Score functions, generalized relative Fisher information and applications

Giuseppe Toscani*

January 19, 2016

Abstract

Generalizations of the linear score function, a well-known concept in theoretical statistics, are introduced. As the Gaussian density and the classical Fisher information are closely related to the linear score, nonlinear (respectively fractional) score functions allow to identify generalized Gaussian densities (respectively Lévy stable laws) as the (unique) probability densities for which the score of a random random variable X is proportional to $-X$. In all cases, it is shown that the variance of the relative to the generalized Gaussian (respectively Lévy) score provides an upper bound for L^1-distance from the generalized Gaussian density (respectively Lévy stable laws). Connections with nonlinear and fractional Fokker–Planck type equations are introduced and discussed.

Keywords: Score function; Fisher information; Fractional calculus; Fokker–Planck type equations.
MSC: Primary 60F05–94A17; Secondary 26A33

1 Introduction

Let X be a random vector with differential probability density functions $f(x)$, with $x \in \mathbb{R}^n$, $n \geq 1$. The Fisher information of X is defined by

$$I(X) = I(f) = \int_{\{f>0\}} \frac{|\nabla f(x)|^2}{f(x)}\, dx. \tag{1}$$

*Department of Mathematics, University of Pavia, via Ferrata 5, 27100 Pavia, Italy. giuseppe.toscani@unipv.it

Likewise, given a pair of random vector X and Y, with differential probability density functions $f(x)$ and $g(x)$, the relative to Y Fisher information of X and Y is defined by

$$I(X|\,Y) = I(f|\,g) = \int_{\{f,\,g>0\}} \left| \frac{\nabla f(x)}{f(x)} - \frac{\nabla g(x)}{g(x)} \right|^2 f(x)\,dx, \tag{2}$$

Both the Fisher and the relative Fisher information appear to be very useful in many applications, which range from information theory [13, 23] to probability theory [3, 4, 15, 16] and kinetic theory [25, 26], where (2), in connection with Fokker–Planck type equations, is commonly referred as relative entropy production.

These notions are closely linked to the concept of linear score function, mostly used in theoretical statistics [8, 21]. Given a random vector X with differential probability density function $f(x)$, its linear score is given by

$$\rho(X) = \frac{\nabla f(X)}{f(X)}. \tag{3}$$

Indeed, for a random vector with zero expectation, the Fisher information (1) is just the variance of $\rho(X)$. Analogously, given the pair of random vectors X and Y, the linear score function of the pair relative to X is represented by

$$\tilde\rho(X) = \frac{\nabla f(X)}{f(X)} - \frac{\nabla g(X)}{g(X)}. \tag{4}$$

Hence the relative (to Y) Fisher information between X and Y with zero expectation is just the variance of $\tilde\rho(X)$. This notion is satisfying because it represents the variance of some error due to the mismatch between the prior distribution f supplied to the estimator and the actual distribution g. Obviously, whenever f and g are identical, then the relative score and the relative Fisher information are equal to zero.

In connection with the linear score (2), Gaussian variables are easily distinguished from others. Let $z_\sigma(x)$ denote the Gaussian density in $\mathbb{R}^n$ with zero mean and variance $n\sigma$

$$z_\sigma(x) = \frac{1}{(2\pi\sigma)^{n/2}} \exp\left(-\frac{|x|^2}{2\sigma}\right). \tag{5}$$

Then a Gaussian random vector Z of density z_σ is uniquely defined by the identity

$$\rho(Z) = -Z/\sigma.$$

Also, the relative (to X) score function of X and Z takes the simple expression

$$\tilde\rho(X) = \frac{\nabla f(X)}{f(X)} + \frac{X}{\sigma}, \tag{6}$$

so that the relative to Z Fisher information of X and Z is given by

$$I(X|\,Z) = I(f|\,z_\sigma) = \int_{\{f>0\}} \left| \frac{\nabla f(x)}{f(x)} + \frac{x}{\sigma} \right|^2 f(x)\,dx. \tag{7}$$

This particular property of Gaussian density with respect to the linear score, suggested how to modify the definition to distinguish stable laws from others [27]. The fractional linear score of a random variable X of density f is defined by

$$\rho_\lambda(X) = \frac{\mathcal{D}_{\lambda-1} f(X)}{f(X)}, \tag{8}$$

where $\mathcal{D}_\alpha f(x)$, $0 < \alpha < 1$ is the fractional derivative of order α of $f(x)$ (cf. the Appendix for the definition). Starting from (8), the definition of relative (to Z_λ) fractional Fisher information of a random variable X follows [27]. Let Z_λ denote a Lévy stable law of density $\omega_\lambda(x)$, $x \in \mathbb{R}$, expressed by its Fourier transform

$$\widehat{\omega}_\lambda(\xi) = e^{-|\xi|^\lambda}, \tag{9}$$

with $1 < \lambda < 2$. For a given random variable X *sufficiently close* to Z_λ, the relative (to Z_λ) fractional Fisher information of X is expressed by the formula

$$I_\lambda(X|\,Z_\lambda) = I_\lambda(f|\,\omega_\lambda) = \int_{\{f>0\}} \left(\frac{\mathcal{D}_{\lambda-1} f(x)}{f(x)} - \frac{\mathcal{D}_{\lambda-1}\omega_\lambda(x)}{\omega_\lambda(x)} \right)^2 f(x)\,dx, \tag{10}$$

The fractional score (8) and the relative fractional Fisher information (10) are obtained from (3) (respectively (2)), by substituting the standard derivative with the fractional derivative of order $\lambda - 1$, which is such that $0 < \lambda - 1 < 1$ for $1 < \lambda < 2$. As the linear score function $f'(X)/f(X)$ of a random variable X with a (smooth) probability density f identifies Gaussian variables as the unique random variables for which the linear score is proportional to $-X$, Lévy symmetric stable laws of order λ are now identified as the unique random variables Y for which the new defined linear fractional score is proportional to $-Y$ (cf. Section 3). Consequently, the relative (to Z_λ) fractional Fisher information (10) can be equivalently written as

$$I_\lambda(X|\,Z_\lambda) = \int_{\{f>0\}} \left(\frac{\mathcal{D}_{\lambda-1} f(x)}{f(x)} + \frac{x}{\lambda} \right)^2 f(x)\,dx. \tag{11}$$

The analysis of [27] made evident that, given a suitable score function, the identification of the random variables X which possess a score function of the form $-CX$ allows to obtain a variance of the relative score (a relative Fisher information) with extremely good properties with respect to convolutions, that provide in addition a control of various distances between densities (typically $L^1(\mathbb{R}^n)$-distance).

In this paper, we extend the notion of score to cover situations different from the ones described by Gaussian and Lévy random variables. The new definition of nonlinear score of Section 2 allows to identify generalized Gaussians, say W, as the unique random variables for which the nonlinear score is proportional to $-W$. Last, in Section 4 we will review various results concerned with relative linear and nonlinear Fisher information at the light of this connection with the notion of score functions.

2 Scores and generalized Gaussians

In the rest of this paper, if not explicitly quoted, and without loss of generality, we will always assume that any random vector X we will consider is centered, i.e. $E(X) = 0$, where as usual $E(\cdot)$ denotes mathematical expectation.

Let X be a random vector with a differentiable probability distribution with density function $f(x)$, $x \in \mathbb{R}^n$, $n \geq 1$, depending on a parameter $\theta \in \mathbb{R}^n$. Then the function

$$L(\theta; x) = f_\theta(x),$$

considered as a function of θ, is called the likelihood function (of θ, given the outcome x of X). For many applications, the natural logarithm of the likelihood function, (the log-likelihood), is more convenient to work with.

In theoretical statistics, the score or efficient score [8, 21] is the gradient, with respect to the parameter θ, of the log-likelihood. The score $\rho_L(X)$ can be found through the chain rule

$$\rho_L(\theta, X) = \frac{1}{L(\theta; X)} \nabla_\theta L(\theta; X). \tag{12}$$

Thus the score indicates the sensitivity of $L(\theta; X)$ (its gradient normalized by its value). In older literature, the term *linear score* refers to the score evaluated with respect to an infinitesimal translation of a given density. In this case, the likelihood of an observation is given by a density of the form $L(\theta; X) = f(X + \theta)$. According to this definition, given a random vector X in $\mathbb{R}^n$, $n \geq 1$, distributed with a differentiable probability density function $f(x)$, its linear score ρ (at $\theta = 0$) is expressed by

$$\rho(X) = \frac{\nabla f(X)}{f(X)}. \tag{13}$$

The linear score has zero mean, and its variance is just the Fisher information (1) of X. It can be easily verified that a Gaussian random vector Z of density z_σ given by (5) satisfies the equality

$$\rho(Z) = -Z/\sigma. \tag{14}$$

Vice versa, equality (14) uniquely characterizes Gaussian random vectors. Thus, Gaussian vectors Z are uniquely characterized by a linear score $\rho(Z)$ proportional to $-Z$.

Among other possible extensions, one can generalize the notion of *linear score* by evaluating it with respect to an x-dependent infinitesimal translation of the underlying density. In this case, given a (nonnegative) function $\Phi(x)$, the likelihood of an observation takes the form $L(\theta; X) = f(X + \Phi(X)\,\theta)$. According to this X-*dependent* translation, given a random vector X in $\mathbb{R}^n$, $n \geq 1$, distributed with a differentiable probability density function $f(x)$, its Φ-score ρ_Φ (at $\theta = 0$) reads

$$\rho_\Phi(X) = \frac{\Phi(X)\nabla f(X)}{f(X)}. \tag{15}$$

A leading example is obtained by choosing $\Phi(x) = pf^{p-1}(x)$, with $p > (n-2)/n$. In this case the score ρ_p, $p > 0$ takes the form

$$\rho_p(X) = \frac{\nabla f^p(X)}{f(X)}. \tag{16}$$

As previously discussed in the classical case, it is natural to look for random vectors X characterized by a score $\rho_p(X)$ proportional to $-X$. Without loss of generality, let us set the constant of proportionality equal to one. Consequently, we have to look for random vectors X with smooth probability density f which satisfy the equality

$$\frac{\nabla f^p(X)}{f(X)} = \frac{p}{p-1}\nabla f^{p-1}(X) = -X = -\frac{1}{2}\nabla |X|^2. \tag{17}$$

Equality (17) is satisfied if and only if, for a given constant D

$$f^{p-1}(X) + \frac{p-1}{2p}\,|X|^2 = D^2. \tag{18}$$

Let $p < 1$. Then, the nonnegative function

$$z_p(x) = \frac{1}{\left(D^2 + \frac{p-1}{2p}|x|^2\right)^{1/(1-p)}}, \qquad p < 1 \tag{19}$$

is a solution to (18). Clearly, since $p > (n-2)/n$ this function has a bounded integral over $\mathbb{R}^n$, that decreases with respect to D^2 from $+\infty$ to zero. Hence, (19) is a probability density function for a suitable choice of the constant D.

If now $p > 1$, let us consider a random vector X such that, for a given constant $D > 0$, $|X| \leq \sqrt{2p/(p-1)}D$. In this case, equality (18) can be equivalent to (17) only on the set of values assumed by X. On this set the function

$$f(x) = \left(D^2 - \frac{p-1}{2p}|x|^2\right)^{1/(p-1)}$$

solves both (18) and (17). For a given function $h(x)$ let $h_+(x)$ define the positive part of h, that is $h_+(x) = \max\{h(x), 0\}$. Then, for a suitable value of the positive constant D the function

$$z_p(x) = \left(D^2 - \frac{p-1}{2p}|x|^2\right)_+^{1/(p-1)}, \qquad p > 1 \tag{20}$$

is a nonnegative probability density function solving (17) on the set of values assumed by the random vector X. Therefore, if W_p is a random vector with density z_p, it holds

$$\frac{\nabla z_p^p(W_p)}{z_p(W_p)} = -W_p. \tag{21}$$

Analogously to the linear case, given the pair of random vectors X and Y, the (non linear) score function of the pair relative to X is now represented by

$$\tilde{\rho}_p(X) = \frac{\nabla f^p(X)}{f(X)} - \frac{\nabla g^p(X)}{g(X)}. \tag{22}$$

In this case the relative (to Y) generalized Fisher information between X and Y, namely the variance of $\tilde{\rho}_p(X)$ is given by

$$I_p(X|\,Y) = I_p(f|\,g) = \int_{\{f,\,g>0\}} \left|\frac{\nabla f^p(x)}{f(x)} - \frac{\nabla g^p(x)}{g(x)}\right|^2 f(x)\,dx. \tag{23}$$

If $p < 1$, the previous computations show that that, with respect to the generalized Gaussian W_p, the relative Fisher information takes the simple form

$$I_p(X|\,W_p) = I_p(f|\,z_p) = \int_{\{f>0\}} \left|\frac{\nabla f^p(x)}{f(x)} + x\right|^2 f(x)\,dx. \tag{24}$$

This generalized Fisher information was at the basis of the study of convergence towards equilibrium of the solution of nonlinear diffusion equations [7, 10].

3 Fractional scores and Lévy densities

The concept of linear score has been extended in [27] to cover fractional derivatives. Given a random variable X in $\mathbb{R}$ distributed with a probability density function $f(x)$ that has a well-defined fractional derivative of order α, with $0 < \alpha < 1$, its linear fractional score, denoted by $\rho_{\alpha+1}$ is given by

$$\rho_{\alpha+1}(X) = \frac{\mathcal{D}_\alpha f(X)}{f(X)}. \tag{25}$$

Thus the linear fractional score indicates the non local (fractional) sensitivity of $f(X+\theta)$ at $\theta = 0$ (its fractional derivative normalized by its value). The fractional score of X is linear in X if and only if X is a Lévy distribution of order $\alpha + 1$. Indeed, for a given positive constant C, the identity

$$\rho_{\alpha+1}(X) = -CX,$$

is verified if and only if, on the set $\{f > 0\}$

$$\mathcal{D}_\alpha f(x) = -Cxf(x). \tag{26}$$

Passing to Fourier transform, this identity yields

$$i\xi|\xi|^{\alpha-1}\widehat{f}(\xi) = -iC\frac{\partial \widehat{f}(\xi)}{\partial \xi},$$

and from this follows

$$\widehat{f}(\xi) = \widehat{f}(0)\exp\left\{-\frac{|\xi|^{\alpha+1}}{C(\alpha+1)}\right\}. \tag{27}$$

Finally, by choosing $C = (\alpha+1)^{-1}$, and imposing that $f(x)$ is a probability density function (i.e. by fixing $\widehat{f}(\xi = 0) = 1$), we obtain that the Lévy stable law of order $\alpha + 1$ is the unique probability density solving (26).

It is important to remark that, unlike in the case of the linear score, the variance of the fractional score is in general unbounded. One can easily realize this by looking at the variance of the fractional score in the case of a Lévy variable. For a Lévy variable, in fact, the variance of the fractional score coincides with a multiple of its variance, which is unbounded [12, 17]. For this reason, a consistent definition in this case is represented by the relative fractional score (10). Thanks to (27), a Lévy random variable of density z_λ, with $1 < \lambda < 2$ is uniquely defined by a linear fractional score function

$$\rho_\lambda(Z_\lambda) = -\frac{Z_\lambda}{\lambda}.$$

Then, the relative (to X) fractional score function of X and Z_λ assumes the simple expression

$$\tilde{\rho}_\lambda(X) = \frac{\mathcal{D}_{\lambda-1}f(X)}{f(X)} + \frac{X}{\lambda}, \tag{28}$$

which induces a (relative to the Lévy) fractional Fisher information (in short λ-Fisher relative information)

$$I_\lambda(X|Z_\lambda) = I_\lambda(f) = \int_{\{f>0\}} \left(\frac{\mathcal{D}_{\lambda-1}f(x)}{f(x)} + \frac{x}{\lambda}\right)^2 f(x)\,dx. \tag{29}$$

The fractional Fisher information is always greater or equal than zero, and it is equal to zero if and only if X is a Lévy symmetric stable distribution of order λ. At difference with the relative standard relative Fisher information, I_λ is well-defined any time that the the random variable X has a probability density function which is suitably closed to the Lévy stable law (typically lies in a subset of the domain of attraction). If $\mathcal{P}_\lambda$ denotes the set of probability density functions such that $I_\lambda(f) < +\infty$, we showed in [27] that $\mathcal{P}_\lambda$ is not empty, and contains important classes of densities, like Linnik's probability density [19, 20].

4 Relative scores and L^1-distances

The discussion of the previous sections enlightens an interesting property of score functions. As it happens for the standard linear score in $\mathbb{R}^n$, where $\rho(X)$ is proportional to $-X$ only when X is a Gaussian random vector, all the previously defined generalized score functions identify uniquely a density $\bar f$ which is such that the corresponding score evaluated at the random variable $\bar X$ of density $\bar f$ is proportional to $-\bar X$. In these cases, for a given positive constant γ the vanishing of the relative score function

$$\tilde\rho(X) = \rho(X) + \gamma X \tag{30}$$

uniquely identifies the densities $\bar X$ with score proportional to $-\bar X$. Consequently, and only for these densities, the variance of $\tilde\rho(X)$ is equal to zero.

For any other density, the size of the (positive) variance of (30) intuitively furnishes a measure of how the probability density f of X differs from the density $\bar f$ of $\bar X$. In many cases, indeed, it has been proven that the variance of the relative score (30), namely the generalized relative (to $\bar X$) Fisher information of X bounds from above the L^1-distance between f and $\bar f$.

In the case of the generalized Gaussian described in Section 2, these results have been usually derived by studying the convergence towards equilibrium of nonlinear diffusion equations by entropy methods, an idea which was developed first in [7, 10], and subsequently improved in a number of papers [1, 2, 5, 6, 11]. In short, consider the solution of the nonlinear Fokker–Planck equation in $\mathbb{R}^n$ with exponent $p \in (0,1)$, $p > (n-2)/n$, given by

$$\frac{\partial u}{\partial t} = \nabla \cdot [\nabla u^p + x\, u] \quad t > 0\,, \tag{31}$$

with initial datum a probability density function $u(t=0,\cdot) = u_0$. Since equation (31) can be rewritten in the equivalent form

$$\frac{\partial u}{\partial t} = \nabla \cdot \left[u \left(\frac{\nabla u^p}{u} + x \right) \right], \tag{32}$$

the vanishing of the (nonlinear) relative score

$$\frac{\nabla u^p(X)}{u(X)} + X$$

identifies the stationary solution to (31). Let u_∞ be the unique stationary solution such that $\int_{\mathbb{R}^d} u\,dx = \int_{\mathbb{R}^d} u_\infty\,dx = 1$. Then, by the result of Section 2 $u_\infty(x) = z_p(x)$, where z_p is given by (19).

It has been established in [1, 2] that, if u is a solution of (31), the *relative entropy* (or *free energy*)

$$F_p[u|u_\infty] := \frac{1}{p-1}\int_{\mathbb{R}^d}\left[u^p - u_\infty^p - p\,u_\infty^{p-1}\,(u-u_\infty)\right]dx$$

decays according to

$$\frac{d}{dt}\,F_p[u(\cdot,t)|u_\infty] = -I_p(u(\cdot,t)|u_\infty)$$

where I_p is the *entropy production term* or *relative Fisher information*, given by (24). If $m \in [n/(n+2), 1)$, according to [2], these two functionals are related by a Gagliardo-Nirenberg interpolation inequality, namely

$$F_p[u|u_\infty] \le \frac{1}{4}\,I_p[u|u_\infty]\,. \tag{33}$$

The same inequality holds when $p > 1$ [7, 10]. On the other hand, Csiszar-Kullback type inequalities [7, 10] imply that, for a suitable universal constant $C > 0$

$$F_p[u|u_\infty] \ge C\|u-u_\infty\|^2_{L^1}. \tag{34}$$

Therefore, considering that $u_\infty(x) = z_p(x)$, one concludes with the bound

$$\|u - z_p\|^2_{L^1} \le C I_p[u|z_p]. \tag{35}$$

This shows that, at least for generalized Gaussians and $p > n/(n+2)$, the variance of the relative score bounds the L^1- distance between a general density f and the density of the generalized Gaussian. It is remarkable that this bound is obtained by resorting to the study of the time decay of the relative entropy in a nonlinear Fokker–Planck equation.

In [28] a similar strategy has been applied to obtain a lower bound for the relative fractional Fisher information (11). To this extent, the time decay of the relative (to Z_λ) Shannon entropy, defined by

$$H(X|\,Z_\lambda) = H(f|\,\omega_\lambda) = \int_{\mathbb{R}} f(x)\log\frac{f(x)}{\omega_\lambda(x)}\,dx, \tag{36}$$

was studied along the solution to the Fokker–Planck equation with fractional diffusion

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(D_{\lambda-1}u + \frac{x}{\lambda}u\right), \tag{37}$$

where $1 < \lambda < 2$, with an initial datum $u_o(x)$ belonging to the domain of normal attraction of the Lévy stable law ω_λ. In consequence of the (explicit) time decay of the relative entropy (36), a new inequality relating the relative (to Z_λ) Shannon entropy in terms of the relative (to Z_λ) fractional Fisher information and the standard Fisher information was obtained (the analogous of (33)). This new inequality reads

$$H(X|\,Z_\lambda) \le \lambda\, 2^{1/\lambda}\, \min\{I(X),\, I(Z_\lambda)\}^{1/2}\, I_\lambda(X|\,Z_\lambda)^{1/2}. \tag{38}$$

As before, Csiszar-Kullback inequality [9, 14]

$$\|f - \omega_\lambda\|^2_{L^1} \le 2H(f|\,\omega_\lambda),$$

allows to bound from above the L^1-distance of f and ω_λ in terms of the relative Shannon entropy. Hence, inequality (38) shows that the relative (to Z_λ) fractional Fisher information of a random variable X with density f in the domain of attraction of I_λ provides a L^1-bound for the L^1-distance between f and ω_λ.

5 Conclusions

In this note, we emphasized the notion of relative score functions, in connection with their applications as possible estimators of L^1-distances between densities. The main result here is that the variance of the relative score $\tilde\rho(X) = \rho(X) + \gamma X$ provides in significant cases explicit bounds on the L^1-distance between the probability density of X and the (unique) density for which the relative score vanishes. A collection of previous results, mainly connected with the study of convergence to equilibrium of linear and nonlinear Fokker–Planck equation shows that this property holds true for generalized Gaussians and Lévy stable laws. In the linear score case, however, further results are available. In fact, it can be proven directly, without resorting to methods directly linked to the standard Fokker–Planck equation, that the relative (to the Gaussian) Fisher information bounds from above the Hellinger distance between the Gaussian density and a density f (cf. [16] and the references therein). Then, since Hellinger distance between probability densities is stronger than the L^1-distance, inequality (35) follows with $p = 1$. This property of the linear score leads to conjecture that there are other possible alternative proofs to obtain for the nonlinear score both lower bounds as in (35), and moreover, stronger bounds in terms of the Hellinger distance.

6 Appendix

In this short appendix we summarize the mathematical notations and the meaning of the fractional derivative. Given a probability density $f(x)$, $x \in \mathbb{R}^n$, we define its *Fourier transform* $\mathcal{F}(f)$ by

$$\mathcal{F}(f)(\xi) = \widehat{f}(\xi) := \int_{\mathbb{R}^n} e^{-i\,\xi\cdot x} f(x)\,dx, \qquad \forall \xi \in \mathbb{R}^n.$$

Let us set $n = 1$. Then, the one-dimensional derivative $\mathcal{D}_\alpha$ is defined as follows. For $0 < \alpha < 1$ we let R_α be the one-dimensional *normalized* Riesz potential operator defined for locally integrable functions by [22, 24]

$$R_\alpha(f)(x) = S(\alpha) \int_{\mathbb{R}} \frac{f(y)\,dy}{|x-y|^{1-\alpha}}.$$

The constant $S(\alpha)$ is chosen to have

$$\widehat{R_\alpha(f)}(\xi) = |\xi|^\alpha \widehat{f}(\xi). \tag{39}$$

Since for $0 < \alpha < 1$ it holds [18]

$$\mathcal{F}|x|^{\alpha-1} = |\xi|^{-\alpha}\pi^{1/2}\Gamma\left(\frac{1-\alpha}{2}\right)\Gamma\left(\frac{\alpha}{2}\right), \tag{40}$$

where, as usual $\Gamma(\cdot)$ denotes the Gamma function, the value of $S(\alpha)$ is given by

$$S(\alpha) = \left[\pi^{1/2}\Gamma\left(\frac{1-\alpha}{2}\right)\Gamma\left(\frac{\alpha}{2}\right)\right]^{-1}.$$

Note that $S(\alpha) = S(1-\alpha)$.

We then define the fractional derivative of order α of a real function f as ($0 < \alpha < 1$)

$$\frac{d^\alpha f(x)}{dx^\alpha} = \mathcal{D}_\alpha f(x) = \frac{d}{dx} R_{1-\alpha}(f)(x). \tag{41}$$

Thanks to (39), in Fourier variables

$$\widehat{\mathcal{D}_\alpha f}(\xi) = i\frac{\xi}{|\xi|}|\xi|^\alpha \widehat{f}(\xi). \tag{42}$$

Acknowledgments: This work has been written within the activities of the National Group of Mathematical Physics of INDAM (National Institute of High Mathematics). The support of the project "Optimal mass transportation, geometrical and functional inequalities with applications", financed by the Minister of University and Research, is kindly acknowledged.

References

[1] A. Blanchet, M. Bonforte, J. Dolbeault, G. Grillo, and J.-L. Vázquez, Hardy-Poincaré inequalities and applications to nonlinear diffusions, *Comptes Rendus Mathématique*, **344** (2007), 431–436.

[2] A. Blanchet, M. Bonforte, J. Dolbeault, G. Grillo, and J. Vázquez, Asymptotics of the fast diffusion equation via entropy estimates, *Arch. Rational Mech. Anal.*, **191** (2009), 347–385.

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

[4] S.G. Bobkov, G.P. Chistyakov, and F. Götze, Fisher information and convergence to stable laws, *Bernoulli* **20** (3) 1620–1646 (2014).

[5] M. Bonforte, J. Dolbeault, G. Grillo, and J.-L. Vázquez, Sharp rates of decay of solutions to the nonlinear fast diffusion equation via functional inequalities. *Proc. Natl. Acad. Sci. USA* **107** (2010), 16459–16464.

[6] M. Bonforte, G. Grillo, and J.-L. Vázquez, Special fast diffusion with slow asymptotics, entropy method and flow on a riemannian manifold, *Arch. Rational Mech. Anal.*, **196** (2010), 631–680.

[7] J.A. Carrillo and G. Toscani, Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity, *Indiana Univ. Math. Journal*, **49** 113–142 (2000).

[8] D.R. Cox, and D.V. Hinkley *Theoretical Statistics*, Chapman & Hall, London 1974.

[9] I. Csiszar, Information-type measures of difference of probability distributions and indirect observations. *Stud. Sci. Math. Hung.*, **2** 299–318 (1967).

[10] M. Del Pino, J. Dolbeault, Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions. *J. Math. Pures Appl.* **81** (2002), 847–875.

[11] J. Dolbeault, G. Toscani, Improved interpolation inequalities, relative entropy and fast diffusion equations. *Annales de l' Institut Henri Poincaré (C) Non Linear Analysis*, **30** (5) (2013) 917–934.

[12] B.V. Gnedenko, and A.N. Kolmogorov, *Limit distributions for sums of independent random variables*, Addison-Wesley, Cambridge, Mass. 1954.

[13] D. Guo, Relative Entropy and Score Function: New Information-Estimation Relationships through Arbitrary Additive Perturbation, in *Proc. IEEE Int. Symp. Inform. Theory* 2009, Seoul, Korea June 2009, 814–818.

[14] S. Kullback, A lower bound for discrimination information in terms of variation. *IEEE Trans. Inf. The.*, **4** 126–127 (1967).

[15] O. Johnson, Entropy inequalities and the central limit theorem, *Stochastic Process. Appl.* **88** 291–304 (2000).

[16] O. Johnson, and A. R. Barron, Fisher information inequalities and the central limit theorem, *Probab. Theory Related Fields* **129** 391–409 (2004).

[17] R. G. Laha, and v. K. Rohatgi, *Probability theory*, John Wiley & Sons, New York-Chichester-Brisbane, Wiley Series in Probability and Mathematical Statistics, 1979.

[18] E.H. Lieb, Sharp constants in the hardy-Littlewood-Sobolev and related inequalities, *Ann. Math.* **118** 349–374 (1983).

[19] Yu.V. Linnik, Linear forms and statistical criteria. II, *Ukrainskii Mat. Zhournal* **5** , 247–290, (1953).

[20] Yu.V. Linnik, Linear forms and statistical criteria. I,II, Selected *Transl. Math. Statist. and Prob.*, **3** Amer. Math. Soc., Providence, R.I., 1–90, (1962).

[21] M. Madiman, and A. Barron, Generalized entropy power inequalities and monotonicity properties of information, *IEEE Trans. Inf. Theory*, **53**, (4) 2317–2329, (2007).

[22] M. Riesz, L'intégrale de Riemann-Liouville et le problème de Cauchy, *Acta Math.* **81** 1–223 (1949).

[23] A.J. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon, *Inf. Contr.* **2**, 101–112, (1959)

[24] E.M. Stein, *Singular integrals and differentiability properties of functions*. Princeton Mathematical Series, Vol. **30**. Princeton University Press, Princeton, 1970.

[25] G. Toscani, Sur l'inegalité logarithmique de Sobolev, *C.R. Acad. Sc. Paris*, **324**, 689–694 (1997).

[26] G. Toscani, Entropy dissipation and the rate of convergence to equilibrium for the Fokker–Planck equation, *Quart. Appl. Math.*, **LVII** 521–541 (1999).

[27] G.Toscani, The fractional Fisher information and the central limit theorem for stable laws. *Ricerche mat.*, in press (2016) , preprint arXiv:1504.07057, (2015).

[28] G.Toscani, Entropy inequalities for stable densities and strengthened central limit theorems, preprint arXiv:1512.05874 (2015).