(x)$ would behave polynomially at infinity contradicting to the integrability of f. Finally,

$$\max_x |f^{(p-1)}(x)| \le \|f^{(p-1)}\|_{\rm TV} \le \sqrt{I^{(p)}(X)}.$$

□

Proposition 3.3. *If* $I^{(p)}(X)$ *is finite, then*

$$\int_{-\infty}^{\infty} |f^{(p)}(x)|^2\,dx \le I^{(p)}(X)^{3/2}.$$

This follows from

$$\int_{-\infty}^{\infty} |f^{(p)}(x)|^2\,dx \le \max_x f(x) \int_{f(x)>0} \frac{|f^{(p)}(x)|^2}{f(x)}\,dx.$$

4. Integrability of derivatives

Applying Proposition 3.2, one may extend its bound on the total variation norm to all derivatives smaller than p (in a certain form). As before, we assume that $p \ge 1$ is an integer.

Proposition 4.1. *If* f *is the density of a random variable* X *with finite* $I^{(p)}(X)$, *then all derivatives* $f^{(k)}$, $1 \le k \le p$, *are integrable functions. Moreover,*

$$\|f^{(k-1)}\|_{\rm TV} \ = \ \int_{-\infty}^{\infty} |f^{(k)}(x)|\,dx \ \le \ C_p\, I^{(p)}(X)^{\frac{k}{2p}} \tag{4.1}$$

with some constants C_p *depending on* p *only. In particular, if* f *is supported on the interval* (a,b), *finite or not, then* $f^{(k-1)}(a+) = f^{(k-1)}(b-) = 0$. *In addition,*

$$\max_x |f^{(k-1)}(x)| \le C_p\, I^{(p)}(X)^{\frac{k}{2p}}. \tag{4.2}$$

Before turning to the proof, let us mention two immediate consequences.

Corollary 4.2. $I^{(p)}(X) > 0$.

Indeed, in the case $I^{(p)}(X) = 0$, we would obtain from (4.1) with $k = 1$ that necessarily $\|f\|_{\rm TV} = 0$. But this is only possible when f is a constant.

Another immediate consequence from Proposition 4.1 concerns the decay of the characteristic function

$$\widehat{f}(t) = \mathbb{E}\, e^{itX} = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx, \quad t \in \mathbb{R}.$$

Corollary 4.3. *If $I^{(p)}(X)$ is finite, then $\widehat{f}(t) = o(|t|^{-p})$ as $|t| \to \infty$.*

For the proof, one may integrate by parts with $t \neq 0$, which gives

$$\begin{aligned} \widehat{f}(t) &= \frac{1}{it} \int_{-\infty}^{\infty} f(x)\, de^{itx} = -\frac{1}{it} \int_{-\infty}^{\infty} e^{itx} f'(x)\, dx \\ &= \ldots = \frac{1}{(it)^p} \int_{-\infty}^{\infty} e^{itx} f^{(p)}(x)\, dx. \end{aligned}$$

Here we used the property that all derivatives $f^{(k)}$ up to order p are integrable and vanishing at infinity for all $k \leq p-1$. Since $f^{(p)}$ is integrable, the last integral tends to zero as $|t| \to \infty$, by the Riemann-Lebesgue lemma.

Lemma 4.4. *For any integrable function u having derivatives up to order $p \geq 2$ (in the Radon-Nikodym sense for the p-th derivative), for all integers $1 \leq k \leq p-1$,*

$$\int_{-\infty}^{\infty} |u^{(k)}(x)|\, dx \leq A_p \int_{-\infty}^{\infty} |u(x)|\, dx + B_p \int_{-\infty}^{\infty} |u^{(p)}(x)|\, dx \tag{4.3}$$

with coefficients A_p and B_p depending on p only.

The integrability of the derivatives $u^{(k)}$ is stated in [4]. The inequality (4.3) can be obtained by the repeated application of its particular case $p = 2$, namely

$$\int_{-\infty}^{\infty} |u'(x)|\, dx \leq \int_{-\infty}^{\infty} |u(x)|\, dx + \frac{2}{3} \int_{-\infty}^{\infty} |u''(x)|\, dx,$$

which is derived for the class $\mathfrak{C}^{(2)}$ in [4], Proposition 5.1.

Proof of Proposition 4.1. The case $k = p$ is governed by Proposition 3.2, so, we may assume that $1 \leq k \leq p-1$ with $p \geq 2$. Let us apply (4.3) to the function $u(x) = f(\lambda x)$ with parameter $\lambda > 0$. Then we get

$$\int_{-\infty}^{\infty} |f^{(k)}(x)|\, dx \ \leq\ A_p\, \lambda^{-k} + B_p\, \lambda^{p-k} \int_{-\infty}^{\infty} |f^{(p)}(x)|\, dx.$$

Optimizing over all λ, this yields

$$\int_{-\infty}^{\infty} |f^{(k)}(x)|\, dx \leq C_p \Big(\int_{-\infty}^{\infty} |f^{(p)}(x)|\, dx\Big)^{k/p}$$

with p-dependent constants C_p. It remains to apply Proposition 3.2. $\square$

5. **Polynomial decay of densities and their derivatives**

If the moment $\beta_{2s} = \mathbb{E}\,|X|^{2s}$ is finite for some real number $s > 0$, then (cf. [5])

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

Moreover,

$$\lim_{|x|\to\infty} (1 + |x|^s)\,f(x) = 0.$$

These results may be generalized, which allows one to control a polynomial decay of densities and their derivatives at infinity.

Proposition 5.1. *If $I^{(p)}(X)$ and β_{2s} are finite for an integer $p \geq 1$ and $s > 0$, then*

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

As a consequence, for all $x \in \mathbb{R}$,

$$|f^{(p-1)}(x)| \leq \frac{c}{1+|x|^s}, \quad c = \big(1 + \sqrt{\beta_{2s}}\big)\sqrt{I^{(p)}(X)}.$$

Moreover,

$$f^{(p-1)}(x) = o\big(|x|^{-s}\big) \ \text{ as } |x| \to \infty.$$

Proof. Put $I = I^{(p)}(X)$. Recall that, by Proposition 3.1, $f^{(p)}(x) \neq 0 \Rightarrow f(x) > 0$ for all points x from a set of full Lebesgue measure. Hence, applying the Cauchy inequality, we have

$$\begin{aligned}\int_{-\infty}^{\infty} |x|^s\,|f^{(p)}(x)|\,dx &= \int_{f(x)>0} |x|^s\,|f^{(p)}(x)|\,dx \\ &= \int_{f(x)>0} \frac{|f^{(p)}(x)|}{\sqrt{f(x)}}\,|x|^s\sqrt{f(x)}\,dx \ \leq \ \sqrt{\beta_{2s} I}.\end{aligned}$$

This proves the first assertion.

Let us combine the obtained inequality with the one of Proposition 3.2. Then we get

$$\int_{-\infty}^{\infty} (1 + |y|^s)\,|f^{(p)}(y)|\,dy \leq \big(1 + \sqrt{\beta_{2s}}\big)\,\sqrt{I}.$$

Restricting the integration on the left-hand side to the half-axis $y \geq x \geq 0$, the left integral can be bounded from below by

$$(1 + |x|^s)\,\varepsilon(x), \quad \text{where} \ \ \varepsilon(x) = \int_x^{\infty} |f^{(p)}(y)|\,dy.$$

Hence, for any $b > x$,

$$\begin{aligned} |f^{(p-1)}(x) - f^{(p-1)}(b)| &= \Big|\int_x^b f^{(p)}(y)\,dy\Big| \\ &\leq \int_x^\infty |f^{(p)}(y)|\,dy \;\leq\; \frac{1}{1+|x|^s}\,\big(1+\sqrt{\beta_{2s}}\big)\sqrt{I}. \end{aligned}$$

Letting $b \to \infty$ and applying the property $f^{(p-1)}(b) \to 0$ (Proposition 3.2), we arrive at the second required inequality. Since $\varepsilon(x) \to 0$ as $x \to \infty$, the last assertion follows as well. □

Proposition 5.2. *If $I^{(p)}(X)$ and β_{2s} are finite for an integer $p \geq 1$ and $s > 0$, then*

$$f^{(p-k)}(x) = o\Big(\frac{1}{|x|^{s-k+1}}\Big), \quad k = 1,\dots,p, \quad s > k-1,$$

as $|x| \to \infty$. Moreover, in the case $k = p$,

$$f(x) = o\Big(\frac{1}{|x|^{s-p+1}}\Big), \quad s \geq p-1.$$

Proof. The case $k = 1$ corresponds to Proposition 5.1:

$$|f^{(p-1)}(y)| \leq \frac{\varepsilon(y)}{1+|y|^s},$$

where $\varepsilon(y) \to 0$ as $|y| \to \infty$. After the repeated integration of this inequality over $y > x \geq 0$, and using $f^{(p-l)}(\infty) = 0$, $1 \leq l \leq p$ (Proposition 4.1), we get, as $x \to \infty$,

$$\begin{aligned} f^{(p-2)}(x) &= o\big(x^{-(s-1)}\big), \qquad p \geq 2,\ s > 1, \\ f^{(p-3)}(x) &= o\big(x^{-(s-2)}\big), \qquad p \geq 3,\ s > 2,\ \dots \\ f^{(p-k)}(x) &= o\big(x^{-(s-(k-1))}\big), \quad p \geq k,\ s > k-1, \end{aligned}$$

which corresponds to the first claim. In the remaining case $k = p$ and $s = p-1$, the second claim $f(x) = o(1)$ holds true according to Proposition 4.1. □

6. **Relative Fisher information of order** p

Given two random variables X and Y with densities f and g from the class $\mathfrak{C}^p$, define the relative Fisher information of an integer order $p \geq 1$ by

$$I^{(p)}(X|Y) = I^{(p)}(f|g) = \int_{-\infty}^\infty \Big|\frac{f^{(p)}(x)}{f(x)} - \frac{g^{(p)}(x)}{g(x)}\Big|^2 f(x)\,dx.$$

This is a natural extension of the classical order $p = 1$ (see also [12] for other extensions).

Of a special interest is the case $Y = Z$ with the standard normal density $g = \varphi$. Then

$$I^{(p)}(X|Z) = I^{(p)}(f|\varphi) = \int_{-\infty}^\infty \Big|\frac{f^{(p)}(x)}{f(x)} - (-1)^p H_p(x)\Big|^2 f(x)\,dx.$$

Since the Chebyshev-Hermite polynomial $H_p(x)$ has degree p, for the last integral to be finite it is natural to require that X have a finite moment $\beta_{2p}(X) = \mathbb{E}\, X^{2p}$. Then, opening the brackets, we get another representation

$$I^{(p)}(X|Z) = I^{(p)}(X) - 2\,(-1)^p \int_{-\infty}^{\infty} f^{(p)}(x) H_p(x)\, dx + \mathbb{E}\, H_p(X)^2.$$

Assuming that $I^{(p)}(X)$ is finite, the above integral is finite according to Proposition 5.1 and may be easily evaluated. Namely, by Proposition 5.2 with $s = p$,

$$f^{(p-k)}(x) = o\Big(\frac{1}{|x|^{p-k+1}}\Big) \quad \text{as } |x| \to \infty, \quad k = 1, \dots, p-1.$$

Hence, integrating by parts and using $H_n'(x) = nH_{n-1}(x)$, we have

$$\begin{aligned} (-1)^p \int_{-\infty}^{\infty} f^{(p)}(x) H_p(x)\, dx &= (-1)^p \int_{-\infty}^{\infty} H_p(x)\, df^{(p-1)}(x) \\ &= (-1)^{p-1} \int_{-\infty}^{\infty} f^{(p-1)}(x)\, dH_p(x) \\ &= (-1)^{p-1}\, p \int_{-\infty}^{\infty} f^{(p-1)}(x) H_{p-1}(x)\, dx. \end{aligned}$$

Repeating the integration by parts, we will arrive at

$$(-1)^p \int_{-\infty}^{\infty} f^{(p)}(x) H_p(x)\, dx = p! \int_{-\infty}^{\infty} f(x) H_0(x)\, dx = p!$$

The latter factorial may also be written as $I^{(p)}(Z) = \mathbb{E}\, H_p(Z)^2$. Let us summarize in the next assertion containing Theorem 1.2.

Proposition 6.1. *If $I^{(p)}(X)$ and $\beta_{2p}(X)$ are finite for an integer $p \geq 1$, then*

$$I^{(p)}(X|Z) = I^{(p)}(X) - 2p! + \mathbb{E}\, H_p(X)^2.$$

In particular,

$$I^{(p)}(X) + \mathbb{E}\, H_p(X)^2 \geq 2p!$$

with equality if and only if X has a standard normal distribution. Therefore,

$$\mathbb{E}\, H_p(X)^2 = \mathbb{E}\, H_p(Z)^2 \;\Rightarrow\; I^{(p)}(X) \geq I^{(p)}(Z).$$

One may generalize this statement by replacing $H_p(x)$ with an arbitrary polynomial $H(x) = x^p + a_{p-1}x^{p-1} + \dots + a_0$ with leading coefficient 1. In this case again

$$\begin{aligned} \int_{-\infty}^{\infty} \Big|\frac{f^{(p)}(x)}{f(x)} - (-1)^p\, H(x)\Big|^2 f(x)\, dx &= I^{(p)}(X) \\ &\quad -2\,(-1)^p \int_{-\infty}^{\infty} f^{(p)}(x) H(x)\, dx + \mathbb{E}\, H(X)^2, \end{aligned}$$

while, integrating by parts, we have

$$\begin{aligned}(-1)^p \int_{-\infty}^{\infty} f^{(p)}(x) H(x)\, dx &= (-1)^p \int_{-\infty}^{\infty} H(x)\, df^{(p-1)}(x) \\ &= (-1)^{p-1} \int_{-\infty}^{\infty} f^{(p-1)}(x)\, dH(x) \\ &= (-1)^{p-1} \int_{-\infty}^{\infty} f^{(p-1)}(x) H'(x)\, dx.\end{aligned}$$

Repeating the integration by parts, we will arrive at

$$(-1)^p \int_{-\infty}^{\infty} f^{(p)}(x) H(x)\, dx = \int_{-\infty}^{\infty} f(x) H^{(p)}(x)\, dx = p!$$

Hence, we arrive at:

Proposition 6.2. *If $I^{(p)}(X)$ and $\beta_{2p}(X)$ are finite for an integer $p \geq 1$, then for any polynomial $H(x) = x^p + a_{p-1} x^{p-1} + \dots + a_0$,*

$$I^{(p)}(X) + \mathbb{E}\, H(X)^2 \geq 2p!$$

7. Lower semi-continuity

We now consider the lower semi-continuity of the Fisher information. In the case $p = 1$, the next statement corresponds to Proposition 3.1 from [5].

Proposition 7.1. *Let $(X_n)_{n \geq 1}$ be a sequence of random variables, and let X be a random variable such that $X_n \Rightarrow X$ weakly in distribution as $n \to \infty$. For any integer $p \geq 1$,*

$$I^{(p)}(X) \ \leq\ \liminf_{n \to \infty} I^{(p)}(X_n). \tag{7.1}$$

Since the general case requires some modifications in the argument used for $p = 1$ (especially in the last steps), we include the proof below.

Proof. Denote by $\mathfrak{P}_p$ the collection of all probability densities f on the real line with finite Fisher information of order p, and let $\mathfrak{P}_p(I)$ denote the subset of all densities which have Fisher information at most I. Since the case $p = 1$ in (7.1) is known, let $p \geq 2$.

For the proof of (7.1), we may assume that $I(X_n) \to I$ as $n \to \infty$ for some finite constant I. Then, for sufficiently large n, and without loss of generality for all $n \geq 1$, the random variables X_n have densities f_n belonging to $\mathfrak{P}_p(I + 1)$. In particular, these densities have derivatives $f_n^{(k)}$ up to order $p - 1$, such that the functions $f_n^{(p-1)}$ are absolutely continuous and have Radon-Nikodym derivatives $f_n^{(p)}$.

According to Proposition 4.1, for every $k = 0, 1, \dots, p - 1$,

$$\|f_n^{(k)}\|_{\mathrm{TV}} + \sup_x |f_n^{(k)}(x)| < C_p(I + 1) \tag{7.2}$$

with a constant C_p depending on p only. By the second Helly theorem (cf. e.g. [K-F]), $f_n^{(k)}(x)$ are convergent pointwise to some functions $g_k(x)$ of bounded total variation along a certain subsequence. For simplicity of notations, let this subsequence be a whole sequence, that is,

$$\lim_{n\to\infty} f_n^{(k)}(x) = g_k(x) \quad \text{for all } x \in \mathbb{R}. \tag{7.3}$$

Due to (7.2), this property can be complemented by the L^1 convergence on bounded intervals (for a proof, cf. [4], Proposition 11.4): For all $a < b$,

$$\lim_{n\to\infty} \int_a^b |f_n^{(k)}(x) - g_k(x)|\, dx = 0. \tag{7.4}$$

Putting $g_0 = g$, we have, in particular, $\lim_{n\to\infty} f_n(x) = g(x)$ and

$$\lim_{n\to\infty} \int_a^b |f_n(x) - g(x)|\, dx = 0, \quad -\infty < a < b < \infty. \tag{7.5}$$

Necessarily, $g(x) \geq 0$ and $\int_{-\infty}^{\infty} g(x)\, dx \leq 1$ (by Fatou's lemma). In fact, $\int_{-\infty}^{\infty} g(x)\, dx = 1$ which follows from the weak convergence of the distributions of X_n. Indeed, the latter implies and is actually equivalent to the property that, for any open set $G \subset \mathbb{R}$,

$$\mathbb{P}(X \in G) \leq \liminf_{n\to\infty} \mathbb{P}(X_n \in G)$$

(cf. e.g. [1]). Given $\varepsilon > 0$, choose an interval $G = (a, b)$ such that $\mathbb{P}(X \in G) > 1 - \varepsilon$, so that

$$\liminf_{n\to\infty} \int_G f_n(x)\, dx > 1 - \varepsilon.$$

By (7.5), the last integrals are convergent to $\int_G g(x)\, dx$. Therefore, $\int_G g(x)\, dx \geq 1 - \varepsilon$ for any $\varepsilon > 0$, hence g is a probability density. Since, the property (7.5) is stronger than the weak convergence, we also conclude that the distribution of X is absolutely continuous with respect to the Lebesgue measure and has density g.

If $1 \leq k \leq p - 1$, from (7.3)-(7.4) it follows that, for all $a, b \in \mathbb{R}$,

$$\int_a^b g_k(x)\, dx = g_{k-1}(b) - g_{k-1}(a). \tag{7.6}$$

This means that g_k represents a Radon-Nikodym derivative of g_{k-1}. In particular, g_{k-1} is continuous, and we conclude that the density g has $p - 2$ continuous derivatives $g^{(k)} = g_k$, $1 \leq k \leq p - 2$. The case $k = p - 1$ in (7.6) similarly implies that $g_{p-1} = g^{(p-1)}$ represents a Radon-Nikodym derivative of $g_{p-2} = g^{(p-2)}$.

Now, by Proposition 3.3,

$$\int_{-\infty}^{\infty} |f_n^{(p)}(x)|^2\, dx \leq C_p (I + 1)^{3/2}. \tag{7.7}$$

Since the unit ball of any separable L^2-space is weakly compact, there is a subsequence of $f_n^{(p)}$ which is weakly convergent to some function $g_p \in L^2(\mathbb{R})$. For simplicity of notations, again let this subsequence be a whole sequence, so that

$$\int_{-\infty}^{\infty} f_n^{(p)}(x) u(x)\, dx \to \int_{-\infty}^{\infty} g_p(x) u(x)\, dx \quad (n \to \infty) \tag{7.8}$$

for any $u \in L^2(\mathbb{R})$. Choosing here the indicator function $u = 1_{(a,b)}$ of a finite interval and applying (7.3) with $k = p-1$, we obtain that

$$g^{(p-1)}(b) - g^{(p-1)}(a) = \int_a^b g_p(x)\,dx.$$

This means that g_p appears as a Radon-Nikodym derivative of g_{p-1}. In particular, g_{p-1} is continuous, and therefore g has $p-1$ continuous derivatives $g^{(k)} = g_k$, $1 \le k \le p-1$. Thus, the function g belongs to the class $\mathfrak{C}^p$ with $g^{(p)} = g_p$ and $I^{(p)}(X) = I^{(p)}(g)$.

Finally, consider the sequence of functions

$$h_n(x,\lambda) = f_n^{(p)}(x)\, e^{-\lambda f_n(x)/2}, \quad x \in \mathbb{R},\ \lambda > 0.$$

They have bounded L^2-norms on the half-plane $\mathbb{R} \times \mathbb{R}_+$, namely

$$\begin{aligned} \|h_n\|^2_{L^2(\mathbb{R}\times\mathbb{R}_+)} &= \int_{-\infty}^{\infty}\int_0^{\infty} h_n(x,\lambda)^2\,dx\,d\lambda \\ &= \int_{f_n(x)>0} \frac{f_n^{(p)}(x)^2}{f_n(x)}\,dx \ = \ I^{(p)}(X_n) \ \le \ I+1. \end{aligned}$$

Here we applied Proposition 3.1, according to which $f_n^{(p)}(x) = 0$ for almost all x on the set where $f_n(x) = 0$.

Let us verify that h_n are weakly convergent in L^2 to the function

$$h(x,\lambda) = g^{(p)}(x)\, e^{-\lambda g(x)/2}$$

on every rectangle $R = [-T,T] \times [\lambda_0,\lambda_1]$ with fixed $T > 0$ and $\lambda_1 > \lambda_0 > 0$. Write

$$\begin{aligned} h_n(x,\lambda) - h(x,\lambda) &= f_n^{(p)}(x)\left(e^{-\lambda f_n(x)/2} - e^{-\lambda g(x)/2}\right) \\ &\quad + \left(f_n^{(p)}(x) - g^{(p)}(x)\right) e^{-\lambda g(x)/2}. \end{aligned} \tag{7.9}$$

Given a Borel measurable function $u \in L^2(\mathbb{R}\times\mathbb{R}_+)$ supported on R, define

$$u_1(x) = \int_{\lambda_0}^{\lambda_1} e^{-\lambda g(x)/2}\, u(x,\lambda)\,d\lambda, \quad x \in \mathbb{R}.$$

It is Borel measurable, supported on $[-T,T]$, and is bounded, since g is continuous (hence bounded on $[-T,T]$). Therefore, by the Fubini theorem and the weak convergence (7.8),

$$\begin{aligned} &\iint_R \left(f_n^{(p)}(x) - g^{(p)}(x)\right) e^{-\lambda g(x)/2}\, u(x,\lambda)\,dx\,d\lambda \\ &\qquad = \int_{-\infty}^{\infty} \left(f_n^{(p)}(x) - g_p(x)\right) u_1(x)\,dx \ \to 0 \qquad (n \to \infty). \end{aligned} \tag{7.10}$$

Next, by (7.3) with $k = 0$, we have $f_n(0) \to g(0)$ as $n \to \infty$. Using the representation

$$(f_n(x) - g(x)) - (f_n(0) - g(0)) = \int_0^x (f_n'(y) - g'(y))\,dy,$$

from (7.4) with $k = 1$ it also follows that

$$\varepsilon_n = \sup_{|x|\le T} |f_n(x) - g(x)| \to 0.$$

Hence

$$|e^{-\lambda f_n(x)/2} - e^{-\lambda g(x)/2}| \le C\varepsilon_n$$

with some constant C (which may depend on T and λ_j). Using Cauchy's inequality, this gives

$$\Big|\iint_R f_n^{(p)}(x)\,\big(e^{-\lambda f_n(x)/2} - e^{-\lambda g(x)/2}\big)\,u(x,\lambda)\,dx\,d\lambda\Big|^2$$

$$\leq\ (C\varepsilon_n)^2\,(\lambda_1-\lambda_0)\int_{-\infty}^{\infty} f_n^{(p)}(x)^2\,dx\iint_R u(x,\lambda)^2\,dx\,d\lambda\ \to\ 0$$

as $n\to\infty$, where we applied (7.7) in the last step. Combining this with (7.10) and returning to (7.9), we conclude that

$$\iint_R (h_n(x,\lambda)-h(x,\lambda))\,u(x,\lambda)\,dx\,d\lambda\ \to 0 \quad (n\to\infty),$$

which means that h_n is weakly convergent to h in the space $L^2(R)$. Therefore

$$\|h\|_{L^2(R)}^2\ \leq\ \liminf_{n\to\infty}\ \|h_n\|_{L^2(R)}^2\ \leq\ \liminf_{n\to\infty} I^{(p)}(X_n)\ =\ I.$$

Thus,

$$\int_{-T}^{T}\int_{\lambda_0}^{\lambda_1} g^{(p)}(x)^2\,e^{-\lambda g(x)}\,dx\,d\lambda\leq I.$$

Letting here $T\to\infty$, $\lambda_1\to\infty$ and $\lambda_0\to 0$, we arrive at (7.1). □

Remark 7.2. On the set $\mathfrak{P}_p(I)$ the weak convergence of the associated probability distributions coincides with the convergence in total variation distance (which corresponds to the convergence of probability densities in the L^1-norm). For the proof, suppose that $X_n \Rightarrow X$ weakly in distribution as $n\to\infty$ with $I^{(p)}(X_n)\leq I$. Then X_n have densities f_n of class $\mathfrak{C}^p$ with $I^{(p)}(f_n)\leq I$. We need to show that X has a density f in the same class such that

$$\int_{-\infty}^{\infty}|f_n(x)-f(x)|\,dx\to 0 \quad (n\to\infty). \tag{7.11}$$

Equivalently, it is sufficient to show that from any prescribed subsequence f_{n_k} one may extract a further subsequence $f_{n_{k_l}}$ which is convergent in L^1 to some density f. Arguing as in the beginning of the proof of Proposition 7.1, we obtain such a subsequence with the property that $f_{n_{k_l}}(x)\to f(x)$ for all $x\in\mathbb{R}$ as $l\to\infty$ for some density f. Applying Scheffe's lemma, this leads to (7.11) for $f_{n_{k_l}}$ and f.

8. **Convex mixtures of probability measures**

We will consider some properties of the Fisher-type information for random variables whose distributions are representable in a natural way as mixture of probability measures (including convolutions). In order to make all statements rigorous and as general as possible, first let us give a few remarks about the notion of mixture.

Denote by $\mathfrak{M}$ the collection of all probability measures on the real line. We treat it as a separable metric space with the topology of weak convergence which may be metrized using the Lévy distance, for example. As such, this space has a canonical Borel σ-algebra generated by the collection of all open subsets of $\mathfrak{M}$.

Lemma 8.1. *For any Borel set $A \subset \mathbb{R}$, the functional $T_A(\nu) = \nu(A)$ is Borel measurable on $\mathfrak{M}$. Moreover, the functional*

$$T_u(\nu) = \int_{-\infty}^{\infty} u\, d\nu$$

is Borel measurble on $\mathfrak{M}$, whenever the function $u : \mathbb{R} \to \mathbb{R}$ is bounded and Borel measurable.

Proof. Consider the collection $\mathfrak{A}$ of all Borel sets $A \subset \mathbb{R}$ such that T_A is Borel measurable on $\mathfrak{M}$. Let us list several basic properties of this functional.

1) For the union A of disjoint Borel sets A_n, we have $T_A = \sum_{n=1}^{\infty} T_{A_n}$.
2) For the monotone limit A of increasing or decreasing Borel sets A_n, $T_A = \lim_{n\to\infty} T_{A_n}$.
3) For the complement $\bar{A} = \mathbb{R} \setminus A$, we have $T_{\bar{A}} = 1 - T_A$.
4) More generally, $T_{A\setminus B} = T_A - T_B$ as long as $B \subset A$.
5) If A is closed, and $\nu_n \to \nu$ weakly in $\mathfrak{M}$, then

$$\limsup_{n\to\infty} \nu_n(A) \le \nu(A).$$

The last property is equivalent to saying that the functional T_A is upper semi-continuous on $\mathfrak{M}$. Hence, it is Borel measurable on $\mathfrak{M}$, that is, $A \in \mathfrak{A}$. Thus, $\mathfrak{A}$ is a monotone class containing all semi-open intervals $(a, b] = (-\infty, b] \setminus (-\infty, a]$, and therefore, this class contains all Borel subsets of the real line.

For the second assertion, first note that if u is simple in the sense that it is a finite linear combination of indicator functions 1_A of Borel sets $A \subset \mathbb{R}$, we are reduced to the previous step. In the general case, if $|u| \le M$, there is a sequence of simple functions u_n with values in $[-M, M]$ such that $u_n(x) \to u(x)$ for all $x \in \mathbb{R}$ as $n \to \infty$. By the Lebesgue dominated convergence theorem, $T_{u_n}(\nu) \to T_u(\nu)$ for any $\nu \in \mathfrak{M}$, implying that T_u is Borel measurable on $\mathfrak{M}$. □

Lemma 8.1 justifies the following:

Definition 8.2. Let π be a Borel probability measure on the space $\mathfrak{M}$. A convex mixture of probability measures on the real line with a mixing measure π is given by

$$\mu(A) = \int_{\mathfrak{M}} \nu(A)\, d\pi(\nu), \quad A \subset \mathbb{R} \text{ (Borel)}. \tag{8.1}$$

Recall that in the space $\mathfrak{M}$ there is a canonical metric defined by the total variation distance $\|\nu - \lambda\|_{\rm TV}$ between probability measures. It generates a stronger topology, and $\mathfrak{M}$ is not separable with respect to this metric (because, for example, $\|\delta_x - \delta_y\|_{\rm TV} = 2$ for all $x, y \in \mathbb{R}$, $x \neq y$). Nevertheless, the balls for this metric are Borel measurable for the weak topology. Indeed, for any signed Borel measure ν_0 on $\mathbb{R}$,

$$\|\nu - \nu_0\|_{\rm TV} = \sup_u |T_u(\nu) - T_u(\nu_0)|,$$

where the supremum may be taken over the set C_0 of all continuous, compactly supported functions u on $\mathbb{R}$ such that $|u| \le 1$. Moreover, this supremum can be restricted to a countable set, since the space C_0 is separable for the supremum-norm. Since for each u in C_0, the functional $\nu \to T_u(\nu)$ is continuous on $\mathfrak{M}$, the functional $\nu \to \|\nu - \nu_0\|_{\rm TV}$ is Borel measurable.

Lemma 8.3. *The collection $\mathfrak{M}_0$ of all absolutely continuous probability measures on the real line (with respect to the Lebesgue measure) represents a Borel set in $\mathfrak{M}$.*

Proof. Recall that a probability measure ν on the real line with distributions function $F(x) = \nu((-\infty, x])$, $x \in \mathbb{R}$, is absolutely continuous, if and only if F is absolutely continuous in the sense of Function Theory: For any $\varepsilon > 0$, there is $\delta > 0$, such that for any finite collection of non-overlapping intervals $(a_i, b_i) \subset \mathbb{R}$, $1 \le i \le n$,

$$\sum_{i=1}^n (b_i - a_i) < \delta \implies \sum_{i=1}^n (F(b_i) - F(a_i)) < \varepsilon.$$

Since F is non-decreasing and right-continuous, here one may additionally require that the endpoints a_i and b_i represent rational numbers. Also, one may replace open intervals in this definition with semi-open intervals $(a_i, b_i]$, leading to the increments $F(b_i) - F(a_i-)$. Define

$$\mathfrak{A} = \Big\{A = \bigcup_{i=1}^n (a_i, b_i] : a_1 < b_1 \le \dots \le a_n < b_n,\ a_i, b_i \in \mathbb{Q},\ n \ge 1\Big\}$$

and rewrite the definition of the absolute continuity of ν as the property that, for any $\varepsilon > 0$, there is $\delta > 0$ such that, for any $A \in \mathfrak{A}$, $\lambda(A) \le \delta \Rightarrow \nu(A) \le \varepsilon$, where λ denotes the Lebesgue measure. In terms of the functional

$$M_\delta(\nu) = \sup\big\{\nu(A) : A \in \mathfrak{A},\ \lambda(A) \le \delta\big\},$$

this is equivalent to saying that

$$M(\nu) \equiv \inf_{\delta > 0} M_\delta(\nu) = \inf_k M_{1/k}(\nu) = 0.$$

Now, the crucial point is that the collection $\mathfrak{A}$ is countable. Applying Lemma 8.1, we conclude that every functional M_δ is Borel measurable on $\mathfrak{M}$ as the supremum of countably many Borel measurable functionals. Therefore, M is Borel measurable as well, and the set $\mathfrak{M}_0$ is described as the pre-image $M^{-1}(\{0\})$. □

Denote by $\mathfrak{P}$ the collection of all (probability) densities f on the real line. It is a closed convex subset of $L^1(\mathbb{R})$ with respect to the usual L^1-metric. With every f in $\mathfrak{P}$ we associate the probability measure μ_f with this density. By Lemma 8.3, the collection $\mathfrak{M}_0 = \{\mu_f : f \in \mathfrak{P}\}$ represents a Borel set in $\mathfrak{M}$. One can thus identify $\mathfrak{P}$ and $\mathfrak{M}_0$ by means of the bijective map $f \to \mu_f$.

Lemma 8.4. *The Borel σ-algebra in $\mathfrak{P}$ induced from $L^1(\mathbb{R})$ coincides with the Borel σ-algebra in $\mathfrak{M}_0$ induced from $\mathfrak{M}$.*

Proof. Given a sequence f_n and f in $\mathfrak{P}$, the weak convergence $\mu_{f_n} \to \mu_f$ in $\mathfrak{M}$ is equivalent to

$$\int_{-\infty}^x f_n(y)\,dy \to \int_{-\infty}^x f(y)\,dy \quad \text{for any } x \in \mathbb{R}$$

(and actually uniformly over all x). It is weaker than the convergence in L^1

$$\|\mu_{f_n} - \mu_f\|_{\rm TV} = \|f_n - f\|_1 = \int_{-\infty}^\infty |f_n(y) - f(y)|\,dy \to 0,$$

which is equivalent to the convergence of the measures in total variation distance. Hence, the Borel σ-algebra in $\mathfrak{M}_0$ induced from $\mathfrak{M}$ is (formally) smaller than the Borel σ-algebra in $\mathfrak{P}$ induced from $L^1(\mathbb{R})$, using the identification of $\mathfrak{P}$ and $\mathfrak{M}_0$.

For the opposite inclusion, first recall that the Borel σ-algebra in $L^1(\mathbb{R})$ is generated by the L^1-balls

$$B = \{f \in L^1(\mathbb{R}) : \|f - f_0\|_1 < r\}, \quad f_0 \in L^1, \ r > 0$$

(since the space L^1 is separable). Hence, it is sufficient to see that any set of the form $D = B \cap \mathfrak{P}$ is Borel measurable in $\mathfrak{M}$ (where we use Lemma 8.3). This is the same as saying that the balls in $\mathfrak{M}$ for the total variation distance are Borel measurable, which has been already explained. □

As a consequence from Lemma 8.4, one may use Definition 8.2 starting from a Borel probability measure π on $\mathfrak{P}$. Following this definition, one can define the convex mixture according to (8.1):

$$\mu(A) = \int_{\mathfrak{P}} \bigg[\int_A g(x)\,dx\bigg]\,d\pi(g), \quad A \subset \mathbb{R} \text{ (Borel)}.$$

This measure belongs to $\mathfrak{M}_0$ and has some density $f(x) = \frac{d\mu(x)}{dx}$ called the (convex) mixture of densities with mixing measure π. For short,

$$f(x) = \int_{\mathfrak{P}} g(x)\,d\pi(g), \quad x \in \mathbb{R}. \tag{8.2}$$

9. **Convexity and continuity along convolutions**

Another general property of the Fisher-type information is its convexity, that is, we have:

Proposition 9.1. *Given probability densities f_i on the real line and weights $\alpha_i > 0$ such that $\sum_{i=1}^n \alpha_i = 1$, we have*

$$I^{(p)}(f) \le \sum_{i=1}^n \alpha_i I^{(p)}(f_i), \quad \text{where } f = \sum_{i=1}^n \alpha_i f_i. \tag{9.1}$$

Proof. This follows from the fact that the function $R(u,v) = u^2/v$ is 1-homogeneous and convex on the upper half-plane $u \in \mathbb{R}$, $v > 0$.

For more details, one may assume that $n = 2$ and $I^{(p)}(f_i) < \infty$, $i = 1,2$. Thus, f_1, f_2 and f belong to the class $\mathfrak{C}^p$ with $f^{(p)} = \alpha_1 f_1^{(p)} + \alpha_2 f_2^{(p)}$. Let G denote the set of all points $x \in \mathbb{R}$ where $f(x) > 0$ and such that the derivatives $f_i^{(p-1)}(x)$ are differentiable at x, so that

$$I^{(p)}(f) = \int_G R\big(f^{(p)}(x), f(x)\big)\,dx.$$

The set G can be decomposed into the three measurable parts

$$\begin{aligned} G_0 &= \{x \in G : f_1(x) > 0, \ f_2(x) > 0\}, \\ G_1 &= \{x \in G : f_1(x) > 0, \ f_2(x) = 0\}, \\ G_2 &= \{x \in G : f_1(x) = 0, \ f_2(x) > 0\}. \end{aligned}$$

On the first part, due to the convexity of R,

$$\int_{G_0} R\big(f^{(p)}(x), f(x)\big)\,dx \,\le\, \alpha_1 \int_{G_0} R\big(f_1^{(p)}(x), f_1(x)\big)\,dx + \alpha_2 \int_{G_0} R\big(f_2^{(p)}(x), f_2(x)\big)\,dx.$$

If $x \in G_1$, then $f(x) = \alpha_1 f_1(x)$ and

$$\int_{G_1} R\big(f^{(p)}(x), f(x)\big)\,dx \,=\, \alpha_1 \int_{G_1} R\big(f_1^{(p)}(x), f_1(x)\big)\,dx.$$

Similarly,

$$\int_{G_2} R\big(f^{(p)}(x), f(x)\big)\,dx \,=\, \alpha_2 \int_{G_2} R\big(f_2^{(p)}(x), f_2(x)\big)\,dx.$$

Summing the last inequality with the last two equalities, we obtain (9.1). □

As a consequence, the collection of all probability densities f on the real line such that $I^{(p)}(f) \le I$ is convex for any value of I.

We need to extend Jensen's inequality (9.1) to arbitrary "continuous" convex mixtures of densities and probability distributions. For this aim, we temporarily employ the notation $I^{(p)}(\mu)$ for $I^{(p)}(X)$, when a random variable X is distributed according to μ.

Proposition 9.1. *If a probability density f is a convex mixture of densities with mixing measure π on $\mathfrak{P}$, then*

$$I^{(p)}(f) \le \int_{\mathfrak{P}} I^{(p)}(g)\,d\pi(g). \tag{9.2}$$

More generally, if a probability measure μ is a convex mixture of probability measures with mixing measure π on $\mathfrak{M}$, then

$$I^{(p)}(\mu) \le \int_{\mathfrak{M}} I^{(p)}(\nu)\,d\pi(\nu). \tag{9.3}$$

The integrals in (9.2)-(9.3) make sense, since the functionals $g \to I^{(p)}(g)$ and $\nu \to I^{(p)}(\nu)$ are lower semi-continuous and hence Borel measurable on $\mathfrak{P}$ and $\mathfrak{M}$, respectively (Proposition 7.1 and Lemma 8.3).

The proof of Proposition 9.1 is similar to the one of Proposition 3.3 in [5] for the case $p = 1$, so we omit it. Note, however, that the argument uses the lower-semi-continuity of the functional $I^{(p)}$.

For the proof of (9.3), one may assume that the integral on the right-hand side is finite. But then $I^{(p)}(\nu) < \infty$ for π-almost all ν, which implies that the mixing measure π is supported on $\mathfrak{M}_0$. In this case, μ belongs to $\mathfrak{M}_0$, and (9.3) is thus reduced to (9.2).

As a consequence of Proposition 9.1, the functional $I^{(p)}$ is monotone under convolutions.

Corollary 9.2. *For all independent random variables X and Z,*

$$I^{(p)}(X + Z) \le I^{(p)}(X). \tag{9.4}$$

Proof. Let ν denote the distribution of X, and let $\nu_z(A) = \nu(A - z)$ be the shift of ν ($z \in \mathbb{R}$). The distribution of $X + Z$ represents the mixture

$$\mu = \int_{-\infty}^{\infty} \nu_z \, dP(z),$$

where P is the distribution of Z. The map $T : \mathbb{R} \to \mathfrak{M}$ defined by $T(z) = \nu_z$ is continuous, so, the image $B = T(\mathbb{R})$ is a σ-compact, hence a Borel set in $\mathfrak{M}$. This map pushes forward P to a Borel probability measure π supported on B. It remains to apply (9.4) and note that $I^{(p)}(\nu_z) = I^{(p)}(\nu)$ for all z. □

Combining Propositions 7.1 and Corollary 9.2, we obtain the continuity property of the functional $I^{(p)}$ for convolved densities as stated in Theorem 1.4: For all independent random variables X and Z,

$$\lim_{\varepsilon \to 0} I^{(p)}(X + \varepsilon Z) = I^{(p)}(X). \tag{9.5}$$

Proof of Theorem 1.4. The distributions of $X + \varepsilon Z$ are weakly convergent to the distribution of X as $\varepsilon \to 0$, so that, by (7.1),

$$I^{(p)}(X) \leq \liminf_{\varepsilon \to 0} I^{(p)}(X + \varepsilon Z).$$

On the other hand, $I^{(p)}(X + \varepsilon Z) \leq I^{(p)}(X)$, by (9.4). Both inequalities lead to (9.5). □

Corollary 9.3. *Suppose that a normal random variable Z is independent of the random variable X. Then the function $\varepsilon \to I^{(p)}(X + \varepsilon Z)$ is finite and non-decreasing in $\varepsilon > 0$.*

Indeed, let $Z \sim N(0, 1)$. By Corollary 9.2 and according to Example 2.1,

$$I^{(p)}(X + \varepsilon Z) \leq I^{(p)}(\varepsilon Z) = p! \, \varepsilon^{-2p}.$$

The monotonicity follows from the fact that the convolution of Gaussian measures is Gaussian.

Remark 9.4. The functional

$$I_p(X) = I_p(f) = \mathbb{E}\, |\rho(X)|^p = \int_{-\infty}^{\infty} \Big|\frac{f'(x)}{f(x)}\Big|^p f(x)\, dx$$

satisfies similar properties as the Fisher information (in the case $p = 1$), such as the lower semi-continuity and the monotonicity

$$I_p(X + Y) \leq \min\{I_p(X), I_p(Y)\}$$

for all for independent summands X and Y. As a consequence, we have the analog of (9.5)

$$\lim_{\varepsilon \to 0} I_p(X + \varepsilon Z) = I_p(X). \tag{9.6}$$

It is shown in [3] that, if $p \geq 1$ is an integer and the random variables $(X_i)_{1 \leq i \leq p+1}$ are independent and have densities of bounded total variation $b_i = I_1(X_i)$, then

$$I_p(X_1 + \dots + X_{p+1}) \leq c_p \, b_1 \dots b_{p+1} \Big(\frac{1}{b_1} + \dots + \frac{1}{b_{p+1}}\Big)$$

with constant $c_p = p^p/(2^p p!)$. Hence, similarly to Corollary 9.3, we have $I_p(X + \varepsilon Z) < \infty$ for all p, assuming that the random variables X and Z are independent, and $Z \sim N(0, 1)$.

10. **Representations in terms of isoperimetric profile**

If a continuous probability density f is supported and positive on the interval $(a,b) \subset \mathbb{R}$, finite or not, the associated distribution may be characterized, up to a shift parameter, by the function (called sometimes the isoperimetric profile)

$$L(t) = f(F^{-1}(t)), \quad 0 < t < 1, \tag{10.1}$$

or equivalently

$$f(x) = L(F(x)), \quad a < x < b. \tag{10.2}$$

This follows from the equality

$$F^{-1}(t_2) - F^{-1}(t_1) = \int_{t_1}^{t_2} \frac{dt}{L(t)}, \quad 0 < t_1, t_2 < 1,$$

where $F^{-1} : (0,1) \to (a,b)$ denotes the inverse of the distribution function $F(x) = \int_a^x f(y)\,dy$, $a < x < b$.

If f is locally absolutely continuous on (a,b) and has a Radon-Nikodym derivative f', both F and F^{-1} will be C^1-smooth functions with absolutely continuous derivatives. Hence, L is also locally absolutely continuous on $(0,1)$. Differentiating (10.2), we obtain $f' = L'(F)f$ a.e. in (a,b), implying that the random variable X with density f has the Fisher information

$$I(X) = \int_a^b L'(F(x))^2 f(x)\,dx = \int_0^1 L'(t)^2\,dt.$$

More generally, the moments of the scores of X are given by

$$I_p(X) = \int_0^1 |L'(t)|^p\,dt. \tag{10.3}$$

Moreover, if f' is locally absolutely continuous on (a,b) and has a Radon-Nikodym derivative f'', then both F and F^{-1} are C^2-smooth with absolutely continuous second order derivatives. Hence, L also has a locally absolutely continuous derivative L'' on $(0,1)$. Starting from (10.1), we get $(L^2)' = 2f'(F^{-1})$ and $(L^2)'' = 2f''(F^{-1})/f(F^{-1})$. This gives:

Proposition 10.1. *Suppose that the density f of the random variable X is supported and positive on an interval, finite or not. If it is of the class $\mathfrak{C}^1$ or $\mathfrak{C}^2$, then respectively*

$$\begin{aligned} I(X) &= \int_0^1 L'(t)^2\,dt, \\ I^{(2)}(X) &= \frac{1}{4}\int_0^1 \big(L^2(t)''\big)^2\,dt = \int_0^1 \big(L'(t)^2 + L(t)L''(t)\big)^2\,dt. \end{aligned} \tag{10.4}$$

Note that, if $I^{(2)}(X)$ is finite, then necessarily

$$\int_0^1 \big(L'(t)^2 + L(t)L''(t)\big)\,dt = \int_0^1 \big(L(t)L'(t)\big)'\,dt = 0.$$

This follows from the property that $f'(a+) = f'(b-) = 0$ according to Proposition 3.2. Indeed, using $LL' = f'(F^{-1})$, we get

$$\int_{t_0}^{t_1} \big(L(t)L'(t)\big)'\,dt = L(t_1)L'(t_1) - L(t_0)L'(t_0) \rightarrow 0 \quad \text{as} \ \ t_0 \downarrow 0, \ t_1 \uparrow 1.$$

There is another representation for the integral in (10.4).

Proposition 10.2. *Suppose that the density $f \in \mathfrak{C}^2$ of the random variable X is supported and positive on an interval, finite or not. Then*

$$I^{(2)}(X) = \int_0^1 \Big(L''(t)^2 L(t)^2 + \frac{1}{3}\,L'(t)^4\Big)\,dt, \tag{10.5}$$

as long as the latter integral is finite, which is equivalent to the finiteness of $I^{(2)}(X)$.

Proof. By (10.4),

$$I^{(2)}(X) = \int_0^1 \Big(L''(t)^2 L(t)^2 + L'(t)^4 + 2L''(t)L'(t)^2 L(t)\Big)\,dt. \tag{10.6}$$

Integrating by parts, we have, for all $0 < t_0 < t_1 < 1$,

$$\begin{aligned}
\int_{t_0}^{t_1} L''(t)L'(t)^2 L(t)\,dt \ &= \ \int_{t_0}^{t_1} L'(t)^2 L(t)\,dL'(t) \\
&= \ L(t_1)L'(t_1)^3 - L(t_0)L'(t_0)^3 - \int_{t_0}^{t_1} L'(t)\,d\,(L'(t)^2 L(t)) \\
&= \ L(t_1)L'(t_1)^3 - L(t_0)L'(t_0)^3 - \int_{t_0}^{t_1} L'(t)^4\,dt - 2\int_{t_0}^{t_1} L''(t)L'(t)^2 L(t)\,dt.
\end{aligned}$$

Equivalently,

$$3\int_{t_0}^{t_1} L''(t)L'(t)^2 L(t)\,dt = L(t_1)L'(t_1)^3 - L(t_0)L'(t_0)^3 - \int_{t_0}^{t_1} L'(t)^4\,dt.$$

If

$$L(t)L'(t)^3 \rightarrow 0 \quad \text{as } t \rightarrow 0 \text{ and } t \rightarrow 1, \tag{10.7}$$

in the limit as $t_0 \rightarrow 0$ and $t_1 \rightarrow 1$ we obtain that

$$\int_0^1 L''(t)L'(t)^2 L(t)\,dt = -\frac{1}{3}\int_0^1 L'(t)^4\,dt.$$

As a result, (10.6) is simplified to (10.5). Note that the requirement (10.7) is equivalent to the property that $\frac{f'(x)^3}{f(x)^2} \rightarrow 0$ as $x \rightarrow a$ and $x \rightarrow b$.

In order to verify (10.7), we apply the Cauchy inequality and use the assumption to get

$$\Big(\int_0^1 L'(t)^2\,L(t)\,|L''(t)|\,dt\Big)^2 \leq \int_0^1 L'(t)^4\,dt \int_0^1 L(t)^2 L''(t)^2\,dt < \infty.$$

This implies that the function $u = LL'^3$ has a bounded total variation on $(0,1)$. Indeed, its derivative

$$u' = L'^4 + 3LL'^2 L''$$

is integrable. Therefore, the limits $c_0 = u(0+)$ and $c_1 = u(1-)$ exist and are finite. Let us show that necessarily $c_0 = c_1 = 0$. Suppose that $c_0 \neq 0$. We have

$$(L(t)^{4/3})' = \frac{4}{3}\,L(t)^{1/3}L'(t) = \frac{4}{3}\,u(t)^{1/3} \,\rightarrow\, \frac{4}{3}\,c_0^{1/3}$$

as $t \to 0$. Since $L(0+) = 0$ and $L(t) > 0$ for $t \in (0,1)$, this implies that $c_0 > 0$ and moreover $L^{4/3}(t) \leq 2c_0^{1/3}t$ for all t small enough, $0 < t \leq t_0$, that is, $L(t) \leq (8c_0)^{1/4}\,t^{3/4}$. This gives

$$L'(t) \sim \frac{c_0^{1/3}}{L(t)^{1/3}} \geq \frac{c'}{t^{1/4}}, \quad 0 < t \leq t_0,$$

with some constant $c' > 0$. As a consequence, the function L'^4 is not integrable on this interval, which contradicts to the assumption. Hence, necessarily $c_0 = 0$, and by a similar argument, $c_1 = 0$ as well. Thus, (10.7) is fulfilled. □

11. Lower bounds for $I^{(2)}$ in terms of I_4 and I

The representation (10.5) may be used for the lower bound on $I^{(2)}$ in terms of I_4 and I, in order to derive the relation (1.5) of Theorem 1.3:

$$I^{(2)}(X) \geq \frac{1}{3}\,I_4(X) \geq \frac{1}{3}\,I(X)^2. \tag{11.1}$$

Proof of Theorem 1.3. First suppose that the conditions of Proposition 10.2 are fulfilled. Then, by (10.5),

$$I^{(2)}(X) \geq \frac{1}{3}\int_0^1 L'(t)^4\,dt \geq \frac{1}{3}\Big(\int_0^1 L'(t)^2\,dt\Big)^2 = \frac{1}{3}\,I(X)^2.$$

Recalling the representation (10.3) for the functionals I_p, (11.1) follows.

For the finiteness of the integral (10.5), we need to assume that $I_4(X)$ is finite together with integrability of the function $(LL'')^2$. In order to give a sufficient condition for this property to hold, write

$$L''(t) \,=\, \frac{d}{dt}\,\frac{f'(F^{-1}(t))}{f(F^{-1}(t))} \,=\, \frac{f''(F^{-1}(t))}{f(F^{-1}(t))^2} - \frac{f'(F^{-1}(t))^2}{f(F^{-1}(t))^3}$$

and

$$L(t)L''(t) \,=\, \frac{f''(F^{-1}(t))}{f(F^{-1}(t))} - \Big(\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}\Big)^2. \tag{11.2}$$

Using $(x+y)^2 \leq 2x^2 + 2y^2$ $(x,y \in \mathbb{R})$, this implies

$$L(t)^2L''(t)^2 \,\leq\, 2\,\Big(\frac{f''(F^{-1}(t))}{f(F^{-1}(t))}\Big)^2 + 2\,\Big(\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}\Big)^4$$

and

$$\int_0^1 L(t)^2L''(t)^2\,dt \,\leq\, 2\,I^{(2)}(X) + 2\,I_4(X). \tag{11.3}$$

Thus, (11.1) is proved provided that the random variable X has a density f of class $\mathfrak{C}^2$ which is positive and is supported on some interval (a,b) and such that $I^{(2)}(X)$ and $I_4(X)$ are finite.

In the general case, the previous step can be applied to the random variables $X_\varepsilon = X + \varepsilon Z$, $\varepsilon > 0$, assuming that $Z \sim N(0,1)$ is independent of X. In this case, all X_ε have positive C^∞-smooth densities with finite $I^{(2)}(X_\varepsilon)$ and $I_4(X_\varepsilon)$, according to Corollary 9.3 and Remark 9.4. Hence we get

$$I^{(2)}(X_\varepsilon) \geq \frac{1}{3}\, I_4(X_\varepsilon) \geq \frac{1}{3}\, I(X_\varepsilon)^2.$$

Letting here $\varepsilon \to 0$ and applying (9.5)-(9.6), we arrive at (11.1). □

Remark 11.1. We can now explain the last assertion in Proposition 10.2 about the convergence of the integral in (10.5). Assuming that the Fisher-type information $I^{(2)}(X)$ is finite and applying (11.1), we conclude that the moment $I_4(X)$ is finite and hence the integral in (11.3) is finite as well. Thus, the integral in (10.5) is finite. Conversely, assuming that this integral is finite, from (11.2) we obtain that

$$\Big(\frac{f''(F^{-1}(t))}{f(F^{-1}(t))}\Big)^2 \ \leq\ 2\, L(t)^2 L''(t)^2 + 2\,\Big(\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}\Big)^4.$$

After the integration of this inequality over $0 < t < 1$, we are led to the desired conclusion

$$I^{(2)}(X) \ \leq\ 2 \int_0^1 L(t)^2 L''(t)^2\, dt + 2\, I_4(X) \ <\ \infty.$$

12. **Stam-type inequality in the case** $p \geq 2$

Recall that the inequality (1.8) of Theorem 1.5 states that, for all $k = 1, \dots, p-1$, $p \geq 2$,

$$\frac{1}{I^{(p)}(X+Y)} \geq \frac{1}{I^{(p)}(X)} + \frac{1}{I^{(p)}(Y)} + \frac{1}{I^{(k)}(X) I^{(p-k)}(Y)} \tag{12.1}$$

whenever the random variables X and Y are independent. In the case $p = 2$, this relation is reduced to

$$\frac{1}{I^{(2)}(X+Y)} \geq \frac{1}{I^{(2)}(X)} + \frac{1}{I^{(2)}(Y)} + \frac{1}{I(X) I(Y)}. \tag{12.2}$$

Let us test it on the normal distributions, that is, for $X \sim N(a_1, \sigma_1^2)$ and $Y \sim N(a_2, \sigma_2^2)$ with $a_1, a_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$. Then $X + Y \sim N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$, so that according to Example 2.1,

$$I(X) = \frac{1}{\sigma_1^2}, \qquad I(Y) = \frac{1}{\sigma_2^2},$$
$$I^{(2)}(X) = \frac{2}{\sigma_1^4}, \qquad I^{(2)}(Y) = \frac{2}{\sigma_2^4}, \qquad I^{(2)}(X+Y) = \frac{2}{(\sigma_1^2 + \sigma_2^2)^2}.$$

In this case, (12.2) becomes the equality

$$\frac{(\sigma_1^2 + \sigma_2^2)^2}{2} = \frac{\sigma_1^4}{2} + \frac{\sigma_2^4}{2} + \sigma_1^2 \sigma_2^2.$$

Proof of Theorem 1.5. One may assume that the random variables X and Y have C^∞-smooth positive densities f and g with finite Fisher information of all orders up to p. Indeed, if (12.1) is established under these conditions, in the general case one may apply this relation to the random variables

$$X_\varepsilon = X + \varepsilon Z_1, \quad Y_\varepsilon = X + \varepsilon Z_2 \quad (\varepsilon > 0),$$

assuming that Z_1 and Z_2 are independent and have a standard normal distribution. Then $X_\varepsilon + Y_\varepsilon = (X+Y) + \varepsilon\sqrt{2}Z$ with $Z \sim N(0,1)$, and (12.1) yields

$$\frac{1}{I^{(p)}(X+Y+\varepsilon\sqrt{2}Z)} \geq \frac{1}{I^{(p)}(X_\varepsilon)} + \frac{1}{I^{(p)}(Y_\varepsilon)} + \frac{1}{I^{(k)}(X_\varepsilon) I^{(p-k)}(Y_\varepsilon)}.$$

Letting $\varepsilon \to \infty$ and applying the continuity property, we arrive at the desired relation (12.1) in full generality.

Under the above assumptions, the density of the sum $X+Y$ represents the convolution

$$h(x) = \int_{-\infty}^{\infty} f(x-y) g(y)\, dy = \int_{-\infty}^{\infty} f(y) g(x-y)\, dy.$$

By Proposition 4.1, all derivatives of f and g are integrable up to order p and are vanishing at infinity up to order $p-1$. Hence the function h is smooth, everywhere positive, and we have similar representations for its derivatives of any order $k \leq p$

$$h^{(k)}(x) = \int_{-\infty}^{\infty} f^{(k)}(x-y) g(y)\, dy = \int_{-\infty}^{\infty} f^{(k)}(y) g(x-y)\, dy.$$

Differentiating this equality $p-k$ times, we obtain a Radon-Nikodym derivative

$$h^{(p)}(x) = \int_{-\infty}^{\infty} f^{(k)}(x-y) g^{(p-k)}(y)\, dy.$$

Hence, given real numbers $\alpha_i \geq 0$ such that $\alpha_0 + \alpha_1 + \dots + \alpha_p = 1$, we have

$$h^{(p)}(x) = \int_{-\infty}^{\infty} \sum_{k=0}^{p} \alpha_k f^{(k)}(x-y) g^{(p-k)}(y)\, dy.$$

Let us introduce the probability measures

$$\frac{d\mu_x(y)}{dy} = \frac{f(x-y)g(y)}{h(x)}, \quad x \in \mathbb{R},$$

and rewrite the above as

$$\frac{h^{(p)}(x)}{h(x)} = \int_{-\infty}^{\infty} \sum_{k=0}^{p} \alpha_k \, \frac{f^{(k)}(x-y)\, g^{(p-k)}(y)}{f(x-y)\, g(y)}\, d\mu_x(y).$$

One may now apply Jensen's inequality, which gives

$$\Big(\frac{h^{(p)}(x)}{h(x)}\Big)^2 \leq \int_{-\infty}^{\infty} \bigg(\sum_{k=0}^{p} \alpha_k \, \frac{f^{(k)}(x-y)\, g^{(p-k)}(y)}{f(x-y)\, g(y)}\bigg)^2 d\mu_x(y),$$

or equivalently

$$\begin{aligned}
\frac{h^{(p)}(x)^2}{h(x)} \quad &\leq \quad \int_{-\infty}^{\infty} \bigg(\sum_{k=0}^{p} \alpha_k \, \frac{f^{(k)}(x-y)\, g^{(p-k)}(y)}{f(x-y)\, g(y)}\bigg)^2 f(x-y) g(y)\, dy \\
&= \quad \sum_{k=0}^{p} \alpha_k^2 \int_{-\infty}^{\infty} \frac{f^{(k)}(x-y)^2\, g^{(p-k)}(y)^2}{f(x-y)\, g(y)}\, dy \\
&\quad + \sum_{k \neq l} \alpha_k \alpha_l \int_{-\infty}^{\infty} \frac{f^{(k)}(x-y) f^{(l)}(x-y)}{f(x-y)}\, \frac{g^{(p-k)}(y) g^{(p-l)}(y)}{g(y)}\, dy. \qquad (12.3)
\end{aligned}$$

Integrating over x, we arrive at

$$I^{(p)}(h) \le \sum_{k=0}^{p} \alpha_k^2\, I^{(k)}(f) I^{(p-k)}(g) + \sum_{k \neq l} \alpha_k \alpha_l\, V_{k,l}(f)\, V_{p-k,p-l}(g), \tag{12.4}$$

where we use the notation

$$V_{k,l}(f) = \int_{-\infty}^{\infty} \frac{f^{(k)}(x) f^{(l)}(x)}{f(x)}\, dx. \tag{12.5}$$

Note that these integrals exist and are finite, since, by Cauchy's inequality,

$$\int_{-\infty}^{\infty} \frac{|f^{(k)}(x) f^{(l)}(x)|}{f(x)}\, dx \le \sqrt{I^{(k)}(X) I^{(l)}(X)} < \infty,$$

and similarly for g. This also justifies the integration with respect to x in (12.3).

If $k = 0$ or $l = 0$, then the integral in (12.5) is vanishing. Indeed, in the case $l = 0$ and $1 \le k \le p$,

$$\begin{aligned} V_{k,0}(f) \ &= \ \int_{-\infty}^{\infty} f^{(k)}(x)\, dx \ = \ \lim_{T \to \infty} \int_{-T}^{T} f^{(k)}(x)\, dx \\ &= \ \lim_{T \to \infty} \big(f^{(k-1)}(T) - f^{(k-1)}(-T)\big) \ = \ 0, \end{aligned}$$

where we applied Proposition 4.1. A similar conclusion applies to g, and as a consequence,

$$V_{k,l}(f)\, V_{p-k,p-l}(g) = 0, \quad \text{if } k = 0, \ k = p, \ l = 0, \ l = p \ \ (k \neq l). \tag{12.6}$$

For the setting of Theorem 1.5, we need to restrict ourselves to the case where $\alpha_j = 0$ whenever $j \neq 0, k, p$ for a fixed $k = 1, \dots, p-1$. Then the second sum in (12.4) contains only 3 terms, which are actually zero, by (12.6). Hence (12.4) is simplified to

$$I^{(p)}(h) \le \alpha_0^2\, I^{(p)}(f) + \alpha_p^2\, I^{(p)}(g) + \alpha_k^2\, I^{(p)}(f) I^{(p-k)}(g).$$

Minimizing the right-hand side over all admissible α_i yields (12.1). $\square$

13. Stam-type inequality with Gaussian components

As we have already mentioned, Theorem 1.5 can be refined in the form of the relation

$$\frac{1}{I^{(p)}(X+Y)} \ \ge \ \sum_{k=0}^{p} \frac{1}{I^{(k)}(X) I^{(p-k)}(Y)}, \tag{13.1}$$

where one of the independent summands has a normal distribution. This is a consequence of a more general assertion which we state as a lemma.

Lemma 13.1. *Let X and Y be independent random variables. Suppose that X has a finite Fisher information $I^{(p)}(X)$ with a density $f \in \mathfrak{C}^p$ such that*

$$V_{k,l}(f) = \int_{-\infty}^{\infty} \frac{f^{(k)}(x) f^{(l)}(x)}{f(x)}\, dx = 0 \quad \textit{for all } k \neq l \ \ (1 \le k, l \le p-1). \tag{13.2}$$

Then the relation (13.1) *holds true.*

Proof. As in the proof of Theorem 1.5, we may assume that Y has a C^∞-smooth positive density g with finite Fisher information of all orders up to p, and that the same is true for X. The density h of the sum $X+Y$ has been already shown to satisfy the relation (12.4), which is simplified under the condition (13.2) to

$$I^{(p)}(h) \, \leq \, \sum_{k=0}^p \alpha_k^2 \, I^{(k)}(f) I^{(p-k)}(g), \quad \alpha_k > 0, \ \alpha_0 + \dots + \alpha_p = 1. \tag{13.3}$$

It remains to minimize the right-hand side over all admissible coefficients α_k. So, consider the quadratic function of the form

$$Q(\alpha_1,\dots,\alpha_p) = A_0\alpha_0^2 + A_1\alpha_1^2 + \dots + A_p\alpha_p^2, \quad \alpha_0 = 1 - \alpha_1 - \dots - \alpha_p,$$

with parameters $A_k > 0$. Its partial derivatives $\partial_{\alpha_k} Q = -2A_0\alpha_0 + 2A_k\alpha_k$ are vanishing if and only if $\alpha_k = \frac{A_0}{A_k}\,\alpha_0$, $k = 1,\dots,p$. Thus, at the point of minimum necessarily

$$\alpha_0\Big(1 + A_0 \sum_{k=1}^p \frac{1}{A_k}\Big) = 1, \quad \text{that is,} \quad \alpha_0 = \frac{1}{A_0}\,\Big(\sum_{k=0}^p \frac{1}{A_k}\Big)^{-1}.$$

From this we find that

$$\alpha_k = \frac{1}{A_k}\,\Big(\sum_{k=0}^s \frac{1}{A_k}\Big)^{-1}, \quad k = 0,1,\dots,p,$$

and

$$Q(\alpha_1,\dots,\alpha_p) = \sum_{k=0}^p \frac{1}{A_k}\,\Big(\sum_{k=0}^p \frac{1}{A_k}\Big)^{-2} = \Big(\sum_{k=0}^p \frac{1}{A_k}\Big)^{-1}.$$

Equivalently, for all $(\alpha_0,\dots,\alpha_p) \in \mathbb{R}^{p+1}$ such that $\alpha_0 + \dots + \alpha_p = 1$,

$$Q(\alpha_1,\dots,\alpha_p)^{-1} \geq \sum_{k=0}^p A_k^{-1}.$$

Hence, (13.1) follows by applying the above inequality with $A_k = I^{(k)}(X) I^{(p-k)}(Y)$. □

Proof of Theorem 1.6. The inequality (13.1) is invariant under all affine transforms $(X,Y) \to (c_1,c_2) + \lambda(X,Y)$. Hence, when verifying (13.2) in the Gaussian case, it is sufficient to consider X having a standard normal distribution with density φ. Since $\varphi^{(k)}(x) = (-1)^k H_k(x)\varphi(x)$, the condition (13.2) is equivalent to the orthogonality of the Chebyshev-Hermite polynomials in the Hilbert space $L^2(\mathbb{R}, \varphi(x)\,dx)$. □

14. **The Gamma distributions**

Let the random variables X_n and X_m be independent and have the Gamma distributions with n and m degrees of freedom (not necessarily integers), that is, with densities

$$f(x) = \frac{x^{n-1}}{\Gamma(n)}\,e^{-x}, \quad g(x) = \frac{x^{m-1}}{\Gamma(m)}\,e^{-x}, \quad x > 0.$$

As was noticed, $I^{(p)}(X_n)$ and $I^{(p)}(X_m)$ are finite if and only if $\min(n,m) > 2p$.

Let us derive the identities (2.1)-(2.3) and then check whether or not the inequality (13.1) is true in the case $p = 3$. Since $X_n + X_m$ has the Gamma distribution with $n+m$ degrees of freedom, this relation becomes

$$\begin{aligned} \frac{1}{I^{(3)}(X_{n+m})} \ &\geq \ \frac{1}{I^{(3)}(X_n)} + \frac{1}{I^{(3)}(X_m)} \\ &\quad + \frac{1}{I(X_n) I^{(2)}(X_m)} + \frac{1}{I^{(2)}(X_n) I(X_m)}. \end{aligned} \tag{14.1}$$

For the computation of the Fisher-type information, we first note that, if $u(x) = P(x)\, e^{-x}$ for a smooth function P, then

$$\begin{aligned} u' \ &= \ (P' - P)\, e^{-x}, \\ u'' \ &= \ (P'' - 2P' + P)\, e^{-x}, \quad u''' \ = \ (P''' - 3P'' + 3P' - P)\, e^{-x}, \end{aligned}$$

so that

$$\begin{aligned} u'^2 \ &= \ (P'^2 + P^2 - 2P'P)\, e^{-x}, \\ u''^2 \ &= \ (P''^2 + 4P'^2 - 4P''P' + P^2 + 2P''P - 4P'P)\, e^{-x}, \\ u'''^2 \ &= \ (P'''^2 + 9P''^2 + 9P'^2 - 6P'''P'' + 6P'''P' - 18P''P')\, e^{-x} \\ &\quad + (P^2 - 2P'''P + 6P''P - 6P'P)\, e^{-x}. \end{aligned}$$

From this, choosing $P(x) = x^{n-1}$, we have

$$\frac{u'^2}{u} \ = \ \Big(\frac{P'^2}{P} + P - 2P'\Big)\, e^{-x} \ = \ \Big((n-1)^2\, x^{n-3} + x^{n-1} - 2(n-1)\, x^{n-2}\Big)\, e^{-x}$$

and

$$\int_0^\infty \frac{u'^2}{u}\, dx = (n-1)^2\, \Gamma(n-2) - \Gamma(n) = \Gamma(n)\, \Big(\frac{n-1}{n-2} - 1\Big) = \Gamma(n)\, \frac{1}{n-2}.$$

Hence

$$I(X_n) = \frac{1}{n-2}, \quad n \geq 2. \tag{14.2}$$

Similarly,

$$\begin{aligned} \frac{u''^2}{u} \ &= \ \Big(\frac{P''^2}{P} + 4\, \frac{P'^2}{P} - 4\, \frac{P''P'}{P} + P + 2P'' - 4P'\Big)\, e^{-x} \\ &= \ \Big((n-1)^2 (n-2)^2\, x^{n-5} + 4(n-1)^2\, x^{n-3} - 4\,(n-1)^2 (n-2)\, x^{n-4} \\ &\quad + x^{n-1} + 2(n-1)(n-2)\, x^{n-3} - 4(n-1)\, x^{n-2}\Big)\, e^{-x} \end{aligned}$$

and

$$\begin{aligned} \int_0^\infty \frac{u''^2}{u}\, dx \ &= \ (n-1)^2 (n-2)^2\, \Gamma(n-4) + 4(n-1)^2\, \Gamma(n-2) + \Gamma(n) \\ &\quad - 4\,(n-1)^2\,(n-2)\, \Gamma(n-3) + 2(n-1)(n-2)\, \Gamma(n-2) - 4(n-1)\Gamma(n-1) \\ &= \ \Gamma(n)\, \Big(\frac{(n-1)(n-2)}{(n-3)(n-4)} - 1 + 4\, \frac{n-1}{n-2} - 4\, \frac{n-1}{n-3}\Big) \\ &= \ \Gamma(n)\, \Big(\frac{(n-1)(n-2)}{(n-3)(n-4)} - 1 - \frac{4(n-1)}{(n-2)(n-3)}\Big). \end{aligned}$$

After simplifications, we arrive at

$$I^{(2)}(X_n) = \frac{2}{(n-3)(n-4)}, \quad n \geq 4. \tag{14.3}$$

Finally, write

$$\begin{aligned}
\frac{u'''^2}{u} &= \Big(\frac{P'''^2}{P} + 9\,\frac{P''^2}{P} + 9\,\frac{P'^2}{P} - 6\,\frac{P'''P''}{P} + 6\,\frac{P'''P'}{P} - 18\,\frac{P''P'}{P}\Big)\, e^{-x} \\
&\quad + (P - 2P''' + 6P'' - 6P')\, e^{-x} \\
&= \Big((n-1)^2(n-2)^2(n-3)^2\, x^{n-7} + 9(n-1)^2(n-2)^2\, x^{n-5} \\
&\quad + 9(n-1)^2\, x^{n-3} - 6\,(n-1)^2(n-2)^2(n-3)\, x^{n-6} \\
&\quad + 6(n-1)^2(n-2)(n-3)\, x^{n-5} - 18(n-1)^2(n-2)\, x^{n-4} + x^{n-1} \\
&\quad - 2(n-1)(n-2)(n-3)\, x^{n-4} + 6(n-1)(n-2)\, x^{n-3} - 6(n-1)\, x^{n-2}\Big)\, e^{-x},
\end{aligned}$$

implying

$$\begin{aligned}
\int_0^\infty \frac{u'''^2}{u}\, dx &= (n-1)^2(n-2)^2(n-3)^2\, \Gamma(n-6) + 9(n-1)^2(n-2)^2\, \Gamma(n-4) \\
&\quad + 9(n-1)^2\, \Gamma(n-2) - 6\,(n-1)^2(n-2)^2(n-3)\, \Gamma(n-5) \\
&\quad + 6(n-1)^2(n-2)(n-3)\, \Gamma(n-4) - 18\,(n-1)^2(n-2)\, \Gamma(n-3) - \Gamma(n) \\
&= \Gamma(n)\, \Big(\frac{(n-1)(n-2)(n-3)}{(n-4)(n-5)(n-6)} + 9\,\frac{(n-1)(n-2)}{(n-3)(n-4)} + 9\,\frac{n-1}{n-2} \\
&\qquad - 6\,\frac{(n-1)(n-2)}{(n-4)(n-5)} + 6\,\frac{n-1}{n-4} - 18\,\frac{n-1}{n-3} - 1\Big).
\end{aligned}$$

Up to the factor $\Gamma(n)$, this is simplified as

$$\begin{aligned}
&\frac{(n-1)(n-2)(n-3)}{(n-4)(n-5)(n-6)} - 1 + 9\,\frac{(n-1)(2n^2-11n+16)}{(n-2)(n-3)(n-4)} - 18\,\frac{n-1}{(n-4)(n-5)} - 18\,\frac{n-1}{n-3} \\
&= \frac{(n-1)(n-2)(n-3)}{(n-4)(n-5)(n-6)} - 1 + 9\,\frac{n(n-1)}{(n-2)(n-3)(n-4)} - 18\,\frac{n-1}{(n-4)(n-5)} \\
&= \frac{(n-1)(n-2)(n-3)}{(n-4)(n-5)(n-6)} - 1 - 9\,\frac{(n-1)(n^2-5n+12)}{(n-2)(n-3)(n-4)(n-5)} \\
&= 3\,\frac{3n^2 - 21\,n + 38}{(n-4)(n-5)(n-6)} - 9\,\frac{(n-1)(n^2-5n+12)}{(n-2)(n-3)(n-4)(n-5)} \\
&= 6\,\frac{n^2+13n+6}{(n-2)(n-3)(n-4)(n-5)(n-6)}.
\end{aligned}$$

Hence

$$I^{(3)}(X_n) = \frac{6\,(n^2+13n+6)}{(n-2)(n-3)(n-4)(n-5)(n-6)}, \quad n \geq 6. \tag{14.4}$$

Suppose that $m = n$. Then (14.1) is simplified to

$$\frac{1}{I^{(3)}(X_{2n})} \ \geq \ \frac{2}{I^{(3)}(X_n)} + \frac{2}{I(X_n) I^{(2)}(X_n)}. \tag{14.5}$$

Using (14.2)-(14.4), the inequality (14.5) is equivalent to

$$\frac{(2n-2)(2n-3)(2n-4)(2n-5)(2n-6)}{6\,(4n^2+26n+6)} \ \geq \ \frac{(n-2)(n-3)(n-4)(n-5)(n-6)}{3\,(n^2+13n+6)} + 2(n-2)\,\frac{(n-3)(n-4)}{2},$$

that is,

$$2\,\frac{(n-1)(2n-3)(n-2)(2n-5)(n-3)}{2n^2+13n+3} \ \geq \ \frac{(n-2)(n-3)(n-4)(n-5)(n-6)}{n^2+13n+6} + 3\,(n-2)(n-3)(n-4),$$

or what is even simpler

$$2\,\frac{(n-1)(2n-3)(2n-5)}{2n^2+13n+3} \geq \frac{(n-4)(n-5)(n-6)}{n^2+13n+6} + 3\,(n-4). \tag{14.6}$$

But this is not true for the first admissible values of n which can be started with $n = 6$. This can be seen in the next table:

	LHS of (14.6)	RHS of (14.6)
$n = 6$	$\frac{70}{17} \sim 4.12$	6
$n = 7$	$\frac{99}{16} \sim 6.19$	$9 + \frac{3}{53} \sim 9.06$
$n = 8$	$\frac{2002}{235} \sim 8.52$	$12 + \frac{4}{29} \sim 12.14$
$n = 9$	$\frac{520}{47} \sim 11.06$	$15 + \frac{5}{17} \sim 15.29$
$n = 10$	$\frac{510}{37} \sim 13.78$	$18 + \frac{30}{59} \sim 18.51$

Let us also look at the value of $V_{1,2}(f_n)$ for the density f_n of X_n and see that it is not vanishing. From the formulas for the first derivatives of $u = Pe^{-x}$ with $P(x) = x^{n-1}$, we find that

$$\begin{aligned} \frac{u'u''}{u} &= \frac{P''P'}{P} - 2\,\frac{P'^2}{P} - P'' + 3P' - P \\ &= (n-1)^2(n-2)\,x^{n-4} - (n-1)(3n-4)\,x^{n-3} + 3(n-1)\,x^{n-2} - x^{n-1}, \end{aligned}$$

so

$$\begin{aligned} \int_0^\infty \frac{u'u''}{u}\,dx &= (n-1)^2(n-2)\,\Gamma(n-3) - (n-1)(3n-4)\,\Gamma(n-2) + 2\,\Gamma(n) \\ &= \Gamma(n)\,\Big(\frac{n-1}{n-3} - \frac{3n-4}{n-2} + 2\Big) = \Gamma(n)\,\frac{2}{(n-2)(n-3)}. \end{aligned}$$

It follows that

$$V_{1,2}(f_n) = \frac{2}{(n-2)(n-3)} > 0, \quad n > 3.$$

Acknowledgement. This work was partially supported by the NSF grant DMS-2154001.

References

[1] Billingsley, P. Convergence of probability measures. Wiley, New York, xii+253, 1968.

[2] Bobkov, S. G. The size of singular component and shift inequalities. Ann. Probab. 27 (1999), no. 1, 416–431.

[3] Bobkov, S. G. Moments of the scores. IEEE Trans. Inform. Theory 65 (2019), no. 9, 5294–5301.

[4] Bobkov, S. G. Upper bounds for Fisher information. Electron. J. Probab. 27 (2022), Paper no. 115, 44 pp.

[5] Bobkov, S. G.; Chistyakov, G. P.; Götze, F. Fisher information and the central limit theorem. Probab. Theory Related Fields 159 (2014), issue 1-2, 1–59.

[6] Dembo, A.; Cover, T. M.; Thomas, J. A. Information-theoretic inequalities. IEEE Trans. Inform. Theory 37 (1991), no. 6, 1501–1518.

[7] Gabetta, E. On a conjecture of McKean with application to Kac's model. Quaderno, Univ. degli Studi di Milano, 1993, 13 pp. Also: Transport Theory and Statistical Physics 24 (1995), issue 1-3, 305–317.

[8] Johnson, O. Information theory and the central limit theorem. Imperial College Press, London, 2004, xiv+209 pp.

[9] 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), 5th edition. "Nauka", Moscow, 1981, 544 pp.

[10] Lions, P.-L.; Toscani, G. A strengthened central limit theorem for smooth densities. J. Funct. Anal. 129 (1995), no. 1, 148–167.

[11] Stam, A. J. Some inequalities satisfied by the quantities of information of Fisher and Shannon. Information and Control 2 (1959), 101–112.

[12] Toscani, G. Score functions, generalized relative Fisher information and applications. Ric. Mat. 66 (2017), no. 1, 15–26.