Upper Bounds for Fisher information[*]

Sergey G. Bobkov[†]

Abstract

Upper bounds are considered for the Fisher information of random vectors in terms of total variation and norms in Sobolev spaces. We also survey and refine a number of known results in this direction.

Keywords: Fisher information; Sobolev spaces.
MSC2020 subject classifications: 60E; 60FEJP.
Submitted to EJP on March 4, 2022, final version accepted on July 27, 2022.

1 Introduction

Given a random vector X in $\mathbb{R}^n$ with density p, its Fisher information is defined by

$$I(X) = I(p) = \int \frac{|\nabla p|^2}{p}\, dx = 4 \int |\nabla \sqrt{p}|^2\, dx. \tag{1.1}$$

This functional is well-defined and finite, when the function $\sqrt{p}$ belongs to the Sobolev space $W_1^2(\mathbb{R}^n)$. In all other cases, one puts $I(X) = \infty$. In the one-dimensional case, the integrals in (1.1) make sense when the density p is locally absolutely continuous and has derivative p' in the Radon-Nikodym sense. One may then write $I(X) = \mathbb{E}\rho(X)^2$ in terms of the score function $\rho = (\log p)'$, also called the logarithmic derivative of p. Of a large interest are also more general functionals (moments of the scores)

$$I_k(X) = \mathbb{E}\,|\rho(X)|^k, \quad k \geq 1.$$

Since the Fisher information appears naturally in many mathematical problems, it is useful to know general conditions which ensure that $I(X)$ is finite. For example, for the applicability of the central limit theorem with respect to the relative Fisher information, one needs to verify that this functional becomes finite when taking several convolutions of densities which might have an infinite Fisher information (such as the uniform distributions on bounded intervals). To this aim, it was shown in [3] that

$$I(X) \leq \frac{1}{2}\,(b_1 b_2 + b_1 b_3 + b_2 b_3) \tag{1.2}$$

in dimension $n = 1$ for the sum $X = X_1 + X_2 + X_3$ of three independent random variables X_j whose densities p_j have finite total variation norms $b_j = \|p_j\|_{\mathrm{TV}}$. Here, adding an independent summand to X may only decrease the Fisher information. On the other hand, it may happen that $I(X_1 + X_2) = \infty$. With similar conclusions, (1.2) was extended in [2] to higher moments of the scores as the relation

$$I_k(X) \leq c_k\, b_1 \ldots b_{k+1} \Big(\frac{1}{b_1} + \cdots + \frac{1}{b_{k+1}} \Big), \quad c_k = \frac{k^k}{2^k k!}, \tag{1.3}$$

for the sum $X = X_1 + \cdots + X_{k+1}$ of $k+1$ independent random variables whose densities p_j have total variation norms $b_j = \|p_j\|_{\mathrm{TV}}$.

The usefulness of such relations is explained by the fact that the total variation norm is much easier tractable. In particular, this norm can be directly related to the characteristic functions of the involved random variables. As a corollary, the following characterization holds in the case where all X_j are independent, have finite absolute moment, and a common characteristic function $f(t) = \mathbb{E}\, e^{itX_j}$, $t \in \mathbb{R}$. Namely, the partial sums

$$S_N = X_1 + \cdots + X_N$$

will have a finite Fisher information for some and then for all large N, if and only if $f(t) = o(t^{-\varepsilon})$ as $t \to \infty$ for some $\varepsilon > 0$. The same conclusion is also true about the moments $I_k(S_N)$ of an arbitrary order k (cf. [2]). One may therefore wonder whether a similar characterization holds in spaces of higher dimensions. Keeping aside this question for a separate discussion, one of the purposes of this note is to extend the relation (1.2) to densities on $\mathbb{R}^n$.

Theorem 1.1. *For the sum $X = X_1 + X_2 + X_3$ of three independent random vectors X_j in $\mathbb{R}^n$ whose densities p_j have finite total variation norms $b_j = \|p_j\|_{\mathrm{TV}}$, we have*

$$I(X) \leq c\, (b_1 b_2 b_3)^{2/3}, \tag{1.4}$$

where $c > 0$ is an absolute constant.

Note that modulo an absolute factor, the expression on the right-hand side is slightly better than the one in (1.2), in view of the arithmetic-geometric inequality. What also looks to be rather remarkable is that the constant in (1.4) is independent of the dimension n (as we will see, one may take $c = 18$).

In general, the total variation norm of an integrable function u on $\mathbb{R}^n$ is defined by

$$\|u\|_{\mathrm{TV}} = \sup \int u(x) \sum_{i=1}^n \frac{\partial w_i}{\partial x_i}\, dx, \tag{1.5}$$

where the supremum is taken over all collections of C_0^∞-smooth functions $w_i : \mathbb{R}^n \to \mathbb{R}$ such that $w_1^2 + \cdots + w_n^2 \leq 1$ pointwise on $\mathbb{R}^n$. This definition leads to the more familiar formula

$$\|u\|_{\mathrm{TV}} = \int |\nabla u(x)|\, dx, \tag{1.6}$$

once u has a weak gradient ∇u. If $u = 1_A$ is an indicator function of a Borel set A in $\mathbb{R}^n$ of finite volume, the expression in (1.5) defines the perimeter $\mathrm{Per}(A)$ of A. It is finite, for example, when A is open, bounded, and has a C^2-smooth boundary (cf. [17], p. 229).

As an example illustrating (1.4), one may consider the random vectors X_j uniformly distributed over sets A_j in $\mathbb{R}^n$ with finite positive volume $v_j = \mathrm{vol}_n(A_j)$ and finite perimeter $P_j = \mathrm{Per}(A_j)$. In this case, the densities of X_j represent normalized indicator functions $p_j = \frac{1}{v_j} 1_{A_j}$, and their total variation norms are given by $b_j = P_j/v_j$. One

can therefore conclude that the sum $X = X_1 + X_2 + X_3$ has finite Fisher information satisfying

$$I(X) \leq c \left(\frac{P_1 P_2 P_3}{v_1 v_2 v_3} \right)^{2/3}.$$

In the one-dimensional situation, the proofs of (1.2)-(1.3) from [2], [3] are based on the application of the Brunn-Minkowski inequality from Convex Geometry which allows one to derive these relations for X_j's uniformly distributed over arbitrary bounded intervals. Another ingredient in the argument is an interesting fact that any probability density p on the real line with finite total variation norm may be represented as a "continuous" convex mixture $p = \int p_t \, d\pi(t)$ of densities of uniform distributions with the property that $\|p\|_{\mathrm{TV}} = \int \|p_t\|_{\mathrm{TV}} \, d\pi(t)$, thus reversing Jensen's inequality for the total variation norm.

However, it is not clear how to push forward this approach in the multidimensional situation. Instead, we refine and employ one result from the theory of differentiable measures due to Uglanov and Bogachev about a general bound on $I(p)$ without assuming that the density p has a convolution structure. A main difficulty in estimating $I(p)$ concerns mostly the one-dimensional case. If a non-negative integrable function p on the real line has 3 continuous derivatives (this class may be enlarged), it was shown in [5], [6] that

$$\int_{-\infty}^{\infty} \frac{p'(x)^2}{p(x)} \, dx \leq 8 \int_{-\infty}^{\infty} |p'(x)| \, dx + 6 \int_{-\infty}^{\infty} |p''(x)| \, dx + 2 \int_{-\infty}^{\infty} |p'''(x)| \, dx. \tag{1.7}$$

In the earlier paper [16], this inequality was stated without proof with existing absolute constants. This relation may be extended to higher dimensions in terms of the corresponding partial derivatives of p. As we will see, the derivatives of the first and second order may actually be eliminated, so that we have:

Theorem 1.2. *For any probability density p on $\mathbb{R}^n$ having continuous partial derivatives up to the third order,*

$$I(p) \leq c n^{1/3} \left(\sum_{i=1}^{n} \int |\partial_{x_i}^3 p(x)| \, dx \right)^{2/3}, \tag{1.8}$$

where $c > 0$ is an absolute constant.

Applying (1.8) to the convolution of 3 densities on $\mathbb{R}^n$, we will derive the relation (1.4).

With this approach in mind, one may wonder whether or not one can obtain similar inequalities for general moments of the scores so that to extend the inequality (1.3) to higher dimensions. In this connection, let us mention that Krugova [13] has extended the inequality (1.7) by proving that

$$I_k(p) = \int_{-\infty}^{\infty} \left(\frac{|p'(x)|}{p(x)} \right)^k p(x) \, dx \leq C_k \int_{-\infty}^{\infty} \left(|p'(x)| + |p''(x)| + |p'''(x)| \right) dx$$

for the region of real orders $1 \leq k < 3$ with some constants C_k depending on k only. But, as is also well-known, such a relation cannot be true for $k \geq 3$, even if we involve higher order derivatives. For example, the density

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

has integrable derivatives of any order, while $I_3(p) = \infty$. Hence, the convolution structure of the distribution of the random variable of X is essential for the bound (1.3) with $k \geq 3$.

Returning to Theorem 1.1, one motivating point for the derivation of multidimensional upper bounds such as (1.4) is the central limit theorem in the i.i.d. model for

the normalized sums $Z_N = S_N/\sqrt{N}$ with respect to the relative Fisher information $I(Z_N\|Z) = I(Z_N) - I(Z) = I(Z_N) - n$ (which we do not discuss here). This functional appears naturally in other limit theorems and bounds as well. For example, of a large interest is the behavior of the relative entropy

$$D(Z_N\|Z) = \int p_N \log(p_N/\varphi)\, dx,$$

where p_N denotes the density of Z_N, and φ is the density of the standard normal random vector Z in $\mathbb{R}^n$. Assuming that the distribution of the random vector X_1 in $\mathbb{R}^n$ is isotropic, has a finite Fisher information, and shares a Poincaré-type inequality

$$\mathrm{Var}(u(X_1)) \leq C\, \mathbb{E}\, |\nabla u(X_1)|^2,$$

it was recently shown by Courtade, Fathi and Pananjady [8] that

$$D(Z_N\|Z) \leq \frac{n(C-1)}{2N} \log\left(1 + \frac{I(X_1\|Z)}{n(C-1)}\, N\right). \tag{1.9}$$

Using (1.4), this bound may be stated under a weaker assumption that X_1 has density with finite total variation b. Applying (1.9) to the normalized sums of $N/3$ independent copies of $\frac{1}{\sqrt{3}}(X_1 + X_2 + X_3)$ with N divisible by 3, we then obtain that

$$D(Z_N\|Z) \leq \frac{3n(C-1)}{2N} \log\left(1 + \frac{18\, b^2}{n(C-1)}\, N\right).$$

This paper consists of two parts. In sections 1–7 we focus on the one-dimensional case and discuss various upper bounds on the Fisher information $I(p)$ both in terms of the second and third derivatives of p, and for several classes of probability distributions (such as compactly supported or unimodal distributions). In Section 7, Theorems 1.1-1.2 are proved for $n = 1$. Sections 8–17 mostly deal with the multidimensional situation. To make the proof of main results rigorous, this case requires a careful analysis of basic concepts from the theory of weak derivatives and Sobolev spaces. Therefore, we include a short reminder of basic definitions and facts in this theory, together with some special results needed for an easy treatment of the Fisher information functional. They are used in particular to rigorously justify some of its important properties such as the lower semi-continuity and convexity.

Contents:
1. Introduction.
2. Functions with bounded second derivative.
3. Decay of densities and their derivatives.
4. Unimodal and quasi-unimodal distributions.
5. Total variation norm via higher order derivatives.
6. The use of the third derivative.
7. Theorems 1.1-1.2 in the one-dimensional case.
8. Weak derivatives.
9. Weak derivatives along single variables.
10. Sobolev spaces.
11. BV-space.
12. Convolution of functions of bounded variation.
13. Fisher information in high dimensions.
14. Lower semi-continuity of Fisher information.
15. Convexity of Fisher information.
16. Upper bounds. Proof of Theorem 1.2.
17. Proof of Theorem 1.1.

2 Functions with bounded second derivative

Definition 2.1. *Given $-\infty \le a < b \le \infty$ and an integer $l \ge 1$, we denote by $\mathfrak{C}^l(a, b)$ the collection of all continuous functions u on the interval (a, b) having continuous derivatives up to order $l - 1$ such that the derivative $u^{(l-1)}$ is (locally) absolutely continuous. Then $u^{(l-1)}$ has a Radon-Nikodym derivative defined almost everywhere on (a, b), which we denote $u^{(l)}$.*

In that case, one may also say that u is $\mathfrak{C}^l$-smooth on (a, b). When the interval coincides with the whole real line, the notation is shortened to $\mathfrak{C}^l$.

The Fisher information

$$I(p) = \int_{-\infty}^{\infty} \frac{p'(x)^2}{p(x)} \, dx = \mathbb{E}\rho(X)^2 \tag{2.1}$$

is well-defined for any probability density p from the class $\mathfrak{C}^1$. If X is a random variable with density p, we have $\mathbb{P}\{p(X) > 0\} = 1$, so the integration in (2.1) may be performed over the set $p(x) > 0$. It may further be restricted to the set of all points x where p is differentiable, with $p(x) > 0$ and $p'(x) \ne 0$. Indeed, $p'(x) \ne 0 \Rightarrow p(x) > 0$ (due to the property $p \ge 0$), while the set where $p'(x) = 0$ and $p(x) > 0$ does not contribute to the integral (2.1).

To get quantitative bounds on $I(p)$, we will consider the classes $\mathfrak{C}^2$ and $\mathfrak{C}^3$ and use the derivatives p'' and p'''.

Proposition 2.2. *Given a $\mathfrak{C}^2$-smooth probability density p, assume that $p(a) = p(b) = 0$ for some numbers $a < b$. Then for all $x \in [a, b]$,*

$$p'(x)^2 \le 2Cp(x), \quad C = \text{ess sup}_{a < x < b} \, p''(x). \tag{2.2}$$

In particular, if p is supported on (a, b), then

$$I(p) \le 2C \, (b - a).$$

One may extend the inequality in (2.2) to the whole real line with a similar constant.

Proposition 2.3. *Given a $\mathfrak{C}^2$-smooth probability density p, we have, for all $x \in \mathbb{R}$,*

$$p'(x)^2 \le 2Cp(x), \quad C = \text{ess sup}_{x \in \mathbb{R}} \, p''(x). \tag{2.3}$$

In particular,

$$\int_{-\infty}^{\infty} p'(x)^2 \, dx \le 2C.$$

The argument is based on two simple calculus lemmas.

Lemma 2.4. *Given a non-negative $\mathfrak{C}^2$-smooth function u on the interval (a, b), finite or not, assume that $u''(x) \le C$ a.e. for some constant C. If u satisfies one of the following two conditions*

(i) *u is non-decreasing with $\liminf_{x \downarrow a} u'(x) = 0$,*
(ii) *u is non-increasing with $\liminf_{x \uparrow b} u'(x) = 0$,*

then $C \ge 0$, and for all $x \in (a, b)$,

$$u'(x)^2 \le 2Cu(x). \tag{2.4}$$

Proof. Under the assumption (i), necessarily $C \ge 0$. Indeed, otherwise the derivative $u'(x)$ would be decreasing which implies that $u'(x) < u'(a+) = 0$ for all $a < x < b$. Hence, the function $u(x)$ itself would be decreasing.

Next, consider the function $\psi(x) = u'(x)^2 - 2Cu(x)$. It is locally absolutely continuous on (a, b) and has a Radon-Nikodym derivative ψ' satisfying

$$\psi'(x) = 2u'(x)\,(u''(x) - C) \le 0, \quad a < x < b \ \ (\text{a.e.})$$

Hence ψ is non-increasing, that is, $\psi(x) \le \psi(y)$ for all $a < y < x < b$. On the other hand,

$$\liminf_{y \downarrow a} \psi(y) = -2Cu(a+) \le 0.$$

Thus $\psi(x) \le 0$, and (2.4) follows.

The scenario as in (ii) is similar; it is reduced to (i) by applying the previous step to the function $x \to u(-x)$ on the interval $(-b, -a)$. $\qquad\square$

Lemma 2.5. *Let a $\mathfrak{C}^2$-smooth function $u \ge 0$ be defined on the interval (a, b), finite or not. If*

$$\liminf_{x \downarrow a} u'(x) = \liminf_{x \uparrow b} u'(x) = 0,$$

and $u''(x) \le C$ a.e. for some constant C, then $C \ge 0$ and (2.4) still holds in (a, b).

Proof. Since the function u' is continuous, the set $U = \{x \in (a, b) : u'(x) > 0\}$ is open and can be decomposed into at most countably many open disjoint intervals (a_k, b_k). If $a_k > a$, then necessarily $u'(a_k) = 0$. By the assumption, we also have $u'(a_k+) = 0$ if $a_k = a$. In both cases, one may apply Lemma 2.4 (i) to the interval (a_k, b_k), and we obtain (2.4) for all $x \in (a_k, b_k)$ with $C \ge 0$.

Similarly, the set $V = \{x \in (a, b) : u'(x) < 0\}$ can be decomposed into open disjoint intervals (a_n, b_n). If $b_k < b$, then $u'(b_k) = 0$, and, by the assumption, $u'(b_k-) = 0$ if $b_k = b$. Applying Lemma 2.4 (ii) to the interval (a_k, b_k), again we obtain the inequality (2.4) for all $x \in (a_k, b_k)$. $\qquad\square$

Proof of Propositions 2.2–2.3. Let p be $\mathfrak{C}^2$-smooth on the real line. First note that necessarily $C \ge 0$ in (2.2). Indeed, if $C < 0$, then p' is decreasing on $[a, b]$. But the assumption $p(a) = 0$ implies $p'(a) = 0$ (since $p \ge 0$), and then we would get that $p(x) < 0$ for all $a < x < b$. By a similar argument, we also have $p'(b) = 0$. Hence, one may apply Lemma 2.5 to the function $u = p$, and (2.2) follows.

Turning to the next proposition, again necessarily $C > 0$. Indeed, if $C \le 0$, then p would be concave on the whole real line, which is impossible for probability densities.

To prove the inequality (2.3), first assume that $p'(x) \to 0$ as $|x| \to \infty$ (this is always fulfilled, as will be shown in the next section). As in the proof of Lemma 2.5, consider the open set $U = \{x \in \mathbb{R} : p'(x) > 0\}$ and decompose it into open disjoint intervals (a_k, b_k). Necessarily, $b_k < \infty$ (otherwise, p would not be integrable). Moreover, if $a_k > -\infty$, then

$$p'(a_k) = p'(b_k) = 0.$$

Hence, one may apply Lemma 2.5 to the interval (a_k, b_k), and we obtain (2.3) for all $a_k \le x \le b_k$. In the case $a_k = -\infty$, Lemma 2.5 is also applicable due to the assumption $p'(-\infty) = 0$. A similar argument allows us to involve the points from the open set $V = \{x \in \mathbb{R} : p'(x) < 0\}$ as well, and we obtain (2.3) on the whole real line.

To remove the assumption on the derivative, consider a random variable X with density p together with an independent variable Z which has a C_0^∞-smooth density q supported on a bounded interval Δ. The convolution of p with density q_ε of the random variable εZ, $\varepsilon > 0$, is given by

$$p_\varepsilon(x) = (p * q_\varepsilon)(x) = \int_{-\infty}^{\infty} q_\varepsilon(x - y)\,p(y)\,dy = \int_\Delta p(x - \varepsilon y)\,q(y)\,dy.$$

This function is C^∞-smooth, and its first two derivatives are given by

$$
p'_\varepsilon(x) = \int_{-\infty}^{\infty} q'_\varepsilon(x-y)\, p(y)\, dy = \int_\Delta p'(x-\varepsilon y)\, q(y)\, dy, \tag{2.5}
$$

$$
p''_\varepsilon(x) = \int_{-\infty}^{\infty} q''_\varepsilon(x-y)\, p(y)\, dy = \int_\Delta p''(x-\varepsilon y)\, q(y)\, dy.
$$

The last equality shows that $p''_\varepsilon(x) \le C$ for any $x \in \mathbb{R}$. Moreover, since for every fixed $x \in \mathbb{R}$,

$$
\sup_{0<\varepsilon<1}\, \sup_{y\in\Delta}\, p(x-\varepsilon y) < \infty, \qquad \sup_{0<\varepsilon<1}\, \sup_{y\in\Delta}\, |p'(x-\varepsilon y)| < \infty,
$$

while $p(x-\varepsilon y) \to p(x)$ and $p'(x-\varepsilon y) \to p'(x)$ as $\varepsilon \to 0$, one may apply the Lebesgue dominated convergence theorem which gives

$$
p(x) = \lim_{\varepsilon\to 0} p_\varepsilon(x), \qquad p'(x) = \lim_{\varepsilon\to 0} p'_\varepsilon(x). \tag{2.6}
$$

In addition, using the property that q'_ε is bounded for any fixed ε, while $q'_\varepsilon(x-y) \to 0$ as $|x| \to \infty$ for every $y \in \mathbb{R}$, from the first equality in (2.5) it also follows that $p'_\varepsilon(x) \to 0$ as $|x| \to \infty$. Hence, one may apply the first step to the density p_ε, and we get that

$$
p'_\varepsilon(x)^2 \le 2C p_\varepsilon(x), \quad x \in \mathbb{R}.
$$

It remains to let $\varepsilon \to 0$ in this inequality and refer to (2.6). $\qquad\square$

Remark 2.6. In the above argument one may also use not necessarily compactly supported smoothing densities such as the standard normal density $q(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$, for example. Note that p and its derivative admit upper bounds $p(x) \le A\,(1+x^2)$, $|p'(x)| \le A\,(1+|x|)$ with some constant $A > 0$, so that

$$
p(x-\varepsilon y) \le A(1 + 2x^2 + 2y^2), \quad |p'(x-\varepsilon y)| \le A(1 + |x| + |y|),
$$

whenever $0 < \varepsilon \le 1$. Thus, for each fixed $x \in \mathbb{R}$, we have integrable majorants for the functions $y \to p(x-\varepsilon y)$ and $y \to p'(x-\varepsilon y)$ with respect to the probability measure $q(y)\, dy$. Hence, the Lebesgue dominated convergence theorem may be applied to obtain the desired relations in (2.3).

3 Decay of densities and their derivatives

Suppose that the constant

$$
C = \operatorname{ess\,sup}_{x\in\mathbb{R}} p''(x) \tag{3.1}
$$

is finite for a given $\mathfrak{C}^2$-smooth probability density p on the real line. This property turns out to be sufficient to bound $p(x)$ and $p'(x)$ in terms of the tails of the distribution function

$$
F(x) = \int_{-\infty}^{x} p(y)\, dy, \quad x \in \mathbb{R},
$$

associated to p. Suppose for a moment that p is everywhere positive, so that $F : \mathbb{R} \to (0,1)$ and its inverse function $F^{-1} : (0,1) \to \mathbb{R}$ represent $\mathfrak{C}^3$-smooth increasing bijections. The function

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

has a smooth continuous derivative

$$
L'(t) = \frac{p'(F^{-1}(t))}{p(F^{-1}(t))},
$$

and we have the identity

$$(L(t)^2)' = 2p'(F^{-1}(t)).$$

An application of the bound

$$|p'(x)| \leq \sqrt{2Cp(x)}$$

of Proposition 2.3 leads to

$$(L(t)^2)' \leq 2\sqrt{2CL(t)}.$$

Thus, the function $y(t) = L(t)^2$ satisfies the differential inequality $y'(t) \leq 2\sqrt{2C}\,y(t)^{1/4}$, which is the same as

$$\left(y(t)^{3/4}\right)' \leq \frac{3}{2}\sqrt{2C}, \qquad 0 < t < 1.$$

After integration over the interval (t_0, t), $0 < t_0 < t < 1$, we get

$$y(t)^{3/4} - y(t_0)^{3/4} \leq \frac{3}{2}\sqrt{2C}\,(t - t_0).$$

Necessarily $\liminf_{x \to -\infty} p(x) = 0$ which is equivalent to $\liminf_{t_0 \to 0} L(t_0) = 0$. Hence, letting t_0 approach zero in a proper way, from the above inequality we obtain that

$$y(t)^{3/4} \leq \frac{3}{2}\sqrt{2C}\,t,$$

that is,

$$L(t) \leq \left(\frac{3}{2}\sqrt{2C}\,t\right)^{2/3}.$$

Simplifying the numerical constant and changing the variable $t = F(x)$, we have been led to the inequality

$$p(x) \leq 2C^{1/3}\,F(x)^{2/3}, \quad x \in \mathbb{R}. \tag{3.2}$$

Now, to remove the assumption that p is positive, one may consider the convolutions p_ε as in the proof of Proposition 2.3, by choosing for q the density of the standard normal law. Hence, the above step yields the bound

$$p_\varepsilon(x) \leq 2C^{1/3}\,F_\varepsilon(x)^{2/3}, \quad x \in \mathbb{R},$$

in terms of the distribution function F_ε associated to p_ε. Here, according to Remark 2.6, one may let $\varepsilon \to 0$, and then we obtain in the limit the inequality (3.2) without any constraints. Moreover, interchanging the role of the points $-\infty$ and ∞, we have a similar bound

$$p(x) \leq 2C^{1/3}\,(1 - F(x))^{2/3}, \quad x \in \mathbb{R}.$$

Once we have established these estimates for $p(x)$, we also obtain similar ones for $p'(x)$, by applying Proposition 2.3. One may now summarize.

Proposition 3.1. *Using the constant C as in (3.1), we have for all $x \in \mathbb{R}$,*

$$p(x) \leq 2C^{1/3}\left(\min(F(x), 1 - F(x))\right)^{2/3} \tag{3.3}$$

and

$$|p'(x)| \leq 3C^{1/6}\left(\min(F(x), 1 - F(x))\right)^{1/3}. \tag{3.4}$$

In particular, $p(x) \to 0$ and $p'(x) \to 0$ as $|x| \to \infty$.

The right-hand sides of (3.3)-(3.4) may further be bounded in terms of absolute moments $\beta_s = \mathbb{E}\,|X|^s$ of a random variable X with density p. Indeed, by Chebyshev's inequality,

$$F(-x) + (1 - F(x)) \leq \frac{\beta_s}{x^s}, \quad x > 0.$$

Corollary 3.2. *Assuming that the constant C and the moment β_s are finite for a real number $s > 0$, we have for all $x \in \mathbb{R}$,*

$$p(x) \leq 2C^{1/3}\beta_s^{2/3}\frac{1}{|x|^{2s/3}},$$

$$|p'(x)| \leq 2C^{1/6}\beta_s^{1/3}\frac{1}{|x|^{s/3}}.$$

In particular, if $\beta_s < \infty$ for some $s > 3$, then p has a bounded total variation.

Another application of Propositions 2.3 and 3.1 concerns an alternative (classical) formula for the Fisher information.

Corollary 3.3. *For any probability density p from the class $\mathfrak{C}^2$ such that the constant C in (3.1) is finite, we have*

$$I(p) = -\int_{p(x)>0} p''(x)\log p(x)\,dx,$$

as long as the function $p''(x)\log p(x)$ is integrable on the set $\{x \in \mathbb{R} : p(x) > 0\}$.

Proof. The open set $U = \{x \in \mathbb{R} : p(x) > 0\}$ can be decomposed into disjoint intervals (a_k, b_k). Necessarily $p(a_k+) = p(b_k-) = 0$ including the cases $a_k = -\infty$ and $b_k = \infty$, by Proposition 3.1. Let $a_k < a < b < b_k$. Since $\log p(x)$ and $p'(x)$ are continuous functions with bounded total variations on $[a, b]$, one may integrate by parts, which gives

$$
\begin{aligned}
-\int_a^b p''(x)\log p(x)\,dx &= -\int_a^b \log p(x)\,dp'(x) \\
&= -p'(b)\log p(b) + p'(a)\log p(a) + \int_a^b \frac{p'(x)^2}{p(x)}\,dx.
\end{aligned}
$$

By Proposition 2.3,

$$|p'(a)\log p(a)| \leq \sqrt{2C}\,\sqrt{p(a)}\,|\log p(a)| \to 0 \quad \text{as} \quad a \to a_k,$$

and similarly $|p'(b)\log p(b)| \to 0$ as $b \to b_k$. Hence, the above formula becomes in the limit

$$-\int_{a_k}^{b_k} p''(x)\log p(x)\,dx = \int_{a_k}^{b_k} \frac{p'(x)^2}{p(x)}\,dx.$$

It remains to perform summation over all k. $\qquad\square$

4 Unimodal and quasi-unimodal distributions

Proposition 2.3 may also be applied to control the Fisher information for a large variety of densities like in the following statement.

Proposition 4.1. *Let p be a probability density of class $\mathfrak{C}^2$ with finite constant*

$$C = \operatorname{ess\,sup}_{x \in \mathbb{R}} p''(x).$$

Suppose that p is non-decreasing on a half-axis $(-\infty, a)$ and is non-increasing on a half-axis (b, ∞) for some $a \leq b$. Then $I(p)$ is finite. Moreover,

$$I(p) \leq 2C(b-a) + 2\sqrt{2C}\left(\sqrt{p(a)} + \sqrt{p(b)}\right). \tag{4.1}$$

Proof. Let $a_0 = \inf\{x : p(x) > 0\}$ and $b_0 = \sup\{x : p(x) > 0\}$. We may assume that $a_0 < a \le b < b_0$. Necessarily $p(a_0+) = p(b_0-) = 0$. By the monotonicity on the interval (a_0, a), the function $\sqrt{p}$ has finite total variation

$$\|\sqrt{p}\|_{\mathrm{TV}(a_0,a)} = \frac{1}{2} \int_{a_0}^{a} \frac{|p'(x)|}{\sqrt{p(x)}} \, dx = \sqrt{p(a)}.$$

Since $|p'(x)| \le \sqrt{2Cp(x)}$, we obtain that

$$\int_{a_0}^{a} \frac{p'^{\,2}}{p} \, dx = \int_{a_0}^{a} \frac{|p'(x)|}{\sqrt{p(x)}} \cdot \frac{|p'(x)|}{\sqrt{p(x)}} \, dx \le 2\sqrt{2Cp(a)}.$$

By a similar argument,

$$\int_{b}^{b_0} \frac{p'^{\,2}}{p} \, dx \le 2\sqrt{2Cp(b)}.$$

Finally, by the upper bound (2.3) from Proposition 2.3,

$$\int_{a}^{b} \frac{p'^{\,2}}{p} \, dx \le 2C(b - a).$$

It remains to add these three estamates. $\qquad\square$

One interesting case in (4.1) is when $a = b$. This corresponds to the so-called unimodal distributions on the real line with mode at the point a. With this in mind, the more general case $a < b$ may be referred to as the class of quasi-unimodal distributions.

In the unimodal case, (4.1) is simplified to

$$I(p) \le 4\sqrt{2Cp(a)}.$$

But then, one can further relax the basic hypothesis on the second derivative.

Proposition 4.2. *Let p be a continuous density of the unimodal distribution with mode at the point a. Suppose that p is $\mathfrak{C}^2$-smooth on the half-axis $(-\infty, a)$ and is $\mathfrak{C}^2$-smooth on the half-axis (a, ∞) with finite*

$$C_0 = \operatorname{ess\,sup}_{x<a} p''(x), \quad C_1 = \operatorname{ess\,sup}_{x>a} p''(x).$$

Then $I(p)$ is finite, and moreover,

$$I(p) \le 2\left(\sqrt{2C_0} + \sqrt{2C_1}\right)\sqrt{p(a)}. \tag{4.2}$$

Proof. Again, let $a_0 = \inf\{x : p(x) > 0\}$ and $b_0 = \sup\{x : p(x) > 0\}$, so that necessarily $a_0 < a < b_0$, by continuity of p. It is also necessary that $p'(a_0+) = p'(b_0-) = 0$. This follows from the fact that p' is continuous on $(-\infty, a)$ and $\lim_{x\to-\infty} p'(x) = 0$ (since otherwise p would not be integrable), and similarly for the second half-axis. Also, by the integrability argument, we have $C_0 \ge 0$ and $C_1 \ge 0$. Hence, we are in position to apply Lemma 2.4 with $u = p$, which yields

$$p'(x)^2 \le 2C_0 p(x) \text{ for } x < a, \qquad p'(x)^2 \le 2C_1 p(x) \text{ for } x > a.$$

It remains to repeat the argument from the proof of Proposition 4.1. $\qquad\square$

As an example, one may consider the symmetric exponential distribution with density

$$p(x) = \frac{1}{2} e^{-|x|}, \quad x \in \mathbb{R}.$$

It satisfies the assumptions of Proposition 4.2 with mode at $a = 0$ and $C_0 = C_1 = \frac{1}{2}$. Hence, by (4.2), $I(p) \le 2\sqrt{2}$. Note that $I(p) = 1$, while Proposition 4.1 is not applicable.

5 Total variation norm via higher order derivatives

To further develop upper bounds on the Fisher information, we have to see how one can estimate the L^1-norm of the first and second derivatives of a smooth density in terms of the L^1-norm of its next third derivative. This is a preliminary step towards Theorem 1.2.

Proposition 5.1. *For any function p in $\mathfrak{C}^2$,*

$$\int_{-\infty}^{\infty} |p'(x)|\,dx \leq \int_{-\infty}^{\infty} |p(x)|\,dx + \frac{2}{3}\int_{-\infty}^{\infty} |p''(x)|\,dx. \tag{5.1}$$

Perhaps, this relation is known (up to factors in front of the integrals). Note that a similar inequality

$$\int_{-\infty}^{\infty} |p'(x)|^2\,dx \leq \frac{1}{2}\int_{-\infty}^{\infty} |p(x)|^2\,dx + \frac{1}{2}\int_{-\infty}^{\infty} |p''(x)|^2\,dx \tag{5.2}$$

for the L^2-norms is obvious. Indeed, under proper integrability assumptions and applying the Plancherel theorem, (5.2) may be rewritten in terms of the Fourier transform

$$\hat{p}(t) = \int_{-\infty}^{\infty} e^{itx}\,p(x)\,dx, \quad t \in \mathbb{R},$$

as

$$\int_{-\infty}^{\infty} |t\hat{p}(t)|^2\,dt \leq \frac{1}{2}\int_{-\infty}^{\infty} |\hat{p}(t)|^2\,dt + \frac{1}{2}\int_{-\infty}^{\infty} |t^2\hat{p}(t)|^2\,dt.$$

This readily holds in view of the pointwise bound $t^2 \leq \frac{1}{2} + \frac{1}{2}\,t^4$.

However, the finiteness of the integrals on the right-hand side of (5.2) does not guarantee that p will have a finite total variation. For example, consider a C^∞-smooth function p on the real line which is vanishing for $x \leq 0$ and such that

$$p(x) = \frac{\sin(x^\alpha)}{x^\beta}, \quad x \geq 1,$$

with parameters $\alpha \geq \beta > 1$. It belongs to $L^1 \cap L^2$, while p' is not integrable ($\|p\|_{\mathrm{TV}} = \infty$). As easy to see, $p'' \in L^2$ if and only if $\alpha < \frac{1}{4}\,(2\beta + 3)$ which may happen when $1 < \beta < \frac{3}{2}$.

Involving higher order derivatives, one may get similar relations in the spirit of the inequality (5.1), like the following ones which we prefer to state in the multiplicative form.

Corollary 5.2. *For any function p in $\mathfrak{C}^3$,*

$$\int_{-\infty}^{\infty} |p'(x)|\,dx \leq \frac{4}{3}\left(\int_{-\infty}^{\infty} |p(x)|\,dx\right)^{2/3}\left(\int_{-\infty}^{\infty} |p'''(x)|\,dx\right)^{1/3},$$

$$\int_{-\infty}^{\infty} |p''(x)|\,dx \leq \frac{4}{3}\left(\int_{-\infty}^{\infty} |p(x)|\,dx\right)^{1/3}\left(\int_{-\infty}^{\infty} |p'''(x)|\,dx\right)^{2/3}.$$

Proof. Applying (5.1) to the functions $p_\lambda(x) = p(\lambda x)$ with parameter $\lambda > 0$, we get

$$\int_{-\infty}^{\infty} |p'(x)|\,dx \leq \frac{1}{\lambda}\int_{-\infty}^{\infty} |p(x)|\,dx + \frac{2\lambda}{3}\int_{-\infty}^{\infty} |p''(x)|\,dx,$$

which may be optimized over λ to yield

$$\int_{-\infty}^{\infty} |p'(x)|\,dx \leq \left(\frac{4}{3}\int_{-\infty}^{\infty} |p(x)|\,dx \int_{-\infty}^{\infty} |p''(x)|\,dx\right)^{1/2}. \tag{5.3}$$

In other words, we arrive at the convexity-type relation

$$a_1 \leq \frac{1}{2} a_0 + \frac{1}{2} a_2 + h, \quad h = \frac{1}{2} \log \frac{4}{3}, \tag{5.4}$$

for the sequence

$$a_k = \log A_k, \quad A_k = \int_{-\infty}^{\infty} |p^{(k)}(x)| \, dx.$$

An application of (5.3) to p' in place of p leads to $a_2 \leq \frac{1}{2} a_1 + \frac{1}{2} a_3 + h$, which, by (5.4), implies $a_1 \leq \frac{1}{2} a_0 + \frac{1}{2} (\frac{1}{2} a_1 + \frac{1}{2} a_3 + h) + h$. That is,

$$a_1 \leq \frac{2}{3} a_0 + \frac{1}{3} a_3 + 2h, \quad \text{or} \quad A_1 \leq e^{2h} A_0^{2/3} A_3^{1/3}.$$

By a similar argument, $a_2 \leq \frac{1}{3} a_0 + \frac{2}{3} a_3 + 2h$, or $A_2 \leq e^{2h} A_0^{1/3} A_3^{2/3}$ which was required. $\square$

Involving further derivatives in a similar manner, we arrive at the following:

Corollary 5.3. *If the function p in $\mathfrak{C}^l$ is integrable and has an integrable derivative $p^{(l)}$ of order $l \geq 2$, then all intermediate derivatives $p^{(k)}$, $1 \leq k \leq l-1$, are integrable as well.*

Proof of Proposition 5.1. First, let us derive an upper bound on the L^1-norm of p' over the unit interval. One may start with the weighted L^1-Poincaré-type inequality

$$\int_0^1 |u(x) - m| \, dx \leq 2 \int_0^1 |u'(x)| \, x(1-x) \, dx, \qquad m = \int_0^1 u(x) \, dx. \tag{5.5}$$

Here, an equality is attained in the asymptotic sense for the indicator function $u = 1_{[0,1/2]}$. To prove it, note that, by Jensen's inequality, the left integral in (5.5) does not exceed

$$
\begin{aligned}
\int_0^1 \int_0^1 |u(x) - u(y)| \, dx \, dy &= 2 \iint_{0<x<z<y<1} \left| \int_x^y u'(z) \, dz \right| dx \, dy \\
&\leq 2 \iint_{0<x<z<y<1} \left[\int_x^y |u'(z)| \, dz \right] dx \, dy \\
&= 2 \int_0^1 |u'(z)| \, z(1-z) \, dz.
\end{aligned}
$$

This proves (5.5). Using this inequality with $u = p'$, we get

$$\int_0^1 |p'(x) - m| \, dx \leq 2 \int_0^1 |p''(x)| \, x(1-x) \, dx,$$

implying

$$\int_0^1 |p'(x)| \, dx \leq |m| + 2 \int_0^1 |p''(x)| \, x(1-x) \, dx, \qquad m = p(1) - p(0). \tag{5.6}$$

Next, we need to derive an upper bound on the increment m of $p(x)$ on $[0,1]$ analogously to the right-hand side in (5.1). By Taylor's integral formula, for all $h \in \mathbb{R}$,

$$p(h) = p(0) + p'(0)h + h^2 \int_0^1 p''(th) \, (1-t) \, dt.$$

Writing this inequality with $-h$ in place of h and averaging, we get

$$p(0) = \frac{p(h) + p(-h)}{2} - \frac{1}{2} \int_0^1 h^2 \left(p''(th) + p''(-th) \right) (1-t) \, dt.$$

Let us now integrate this identity over $0 < h < 1$. This gives another general identity

$$p(0) = \frac{1}{2} \int_{-1}^{1} p(x)\, dx - \frac{1}{4} \int_0^1 p''(x)\,(1-x)^2\, dx - \frac{1}{4} \int_{-1}^0 p''(x)\,(1+x)^2\, dx.$$

Applying it to the function $x \to p(1+x)$, we also get

$$p(1) = \frac{1}{2} \int_0^2 p(x)\, dx - \frac{1}{4} \int_1^2 p''(x)\,(2-x)^2\, dx - \frac{1}{4} \int_0^1 p''(x)\, x^2\, dx.$$

Hence

$$\begin{aligned}
p(1) - p(0) &= \frac{1}{2} \int_1^2 p(x)\, dx - \frac{1}{2} \int_{-1}^0 p(x)\, dx + \frac{1}{4} \int_{-1}^0 p''(x)\,(1+x)^2\, dx \\
&\quad + \frac{1}{4} \int_0^1 p''(x)\,(1-2x)\, dx - \frac{1}{4} \int_1^2 p''(x)\,(2-x)^2\, dx,
\end{aligned}$$

and thus

$$\begin{aligned}
|p(1) - p(0)| &\leq \frac{1}{2} \int_{-1}^0 |p(x)|\, dx + \frac{1}{2} \int_1^2 |p(x)|\, dx + \frac{1}{4} \int_{-1}^0 |p''(x)|\,(1+x)^2\, dx \\
&\quad + \frac{1}{4} \int_0^1 |p''(x)|\,|1-2x|\, dx + \frac{1}{4} \int_1^2 |p''(x)|\,(2-x)^2\, dx.
\end{aligned}$$

Thus, together with (5.6) we arrive at the similar bound

$$\int_0^1 |p'(x)|\, dx \leq \frac{1}{2} \int_{-1}^0 |p(x)|\, dx + \frac{1}{2} \int_1^2 |p(x)|\, dx + \int_{-\infty}^{\infty} |p''(x)|\, w(x)\, dx \qquad (5.7)$$

with

$$w(x) = \begin{cases}
0 & \text{for} \quad x < -1 \ \text{ and } \ x > 2, \\
\frac{1}{4}\,(1+x)^2 & \text{for} \quad -1 \leq x \leq 0, \\
\frac{1}{4}\,|1-2x| + 2x(1-x) & \text{for} \quad 0 \leq x \leq 1, \\
\frac{1}{4}\,(2-x)^2 & \text{for} \quad 1 \leq x \leq 2.
\end{cases}$$

Let us now apply the relation (5.7) to the functions $p(x+k)$ and perform summation over all integers k. This will give

$$\int_{-\infty}^{\infty} |p'(x)|\, dx \leq \int_{-\infty}^{\infty} |p(x)|\, dx + \int_{-\infty}^{\infty} |p''(x)|\, W(x)\, dx \qquad (5.8)$$

with weight function $W(x) = \sum_{k \in \mathbb{Z}} w(x+k)$. One can easily evaluate it using the property that it is 1-periodic. Restricting ourselves to the values $x \in [0,1]$, we have

$$\begin{aligned}
W(x) &= w(x-1) + w(x) + w(x+1) \\
&= \frac{1}{4}\, x^2 + \frac{1}{4}\,(1-x)^2 + \frac{1}{4}\,|1-2x| + 2x(1-x).
\end{aligned}$$

This function is symmetric about the point $x = \frac{1}{2}$, and $W(x) = \frac{1}{2}\,(1 + 2x - 3x^2)$ for $0 \leq x \leq \frac{1}{2}$. The latter expression is maximized at $x = \frac{1}{3}$ with $W(\frac{1}{3}) = \frac{2}{3}$. Thus, $W(x) \leq \frac{2}{3}$ for all $x \in \mathbb{R}$, so that (5.8) yields the desired inequality (5.1). $\qquad \square$

6 The use of the third derivative

The boundedness condition for the second derivative p'' is guaranteed, for example, by the integrability of the next derivative p'''. Hence, some of the previous statements can be made in terms of the L^1-norm of p'''. As a first step towards the one-dimensional

variant of Theorem 1.2, here we prove the following relation. We basically follow the arguments described in the book by Bogachev [6] and employ Corollary 5.2. Let us use the notation

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

for all non-negative functions p from the class $\mathfrak{C}^1$ (even if p is not a probability density).

The next inequality slightly sharpens (1.7).

Proposition 6.1. *For any non-negative function p of class $\mathfrak{C}^3$,*

$$I(p) \leq 2 \int_{-\infty}^{\infty} |p'(x)| \, dx + 4 \int_{-\infty}^{\infty} |p''(x)| \, dx + 2 \int_{-\infty}^{\infty} |p'''(x)| \, dx. \tag{6.1}$$

Proof. One may assume that all integrals on the right-hand side are finite, and p is not identically zero. In particular, p, p' and p'' are bounded on the whole real line.

The set $U = \{x \in \mathbb{R} : p'(x) > 0\}$ is non-empty and open. Let us decompose it into at most countably many disjoint open intervals, and let $\Delta = (a, b)$ be one of these intervals. Necessarily $b < \infty$, $p'(b) = 0$, and $p'(a) = 0$ in the case $a > -\infty$. Since $p'(x)$ is vanishing at infinity (Proposition 3.1), we also have $p'(-\infty) = 0$ if $a = -\infty$.

By the monotonicity, $p(x) > 0$ on Δ, and we may define a positive function

$$v(x) = \frac{p'(x)^2}{p(x)},$$

which is $\mathfrak{C}^2$-smooth on the interval Δ. Put

$$\Delta' = \{x \in \Delta : v(x) \leq 2p''(x)\}, \qquad \Delta'' = \{x \in \Delta : v(x) > 2p''(x)\}.$$

Note that $p'' > 0$ on Δ' and

$$\int_{\Delta'} \frac{p'(x)^2}{p(x)} \, dx \leq 2 \int_{\Delta'} p''(x) \, dx. \tag{6.2}$$

Turning to the set Δ'', note that it is open and can be decomposed into at most countably many disjoint intervals (a_k, b_k). In particular, $v(a_k) = 2p''(a_k)$ in the case $a_k > a$.

The function $v(x)$ has a continuous derivative satisfying

$$v'(x) = \frac{p'(x) \, (2p(x) \, p''(x) - p'(x)^2)}{p(x)^2} < 0, \quad x \in \Delta''.$$

Hence, $v(x)$ is decreasing on every (a_k, b_k). As a consequence, if $a_k > a$,

$$\frac{p'(x)^2}{p(x)} \leq 2p''(a_k), \quad a_k < x < b_k. \tag{6.3}$$

In the case $a_k = a > -\infty$, recall that $p'(a_k) = 0$. Given $\varepsilon > 0$ such that $a + \varepsilon < b_k$, one may apply Lemma 2.4 (i) to the function $u = p$ on the interval $(a, a + \varepsilon)$ which yields

$$p'(x)^2 \leq 2C_\varepsilon p(x), \quad C_\varepsilon = \sup_{a < y < a+\varepsilon} p''(y) \geq 0,$$

for every $x \in (a, a + \varepsilon)$. By the monotonicity of v, it follows that

$$\sup_{a < x < b_k} \frac{p'(x)^2}{p(x)} = \sup_{a < x < a+\varepsilon} \frac{p'(x)^2}{p(x)} \leq 2C_\varepsilon.$$

Letting $\varepsilon \to 0$ and using the continuity of p'', we have $C_\varepsilon \to p''(a)$, so that the above relation leads to (6.3) for this case as well. Note that, by (6.3), necessarily $p''(a_k) > 0$.

As for the remaining case $a_k = a = -\infty$, let us show that it is impossible. Since v is decreasing on the half-axis $(a_k, b_k) = (-\infty, b_k)$, this would imply that, for any $z < b_k$,

$$\int_{-\infty}^z \frac{p'(x)^2}{p(x)}\, dx = \int_{-\infty}^z v(x)\, dx \geq \int_{-\infty}^z v(z)\, dx = \infty.$$

On the other hand, since $|p'(x)| \leq \sqrt{2Cp(x)}$ for all $x \in \mathbb{R}$ with constant $C = \sup_x p''(x)$ (Proposition 2.3) and since $p'(x) > 0$ on U, we have

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

$$\leq 2\sqrt{2C} \int_{-\infty}^z \sqrt{p(x)}'\, dx = 2\sqrt{2Cp(z)} < \infty,$$

where we used the property $p(x) \to 0$ as $x \to -\infty$ (Proposition 3.1).

Thus, necessarily $-\infty < a_k < b_k \leq b < \infty$, $p''(a_k) > 0$, and the inequality (6.3) holds true on every interval (a_k, b_k). Applying it, we get

$$J_k \equiv \int_{a_k}^{b_k} \frac{p'(x)^2}{p(x)}\, dx \leq \sqrt{2p''(a_k)} \int_{a_k}^{b_k} \frac{p'(x)}{\sqrt{p(x)}}\, dx$$

$$= 2\sqrt{2p''(a_k)} \left(\sqrt{p(b_k)} - \sqrt{p(a_k)} \right).$$

Using the simple inequality $(\sqrt{t} - \sqrt{s})^2 \leq t - s$ $(t \geq s)$, it follows that

$$J_k^2 \leq 8p''(a_k)\, (p(b_k) - p(a_k)) = 8p''(a_k) \int_{a_k}^{b_k} p'(x)\, dx.$$

Hence (using $\sqrt{ts} \leq \frac{1}{2} t + \frac{1}{2} s$, $t, s \geq 0$), we get

$$J_k \leq 2 \int_{a_k}^{b_k} p'(x)\, dx + p''(a_k). \tag{6.4}$$

To further estimate the right-hand side, consider two scenarios. If $p''(a_k) \leq 2p''(x)$ in (a_k, b_k), then, by (6.3), whenever $a_k < x < b_k$,

$$\frac{p'(x)^2}{p(x)} \leq 2p''(a_k) \leq 4p''(x),$$

implying

$$J_k \leq 4 \int_{a_k}^{b_k} |p''(x)|\, dx,$$

which is similar to (6.2). In the other case, there is a point $x_k \in (a_k, b_k)$ such that $p''(a_k) > 2p''(x_k)$. Then

$$p''(a_k) < 2\, (p''(a_k) - p''(x_k))$$

$$= 2 \int_{x_k}^{b_k} p'''(x)\, dx \leq 2 \int_{a_k}^{b_k} |p'''(x)|\, dx,$$

which, by (6.4), gives

$$J_k \leq 2 \int_{a_k}^{b_k} p'(x)\, dx + 2 \int_{a_k}^{b_k} |p'''(x)|\, dx.$$

One can unite both scenarios using the formally weaker relation

$$J_k \leq 2 \int_{a_k}^{b_k} |p'(x)| \, dx + 4 \int_{a_k}^{b_k} |p''(x)| \, dx + 2 \int_{a_k}^{b_k} |p'''(x)| \, dx.$$

Let us now perform summation over all k, which leads to

$$\int_{\Delta''} \frac{p'(x)^2}{p(x)} \, dx \leq 2 \int_{\Delta''} |p'(x)| \, dx + 4 \int_{\Delta''} |p''(x)| \, dx + 2 \int_{\Delta''} |p'''(x)| \, dx.$$

Here the right-hand side dominates the one of (6.2). Hence, adding the two inequalities with integrals taken over Δ' and Δ'', we get

$$\int_{\Delta} \frac{p'(x)^2}{p(x)} \, dx \leq 2 \int_{\Delta} |p'(x)| \, dx + 4 \int_{\Delta} |p''(x)| \, dx + 2 \int_{\Delta} |p'''(x)| \, dx.$$

It is time to perform summation over all intervals Δ's contained in the decomposition of U, which leads to the similar bound

$$\int_{U} \frac{p'(x)^2}{p(x)} \, dx \leq 2 \int_{U} |p'(x)| \, dx + 4 \int_{U} |p''(x)| \, dx + 2 \int_{U} |p'''(x)| \, dx.$$

A similar relation holds true when integrating over the set $V = \{x \in \mathbb{R} : p'(x) < 0\}$ in place of U (alternatively, one may apply the previous step to the function $x \to p(-x)$). Adding the two inequalities with integrals over U and V, we then get (6.1). $\qquad\square$

7 Theorems 1.1-1.2 in the one-dimensional case

We are prepared to prove the one-dimensional variant of Theorem 1.2.

Proposition 7.1. *For any probability density $p \in \mathfrak{C}^3$, we have*

$$I(p) \leq 10 \left(\int_{-\infty}^{\infty} |p'''(x)| \, dx \right)^{2/3}. \tag{7.1}$$

Proof. Denote by A the integral in (7.1). We apply Corollary 5.2 to (6.1) to get that

$$I(p) \leq \frac{8}{3} A^{1/3} + \frac{16}{3} A^{2/3} + 2A.$$

This inequality is not invariant under rescaling of the space variable. So, let us apply it to the probability density functions $p_\lambda(x) = \lambda p(\lambda x)$ with parameter $\lambda > 0$. Then we get

$$I(p) \leq \frac{8}{3\lambda} A^{1/3} + \frac{16}{3} A^{2/3} + 2\lambda A.$$

Optimizing over all λ, we arrive at $I(p) \leq c A^{2/3}$ with constant $c = \frac{8}{\sqrt{3}} + \frac{16}{3} < 10$. $\qquad\square$

We now consider a particular case of the inequality (7.1) when p has a convolution structure.

First, let us remind some of the basic properties of this operation. Given integrable functions p_1 and p_2, the convolution

$$(p_1 * p_2)(x) = \int_{-\infty}^{\infty} p_1(x - y) \, p_2(y) \, dy, \quad x \in \mathbb{R}, \tag{7.2}$$

is defined a.e. and represents an integrable function with the L^1-norm

$$\|p_1 * p_2\|_1 \leq \|p_1\|_1 \, \|p_2\|_1. \tag{7.3}$$

Moreover, if p_j are non-negative, the integral in (7.2) is well-defined for any fixed x (although it may be infinite); it does not change when changing p_j on a set of Lebesgue measure zero.

In general, the convolution improves smoothing properties. For example, if p_1 and p_2 are bounded, the function $p = p_1 * p_2$ is bounded and uniformly continuous. In this case, both p_1 and p_2 belong to $L^2(\mathbb{R})$, so do their Fourier transforms

$$\hat{p}_j(t) = \int_{-\infty}^{\infty} e^{itx}\, p_j(x)\, dx, \quad t \in \mathbb{R},$$

by the Plancherel theorem. Hence, the Fourier transform $\hat{p} = \hat{p}_1 \hat{p}_2$ is an integrable function, which implies the desired assertion by applying the inverse Fourier formula.

We will need the following elementary statement.

Lemma 7.2. *If non-negative integrable functions p_j, $1 \leq j \leq l$, are absolutely continuous and have integrable Radon-Nikodym derivatives p_j', the convolution $p = p_1 * \cdots * p_l$ belongs to the class $\mathfrak{C}^l$. Moreover, its derivatives up to order $l-1$ are bounded and integrable, while the l-th Radon-Nikodym derivative of p represents the convolution*

$$p^{(l)} = p_1' * \cdots * p_l' \tag{7.4}$$

with the L^1-norm satisfying

$$\|p^{(l)}\|_1 \leq \|p_1'\|_1 \ldots \|p_l'\|_1. \tag{7.5}$$

Proof. For simplicity, let us consider the case $l = 2$. Put $q_j = p_j'$, $j = 1, 2$. According to (7.2), for any $x \in \mathbb{R}$,

$$p(x) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{x-y} q_1(z)\, dz \right] p_2(y)\, dy = \iint_{z+y \leq x} q_1(z)\, p_2(y)\, dy\, dz.$$

In particular,

$$|p(x)| \leq \|q_1\|_1 \|p_2\|_1 < \infty,$$

so that p is bounded. After change of variable $y = \xi - z$, the last double integral may be rewritten as

$$p(x) = \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} q_1(z)\, p_2(\xi - z)\, dz \right] d\xi = \int_{-\infty}^{x} (q_1 * p_2)(\xi)\, d\xi.$$

This equality shows that p is absolutely continuous and has an integrable Radon-Nikodym derivative $p' = q_1 * p_2$. Thus,

$$p'(x) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{x-y} q_2(z)\, dz \right] q_1(y)\, dy = \iint_{z+y \leq x} q_2(z)\, q_1(y)\, dy\, dz.$$

The last integral is finite and represents a continuous bounded function of x, with

$$|p'(x)| \leq \|q_1\|_1 \|q_2\|_1 < \infty.$$

After the same change of variable $y = \xi - z$, the last double integral may be rewritten as

$$p'(x) = \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} q_2(z)\, q_1(\xi - z)\, dz \right] d\xi = \int_{-\infty}^{x} (q_1 * q_2)(\xi)\, d\xi.$$

This equality shows that p' is absolutely continuous and has an integrable Radon-Nikodym derivative $p'' = q_1 * q_2$. In particular, $p \in \mathfrak{C}^2$. The inequality in (7.5) follows from (7.3).

The general case $l \geq 2$ in (7.4)-(7.5) is similar. $\qquad\square$

In the case $l = 3$, one may combine Lemma 7.2 with Proposition 7.1 to obtain the following consequence from (7.1).

Proposition 7.3. *Given absolutely continuous probability densities p_j, $j = 1, 2, 3$, the convolution $p = p_1 * p_2 * p_3$ belongs to the class $\mathfrak{C}^3$ and has finite Fisher information satisfying*

$$I(p) \leq 10 \, (b_1 b_2 b_3)^{2/3}, \quad b_j = \|p_j'\|_1. \tag{7.6}$$

The inequality (7.6) may be further extended to the class of probability densities of bounded variation, by a suitable approximation. On the real line, the total variation semi-norm of a function p is defined by

$$\|p\|_{\mathrm{TV}} = \sup \sum_{k=1}^{N} |p(x_k) - p(x_{k-1})|, \tag{7.7}$$

where the supremum is taken over all collections of points $x_0 < x_1 < \cdots < x_N$. If this semi-norm is finite, necessarily the limits

$$p(x-) = \lim_{y \uparrow x} p(y), \quad p(x+) = \lim_{y \downarrow x} p(y)$$

exist and are finite for all $x \in \mathbb{R}$. Without loss of generality, one may always assume that the value $p(x)$ is located between these limits. For example, one may require that $p(x+) = p(x)$, that is, p is right-continuous. With this requirement, the value in (7.7) coincides with the so-called essential total variation semi-norm, which is consistent with the definition (1.5).

If p is integrable, then necessarily $p(-\infty) = p(\infty) = 0$. Hence, being restricted to the linear space of all integrable right-continuous functions p of bounded variation, $\|p\|_{\mathrm{TV}}$ represents a norm. If p is absolutely continuous and has a Radon-Nikodym derivative p', then

$$\|p\|_{\mathrm{TV}} = \|p'\|_1 = \int_{-\infty}^{\infty} |p'(x)| \, dx. \tag{7.8}$$

Like the Fisher information, the convolution of p with an arbitrary probability density q does not increase the total variation norm:

$$\|p * q\|_{\mathrm{TV}} \leq \|p\|_{\mathrm{TV}}. \tag{7.9}$$

Proof of Theorem 1.1 ($n = 1$). Let p_j, $j = 1, 2, 3$, be probability densities with finite total variation norms $b_j = \|p_j\|_{\mathrm{TV}}$. Introduce the normal density $\varphi_\varepsilon(x) = \frac{1}{\varepsilon\sqrt{2\pi}} \, e^{-x^2/2\varepsilon^2}$ with mean zero and standard deviation $\varepsilon > 0$ and define the convolutions

$$p_{j,\varepsilon} = p_j * \varphi_\varepsilon, \qquad p_\varepsilon = p_{1,\varepsilon} * p_{2,\varepsilon} * p_{3,\varepsilon} = p * \varphi_{\varepsilon\sqrt{3}}.$$

All these functions represent C^∞-smooth probability densities, so that the relations (7.6) and (7.8) are applicable, which yield

$$
\begin{aligned}
I(p_\varepsilon) &\leq 10 \left(\|p_{1,\varepsilon}\|_{\mathrm{TV}} \|p_{2,\varepsilon}\|_{\mathrm{TV}} \|p_{3,\varepsilon}\|_{\mathrm{TV}} \right)^{2/3} \\
&\leq 10 \left(\|p_1\|_{\mathrm{TV}} \|p_2\|_{\mathrm{TV}} \|p_3\|_{\mathrm{TV}} \right)^{2/3},
\end{aligned}
$$

where we made use of (7.9) on the last step. It remains to apply the lower semi-continuity of the Fisher information (cf. [3], Proposition 3.1, or Section 14 below): If random variables ξ_k are convergent weakly in distribution to a random variable ξ, then

$$I(\xi) \leq \liminf_{k \to \infty} I(\xi_k).$$

As a consequence, $I(p) = \lim_{\varepsilon \to 0} I(p_\varepsilon)$. Thus, Theorem 1.1 is proved with $c = 10$. $\qquad\square$

8 Weak derivatives

In spaces of higher dimensions, the general theory about the Fisher information is somewhat different than in dimension one. For example, for the finiteness of $I(p)$, the density p does not need be bounded and continuous anymore, in contrast with the one-dimensional case. We refer an interested reader to [4] for related issues.

Upper bounds on the Fisher information for probability densities on $\mathbb{R}^n$ may be explored in appropriate Sobolev spaces. A main approach to the definition of Sobolev spaces is based on the integration by parts formula. Let us recall some basic notations and facts in this theory and give some additional remarks (for background we refer to [17], [10]). As usual, $C_0^\infty(\mathbb{R}^n)$ denotes the space of all compactly supported functions w on $\mathbb{R}^n$ that have continuous partial derivatives of all orders.

Definition 8.1. *Let $\alpha = (\alpha_1, \dots, \alpha_n)$ be a multi-index with $|\alpha| = \alpha_1 + \cdots + \alpha_n$ (α_j are non-negative integers). Given a locally integrable function u on $\mathbb{R}^n$, a locally integrable function v on $\mathbb{R}^n$ is called a generalized α-th derivative of u, and we write $v = D^\alpha u$, if*

$$\int u\, D^\alpha w\, dx = (-1)^{|\alpha|} \int vw\, dx \qquad (8.1)$$

for all $w \in C_0^\infty(\mathbb{R}^n)$.

Here and elsewhere $D^\alpha w = \partial_{x_1}^{\alpha_1} \dots \partial_{x_n}^{\alpha_n} w$ denotes the corresponding partial derivative. We are especially interested in partial derivatives along one variable only and then we also write

$$D_i^l u = \partial_{x_i}^l u \ (i = 1, \dots, n), \quad D_i u = \partial_{x_i} u.$$

If $v = D^\alpha u$ exists, this generalized partial derivative is defined uniquely up to a set of Lebesgue measure zero. Of course, if u has continuous usual partial derivatives of orders up to $|\alpha|$, the generalized α-th derivative exists and may be chosen to be the usual one.

Like in the usual differentiation, generalized derivatives are commutative with respect to α and have a semi-group structure: If u has a generalized α-th derivative $D^\alpha u$, which in turn has a generalized β-th derivative $v = D^\beta D^\alpha u$, then u has a generalized $(\alpha + \beta)$-th derivative v. That is, $D^{\alpha+\beta} = D^\beta D^\alpha = D^\alpha D^\beta$ for all multi-indices α and β.

In dimension $n = 1$ with $\alpha = 1$, Definition 8.1 returns us to the setting of locally absolutely continuous functions for which Radon-Nikodym derivatives serve as generalized derivatives. More generally, one may give the following characterization.

Proposition 8.2. *A locally integrable function u on the real line has a generalized l-th derivative v of an integer order $l \geq 1$, if and only if $u = \tilde{u}$ a.e. for some $\tilde{u}$ from the class $\mathfrak{C}^l$. In this case, $v = \tilde{u}^{(l)}$ a.e.*

In the case $l = 1$, Proposition 8.2 is thus telling us that a locally integrable function on the real line has a generalized derivative, if and only if after a modification on a set of Lebesgue measure zero, it will be locally absolutely continuous.

In the proof we involve the so-called regularized functions which are commonly used for approximation of Sobolev functions. Let $\omega \in C_0^\infty(\mathbb{R}^n)$ be non-negative and compactly supported, with $\int \omega\, dx = 1$. So, it is a probability density. The probability densities

$$\omega_\varepsilon(x) = \varepsilon^{-n}\, \omega(x/\varepsilon), \quad x \in \mathbb{R}^n \ (\varepsilon > 0),$$

are also compactly supported and belong to the class C_0^∞. They are called regularizers.

Definition 8.3. *Given a locally integrable function u on $\mathbb{R}^n$, define the convolutions*

$$\begin{aligned} u_\varepsilon(x) &= (u * \omega_\varepsilon)(x) = \int u(x-y)\omega_\varepsilon(y)\, dy \\ &= \int \omega_\varepsilon(x-y)u(y)\, dy, \quad x \in \mathbb{R}^n. \end{aligned} \qquad (8.2)$$

They are called the regularized functions.

These functions are locally integrable, belong to the class $C^\infty(\mathbb{R}^n)$, and we have the commutativity $(\omega_{\varepsilon_1})_{\varepsilon_2} = (\omega_{\varepsilon_2})_{\varepsilon_1}$, $\varepsilon_j > 0$. Let us list a few elementary basic properties of the regularized functions, in which the choice of the regulizer ω is irrelevant.

Lemma 8.4. *Let u be a locally integrable function on $\mathbb{R}^n$ and α be a multi-index.*

1) *$u_\varepsilon(x) \to u(x)$ as $\varepsilon \to 0$ at every Lebesgue point x of u, hence almost everywhere.*
2) *We have $D^\alpha u_\varepsilon = u * D^\alpha \omega_\varepsilon$.*
3) *Moreover, $D^\alpha u_\varepsilon = (D^\alpha u)_\varepsilon$ provided that u has a generalized derivative $D^\alpha u$.*
4) *In this case, $D^\alpha u_\varepsilon(x) \to D^\alpha u(x)$ almost everywhere (by combining 1) and 3)).*

Lemma 8.5. *Given a locally integrable function u on $\mathbb{R}^n$ and a multi-index α with $l = |\alpha|$, suppose that, for any $w \in C_0^\infty(\mathbb{R}^n)$,*

$$\int u D^\alpha w \, dx = 0. \tag{8.3}$$

Then $u = \tilde{u}$ a.e. for some polynomial $\tilde{u}(x_1, \ldots, x_n)$ in n real variables of degree at most $l - 1$.

Proof. Starting from (8.3), we obtain a similar equality for the regularized functions, i.e.

$$\int u_\varepsilon D^\alpha w \, dx = 0. \tag{8.4}$$

Indeed, according to (8.2) and applying Fubini's theorem so as to justify the change of the order of integration, we see that the above integral is equal to

$$\int \left[\int u(x-y)\omega_\varepsilon(y) \, dy \right] D^\alpha w(x) \, dx = \int \left[\int u(x-y) D^\alpha w(x) \, dx \right] \omega_\varepsilon(y) \, dy.$$

But, by (8.3), the inner integral on the right-hand side is vanishing for any fixed $y \in \mathbb{R}^n$.

Now, since u_ε is C^∞-smooth, one may integrate in (8.4) by parts and conclude that $\int w D^\alpha u_\varepsilon \, dx = 0$ for all $w \in C_0^\infty(\mathbb{R}^n)$. This implies that $D^\alpha u_\varepsilon(x) = 0$ for all $x \in \mathbb{R}^n$, which is only possible when u_ε is a polynomial of degree at most $d = l - 1$. It remains to apply the property 1) and note that the pointwise limit of polynomials of degree at most d is a polynomial of degree at most d. $\square$

Proof of Poposition 8.2. In one direction (the sufficiency part), one may assume that $\tilde{u} = u$, so that the function u belongs to the class $\mathfrak{C}^l$. In particular, its l-th derivative $u^{(l)}$, being understood in the Radon-Nikodym sense, is locally integrable on the real line. Hence

$$u^{(l-1)}(b) - u^{(l-1)}(a) = \int_a^b u^{(l)}(x) \, dx \quad \text{for all } a < b,$$

which implies that $u^{(l-1)}$ has a bounded variation on every bounded interval. Then one may integrate by parts to get that

$$\int_{-\infty}^\infty w(x) \, u^{(l)}(x) \, dx = - \int_{-\infty}^\infty w'(x) \, u^{(l-1)}(x) \, dx$$

for all $w \in C_0^\infty(\mathbb{R})$. Since the first $l-1$ derivatives of u are continuous, one may further integrate by parts which leads to the desired equality

$$\int_{-\infty}^\infty w(x) \, u^{(l)}(x) \, dx = (-1)^l \int_{-\infty}^\infty w^{(l)}(x) \, u(x) \, dx.$$

According to (8.1), this shows that $u^{(l)}$ serves as an l-th generalized derivative for u.

Arguing in the opposite direction, suppose that, for some locally integrable function $v(x)$, we have

$$\int_{-\infty}^{\infty} w^{(l)}(x)\, u(x)\, dx = (-1)^l \int_{-\infty}^{\infty} v(x)\, w(x)\, dx \tag{8.5}$$

for all $w \in C_0^{\infty}(\mathbb{R})$. Introduce the integration operators

$$Tv(x) = \int_0^x v(y)\, dy, \quad T^k = T(T^{k-1}), \quad k = 1, \dots, l, \tag{8.6}$$

and note that the function $T^k v$ belongs to $\mathfrak{C}^k$ and has a k-th Radon-Nikodym derivative v, as long as v is locally integrable. In particular, Tv is locally absolutely continuous and has v as its Radon-Nikodym derivative. Hence, one may integrate by parts to get that

$$\int_{-\infty}^{\infty} v(x)\, w(x)\, dx = -\int_{-\infty}^{\infty} Tv(x)\, w'(x)\, dx$$

for any $w \in C_0^{\infty}(\mathbb{R})$. By repeated integration by parts, we obtain that

$$\int_{-\infty}^{\infty} v(x)\, w(x)\, dx = (-1)^l \int_{-\infty}^{\infty} T^l v(x)\, w^{(l)}(x)\, dx.$$

In view of (8.5), this gives

$$\int_{-\infty}^{\infty} w^{(l)}(x) \left(u(x) - T^l v(x) \right) dx = 0$$

for all $w \in C_0^{\infty}(\mathbb{R})$. We are in position to apply Lemma 8.5 (in dimension $n = 1$) and conclude that $u - T^l v = Q$ a.e. for some polynomial Q of degree at most $d = l - 1$. Then $\tilde{u} = T^l v + Q$ belongs to $\mathfrak{C}^l$, has an l-th generalized derivative v, and is equal to u a.e. $\quad\square$

9 Weak derivatives along single variables

A similar characterization about the generalized partial derivatives D_i^l, $1 \le i \le n$, also holds in the n-dimensional case. Fix an integer $l \ge 1$.

Proposition 9.1. *A locally integrable function u on $\mathbb{R}^n$, $n \ge 2$, has a generalized partial derivative $v = D_i^l$, if and only if $u = \tilde{u}$ a.e. for some Borel measurable function $\tilde{u}$ such that, for almost all points $(x_j)_{j \ne i} \in \mathbb{R}^{n-1}$, the function*

$$f(x_i) = \tilde{u}(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n)$$

belongs to the class $\mathfrak{C}^l$ and has an l-th Radon-Nikodym derivative $f^{(l)}$ which is Borel measurable and locally integrable on $\mathbb{R}^n$. In this case, $v = f^{(l)}$ a.e.

Proof. We apply an argument as in the proof of Proposition 8.2, with a few modifications. In the sufficiency direction, assume that $\tilde{u} = u$. Write $x = (x_i, \bar{x})$ with $x_i \in \mathbb{R}$, $\bar{x} \in \mathbb{R}^{n-1}$, and let $E_i \subset \mathbb{R}^{n-1}$ be an exceptional null set of collections $\bar{x} = (x_j)_{j \ne i}$. By the assumption, for every fixed $\bar{x}$ outside this set,

1) the function $f(x_i)$ belongs to the class $\mathfrak{C}^l$;

2) there is a representative for its l-th Radon-Nikodym derivative $f^{(l)}(x_i) = v(x)$, which defines a Borel measurable, locally integrable function on $\mathbb{R}^n$.

The first property allows us to perform the repeated integration by parts along the i-th coordinate to get that, for any $w \in C_0^{\infty}(\mathbb{R}^n)$,

$$\int_{-\infty}^{\infty} w(x_i, \bar{x})\, f^{(l)}(x_i)\, dx_i = (-1)^l \int_{-\infty}^{\infty} D_i^l w(x_i, \bar{x})\, u(x_i, \bar{x})\, dx_i, \quad \bar{x} \notin E_i.$$

By property 2), and since u is Borel measurable and locally integrable, both sides of this equality represent integrable functions on $\mathbb{R}^{n-1}$. Using Fubini's theorem, one may integrate over $\bar{x}$ to get

$$\int_{\mathbb{R}^{n-1}} \int_{-\infty}^{\infty} w(x_i, \bar{x}) \, f^{(l)}(x_i) \, dx_i \, d\bar{x} = (-1)^l \int_{\mathbb{R}^{n-1}} \int_{-\infty}^{\infty} D_i^l w(x_i, \bar{x}) \, u(x_i, \bar{x}) \, dx_i \, d\bar{x},$$

that is,

$$\int_{\mathbb{R}^n} w(x) \, f^{(l)}(x_i) \, dx = (-1)^l \int_{\mathbb{R}^n} D_i^l w(x) \, u(x) \, dx.$$

Hence, according to (8.1), the function $v(x) = f^{(l)}(x_i)$ serves as an l-th generalized derivative for u.

For an opposite direction, assume that a generalized partial derivative $v = D_i^l u$ exists, so that it may be chosen to be Borel measurable and locally integrable. Then, the function $x_i \to v(x_i, \bar{x})$ is Borel measurable for all $\bar{x} \in \mathbb{R}^{n-1}$, and is locally integrable on the real line for almost all $\bar{x} \in \mathbb{R}^{n-1}$. Indeed, since for all integers $m, N \geq 1$,

$$\int_{B_N} \left[\int_{-m}^m |v(x_i, \bar{x})| \, dx_i \right] d\bar{x} < \infty, \quad B_N = [-N, N]^{n-1},$$

we conclude that the set

$$A_{m,N} = \left\{ \bar{x} \in B_N : \int_{-m}^m |v(x_i, \bar{x})| \, dx_i < \infty \right\}$$

is Borel measurable and has a full Lebesgue measure inside the cube B_N. Hence the set

$$A = \bigcup_{N=1}^{\infty} \bigcap_{m=1}^{\infty} A_{m,N}$$

has a full Lebesgue measure on $\mathbb{R}^{n-1}$. But it contains exactly those points $\bar{x} \in \mathbb{R}^{n-1}$ for which the function $x_i \to v(x_i, \bar{x})$ is integrable on all bounded intervals on the real line.

Next, we employ the integration operators (8.6) which are applied along the i-th coordinate: For $\bar{x} \in A$, put

$$Tv(x_i, \bar{x}) = \int_0^{x_i} v(y, \bar{x}) \, dy, \quad T^k = T(T^{k-1}), \quad k = 1, \ldots, l.$$

Clearly, the function Tv is finite, Borel measurable on $\mathbb{R} \times A$, and locally integrable, since

$$\int_{B_N} \left[\int_{-m}^m |Tv(x_i, \bar{x})| \, dx_i \right] d\bar{x} \leq m \int_{B_N} \left[\int_{-m}^m |v(x_i, \bar{x})| \, dx_i \right] d\bar{x} < \infty$$

for all $m, N \geq 1$. The same conclusions are also true about all functions $T^k v$.

Moreover, since $x_i \to v(x_i, \bar{x})$ is locally integrable, the function $x_i \to T^k v(x_i, \bar{x})$ belongs to the class $\mathfrak{C}^k$ and has $v(x_i, \bar{x})$ as a generalized derivative of order k. In particular, $Tv(x_i, \bar{x})$ is locally absolutely continuous with respect to x_i, and one may integrate by parts with respect to this variable to get that, for any $w \in C_0^\infty(\mathbb{R}^n)$,

$$\int_{-\infty}^{\infty} v(x_i, \bar{x}) \, w(x_i, \bar{x}) \, dx_i = - \int_{-\infty}^{\infty} Tv(x_i, \bar{x}) \, D_i w(x_i, \bar{x}) \, dx_i.$$

By repeated integration by parts, we obtain that

$$\int_{-\infty}^{\infty} v(x_i, \bar{x}) \, w(x_i, \bar{x}) \, dx_i = (-1)^l \int_{-\infty}^{\infty} T^l v(x_i, \bar{x}) \, D_i^l w(x_i, \bar{x}) \, dx_i.$$

Here, the integrands represent Borel measurable, integrable functions. Hence, it is possible to integrate both sides according to the Fubini theorem, and then we arrive at

$$\int v(x)\, w(x)\, dx = (-1)^l \int T^l v(x)\, D_i^l w(x)\, dx.$$

Applying (8.1), which is our hypothesis, this yields

$$\int D_i^l w(x)\, \big(u(x) - T^l v(x)\big)\, dx = 0$$

for all $w \in C_0^\infty(\mathbb{R})$. We are in position to apply Lemma 8.5 and conclude that $u - T^l v = Q$ a.e. for some polynomial Q in n real variables of degree at most $l - 1$. It remains to put $\tilde{u}(x) = T^l v(x) + Q(x)$. $\qquad\square$

Returning to Definition 8.1 with an arbitrary α such that $|\alpha| = 1$, we obtain n partial derivatives $v = \partial_{x_i} u = D_i u$, and one may speak about the generalized gradient

$$\nabla u = (\partial_{x_1} u, \ldots, \partial_{x_n} u)$$

and its Euclidean length $|\nabla u|$. Thus, according to (8.1), for all $w \in C_0^\infty(\mathbb{R}^n)$,

$$\int u\, \partial_{x_i} w\, dx = -\int w\, \partial_{x_i} u\, dx, \quad i = 1, \ldots, n. \tag{9.1}$$

Let us formulate Proposition 9.1 once more in this particular case.

Corollary 9.2. *A locally integrable function u on $\mathbb{R}^n$ has a generalized gradient, if and only if after a modification on a set of measure zero the modified (Borel measurable) function $\tilde{u}$ is locally absolutely continuous on almost all lines parallel to the coordinate axes and have partial Radon-Nikodym derivatives that are locally integrable on $\mathbb{R}^n$.*

This characterization may be used to derive the following assertion which will be needed to correctly introduce the Fisher information.

Corollary 9.3. *If a non-negative locally integrable function u on $\mathbb{R}^n$ has a generalized gradient with partial derivative $D_i u$, then the sets*

$$E_i = \{x \in \mathbb{R}^n : u(x) = 0,\ |D_i u(x)| > 0\}, \quad i = 1, \ldots, n,$$

have Lebesgue measure zero.

Proof. We may assume that u is properly modified so that $u = \tilde{u}$. Then, using the previous notation $x = (x_i, \bar{x})$, we have that, for all $\bar{x}$ except for a null set $A_i \subset \mathbb{R}^{n-1}$, the function $u_i(x_i) = u(x_i, \bar{x})$ is a.e. differentiable, and its derivative $u_i'(x_i)$ serves as a generalized partial derivative $D_i u(x)$. Since $u_i \geq 0$, it follows that $u_i(x_i) = 0 \Rightarrow u_i'(x_i) = 0$ at every point of differentiability (similarly to dimension one). Hence, the set

$$E_i(\bar{x}) = \{x_i \in \mathbb{R} : u_i(x_i) = 0,\ |u_i'(x_i)| > 0\}, \qquad \bar{x} \notin A_i,$$

has Lebesgue measure zero on the real line. By Fubini's theorem,

$$\mathrm{mes}_n(E_i) = \int_{\mathbb{R}^{n-1}} \mathrm{mes}_1(E_i(\bar{x}))\, d\bar{x} = 0,$$

where mes_n stands for the Lebesgue measure on $\mathbb{R}^n$. $\qquad\square$

Locally Lipschitz functions. If u has a finite Lipschitz semi-norm in some neighborhood of any point, it is almost everywhere differentiable, so that it has a usual gradient $\nabla u(x)$ a.e. (by Rademacher's theorem, cf. [17], p. 50). Such functions are locally absolutely continuous along every line, and the usual gradient also serves as a generalized gradient. As a representative of the modulus of the gradient, one may take a locally finite function

$$|\nabla u(x)| = \limsup_{y \to x} \frac{|u(x) - u(y)|}{|x - y|}, \quad x \in \mathbb{R}^n.$$

10 Sobolev spaces

Here and elsewhere, we use the usual notation for the space $L^s(\mathbb{R}^n)$ of all measurable functions u on $\mathbb{R}^n$ with finite norm

$$\|u\|_s = \left(\int |u(x)|^s \, dx \right)^{1/s}, \quad s \geq 1.$$

Given an integer $l \geq 1$, the Sobolev space with parameters (l, s) is defined as

$$W_l^s(\mathbb{R}^n) = \{u \in L^s(\mathbb{R}^n) : D^\alpha u \in L^s(\mathbb{R}^n) \text{ for all } 1 \leq |\alpha| \leq l\}.$$

It is a Banach space endowed with the norm

$$\|u\|_{W_l^s} = \left(\sum_{0 \leq |\alpha| \leq l} \|D^\alpha u\|_s^s \right)^{1/s}.$$

When $s = 2$ we obtain a Hilbert space.

Thus, $W_1^s(\mathbb{R}^n)$ contains all functions u in $L^s(\mathbb{R}^n)$ that have a generalized gradient ∇u such that $|\nabla u|$ belongs to the same space $L^s(\mathbb{R}^n)$.

Characterizations. A function u in $L^s(\mathbb{R}^n)$ belongs to $W_1^s(\mathbb{R}^n)$, if and only if after a modification on a set of measure zero the modified function $\tilde{u}$ is locally absolutely continuous on almost all lines parallel to the coordinate axes whose partial Radon-Nikodym derivatives belong to $L^s(\mathbb{R}^n)$, cf. [17], pp. 44-45. This characterization is also a consequence of Corollary 9.2, which in turn is a particular case of a more general Proposition 9.1. Thus, the generalized partial derivatives $\partial_{x_i} \tilde{u}(x)$, $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, with respect to the i-th coordinate may be understood in the usual sense for almost all collections $(x_j)_{j \neq i}$.

If $s > 1$, a function u from $L^s(\mathbb{R}^n)$ belongs to $W_1^s(\mathbb{R}^n)$, if and only if

$$\sup_{h \neq 0} \frac{\|u(x + h) - u(x)\|_s}{|h|} < \infty.$$

For $s = 2$, we have

$$\|u\|_{W_1^2}^2 = \int \left(|u|^2 + |\nabla u|^2 \right) dx.$$

In this case, another characterization can be given in terms of the Fourier transform

$$\hat{u}(t) = \int e^{i\langle t, x \rangle} u(x) \, dx, \quad t \in \mathbb{R}^n, \ u \in L^2(\mathbb{R}^n),$$

which is well-defined as an element of $L^2(\mathbb{R}^n)$. Namely, u belongs to the Sobolev space $W_1^2(\mathbb{R}^n)$, if and only if the function $(1 + |t|) \hat{u}(t)$ belongs to $L^2(\mathbb{R}^n)$. Moreover, by the Plancherel theorem,

$$\|u\|_{W_1^2}^2 = \frac{1}{(2\pi)^n} \int (1 + |t|^2) \, |\hat{u}(t)|^2 \, dt.$$

This way $W_1^2(\mathbb{R}^n)$ may be identified with the usual L^2-space over the measure $(1 + |t|^2)\,dt$.

Example. In contrast with the one-dimensional case, the elements of the Sobolev space $W_1^s(\mathbb{R}^n)$ in dimension $n \geq 2$ do not need be bounded and continuous. One may consider the example of the unbounded (near zero) function

$$u(x) = \frac{\log |x|}{(1 + |x|^2)^n},$$

which belongs to all $L^s(\mathbb{R}^n)$ and has usual partial derivatives $\partial_{x_i} u(x)$ for all $x \neq 0$. These derivatives are integrable with any power $s < n$.

Sobolev inequalities. It is a well-known classical fact that any function u in $W_1^1(\mathbb{R}^n)$ belongs to $L^{\frac{n}{n-1}}(\mathbb{R}^n)$. Moreover,

$$\|u\|_{\frac{n}{n-1}} \leq C_n \|\nabla u\|_1 \tag{10.1}$$

with a constant C_n independent of u. If it is optimal, an equality in (10.1) is attained asymptotically when u approaches indicator functions of the Euclidean balls.

The inequality (10.1) may be extended to the W_1^s-space with $1 \leq s < n$ as the relation

$$\|u\|_{s^*} \leq C_n(s) \|\nabla u\|_s,$$

where $s^* = \frac{ns}{n-s}$ is the so-called Sobolev conjugate.

Elements of $W_l^s(\mathbb{R}^n)$ are called Sobolev functions. They can be well approximated by smooth functions using any regularizer ω and associated regularized functions which we discussed before, cf. Definition 8.3. Let us extend the list of basic properties given in Lemma 8.4 by the following. Below, α denotes an arbitrary multi-index and $s \geq 1$.

Proposition 10.1. *Let u be a locally integrable function on $\mathbb{R}^n$.*

1) $\|u_\varepsilon - u\|_{L^s(\Omega)} \to 0$ *as $\varepsilon \to 0$ on every bounded Borel set Ω in $\mathbb{R}^n$.*

2) *If $u \in L^s(\mathbb{R}^n)$, then $u_\varepsilon \in L^s(\mathbb{R}^n)$ as well, and moreover,*

$$\|u_\varepsilon\|_s \leq \|u\|_s, \quad \|u_\varepsilon - u\|_s \to 0 \ \ as \ \ \varepsilon \to 0.$$

3) *In addition, u_ε is bounded:*

$$\|u_\varepsilon\|_\infty \leq C\varepsilon^{-n/s} \|u\|_s, \quad C = \|\omega\|_{s'}, \ s' = \frac{s}{s-1}.$$

4) *If u has a generalized α-th derivative $D^\alpha u \in L^s(\mathbb{R}^n)$, then*

$$\|D^\alpha u_\varepsilon\|_s \leq \|D^\alpha u\|_s, \qquad \|D^\alpha u_\varepsilon - D^\alpha u\|_s \to 0 \ \ as \ \ \varepsilon \to 0.$$

5) *Hence, if $u \in W_l^s(\mathbb{R}^n)$ and $|\alpha| \leq l$, then*

$$\|D^\alpha u_\varepsilon\|_{W_l^s} \leq \|D^\alpha u\|_{W_l^s}, \qquad \|D^\alpha u_\varepsilon - D^\alpha u\|_{W_l^s} \to 0.$$

6) *In addition,*

$$\|D^\alpha u_\varepsilon\|_\infty \leq C\varepsilon^{-n/s} \|D^\alpha u\|_s.$$

7) *If $u \in W_1^s(\mathbb{R}^n)$, then with some constant C depending on ω only, we have*

$$\|u_\varepsilon - u\|_s \leq C\varepsilon \|\nabla u\|_s.$$

The property 5) shows that $C^\infty(\mathbb{R}^n) \cap W_l^s(\mathbb{R}^n)$ is dense in $W_l^s(\mathbb{R}^n)$.

All properties and their proofs are rather standard. For illustration, let us explain the inequality in 7). By Definition 8.3, assuming that u is smooth and that ω is supported on the ball of radious r, we have, by Hölder's inequality,

$$
\begin{aligned}
u_\varepsilon(x) - u(x) &= \int (u(x - \varepsilon y) - u(x))\,\omega(y)\,dy \\
&= -\varepsilon \int_{|y|<r} \left[\int_0^1 \langle \nabla u(x - \varepsilon t y), y \rangle \, dt \right] \omega(y)\,dy,
\end{aligned}
$$

implying

$$
|u_\varepsilon(x) - u(x)|^s \leq C \varepsilon^s \int_{|y|<r} \int_0^1 |\nabla u(x - \varepsilon t y)|^s \, dt\,dy.
$$

Hence

$$
\|u_\varepsilon - u\|_s^s \leq C' \varepsilon^s \int |\nabla u(x)|^s \, dx.
$$

On this step, the smoothness condition may be removed by approximation: Let us apply this relation with u_δ in place of u, and then using the commutativity of the regularization, we get

$$
\|(u_\varepsilon)_\delta - u_\delta\|_s^s \leq C' \varepsilon^s \int |\nabla u_\delta(x)|^s \, dx.
$$

Letting $\delta \downarrow 0$, it remains to refer to the properties 2) and 5).

We now extend Lemma 7.2 to the multi-dimensional setting.

Proposition 10.2. *If the functions $u_1, \dots, u_l$ belong to $W_1^1(\mathbb{R}^n)$, the convolution $u = u_1 * \cdots * u_l$ has integrable generalized partial derivatives along every coordinate up to order l. Moreover,*

$$
D_i^l u = D_i u_1 * \cdots * D_i u_l, \quad i = 1, \dots, l, \tag{10.2}
$$

with

$$
\|D_i^l u\|_1 \leq \|D_i u_1\|_1 \dots \|D_i u_l\|_1. \tag{10.3}
$$

Proof. To sumplify notations, write $D = D_i$, $D^l = D_i^l$. One may argue by induction on l. Write $u = v * u_l$ with $v = u_1 * \cdots * u_{l-1}$, $l \geq 2$. By the induction hypothesis, v has integrable generalized partial derivatives along every coordinate up to order $l - 1$. Moreover,

$$
D^{l-1} v = D u_1 * \cdots * D u_{l-1}. \tag{10.4}
$$

The convolution

$$
u(x) = \int u_l(x - y) v(y)\,dy, \quad x \in \mathbb{R}^n,
$$

is an integrable function. Given a C_0^∞-function w on $\mathbb{R}^n$, integrating by parts, we have

$$
\begin{aligned}
\int u(x) D^l w(x)\,dx &= \int \left[\int u_l(x - y) D^l w(x)\,dx \right] v(y)\,dy \\
&= -\int \left[\int D u_l(x - y) D^{l-1} w(x)\,dx \right] v(y)\,dy \\
&= -\int \left[\int D u_l(x') D^{l-1} w(x' + y)\,dx' \right] v(y)\,dy.
\end{aligned}
$$

Changing the order of integration and using the induction hypothesis, the latter double integral is equal to

$$\int \left[\int v(y) D^{l-1} w(x' + y) \, dy \right] Du_l(x') \, dx'$$

$$= (-1)^{l-1} \int \left[\int D^{l-1} v(y) \, w(x' + y) \, dy \right] Du_l(x') \, dx'$$

$$= (-1)^{l-1} \int w(x) \, (D^{l-1} v * Du_l)(x) \, dx.$$

Hence u has a generalized derivative $D^l u = D^{l-1} v * Du_l$. It remains to recall (10.4), and we arrive at (10.2)-(10.3). □

11 BV-space

Definition 11.1. *An integrable function u on $\mathbb{R}^n$ is said to be a function of bounded variation, if for some signed Borel measures μ_i on $\mathbb{R}^n$, we have*

$$\int u \, \partial_{x_i} w \, dx = - \int w \, d\mu_i, \quad i = 1, \dots, n, \tag{11.1}$$

for all $w \in C_0^\infty(\mathbb{R}^n)$.

The generalized gradient of u is then defined as a vector-valued measure $\mu = (\mu_1, \dots, \mu_n)$, whose total variation $\|\mu\|_{\mathrm{TV}}$ in the sense of Measure Theory is denoted $\|u\|_{\mathrm{TV}}$ in the sense of Theory of Functions. Using (11.1) and the notation $\mathrm{div}\, w = \sum_{i=1}^n \partial_{x_i} w_i$, it follows that

$$\|u\|_{\mathrm{TV}} = \|\mu\|_{\mathrm{TV}} = \sup \sum_{k=1}^N |\mu(A_k)| = \sup \left| \int \langle w, d\mu \rangle \right|$$

$$= \sup \left| \int \sum_{i=1}^n w_i \, d\mu_i \right| = \sup \left| \int u \, \mathrm{div}\, w \, dx \right|, \tag{11.2}$$

where the first supremum is running over all partitions of $\mathbb{R}^n$ into Borel sets $A_1, \dots, A_N$, and the next ones are taken over all C_0^∞-smooth maps $w = (w_1, \dots, w_n) : \mathbb{R}^n \to \mathbb{R}^n$ such that $|w(x)| \le 1$ for all $x \in \mathbb{R}^n$. Thus, one arrives at the formula (1.5).

Put

$$BV(\mathbb{R}^n) = \{ u \in L^1(\mathbb{R}^n) : \|u\|_{\mathrm{TV}} < \infty \},$$

which is a Banach space endowed with the norm

$$\|u\|_{BV} = \|u\|_1 + \|u\|_{\mathrm{TV}}.$$

Thus, an integrable function u belongs to $BV(\mathbb{R}^n)$, if and only if the last supremum in (11.2) is finite.

Note that $BV(\mathbb{R}^n)$ contains $W_1^1(\mathbb{R}^n)$, and the total variation norm for elements of this Sobolev space is simplified to (1.6), i.e.

$$\|u\|_{\mathrm{TV}} = \|\nabla u\|_1. \tag{11.3}$$

This assertion can be strengthened: Suppose that an integrable function u on $\mathbb{R}^n$ has a generalized gradient $\nabla u = (\partial_{x_1} u, \dots, \partial_{x_n} u)$. Then u belongs to $BV(\mathbb{R}^n)$, if and only if it belongs to $W_1^1(\mathbb{R}^n)$, in which case (11.3) holds true. Indeed, returning to (11.2) and

integrating by parts, we have

$$\sup \left| \int u \operatorname{div} w \, dx \right| \quad = \quad \sup \left| \int \sum_{i=1}^{n} w_i \, \partial_{x_i} u \, dx \right|$$

$$= \quad \sup \left| \int \langle w, \nabla u \rangle \, dx \right| = \int |\nabla u| \, dx,$$

where as before the supremum is running over all C_0^∞-maps $w = (w_1, \ldots, w_n) : \mathbb{R}^n \to \mathbb{R}^n$ such that $|w| \le 1$ pointwise on $\mathbb{R}^n$.

What will be important for us is that the regularization does not increase the total variation and BV-norm, in full analogy with L^s- and W_l^s-norms (cf. also [17]).

Proposition 11.2. *For any function u in $BV(\mathbb{R}^n)$, the regularized functions u_ε defined in* (8.2) *satisfy*

$$\|u_\varepsilon\|_{\mathrm{TV}} \le \|u\|_{\mathrm{TV}}, \quad \|u_\varepsilon\|_{BV} \le \|u\|_{BV}. \tag{11.4}$$

Proof. Let $w = (w_1, \ldots, w_n)$ be an arbitrary C_0^∞-smooth map participating in the last supremum in (11.2). We first notice that

$$\int u_\varepsilon \operatorname{div} w \, dx = \int u \operatorname{div} \psi \, dx, \tag{11.5}$$

where the map $\psi = (\psi_1, \ldots, \psi_n)$ has components

$$\psi_i(x) = \int w_i(x - y) \, \omega_\varepsilon(-y) \, dy, \quad x \in \mathbb{R}^n.$$

They are C_0^∞-smooth and represent the regularized functions for w_i by means of the regulizer $\tilde{\omega}(x) = \omega(-x)$. Since $\tilde{\omega}_\varepsilon$ represents a probability density, an application of the Cauchy inequality yields

$$\sum_{i=1}^{n} \psi_i(x)^2 \le \sum_{i=1}^{n} \int w_i(x - y)^2 \, \tilde{\omega}_\varepsilon(y) \, dy \le 1$$

due to the condition $|w(x)| \le 1$. Thus, ψ is one of the maps participating in the last supremum in (11.2), so that the right-hand side in (11.5) may not exceed $\|u\|_{\mathrm{TV}}$. Taking the supremum in (11.5) over all admissible maps w, we arrive at the first inequality in (11.4). Using $\|u_\varepsilon\|_1 \le \|u\|_1$, we also obtain the second one. $\qquad\square$

The norm in $BV(\mathbb{R}^n)$ is lower semi-continuous. More precisely, the following holds.

Proposition 11.3. *Suppose that u_k belong to $BV(\mathbb{R}^n)$ and have bounded norms in this space. If $\|u_k - u\|_{L^1(\Omega)} \to 0$ as $k \to \infty$ on all balls Ω in $\mathbb{R}^n$ for some locally integrable function u, then u belongs to $BV(\mathbb{R}^n)$ with*

$$\|u\|_{\mathrm{TV}} \le \liminf_{k \to \infty} \|u_k\|_{\mathrm{TV}}, \tag{11.6}$$

and similarly for the BV-norm.

Proof. One may assume additionally that $u_k \to u$ a.e. Then, for any admissible map w,

$$\int u \operatorname{div} w \, dx = \lim_{k \to \infty} \int u_k \operatorname{div} w \, dx \le \liminf_{k \to \infty} \|u_k\|_{\mathrm{TV}}.$$

Taking the supremum over all w as in (11.2), we arrive at (11.6). Since, by Fatou's lemma,

$$\int |u| \, dx \le \liminf_{k \to \infty} \int |u_k| \, dx,$$

we get the lower semi-continuity for the BV-norm as well. $\qquad\square$

This property is used in the proof of the following variant of the compactness theorem. In a slightly different way it is mentioned in Remark on p. 146, [10].

Proposition 11.4. *Let $(u_k)_{k\geq 1}$ be elements of $W_1^1(\mathbb{R}^n)$ such that $\sup_k \|u_k\|_{W_1^1} < \infty$. There exists a subsequence u_{k_j} which is a.e. convergent to a function u in $BV(\mathbb{R}^n)$ with the property that $\|u_{k_j} - u\|_{L^1(\Omega)} \to 0$ as $j \to \infty$ on all balls Ω in $\mathbb{R}^n$.*

Proof. By property 7) with $s = 1$ in Proposition 10.1, the regularized functions $v_k = (u_k)_\varepsilon$ satisfy

$$\|v_k - u_k\|_1 \leq C\varepsilon \|u_k\|_{W_1^1}.$$

Hence, given $\delta > 0$, we may choose $\varepsilon > 0$ such that for all $k \geq 1$

$$\|v_k - u_k\|_1 < \delta. \tag{11.7}$$

Moreover, by properties 3) and 6) with $s = 1$, v_k are bounded and have bounded gradients uniformly over all k. Hence, these functions are bounded and equicontinuous. Applying the Arzela-Ascoli theorem, for every integer $N \geq 1$, one can find a subsequence v_{k_j} which converges uniformly on the ball $\Omega : |x| < N$. Moreover, using the diagonal argument, one may find a subsequence v_{k_j} which converges pointwise on $\mathbb{R}^n$ and uniformly on all balls. In particular,

$$\|v_{k_j} - v_{k_i}\|_{L^1(\Omega)} \to 0 \quad \text{as} \ \ i, j \to \infty$$

on every ball Ω in $\mathbb{R}^n$. Hence, by (11.7), for all sufficiently large i and j,

$$
\begin{aligned}
\|u_{k_j} - u_{k_i}\|_{L^1(\Omega)} \ \leq \ & \|u_{k_j} - v_{k_j}\|_{L^1(\Omega)} + \|v_{k_j} - v_{k_i}\|_{L^1(\Omega)} \\
& + \|v_{k_i} - u_{k_i}\|_{L^1(\Omega)} \ < \ 3\delta.
\end{aligned}
$$

Applying this conclusion to the values $\delta = 2^{-m}$ and using the diagonal argument, we obtain a further subsequence u_{k_j} such that

$$\|u_{k_j} - u_{k_i}\|_{L^1(\Omega)} \to 0 \quad \text{as} \ \ i, j \to \infty.$$

Thus, u_{k_j} is a Cauchy sequence in $L^1(\Omega)$ and has a limit u belonging to this space so that $\|u_{k_j} - u\|_{L^1(\Omega)} \to 0$ as $j \to \infty$. As before, one may find a further subsequence, say u_{k_j}, which converges to a locally integrable function u a.e. on the whole space $\mathbb{R}^n$, with the property that $\|u_{k_j} - u\|_{L^1(\Omega)} \to 0$ as $j \to \infty$ for all balls Ω. By the Fatou lemma,

$$\|u\|_1 \leq \liminf_{j\to\infty} \|u_{k_j}\|_1 \leq \sup_k \|u_k\|_{W_1^1} < \infty,$$

so that u is integrable. Moreover, by (11.6), using $\|u_k\|_{\mathrm{TV}} = \|\nabla u_k\|_1$, we also conclude that $\|u\|_{\mathrm{TV}} < \infty$. Hence u belongs to $BV(\mathbb{R}^n)$. □

12 Convolution of functions of bounded variation

We are prepared to prove a multi-dimensional variant of Lemma 7.2 for the BV-space. For simplicity, we consider the convolution of two functions only.

Proposition 12.1. *Given functions p_1, p_2 in $BV(\mathbb{R}^n)$, the convolution $p = p_1 * p_2$ belongs to $W_1^1(\mathbb{R}^n)$. Moreover, its generalized partial derivatives are functions of bounded variation satisfying*

$$\|\partial_{x_i} p\|_{\mathrm{TV}} \leq n \|p_1\|_{\mathrm{TV}} \|p_2\|_{\mathrm{TV}}, \quad i = 1, \dots, n. \tag{12.1}$$

Proof. Define the regularized functions $u = (p_1)_\varepsilon$, $v = (p_2)_\varepsilon$ with parameter $\varepsilon > 0$ and consider the convolutions

$$w(x) = \int u(x - y)\, v(y)\, dy.$$

We have

$$\|w\|_1 \leq \|u\|_1 \|v\|_1 \leq \|p_1\|_1 \|p_2\|_1,$$

where we used the property that the regularization does not increase the L^1-norm.

Since u, v and their partial derivatives are integrable and C^∞-smooth, the same is true for w. Moreover, using the notation for partial derivatives $D_j = \partial_{x_j}$, $D_{ij} = \partial^2_{x_i x_j}$, we have

$$D_j w(x) = \int D_j u(x - y)\, v(y)\, dy = \int u(x - y)\, D_j v(y)\, dy,$$

$$D_{ij} w(x) = \int D_i u(x - y)\, D_j v(y)\, dy,$$

that is,

$$D_j w = (D_j u) * v = u * (D_j v), \quad D_{ij} w = (D_i u) * (D_j v).$$

This implies

$$\begin{aligned}
\int |D_j w|\, dx &\leq \int |D_j u|\, dx \int |v|\, dx \\
&\leq \int |\nabla u|\, dx \int |v|\, dx = \|u\|_{\mathrm{TV}} \|v\|_1 \leq \|p_1\|_{\mathrm{TV}} \|p_2\|_1
\end{aligned}$$

and

$$\begin{aligned}
\int |D_{ij} w|\, dx &\leq \int |D_i u|\, dx \int |D_j v|\, dx \\
&\leq \int |\nabla u|\, dx \int |\nabla v|\, dx \leq \|p_1\|_{\mathrm{TV}} \|p_2\|_{\mathrm{TV}},
\end{aligned}$$

since the regularization does not increase the total variation norm (Proposition 11.2). It follows that

$$\|D_i w\|_{\mathrm{TV}} \leq \sum_{j=1}^n \int |D_{ij} w|\, dx \leq n \|p_1\|_{\mathrm{TV}} \|p_2\|_{\mathrm{TV}}. \tag{12.2}$$

Thus, the functions $D_i w$ have bounded W_1^1-norms uniformly over all $\varepsilon > 0$. Hence, we are in position to apply Proposition 11.4: There exists a sequence $\varepsilon = \varepsilon_k \to 0$ and functions q_i in $BV(\mathbb{R}^n)$, $1 \leq i \leq n$, such that the partial derivatives $D_i w_k$ for the corresponding functions $w = w_k$ are convergent to q_i a.e. and have the property that

$$\|D_i w_k - q_i\|_{L^1(\Omega)} \to 0 \quad \text{as } k \to \infty$$

on all balls Ω in $\mathbb{R}^n$. By (12.2), and applying the lower semi-continuity property (11.6), we conclude that all q_i are functions of bounded variation satisfying

$$\|q_i\|_{\mathrm{TV}} \leq \liminf_{k \to \infty} \|D_i w_k\|_{\mathrm{TV}} \leq n \|p_1\|_{\mathrm{TV}} \|p_2\|_{\mathrm{TV}}. \tag{12.3}$$

We claim that the function q_i represents a generalized derivative $\partial_{x_i} p = D_i p$, and then (12.3) yields the desired relation (12.1). As required in (9.1), we need to show that

$$\int p\, D_i \psi\, dx = - \int \psi\, q_i\, dx \tag{12.4}$$

for all $\psi \in C_0^\infty(\mathbb{R}^n)$. Indeed, by the definition,

$$\int w_k \, D_i\psi \, dx = -\int \psi \, D_i w_k \, dx. \tag{12.5}$$

But the regularized functions $u = u_k$ and $v = v_k$ are convergent in L^1 to p_1 and p_2 respectively, by property 2) in Proposition 10.1, and thus $\|w_k - p\|_1 \to 0$ as $k \to \infty$. Hence, the left integrals in (12.5) are convergent as $k \to \infty$ to the left integral in (12.4). The same is true about the right integrals, since $D_i w_k$ are convergent to q_i locally in L^1, while ψ is compactly supported. This shows that p has the generalized gradient $\nabla p = (q_1, \dots, q_n)$. $\qquad \square$

Example. In dimension $n = 1$, let $p_1 = p_2$ be the density of the uniform distribution on the interval $[-\frac{1}{2}, \frac{1}{2}]$. It is a function of bounded variation with total variation norm $\|p_1\|_{\mathrm{TV}} = \|p_2\|_{\mathrm{TV}} = 2$. The convolution

$$p(x) = (p_1 * p_2)(x) = (1 - |x|)^+, \quad x \in \mathbb{R},$$

represents the density of the so-called triangle distribution. It is absolutely continuous and its generalized (Radon-Nikodym) derivative

$$p'(x) = -\mathrm{sign}(x) \, 1_{[-1,1]}(x)$$

is a function of bounded variation with total variation norm $\|p\|_{\mathrm{TV}} = 4$. Hence, (12.1) becomes an equality. Note that Lemma 7.2 is not applicable in this case.

13 Fisher information in high dimensions

Given a probability density p on $\mathbb{R}^n$, the first basic formula in (1.1)

$$I(p) = \int_{p(x)>0} \frac{|\nabla p|^2}{p} \, dx = \sum_{i=1}^n \int_{p(x)>0} \frac{(\partial_{x_i} p)^2}{p} \, dx \tag{13.1}$$

makes sense, once the function p has a generalized gradient

$$\nabla p = (\partial_{x_1} p, \dots, \partial_{x_n} p)$$

as in the definition (9.1). Here the partial derivatives are required to be locally integrable functions. Moreover, they have to be integrable for the finiteness of the Fisher information.

Proposition 13.1. *If $I(p)$ is finite, then $p \in W_1^1(\mathbb{R}^n)$ and*

$$\int |\nabla p| \, dx \leq \sqrt{I(p)}. \tag{13.2}$$

Moreover, $p^{1/2} \in W_1^2(\mathbb{R}^n)$ and

$$I(p) = 4 \int |\nabla p^{1/2}|^2 \, dx. \tag{13.3}$$

Conversely, if $p^{1/2} \in W_1^2(\mathbb{R}^n)$, then p has a generalized gradient, and the integrals in (13.1) and (13.3) coincide.

The proof is based on the chain rule formula and the next general characterization. Suppose that we are given a continuous function $T : [0, \infty) \to \mathbb{R}$ which has a continuous derivative $T'(t)$ for $t > 0$ such that $\sup_{t \geq t_0} |T'(t)| < \infty$ for any $t_0 > 0$.

Lemma 13.2. *Let p be a non-negative, locally integrable function on $\mathbb{R}^n$ having a generalized partial derivative $D_i p$. Then $u = T(p)$ has a generalized derivative $D_i u$, if and only if the function $T'(p) D_i p \, 1_{\{p>0\}}$ is locally integrable. In that case,*

$$D_i u = T'(p) D_i p \, 1_{\{p>0\}} \ a.e.$$

Proof. In the one-dimensional case, this assertion may be refined. Let the function p be locally absolutely continuous on the real line and have a Radon-Nikodym derivative p'. By continuity, the set $\{x \in \mathbb{R} : p(x) > 0\}$ is open and may be decomposed into at most countably many intervals (a_k, b_k), finite or not. Then, on every (a_k, b_k), $T(p)$ is locally absolutely continuous and has a Radon-Nikodym derivative $T(p)' = T'(p) \, p'$. Indeed, the assumption on the local absolute continuity of p is equivalent to the property that, for all $\alpha < \beta$ and $\varepsilon > 0$, there is $\delta > 0$ such that

$$\sum_{l=1}^{N} (y_l - x_l) < \delta \ \Rightarrow \ \sum_{l=1}^{N} |p(y_l) - p(x_l)| < \varepsilon$$

for any collection of non-overlapping intervals (x_l, y_l) inside $[\alpha, \beta]$ (cf. e.g. [12]). If this segment is contained in (a_k, b_k), then p is bounded away from zero, that is, $p(x) \geq t_0 > 0$ for all $x \in [\alpha, \beta]$. This implies that $|T(p(y_l)) - T(p(x_l))| \leq A \, |p(y_l) - p(x_l)|$ with some constant $A = A(t_0)$, so that

$$\sum_{l=1}^{N} (y_l - x_l) < \delta \ \Rightarrow \ \sum_{l=1}^{N} |T(p(y_l)) - T(p(x_l)| < A\varepsilon.$$

Hence, by the same characterization, $u = T(p)$ is locally absolutely continuous on (a_k, b_k) and has a finite derivative q, which exists almost everywhere on this interval. But, since $p(x)$ has a finite derivative $p'(x)$ for almost all $x \in \mathbb{R}$, $u(x)$ has derivative $T'(p(x)) \, p'(x)$ for almost all $x \in (a_k, b_k)$. This shows that $q = T'(p) \, p'$ a.e., thus proving the claim.

Turning to the general case $n \geq 2$, note that $|T(t)| \leq c \, (1 + t)$ for all $t \geq 0$ with some constant $c \geq 0$. Hence $u = T(p)$ is locally integrable. Without loss of generality, let $T(0) = 0$.

We may assume that p is modified as in Proposition 9.1 for $l = 1$, with a Borel measurable, locally integrable, generalized derivative $D_i p$. Thus, using the notation $x = (x_i, \bar{x})$, $x_i \in \mathbb{R}$, $\bar{x} \in \mathbb{R}^{n-1}$, the function $x_i \to p(x)$ is locally absolutely continuous and has a Radon-Nikodym derivative $x_i \to D_i p(x)$ for all $\bar{x}$ except for a null set $E_i \subset \mathbb{R}^{n-1}$. Given such a point $\bar{x}$, the set $U(\bar{x}) = \{x_i \in \mathbb{R} : p(x) > 0\}$ is open and may be decomposed into at most countably many intervals (a_k, b_k), finite or not. According to the one-dimensional claim, the function $x_i \to u(x)$ is locally absolutely continuous and has a Radon-Nikodym derivative $D_i u(x) = T'(p(x)) \, D_i p(x)$ on every such interval $a_k < x_i < b_k$. Hence, given a C_0^∞-function w on $\mathbb{R}^n$, on any subinterval $[\alpha, \beta] \subset (a_k, b_k)$ one may integrate by parts along the x_i-coordinate to get

$$\int_\alpha^\beta u(x_i, \bar{x}) \, D_i w(x_i, \bar{x}) \, dx_i \ = \ u(\beta, \bar{x}) \, w(\beta, \bar{x}) - u(\alpha, \bar{x}) \, w(\alpha, \bar{x})$$

$$- \int_\alpha^\beta T'(p(x_i, \bar{x})) D_i p(x_i, \bar{x}) \, w(x_i, \bar{x}) \, dx_i. \qquad (13.4)$$

First assume that $T'(p) D_i p \, 1_{\{p>0\}}$ is locally integrable on $\mathbb{R}^n$. Then, this function will be locally integrable with respect to x_i for almost all $\bar{x}$ except for a null set E_i' containing E_i (cf. proof of Proposition 9.1). That is, for any bounded interval $[a, b]$,

$$\int_a^b |T'(p(x_i, \bar{x})) D_i p(x_i, \bar{x})| \, 1_{\{p(x_i, \bar{x}) > 0\}} \, dx_i < \infty, \quad \bar{x} \notin E_i'. \qquad (13.5)$$

Necessarily $p(a_k, \bar{x}) = p(b_k, \bar{x}) = 0$, and thus $u(a_k, \bar{x}) = u(b_k, \bar{x}) = 0$, as long as the endpoints a_k and b_k are finite. Hence, letting $\alpha \to a_k$ and $\beta \to b_k$ in (13.4), we get

$$\int_{a_k}^{b_k} u(x_i, \bar{x}) \, D_i w(x_i, \bar{x}) \, dx_i = -\int_{a_k}^{b_k} T'(p(x_i, \bar{x})) D_i p(x_i, \bar{x}) \, w(x_i, \bar{x}) \, dx_i. \qquad (13.6)$$

But this equality also holds when a_k and/or b_k are infinite, since $w(\alpha, \bar{x})$ and $w(\beta, \bar{x})$ are vanishing for α and β being sufficiently large. Due to (13.5), one may perform summation over all k in (13.6), and then we arrive at

$$\int_{U(\bar{x})} u(x_i, \bar{x}) \, D_i w(x_i, \bar{x}) \, dx_i = -\int_{U(\bar{x})} T'(p(x_i, \bar{x})) D_i p(x_i, \bar{x}) \, w(x_i, \bar{x}) \, dx_i$$

with an arbitrary point $\bar{x}$ outside E_i' (since the integrands on both sides are supported on a bounded set). Here, the left integral does not change when extending the integration over the whole real line, so,

$$\int_{-\infty}^{\infty} u(x_i, \bar{x}) \, D_i w(x_i, \bar{x}) \, dx_i = -\int_{-\infty}^{\infty} T'(p(x_i, \bar{x})) D_i p(x_i, \bar{x}) \, 1_{\{p(x_i, \bar{x}) > 0\}} \, w(x_i, \bar{x}) \, dx_i.$$

Using Fubini's theorem, this equality may now be integrated over $\bar{x}$, and we obtain

$$\int_{\mathbb{R}^n} u(x) \, D_i w(x) \, dx = -\int_{\mathbb{R}^n} T'(p(x)) D_i p(x) \, 1_{\{p(x) > 0\}} \, w(x) \, dx.$$

This means that $T'(p(x)) D_i p(x) \, 1_{\{p(x) > 0\}}$ serves as a generalized partial derivative for $u(x)$.

Conversely, suppose that $u = T(p)$ has a generalized partial derivative q_i; in particular, it is locally integrable. As we have already noticed, for any $\bar{x}$ outside E_i, the function $x_i \to u(x)$ has a Radon-Nikodym derivative $T'(p) D_i p$ on every interval (a_k, b_k). By Proposition 9.1, for almost all $\bar{x}$, we have $q_i = T'(p) D_i p$ for almost all $x_i \in (a_k, b_k)$. Therefore, this equality holds true a.e. on the whole set $p(x) > 0$. As a consequence, the function $T'(p) D_i p \, 1_{\{p > 0\}}$ is locally integrable. $\qquad \square$

Proof of Proposition 13.1. Suppose that p has a generalized gradient ∇p with finite $I(p)$. By Corollary 9.3, the left integral in (13.2) may be restricted to the set $\{p(x) > 0\}$. Hence, applying Cauchy's inequality, we have

$$\int |\nabla p| \, dx = \int_{p(x) > 0} \frac{|\nabla p|}{\sqrt{p}} \sqrt{p} \, dx \leq \left(\int_{p(x) > 0} \frac{|\nabla p|^2}{p} \, dx \right)^{1/2},$$

which yields (13.2). Thus, $p \in W_1^1(\mathbb{R}^n)$.

In order to derive (13.3), we apply Lemma 13.2 with the function $T(t) = \sqrt{t}$, $t \geq 0$. Since $\frac{D_i p}{\sqrt{p}} 1_{\{p > 0\}}$ is square integrable, we conclude that the function $\sqrt{p}$ has a generalized partial derivative $q_i = D_i \sqrt{p} = \frac{D_i p}{2\sqrt{p}} 1_{\{p > 0\}}$. It follows that

$$\int_{p(x) > 0} \frac{(D_i p)^2}{p} \, dx = 4 \int q_i^2 \, dx.$$

Summing over all $i \leq n$ and recalling (13.1), we arrive at the representation (13.3).

The converse statement is similar. $\qquad \square$

Denote by $\mathfrak{P}_n$ the collection of all probability densities on $\mathbb{R}^n$. According to (13.2), for every $I > 0$, the set

$$\mathfrak{R}_n(I) = \left\{ |\nabla p| : p \in \mathfrak{P}_n, \ I(p) \leq I \right\}$$

is bounded in $L^1(\mathbb{R}^n)$. The next statement refines this property.

Proposition 13.3. *The elements in $\mathfrak{R}_n(I)$ are uniformly integrable on $\mathbb{R}^n$.*

Proof. Since $\mathfrak{R}_n(I)$ is bounded in $L^1(\mathbb{R}^n)$, we need to show that, for any $\varepsilon > 0$, there is $\delta > 0$ such that for any Borel set $A \subset \mathbb{R}^n$,

$$|A| < \delta \ \Rightarrow \ \int_A |\nabla p| \, dx < \varepsilon \ \text{ for all } p \in \mathfrak{R}_n(I). \tag{13.7}$$

Here we denote by $|A| = \mathrm{mes}_n(A)$ the n-dimensional volume of A.

If $n = 1$, the assertion is obvious. Indeed, according to (13.2), p has a bounded total variation, implying that $p(x) \to 0$ as $|x| \to \infty$. Moreover,

$$p(x) = \int_{-\infty}^{x} p'(y) \, dy \le \|p'\|_1 \le \sqrt{I},$$

so that p is bounded. As a consequence,

$$\int_{-\infty}^{\infty} p'(x)^2 \, dx = \int_{p(x)>0} \frac{p'(x)^2}{p(x)} \, p(x) \, dx \le I^{3/2}.$$

Here we used the property that $p(x) = 0 \Rightarrow p'(x) = 0$ for all points x where p is differentiable. Thus, $\mathfrak{R}_1(I)$ is also bounded in $L^2(\mathbb{R})$, which is sufficient for (13.7).

Now, let $n \ge 2$. Recall that the set $E = \{x \in \mathbb{R}^n : p(x) = 0, \ |\nabla p(x)| > 0\}$ has Lebesgue measure zero (Corollary 9.3). Hence, by Cauchy's inequality,

$$\begin{aligned}
\int_A |\nabla p| \, dx &= \int_{A \setminus E} |\nabla p| \, dx \\
&= \int_{A \setminus E} \frac{|\nabla p|}{\sqrt{p}} \sqrt{p} \, dx \ \le \ I^{1/2} \left(\int_A p(x) \, dx \right)^{1/2}.
\end{aligned}$$

On the other hand, using Hölder's inequality with exponents $\frac{n}{n-1}$ and n, and then applying Sobolev's inequality (10.1) together with (13.2) once more, one can bound the last integral by

$$\|p\|_{\frac{n}{n-1}} \, |A|^{1/n} \ \le \ C_n \, \|\nabla p\|_1 \, |A|^{1/n} \ \le \ C_n \sqrt{I} \, |A|^{1/n}.$$

As a result,

$$\int_A |\nabla p| \, dx \ \le \ C_n^{1/2} I^{3/4} \, |A|^{1/2n}.$$

This relation readily implies (13.7). $\qquad\square$

14 Lower semi-continuity of Fisher information

The next important property indicates that the Fisher information represents a lower semi-continuous functional on $W_1^1(\mathbb{R}^n)$.

Proposition 14.1. *Given probability densities p, p_k in $W_1^1(\mathbb{R}^n)$ such that $\|p_k - p\|_{W_1^1} \to 0$, we have*

$$I(p) \le \liminf_{k \to \infty} I(p_k). \tag{14.1}$$

Proof. Since p_k and ∇p_k are respectively convergent to p and ∇p in $L^1(\mathbb{R}^n)$, one may choose a subsequence of p_k which is convergent a.e. with a similar convergence property for generalized gradients. So, we may assume that $p_k \to p$ and $\nabla p_k \to \nabla p$ a.e. as $k \to \infty$. But then

$$\lim_{k \to 0} \frac{|\nabla p_k(x)|^2}{p_k(x)} \, 1_{\{p_k(x)>0\}} \ = \ \frac{|\nabla p(x)|^2}{p(x)}$$

a.e. on the set $p(x) > 0$. It remains to apply Fatou's lemma. $\qquad\square$

Proposition 14.1 may be considerably sharpened by weakening the basic hypothesis on the convergence in $W_1^1(\mathbb{R}^n)$.

Proposition 14.2. *Given random vectors X, X_k ($k \geq 1$) with values in $\mathbb{R}^n$, suppose that $X_k \Rightarrow X$ weakly in distribution as $k \to \infty$. Then*

$$I(X) \leq \liminf_{k \to \infty} I(X_k). \tag{14.2}$$

Proof. It is sufficient to prove the following: For any subsequence of X_k, one can extract a further subsequence for which the relation (14.2) holds true, even with limsup in place of liminf. For simplicity of notations, let the first subsequence be the whole sequence of positive integers. By the assumption on the weak convergence, $\int v \, d\mu_k \to \int v \, d\mu$ as $k \to \infty$ for any bounded continuous function v on $\mathbb{R}^n$, where μ and μ_k denote the distributions of X and X_k respectively.

One may assume that $I(X_k) \leq I < \infty$ for all k. In this case, X_k have absolutely continuous distributions on $\mathbb{R}^n$ whose densities p_k lie in the Sobolev space $W_1^1(\mathbb{R}^n)$ and have Fisher information $I(p_k)$ bounded by I. By Proposition 13.1, p_k have bounded norms in $W_1^1(\mathbb{R}^n)$, so that we may apply Proposition 11.4. Thus, some subsequence p_{k_j} is a.e. convergent to a function $p \in BV(\mathbb{R}^n)$ with the property that

$$\|p_{k_j} - p\|_{L^1(\Omega)} \to 0 \quad \text{as } j \to \infty \tag{14.3}$$

on bounded Borel sets Ω in $\mathbb{R}^n$. The latter ensures that p represents a probability density on $\mathbb{R}^n$. Indeed, necessarily $p(x) \geq 0$ a.e. as p is a pointwise limit of non-negative functions. In addition, since $\int_\Omega p_{k_j} \to \int_\Omega p$, we have $\int_\Omega p \leq 1$ for all bounded Borel sets Ω in $\mathbb{R}^n$, hence $\int_{\mathbb{R}^n} p \leq 1$. For an opposite inequality, choose a bounded open set such that $\mathbb{P}\{X \in \Omega\} > 1 - \varepsilon$ for a given $\varepsilon > 0$. Since X_k are convergent weakly in distribution, we obtain that

$$\liminf_{j \to \infty} \mathbb{P}\{X_{k_j} \in \Omega\} \geq \mathbb{P}\{X \in \Omega\} > 1 - \varepsilon$$

(cf. e.g. [1]). On the other hand, by (14.3),

$$\mathbb{P}\{X_{k_j} \in \Omega\} = \int_\Omega p_{k_j}(x) \, dx \to \int_\Omega p(x) \, dx.$$

Hence $\int_\Omega p \geq 1 - \varepsilon$ for any $\varepsilon > 0$, so that $\int_{\mathbb{R}^n} p = 1$.

In particular, assuming again for simplicity of notations that k_j is the whole sequence of positive integers, we get (by applying Scheffe's lemma) that

$$\int |p_k - p| \, dx \to 0 \quad \text{as } k \to \infty. \tag{14.4}$$

This means that μ_k are convergent in total variation norm to the probability measure with density p. Consequently, the measure μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$ and has density p. In addition, the weak convergence is strengthened as the property (14.4) which implies that

$$\int p_k v \, dx \to \int p v \, dx \tag{14.5}$$

as $k \to \infty$ with an arbitrary bounded measurable function v on $\mathbb{R}^n$.

The uniform integrability property as in Proposition 13.3 allows us to apply the Dunford-Pettis compactness criterion for the space L^1 over finite measures (cf. [9], p. 20). It implies that the set $\mathfrak{R}_n(I)$ is pre-compact and also sequentially pre-compact in $L^1(\Omega)$ with respect to the weak $\sigma(L^1, L^\infty)$-topology for any bounded Borel set Ω in $\mathbb{R}^n$. Hence

the same is true for the collection of generalized partial derivatives $D_i p = \partial_{x_i} p$ with $p \in \mathfrak{P}_n(I)$, $i = 1, \ldots, n$. As a consequence, there is a subsequence of p_k along which $D_i p_k$ are weakly convergent in $L^1(\Omega)$ to some $q_i \in L^1(\Omega)$. Clearly, these limit functions may be chosen to be common for all such Ω's. Thus, assuming again that the subsequence is the whole sequence, we obtain that

$$\int w\, D_i p_k\, dx \to \int w q_i\, dx \quad \text{as } k \to \infty,\ i = 1, \ldots, n,$$

for any bounded measurable function w on $\mathbb{R}^n$ with a compact support. But, restricting ourselves to $w \in C_0^\infty(\mathbb{R}^n)$, the above left integrals are equal to $-\int p_k\, D_i w\, dx$, according to the definition (9.1) of the weak derivative of p_k, and they converge to $-\int p\, D_i w\, dx$, by (14.5) with $v = D_i w$. Thus

$$\int p\, D_i w\, dx = -\int w q_i\, dx.$$

This shows that $q_i = D_i p$ serve as generalized partial derivatives for p, so that p has a generalized gradient $\nabla p = (q_1, \ldots, q_n)$. Moreover, since p belongs to $BV(\mathbb{R}^n)$, we get

$$\|p\|_{\mathrm{TV}} = \int |\nabla p|\, dx = \int (q_1^2 + \cdots + q_n^2)^{1/2}\, dx < \infty.$$

Therefore, all $D_i p$ are integrable, and we may conclude that $p \in W_1^1(\mathbb{R}^n)$. In addition,

$$\int w\, D_i p_k\, dx \to \int w\, D_i p\, dx \quad \text{as } k \to \infty \tag{14.6}$$

for any bounded measurable function w on $\mathbb{R}^n$ with a compact support.

Now, as we have observed in Proposition 13.1 and Lemma 13.2, the functions $\sqrt{p_k}$ belong to $W_1^2(\mathbb{R}^n)$ and have partial generalized derivatives

$$\psi_{i,k} \equiv D_i \sqrt{p_k} = \frac{D_i p_k}{2\sqrt{p_k}}\, 1_{\{p_k > 0\}},$$

whose L^2-norms are bounded by $\sqrt{I}$. Since the balls in $L^2(\mathbb{R}^n)$ are weakly compact, one can extract a subsequence of $(\psi_{i,k})_{k \geq 1}$ which is weakly convergent to some function $\psi_i \in L^2(\mathbb{R}^n)$. That is, we may assume that, for any $i = 1, \ldots, n$,

$$\int v\, \psi_{i,k}\, dx \to \int v\, \psi_i\, dx \quad \text{as } k \to \infty$$

with an arbitrary function $v \in L^2(\mathbb{R}^n)$. Then, more generally, we have

$$\int v_k\, \psi_{i,k}\, dx \to \int v\, \psi_i\, dx \quad \text{as } k \to \infty, \tag{14.7}$$

as long as $v_k \to v$ in $L^2(\mathbb{R}^n)$. Indeed, by the Cauchy and triangle inequalities,

$$\left| \int v_k\, \psi_{i,k}\, dx - \int v\, \psi_i\, dx \right| \leq \int |v_k - v|\, |\psi_{i,k}|\, dx + \left| \int v\, \psi_{i,k}\, dx - \int v\, \psi_i\, dx \right|$$

$$\leq \sqrt{I}\, \|v_k - v\|_2 + \left| \int v\, \psi_{i,k}\, dx - \int v\, \psi_i\, dx \right| \to 0.$$

Thus, the property (14.7) holds true. Let us choose here $v_k = w\sqrt{p_k}$ and $v = w\sqrt{p}$ with an arbitrary bounded measurable function w on $\mathbb{R}^n$. Applying the pointwise inequality $(\sqrt{p_k} - \sqrt{p})^2 \leq |p_k - p|$, from (14.4) we get

$$\|v_k - v\|_2^2 = \int w^2\, (\sqrt{p_k} - \sqrt{p})^2\, dx \leq \int w^2\, |p_k - p|\, dx \to 0$$

as $k \to 0$. Hence, (14.7) is applicable and gives

$$\frac{1}{2} \int w \, D_i p_k \, 1_{\{p_k > 0\}} \, dx = \frac{1}{2} \int w \, D_i p_k \, dx \to \int w \sqrt{p} \, \psi_i \, dx,$$

where we also used Corollary 9.3 so as to remove the indicator function. Comparing this with (14.6), we obtain that the equality

$$\frac{1}{2} \int w \, D_i p \, dx = \frac{1}{2} \int w \, D_i p \, 1_{\{p > 0\}} \, dx = \int w \sqrt{p} \, \psi_i \, dx$$

holds true for any bounded measurable function w on $\mathbb{R}^n$ with a compact support. But this is only possible when $\frac{1}{2} D_i p = \sqrt{p} \, \psi_i$ a.e., so

$$\psi_i = \frac{D_i p}{2\sqrt{p}} \, 1_{\{p > 0\}} = D_i \sqrt{p}$$

a.e. on the set $p(x) > 0$.

Finally, since $\psi_{i,k} \, 1_{\{p > 0\}}$ are weakly convergent to $\psi_i \, 1_{\{p > 0\}}$ in $L^2(\mathbb{R}^n)$, by the lower semi-continuity of the norm with respect to the weak topology, we have

$$\|\psi_i \, 1_{\{p > 0\}}\|_2 \le \liminf_{k \to \infty} \|\psi_{i,k} \, 1_{\{p > 0\}}\|_2 \le \liminf_{k \to \infty} \|\psi_{i,k}\|_2.$$

But this is the same as

$$\int \left(\frac{D_i p}{2\sqrt{p}}\right)^2 1_{\{p > 0\}} \, dx \le \liminf_{k \to \infty} \int \left(\frac{D_i p_k}{2\sqrt{p_k}}\right)^2 1_{\{p_k > 0\}} \, dx.$$

Summing these inequalities over $i \le n$, we arrive at the relations (14.1)-(14.2). $\qquad \square$

15 Convexity of Fisher information

Recall that the collection $\mathfrak{P}_n$ of all probability densities on $\mathbb{R}^n$ represents a convex closed set in $L^1(\mathbb{R}^n)$. Another general property of the Fisher information is its convexity,

$$I(p) \le \sum_{k=1}^{N} \alpha_k I(p_k), \quad p_k \in \mathfrak{P}_n, \tag{15.1}$$

where $p = \sum_{k=1}^{N} \alpha_k p_k$ with arbitrary weights $\alpha_k > 0$ such that $\sum_{k=1}^{N} \alpha_k = 1$. This follows from the convexity of the function $R(u, v) = u^2/v$ in the upper half-plane $u \in \mathbb{R}$, $v > 0$. Moreover, Cohen [7] showed that the inequality in (15.1) is strict.

As a consequence, the collection $\mathfrak{P}_n(I)$ of all probability densities on $\mathbb{R}^n$ with Fisher information not exceeding a fixed number I represents a convex closed subset of $\mathfrak{P}_n$.

We need to extend Jensen's inequality (15.1) to arbitrary convex mixtures of probability densities. In order to formulate this more precisely, let us recall the definition of mixtures. For any Borel set A in $\mathbb{R}^n$, the linear functional $q \to \int_A q(x) \, dx$ is continuous on $L^1(\mathbb{R}^n)$ and takes values in $[0, 1]$ when $q \in \mathfrak{P}_n$. So, given a Borel probability measure π on $\mathfrak{P}_n$, one may introduce the Borel probability measure on $\mathbb{R}^n$ by virtue of the formula

$$\mu(A) = \int_{\mathfrak{P}_n} \left[\int_A q(x) \, dx\right] d\pi(q). \tag{15.2}$$

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

$$p = \int_{\mathfrak{P}_n} q \, d\pi(q). \tag{15.3}$$

Proposition 15.1. *For the probability density p as above,*

$$I(p) \leq \int_{\mathfrak{P}_n} I(q)\, d\pi(q). \tag{15.4}$$

Proof. Note that the integral in (15.4) makes sense, since the functional $q \to I(q)$ is lower semi-continuous and hence Borel measurable on $\mathfrak{P}_n$ (Proposition 14.1). We may assume that this integral is finite, so that π is supported on the convex (Borel measurable) set $\mathfrak{P}_n(\infty) = \cup_I \mathfrak{P}_n(I)$.

Step 1. Suppose that the measure π is supported on some convex compact set K contained in $\mathfrak{P}_n(I)$. We apply the following general theorem (cf. e.g. Meyer [14], Chapter XI, Theorem T7): If a function $I : K \to \mathbb{R}$ is convex and lower semi-continuous on a convex compact set K in a locally convex space E, then it admits the representation

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

where L denotes the family of all continuous affine functionals on E such that $l(q) < I(q)$ for all $q \in K$. In our particular case with $E = L^1(\mathbb{R}^n)$, any such functional acts on probability densities as $l(q) = \int \psi q\, dx$ with some bounded measurable function ψ on $\mathbb{R}^n$. Hence

$$I(q) = \sup_{\psi \in \Psi} \int \psi(x)\, q(x)\, dx, \quad q \in K,$$

for some family Ψ of bounded measurable functions ψ on $\mathbb{R}^n$. As a consequence, by the definition (15.2) for the measure μ with density p,

$$
\begin{aligned}
\int_{\mathfrak{P}_n} I(q)\, d\pi(q) \;\geq\; & \sup_{\psi \in \Psi} \int_{\mathfrak{P}_n} \left[\int \psi(x)\, q(x)\, dx \right] d\pi(q) \\
=\; & \sup_{\psi \in \Psi} \int \psi(x)\, p(x)\, dx \;=\; I(p),
\end{aligned}
$$

which is the desired inequality (15.4).

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

$$p_l(x) = \int_{K_l} q(x)\, d\pi_l(q), \tag{15.5}$$

which is a density of some probability measure μ_l on $\mathbb{R}^n$ as in the definition (15.3). Then, for any Borel measurable function f on $\mathbb{R}^n$ such that $|f| \leq 1$, we have

$$
\begin{aligned}
\int f\, d\mu_l - \int f\, d\mu \;=\; & \int_{\mathfrak{P}_n} \left[\int f(x)\, q(x)\, dx \right] d(\pi_l(q) - \pi(q)) \\
\leq\; & \|\pi_l - \pi\|_{\mathrm{TV}} \;=\; 2(1 - c_l).
\end{aligned}
$$

Taking the supremum over all admissible f, we get

$$\int |p_l - p|\, dx \leq 2(1 - c_l) \to 0 \quad \text{as}\ \ l \to \infty,$$

which means that p_l are convergent to p in $L^1(\mathbb{R}^n)$. Hence the relation (14.1) holds: $I(p) \leq \liminf_{l \to \infty} I(p_l)$. On the other hand, by the previous step,

$$I(p_l) \leq \int_{K_l} I(q) \, d\pi_l(q) = \frac{1}{c_l} \int_{K_l} I(q) \, d\pi(q) \to \int_{\mathfrak{P}_n(I)} I(q) \, d\pi(q) \qquad (15.6)$$

as $l \to \infty$, and we obtain (15.4).

Step 3. In the general case, we may apply Step 2 to the normalized restrictions π_l of π to the sets $K_l = \mathfrak{P}_n(l)$. Again, for the densities π_l defined as in (15.5), we obtain (15.6), where $\mathfrak{P}_n(I)$ should be replaced with $\mathfrak{P}_n(\infty)$. Another application of the lower semi-continuity of the Fisher information finishes the proof. $\qquad \square$

Corollary 15.2. *If a probability density p on $\mathbb{R}^n$ has a generalized gradient, then the regularized probability densities p_ε defined in* (8.2) *satisfy*

$$I(p_\varepsilon) \leq I(p), \quad \lim_{\varepsilon \to 0} I(p_\varepsilon) = I(p).$$

In fact, this approximation property may be generalized similarly to the setting of Proposition 14.2.

Corollary 15.3. *Given independent random vectors X and Z with values in $\mathbb{R}^n$, for the random vectors $X_\varepsilon = X + \varepsilon Z$, $\varepsilon \in \mathbb{R}$, we have*

$$I(X_\varepsilon) \leq I(X), \quad \lim_{\varepsilon \to 0} I(X_\varepsilon) = I(X). \qquad (15.7)$$

Proof. For the first claim in (15.7), we may assume that $I(X)$ is finite, so that X has an absolutely continuous distribution with density p having a generalized gradient. In this case, X_ε has a density p_ε representing a convex mixture of probability densities of the form $q_h(x) = p(x - h)$, $h \in \mathbb{R}^n$. Since $I(q_h) = I(p)$, we obtain the inequality in (15.7) by applying Proposition 15.1. The second claim in (15.7) is obtained by applying Proposition 14.2. Since $X_\varepsilon \Rightarrow X$ weakly in distribution as $\varepsilon \to 0$, it follows that

$$I(X) \leq \liminf_{\varepsilon \to 0} I(X_\varepsilon). \qquad \square$$

The inequality in (15.7) can be sharpened by virtue of the Stam inequality

$$\frac{1}{I(X + Y)} \geq \frac{1}{I(X)} + \frac{1}{I(Y)}, \qquad (15.8)$$

which holds whenever the random vectors X and Y in $\mathbb{R}^n$ are independent (cf. [15], [11]).

One interesting case in Corollary 15.3 is when Z has a standard normal distribution. Combining both claims in (15.7), we then obtain that the function $\varepsilon \to I(X_\varepsilon)$ is monotone. This choice of smoothing allows one to reduce various relations about the Fisher information $I(p)$ such as (15.8) to the case of C^∞-smooth densities p. Here is another example.

Corollary 15.4. *Given a random vector X in $\mathbb{R}^n$, we have $I(U(X)) = I(X)$ for any linear orthogonal map $U : \mathbb{R}^n \to \mathbb{R}^n$.*

Indeed, by (13.1), $I(U(X)) = I(X)$ as long as X has a C^∞-smooth density. Hence, in the general case, we have $I(U(X) + \varepsilon Z) = I(X + \varepsilon Z)$, where Z is a standard normal random vector in $\mathbb{R}^n$, independent of X (since $U(Z)$ is standard normal). It remains to apply Corollary 15.3.

16 Upper bounds. Proof of Theorem 1.2

We are now prepared to extend several upper bounds for the Fisher information

$$I(p) = \int \frac{|\nabla p(x)|^2}{p(x)} \, dx = \sum_{i=1}^{n} \int \frac{(\partial_{x_i} p(x))^2}{p(x)} \, dx$$

from the one-dimesnional case to higher dimensions (as before, one may adopt the agreement that $0/0 = 0$). The first upper bounds were developed for densities from the classes $\mathfrak{C}^l$ on the real line with $l = 2$ and $l = 3$. Analogously, one may say that p belongs to $\mathfrak{C}^l(\mathbb{R}^n)$, if for for any $i = 1, \ldots, n$, for almost all $(x_j)_{j \neq i} \in \mathbb{R}^{n-1}$, the function

$$x_i \to p(x) = p(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) \tag{16.1}$$

belongs to $\mathfrak{C}^l(\mathbb{R})$ with an additional requirement that the l-th derivative $\partial_{x_i}^l p(x)$ with respect to x_i in the Radon-Nikodym sense is locally integrable on $\mathbb{R}^n$. According to Proposition 9.1, such densities describe representatives of functions p on $\mathbb{R}^n$ having generalized partial derivatives $D_i^l p = \partial_{x_i}^l p$. With this definition, Propositions 2.2-2.3 yield:

Proposition 16.1. *If the probability density p belongs to the class $\mathfrak{C}^2(\mathbb{R}^n)$ and is supported on a bounded, open, convex set Ω in $\mathbb{R}^n$, then*

$$I(p) \leq 2Cn \, |\Omega|, \quad C = \max_{1 \leq i \leq n} \text{ess sup}_x \, \partial_{x_i}^2 p(x).$$

Here $|\Omega| = \text{mes}_n(\Omega)$ denotes the n-dimensional volume of Ω.

Proof. Note that the inequality (2.2) is homogeneous in p. Applying it to the function in (16.1) with fixed $\bar{x} = (x_j)_{j \neq i} \in \mathbb{R}^{n-1}$, we obtain that

$$\frac{(\partial_{x_i} p(x))^2}{p(x)} \leq 2C_i, \quad C_i = \text{ess sup}_x \, \partial_{x_i}^2 p(x).$$

By the convexity assumption, the section

$$\Omega(\bar{x}) = \{x_i \in \mathbb{R} : (x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) \in \Omega\}$$

represents a certain interval (a_i, b_i) depending on $\bar{x}$. Let us integrate the last inequality over this section to get that

$$\int_{a_i}^{b_i} \frac{(\partial_{x_i} p(x))^2}{p(x)} \, dx_i \leq 2C_i(b_i - a_i) \leq 2C(b_i - a_i).$$

The next integration over $\bar{x}$ gives, by Fubini's theorem,

$$\int_\Omega \frac{(\partial_{x_i} p(x))^2}{p(x)} \, dx \leq 2C \int_{\mathbb{R}^{n-1}} (b_i - a_i) \, d\bar{x} = 2C \, |\Omega|.$$

It remains to perform summation over $i \leq n$. $\qquad\square$

Let us now turn to the multidimensional variant of Proposition 7.1, i.e. Theorem 1.2. It may be stated for a slightly more general class of densities as the following.

Proposition 16.2. *If the probability density p belongs to the class $\mathfrak{C}^3(\mathbb{R}^n)$, then*

$$I(p) \leq 10 \, n^{1/3} \left(\sum_{i=1}^{n} \int |\partial_{x_i}^3 p(x)| \, dx \right)^{2/3}. \tag{16.2}$$

Moreover, this inequality holds true as long as p has generalized partial derivatives $\partial_{x_i}^3 p$.

Proof. Applying (7.1) to the functions in (16.1) together with the arithmetic-geometric inequality, we have

$$
\frac{1}{c} \int_{-\infty}^{\infty} \frac{(\partial_{x_i} p(x))^2}{p(x)} \, dx_i \quad \leq \quad \Big(\int_{-\infty}^{\infty} p(x) \, dx_i \Big)^{1/3} \Big(\int_{-\infty}^{\infty} |\partial_{x_i}^3 p(x)| \, dx \Big)^{2/3}
$$

$$
\leq \quad \frac{1}{3} \int_{-\infty}^{\infty} p(x) \, dx_i + \frac{2}{3} \int_{-\infty}^{\infty} |\partial_{x_i}^3 p(x)| \, dx
$$

with constant $c = 10$, which holds for almost all $(x_j)_{j \neq i} \in \mathbb{R}^{n-1}$. The integration over the remaining variables gives a similar relation

$$
\frac{1}{c} \int \frac{(\partial_{x_i} p(x))^2}{p(x)} \, dx \leq \frac{1}{3} + \frac{2}{3} \int |\partial_{x_i}^3 p(x)| \, dx.
$$

Performing summation over all $i \leq n$, we arrive at

$$
\frac{1}{c} I(p) \leq \frac{n}{3} + \frac{2}{3} \sum_{i=1}^{n} \int |\partial_{x_i}^3 p(x)| \, dx.
$$

Moreover, applying this to densities $p_\lambda(x) = \lambda^n p(\lambda x)$ with parameter $\lambda > 0$, we get

$$
\frac{1}{c} I(p) \leq \frac{n}{3\lambda^2} + \frac{2}{3} \lambda B, \quad B = \sum_{i=1}^{n} \int |\partial_{x_i}^3 p(x)| \, dx.
$$

It remains to optimize this inequality over all λ, which leads to $\frac{1}{c} I(p) \leq n^{1/3} B^{2/3}$. $\quad \square$

17 Proof of Theorem 1.1

Using the regularization operation, we may assume that the densities p_j are C^∞-smooth, so that

$$
b_j = \|p_j\|_{\mathrm{TV}} = \int |\nabla p_j(x)| \, dx < \infty, \quad j = 1, 2, 3 \tag{17.1}
$$

(to give more details, let us return to this reduction in the end of the proof).

The convolution $p = p_1 * p_2 * p_3$ is defined by

$$
p(x) = \iint p_1(x - y - z) p_2(y) p_3(z) \, dy \, dz, \quad x \in \mathbb{R}^n.
$$

By Proposition 12.1, the density p belongs to the Sobolev space $W_1^1(\mathbb{R}^n)$. In fact, its generalized partial derivatives may be expressed as convolutions. Indeed, by Proposition 10.2, for every $i = 1, \ldots, n$, p has a generalized derivative of the third order

$$
D_i^3 p = D_i p_1 * D_i p_2 * D_i p_3,
$$

so that

$$
\int |D_i^3 p(x)| \, dx \leq \|D_i p_1\|_1 \|D_i p_2\|_1 \|D_i p_3\|_1.
$$

We are in position to apply Proposition 16.2 which yields

$$
I(p) \leq c_0 n^{1/3} \Big(\sum_{i=1}^{n} \|D_i p_1\|_1 \|D_i p_2\|_1 \|D_i p_3\|_1 \Big)^{2/3}
$$

with constant $c_0 = 10$, or equivalently,

$$
I(p)^{3/2} \leq c_0^{3/2} \sqrt{n} \sum_{i=1}^{n} \| \langle \nabla p_1, e_i \rangle \|_1 \| \langle \nabla p_2, e_i \rangle \|_1 \| \langle \nabla p_3, e_i \rangle \|_1, \tag{17.2}
$$

where $e_1, \ldots, e_n$ denote the canonical basis in $\mathbb{R}^n$. But, as emphasized in Corollary 15.4, the functional $I(p)$ does not depend on the choice of the orthonormal basis, in the sense that $I(p) = I(p_U)$ for any linear othogonal map $U : \mathbb{R}^n \to \mathbb{R}^n$, where $p_U(x) = p(U(x))$. Hence, the inequality (17.2) remains to hold for any orthonormal basis $E = (e_1, \ldots, e_n)$ in $\mathbb{R}^n$.

Every such collection E may be viewed as an element of the orthogonal group $O(n)$, which is equipped with the Haar probability measure π. The map $E \to e_i$ pushes forward π to the uniform distribution σ_{n-1} on the unit sphere $S^{n-1} = \{\theta \in \mathbb{R}^n : |\theta| = 1\}$. Hence, averaging (17.2) over the measure π, we arrive at the relation

$$I(p)^{3/2} \leq (c_0 n)^{3/2} \int L(\theta)\, d\sigma_{n-1}(\theta), \qquad (17.3)$$

where

$$L(\theta) = \iiint |\langle \nabla p_1(x), \theta \rangle|\, |\langle \nabla p_2(y), \theta \rangle|\, |\langle \nabla p_3(x), \theta \rangle|\, dx\, dy\, dz, \quad \theta \in S^{n-1}.$$

In order to estimate the integral in (17.3), it is sufficient to bound the L^3-norm of the linear functional $f(\theta) = \langle v, \theta \rangle$ with $v \in S^{n-1}$ over the measure σ_{n-1} via its L^4-norm. As is well-known, if $Z = (Z_1, \ldots, Z_n)$ is a standard normal random vector in $\mathbb{R}^n$, then $Z/|Z|$ is independent of $|Z|$ and is uniformly distributed on the sphere. Therefore, $\langle v, Z \rangle /|Z|$ has the same distribution as $f(\theta)$. In addition, by independence, and since $\langle v, Z \rangle \sim N(0,1)$,

$$3 = \mathbb{E}\, \langle v, Z \rangle^4 = \mathbb{E}\, \Big(\frac{\langle v, Z \rangle}{|Z|} \Big)^4 |Z|^4 = \mathbb{E}\, f(\theta)^4\, \mathbb{E}\, |Z|^4.$$

Here, using the independence of Z_i, we also have

$$\mathbb{E}\, |Z|^4 = (\mathbb{E}\, |Z|^2)^2 + \mathrm{Var}(|Z|^2) = n^2 + n\mathrm{Var}(Z_1^2) = n^2 + 2n,$$

that is,

$$\mathbb{E}\, f(\theta)^4 = \frac{3}{n^2 + 2n}.$$

This gives

$$\|f\|_{L^3(\sigma_{n-1})} \leq \|f\|_{L^4(\sigma_{n-1})} = \Big(\frac{3}{n^2 + 2n} \Big)^{1/4} \leq \frac{3^{1/4}}{\sqrt{n}}.$$

Hence, for any 3 linear functionals $f_j(\theta) = \langle v_j, \theta \rangle$, $v_j \in \mathbb{R}^n$, we have, by Hölder's inequality,

$$
\begin{aligned}
\int |f_1(\theta) f_2(\theta) f_3(\theta)|\, d\sigma_{n-1}(\theta) &\leq \|f_1\|_{L^3(\sigma_{n-1})} \|f_2\|_{L^3(\sigma_{n-1})} \|f_3\|_{L^3(\sigma_{n-1})} \\
&\leq \frac{3^{3/4}}{n^{3/2}}\, |v_1|\, |v_2|\, |v_3|.
\end{aligned}
$$

Applying this with $v_j = \nabla p_j(x)$, (17.3) is simplified to

$$I(p)^{3/2} \leq c_0^{3/2}\, 3^{3/4} \iiint |\nabla p_1(x)|\, |\nabla p_2(y)|\, |\nabla p_3(z)|\, dx\, dy\, dz,$$

which is the same as

$$I(p) \leq c\, (b_1 b_2 b_3)^{2/3} \qquad (17.4)$$

wiith constant $c = c_0 \sqrt{3} < 18$, according to (17.1).

As the last step, we use the regularized densities $p_{j,\varepsilon} = (p_j)_\varepsilon$, $\varepsilon > 0$. By the assumption, the densities p_j of X_j have finite total variation norms $b_j = \|p_j\|_{\mathrm{TV}}$. By Proposition 12.1, the random vector $X = X_1 + X_2 + X_3$ has an absolutely continuous distribution,

whose density p belongs to the Sobolev space $W_1^1(\mathbb{R}^n)$. Let $X_\varepsilon = X_{1,\varepsilon} + X_{2,\varepsilon} + X_{3,\varepsilon}$ with independent summands $X_{j,\varepsilon}$ having densities $p_{j,\varepsilon}$. By the previous step (17.4),

$$I(X_\varepsilon) \leq c \left(\|p_{1,\varepsilon}\|_{\mathrm{TV}} \|p_{2,\varepsilon}\|_{\mathrm{TV}} \|p_{3,\varepsilon}\|_{\mathrm{TV}} \right)^{2/3}.$$

As we know, cf. (11.4), the total variation norm may only decrease under regularization, so that we get

$$I(X_\varepsilon) \leq c \left(\|p_1\|_{\mathrm{TV}} \|p_2\|_{\mathrm{TV}} \|p_3\|_{\mathrm{TV}} \right)^{2/3}. \tag{17.5}$$

On the other hand, according to Corollaries 15.2-15.3, the Fisher information is continuous under regularization, that is, $I(X_\varepsilon) \to I(X)$ as $\varepsilon \to 0$. Hence, (17.5) yields in the limit the desired relation (1.4) with constant $c = 18$. $\qquad\square$

References

[1] Billingsley, P. Convergence of probability measures. John Wiley & Sons, Inc., New York-London-Sydney 1968 xii+253 pp. MR0233396

[2] Bobkov, S. G. Moments of the scores. IEEE Trans. Inform. Theory 65 (2019), no. 9, 5294–5301. MR4009234

[3] Bobkov, S. G.; Chistyakov, G. P.; Götze, F. Fisher information and the central limit theorem. Probab. Theory Relat. Fields 159 (2014), 1–59. MR3201916

[4] Bobkov, S. G.; Roberto, C. Entropic isoperimetric inequalities. Preprint (2021). To appear in: High Dimensional Probability Proceedings, vol. 9, Progress in Probability, Birkhäuser/Springer.

[5] Bogachev, V. I. Remarks on integration by parts in infinite dimension. Selected papers from the 21st Winter School on Abstract Analysis (Poděbrady, 1993). Acta Univ. Carolin. Math. Phys. 34 (1993), no. 2, 11–29. MR1282962

[6] Bogachev, V. I. Differentiable measures and the Malliavin calculus. Mathematical Surveys and Monographs, 164. American Mathematical Society, Providence, RI, 2010. xvi+488 pp. MR2663405

[7] Cohen, M. The Fisher information and convexity. IEEE Trans. Inform. Theory 14 (1968), 591–592.

[8] Courtade, T. A.; Fathi, M.; Pananjady, A. Existence of Stein kernels under a spectral gap, and discrepancy bounds. Ann. Inst. Henri Poincaré Probab. Stat. 55 (2019), no. 2, 777–790. MR3949953

[9] Dunford, N.; Schwartz, J. T. Linear Operators. I. General Theory. With the assistance of W. G. Bade and R. G. Bartle. Pure and Applied Mathematics, vol. 7, Interscience Publishers, Inc., New York; Interscience Publishers, Ltd., London 1958 xiv+858 pp. MR0117523

[10] Evans, L. C.; Gariepy, R. F. Measure theory and fine properties of functions. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992. viii+268 pp. MR1158660

[11] Johnson, O. Information theory and the central limit theorem. Imperial College Press, London, 2004. xiv+209 pp. MR2109042

[12] Kolmogorov, A. N.; Fomin, S. V. Introductory real analysis. Translated from the second Russian edition and edited by Richard A. Silverman. Corrected reprinting. Dover Publications, Inc., New York, 1975. xii+403 pp. MR0377445

[13] Krugova, E. P. On the integrability of logarithmic derivatives of measures. (Russian) Mat. Zametki 53 (1993), no. 5, 76–86; translation in Math. Notes 53 (1993), no. 5-6, 506–512. MR1325617

[14] Meyer, P.-A. Probability and potentials. Blaisdell Publishing Co. [Ginn and Co.], Waltham, Mass.-Toronto, Ont.-London 1966 xiii+266 pp. MR0205288

[15] Stam, A. J. Some inequalities satisfied by the quantities of information of Fisher and Shannon. Information and Control 2 (1959), 101–112. MR0109101

[16] Uglanov, A.V. The quotient of smooth measures is a smooth function. (Russian) Izv. Vyssh. Uchebn. Zaved. Mat. 1989, no. 9, 72–76; translation in Soviet Math. (Iz. VUZ) 33 (1989), no. 9, 71–75. MR1026829

[17] Ziemer, W. P. Weakly differentiable functions. Sobolev spaces and functions of bounded variation. Graduate Texts in Mathematics, 120. Springer-Verlag, New York, 1989. xvi+308 pp. MR1014685