Information-theoretical inequalities for stable densities

Giuseppe Toscani

Department of Mathematics
University of Pavia, Italy

Nonlocal PDEs and Applications to Geometry, Physics and Probability
Trieste, May 24, 2017

Outline

Outline

1. Entropy and the central limit theorem
   - A short history
   - The fractional Fisher information
   - Monotonicity of the fractional Fisher information
2. Inequalities for relative entropy
   - A logarithmic type Sobolev inequality
   - Convergence results in relative entropy
   - References

- The entropy functional (or Shannon's entropy) of the random vector X in $\mathbb{R}^n$

$$H(X) = H(f) = -\int_{\mathbb{R}^n} f(x) \log f(x)\, dx.$$

- The entropy power inequality Shannon (1948); Stam (1959). If X, Y are independent random vectors

$$e^{\frac{2}{n}H(X+Y))} \geq e^{\frac{2}{n}H(X)} + e^{\frac{2}{n}H(Y)}.$$

- For a Gaussian random vector N_σ with covariance σI.

$$e^{\frac{2}{n}H(N_\sigma))} = 2\pi\sigma e.$$

- If X, Y are independent Gaussian random vectors (with proportional covariances) there is equality in the entropy power inequality.
- The proof is based on Fisher information bounds and on the relationship between entropy and Fisher information

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

- Strong connections of entropy power inequality with the central limit theorem
- Consider the law of (X_i i.i.d.)

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}, \qquad n \geq 1.$$

- Application of the entropy power inequality shows that

$$H(S_2) = H\left(\frac{X_1 + X_2}{\sqrt{2}}\right) \geq H(S_1).$$

- The entropy is increasing at least along the subsequence S_{2^k}.

- The sequence S_n is such that, if the X_i are centered, mass, mean and variance are preserved.
- Like in kinetic theory, where relaxation to equilibrium in the Boltzmann equation can be viewed as a consequence of the increasing of entropy , one could conjecture that $H(S_n)$ is monotonically increasing in n.
- Difficult to prove that

$$H(S_3) \geq H(S_2).$$

- The problem remained open up to 2002.
- Monotonicity verified by Artstein, Ball, Barthe, Naor (2002). Simpler proof in Madiman, Barron (2007).

- In kinetic theory, entropy decays towards the equilibrium density with a certain rate.
- There is a decay rate of $H(S_n)$ towards $H(N_\sigma)$?
- Important to quantify the entropy jump

$$H\left(\frac{X_1 + X_2}{\sqrt{2}}\right) - H(X_1) \geq 0$$

- Recent results Ball, Barthe, Naor (2003), Carlen, Soffer (2011), Ball, Nguyen (2012) for log-concave densities.

- The heat equation in the whole space $\mathbb{R}^n$

$$\frac{\partial u}{\partial t} = \kappa \Delta u, \qquad u(x, t = 0) = f(x)$$

relates Shannon's entropy and Fisher information.

- McKean McKean(1965) , computed the evolution in time of the subsequent derivatives of the entropy functional $H(u(t))$. At the first two orders, with $\kappa = 1$

$$I(f) = \left.\frac{d}{dt}\right|_{t=0} H(u(t)); \quad J(f) = -\frac{1}{2}\left.\frac{d}{dt}\right|_{t=0} I(u(t)).$$

- The functional $J(X)$ is given by

$$J(X) = J(f) = \sum_{i,j=1}^{n} \int_{\{f>0\}} [\partial_{ij}(\log f)]^2 f\, dx =$$

$$\sum_{i,j=1}^{n} \int_{\{f>0\}} \left[\frac{\partial_{ij} f}{f} - \frac{\partial_i f \partial_j f}{f^2} \right]^2 f\, dx.$$

- The functionals $J(X)$ and $I(X)$ are related. It is known that

$$J(X) \geq \frac{I^2(X)}{n}.$$

- Fisher information satisfies the inequality $(a, b > 0)$

$$I(X+Y) \le \frac{a^2}{(a+b)^2} I(X) + \frac{b^2}{(a+b)^2} I(Y)$$

- Optimizing over a and b one obtains Stam's Fisher information inequality

$$\frac{1}{I(X+Y)} \ge \frac{1}{I(X)} + \frac{1}{I(Y)}.$$

- Note that for the Gaussian random vector $I(N_\sigma) = n/\sigma$. Hence, equality holds if and only X and Y are Gaussian random vectors with proportional covariance matrices.

- Entropy power inequality implies isoperimetric inequality for entropies. If N is a Gaussian random vector with covariance I, for $t > 0$

$$e^{\frac{2}{n}H(X+2tN))} \geq e^{\frac{2}{n}H(X)} + e^{\frac{2}{n}H(2tN)} = e^{\frac{2}{n}H(X)} + 4t\pi e.$$

- This implies

$$\frac{e^{\frac{2}{n}H(X+2tN))} - e^{\frac{2}{n}H(X)}}{t} \geq 4\pi e.$$

- Letting $t \to 0$

$$I(X)e^{\frac{2}{n}H(X)} \geq 2\pi en.$$

- The isoperimetric inequality for entropies implies logarithmic Sobolev inequality with a remainder [G.T. (2013) Rend. Lincei. Mat. Appl.] .
- Same strategy in Dembo(1989), (cf. Villani(2000)). If N is a Gaussian random vector with covariance I, for $t > 0$

$$1/I(X + 2tN) \geq 1/I(X) + 1/I(2tN) = 1/I(X) + \frac{2t}{n}.$$

- This implies

$$\frac{1/I(X + 2tN) - 1/I(X)}{t} \geq \frac{2}{n}.$$

- Letting $t \to 0$ gives the inequality

$$\frac{1}{I^2(X)} J(X) \geq \frac{1}{n}.$$

- The inequality part of the proof of the concavity of entropy power Costa(1985). If N is a Gaussian random vector with covariance I, the entropy power

$$e^{\frac{2}{n}H(X+tN)}$$

is concave in t.

$$\frac{d^2}{dt^2}e^{\frac{2}{n}H(X+tN)} \leq 0.$$

- Concavity of entropy power generalized to Renyi entropies G.T. and Savaré (2014).

- The central limit theorem for stable laws studies convergence of the law of (X_i i.i.d.)

$$T_n = \frac{X_1 + X_2 + \cdots + X_n}{n^{1/\lambda}}, \qquad n \geq 1.$$

- If the random variable X_i lies in the domain of attraction of the Lévy symmetric stable variable Z_λ, the law of T_n converges weakly to the law of Z_λ.
- A Lévy symmetric stable law L_λ defined in Fourier by

$$\widehat{L}_\lambda(\xi) = e^{-|\xi|^\lambda}.$$

- While the Gaussian density is related to the linear diffusion equation, Lévy distributions are related to linear fractional diffusion equations.

- In the classical central limit theorem the monotonicity of Shannon's entropy of S_n,

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n^{1/2}}, \qquad n \geq 1.$$

is a consequence of the monotonicity of Fisher information of S_n Madiman, Barron (2007).

- Main idea is to introduce the definition of score (used in theoretical statistics). Given an observation X, with law $f(x)$, the linear score $\rho(X)$ is given by

$$\rho(X) = \frac{f'(X)}{f(X)}$$

- The linear score has zero mean, and its variance is just the Fisher information.

- Given X and Y with differentiable density functions f (respectively g), the score function of the pair relative to X is represented by

$$\tilde{\rho}(X) = \frac{f'(X)}{f(X)} - \frac{g'(X)}{g(X)}.$$

- In this case, the relative to X Fisher information between X and Y is just the variance of $\tilde{\rho}(X)$.
- A centered Gaussian random variable Z_σ of variance σ is uniquely defined by the score function

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

- The relative (to X) score function of X and Z_σ

$$\tilde{\rho}(X) = \frac{f'(X)}{f(X)} + \frac{X}{\sigma}.$$

- The (relative to the Gaussian) Fisher information

$$\tilde{I}(X) = \tilde{I}(f) = \int_{\{f>0\}} \left(\frac{f'(x)}{f(x)} + \frac{x}{\sigma} \right)^2 f(x)\, dx.$$

- $\tilde{I}(X) \geq 0$, while $\tilde{I}(X) = 0$ if (and only if) X is a centered Gaussian variable of variance σ
- The concept of linear score can be naturally extended 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 $0 < \alpha < 1$, the linear fractional score

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

- The (relative to the Gaussian) Fisher information

$$\tilde{I}(X) = \tilde{I}(f) = \int_{\{f>0\}} \left(\frac{f'(x)}{f(x)} + \frac{x}{\sigma} \right)^2 f(x)\, dx.$$

- $\tilde{I}(X) \geq 0$, while $\tilde{I}(X) = 0$ if (and only if) X is a centered Gaussian variable of variance σ
- The concept of linear score can be naturally extended 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 $0 < \alpha < 1$, the linear fractional score

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

- The interest in fractional calculus after the reading of [Caffarelli, Vazquez (2011) Arch. Ration. Mech. Anal.], who studied a nonlinear porous medium flow with fractional potential pressure.
- To fix notations, for $0 < \alpha < 1$, we let R_α be the one-dimensional *normalized* Riesz potential operator

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

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

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

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

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

- In Fourier variables

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

- Differently from the classical case, the fractional score of X is linear in X if and only if X is a Lévy distribution of order $\alpha + 1$.

- For a given positive constant C, the identity

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

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

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

- Passing to Fourier transform, this identity yields

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

Consequently

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

- Arranging constants, we show that, if Z_λ is a Lévy distribution of density L_λ $(1 < \lambda < 2)$

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

- The relative (to X) fractional score function of X and Z_λ assumes the simple expression

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

- The (relative to the Lévy) fractional Fisher information (in short λ-Fisher relative information) is then defined

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

- 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). We will define by $\mathcal{P}_\lambda$ the set of probability density functions such that $I_\lambda(f) < +\infty$
- The concept of fractional score can be generalized. For $\upsilon > 0$

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

This leads to the relative fractional Fisher information $I_{\lambda,\upsilon}(X)$

- The following Lemma will be useful

Lemma

Let X_1 and X_2 be independent random variables with smooth densities, and let $\rho^{(1)}$ (respectively $\rho^{(2)}$) denote their fractional scores. Then, for each constant λ, with $1 < \lambda < 2$, and each positive constant δ, with $0 < \delta < 1$, the relative fractional score function of the sum $X_1 + X_2$ can be expressed as

$$\tilde{\rho}_\lambda(x) = E\left[\delta\, \tilde{\rho}^{(1)}_{\lambda,\delta}(X_1) + (1-\delta)\, \tilde{\rho}^{(2)}_{\lambda,1-\delta}(X_2) \middle| X_1 + X_2 = x\right].$$

- This Lemma has several interesting consequences.

- Since the norm of the relative fractional score is not less than that of its projection (i.e. by the Cauchy–Schwarz inequality)

$$I_\lambda(X_1+X_2) = E\left[\tilde{\rho}_\lambda^2(X_1+X_2)\right] \leq \delta^2 I_{\lambda,\delta}(X_1) + (1-\delta)^2 I_{\lambda,1-\delta}(X_2).$$

- For X such that one of the two sides is bounded, and positive constant υ, the following identity holds

$$I_{\lambda,\upsilon}(\upsilon^{1/\lambda}X) = \upsilon^{-2(1-1/\lambda)} I_\lambda(X).$$

- This relation implies the following

Theorem

Let X_j, $j = 1, 2$ be independent random variables such that their relative fractional Fisher information functions $I_\lambda(X_j)$, $j = 1, 2$ are bounded for some λ, with $1 < \lambda < 2$. Then, for each constant δ with $0 < \delta < 1$, $I_\lambda(\delta^{1/\lambda} X_1 + (1-\delta)^{1/\lambda} X_2)$ is bounded, and

$$I_\lambda(\delta^{1/\lambda} X_1 + (1-\delta)^{1/\lambda} X_2) \le \delta^{2/\lambda} I_\lambda(X_1) + (1-\delta)^{2/\lambda} I_\lambda(X_2).$$

Moreover, there is equality if and only if, up to translation, both X_j, $j = 1, 2$ are Lévy variables of exponent λ.

- The result is the analogous of the Blachman–Stam inequality for the standard relative Fisher information.

- The next ingredient in the proof of monotonicity deals with the so-called variance drop inequality Hoeffding (1948).
- Let $[n]$ denote the index set $\{1, 2, \ldots, n\}$, and, for any $\mathbf{s} \subset [n]$, let $X_{\mathbf{s}}$ stand for the collection of random variables $(X_i : i \in \mathbf{s})$, with the indices taken in their natural increasing order. Then

Theorem

Let the function $\Phi : \mathbb{R}^m \to \mathbb{R}$, with $1 \le m \in \mathbb{N}$, be symmetric in its arguments, and suppose that $E\left[\Phi(X_1, X_2, \ldots, X_m)\right] = 0$. Define

$$U(X_1, X_2, \ldots, X_n) = \frac{m!(n-m)!}{n!} \sum_{\{\mathbf{s} \subset [n] : |\mathbf{s}| = m\}} \Phi\left(X_{\mathbf{s}}\right).$$

Then

$$E\left[U^2\right] \le \frac{m}{n} E\left[\Phi^2\right].$$

- This quantifies the reduction.

- We apply the variance drop inequality of Hoeffding to the relative score $\tilde{\rho}(T_n)$.
- The following theorem holds true

Theorem

Let T_n denote the sum

$$T_n = \frac{X_1 + X_2 + \cdots + X_n}{n^{1/\lambda}},$$

where the random variables X_j are independent copies of a centered random variable X with bounded relative λ-Fisher information, $1 < \lambda < 2$. Then, for each $n > 1$, the relative λ-Fisher information of T_n is decreasing in n, and the following bound holds

$$I_\lambda\left(T_n\right) \leq \left(\frac{n-1}{n}\right)^{(2-\lambda)/\lambda} I_\lambda\left(T_{n-1}\right).$$

- At difference with the classical entropic central limit theorem, this quantifies the decay.

$$I_\lambda(T_n) \le \left(\frac{1}{n}\right)^{(2-\lambda)/\lambda} I_\lambda(X).$$

- There is convergence in relative λ-Fisher information sense at rate $1/n^{(2-\lambda)/\lambda}$.
- A strong difference between the classical central limit theorem and the central limit theorem for stable laws. In the classical central limit theorem , a very large domain of attraction with a very low convergence in relative Fisher(only monotonicity is guaranteed).
- In this case the domain of attraction is very restricted (only distribution which has the same tails at infinity of the Lévy stable law), but the attraction in terms of the relative fractional Fisher information is very strong.

- The leading example of a function which belongs to the domain of attraction of the λ-stable law is the so-called *Linnik distribution*

$$\widehat{p}_\lambda(\xi) = \frac{1}{1+|\xi|^\lambda}.$$

- For all $0 < \lambda \le 2$, this function is the characteristic function of a symmetric probability distribution. In addition, when $\lambda > 1$, $\widehat{p}_\lambda \in L^1(\mathbb{R})$, which, by applying the inversion formula, shows that p_λ is a probability density function.
- Linnik distribution belongs to the domain of attraction of the fractional Fisher information. How large is this domain (compared to the domain of attraction of the λ-stable law)?
- As in the classical case convergence in relative fractional Fisher information implies convergence in $L^1(\mathbb{R})$?

- Let us consider the Fokker–Planck equation with fractional diffusion

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

where $1 < \lambda < 2$,

- The initial datum $\varphi(x)$ belongs to the domain of normal attraction of the Lévy stable law ω of parameter λ, defined by

$$\widehat{\omega}(\xi) = \epsilon^{-|\xi|^\lambda}.$$

- $\omega(x)$ results to be a stationary solution of the Fokker–Planck equation.

- Given a random variable X of density $h(x)$, and a constant $a > 0$, let us denote by $h_a(x)$ the probability density of aX.
- Let Y a random variable with density φ, and let Z_λ be a Lévy variable independent of Y, such that $1 < \lambda < 2$, of density $\omega(x)$. For a given $t > 0$ we define

$$X_t = \alpha(t)Y + \beta(t)Z_\lambda,$$

where

$$\alpha(t) = e^{-t/\lambda}, \qquad \beta(t) = (1 - e^{-t})^{1/\lambda}.$$

- It holds

$$\alpha^\lambda(t) + \beta^\lambda(t) = 1.$$

- The random variable X_t, $t > 0$, has a density given by the convolution product

$$f(x,t) = \varphi_{\alpha(t)} * \omega_{\beta(t)}(x),$$

- Immediate to show that $f(x,t)$ solves the fractional Fokker-Planck equation with initial value $f(x, t = 0) = \varphi(x)$.
- Similarly to the classical Fokker–Planck equation, where the solution interpolates continuously between the initial datum and the Gaussian density, here the solution to the Fokker–Planck equation with fractional diffusion interpolates continuously between the initial datum φ and the Lévy density L of order λ.

- Passing to Fourier transform, we obtain that $\widehat{f}(\xi, t)$ solves the equation

$$\frac{\partial \widehat{f}}{\partial t} = -|\xi|^{\lambda}\widehat{f}(\xi, t) - \frac{\xi}{\lambda}\frac{\partial \widehat{f}(\xi, t)}{\partial \xi}.$$

- Integrating this equation along characteristics gives

$$\widehat{f}(\xi, t) = \widehat{\varphi}\left(\xi e^{-t/\lambda}\right) e^{-|\xi|^{\lambda}(1-e^{-t})}.$$

- The Lévy density ω is invariant under scaled convolutions

$$\omega(x) = \omega_{\alpha(t)} * \omega_{\beta(t)}(x).$$

- Let us consider the relative (to the Lévy density) entropy of X_t

$$H(X_t|\, Z_\lambda) = H(f(t)|\,\omega) = \int_{\mathbb{R}} f(x,t) \log \frac{f(x,t)}{\omega(x)}\, dx.$$

- Then

Theorem

Let the initial density φ be such that $H(\varphi|\,\omega)$ is finite. Then, if $f(x,t)$ is the solution to the fractional Fokker–Planck equation, the relative entropy $H(f(t)|\,\omega)$ is monotonically decreasing in time. In addition, if the density φ belongs to the domain of normal attraction of Z_λ, as time goes to infinity

$$\lim_{t\to\infty} H(f(t)|\,\omega) = 0.$$

- Assume that the density φ belongs to the domain of normal attraction of Z_λ, with bounded relative fractional Fisher information $I_\lambda(\varphi)$
- We write the fractional Fokker–Planck equation in the form

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial x}\left[f \left(\frac{D_{\lambda-1} f}{f} - \frac{D_{\lambda-1}\,\omega}{\omega} \right) \right].$$

- It holds $H(f(t)|\,\omega)$ is non increasing

$$\frac{d}{dt}H(f(t)|\,\omega) = \frac{d}{dt}\int_{\mathbb{R}} f(x,t)\log\frac{f(x,t)}{\omega(x)}\,dx =$$

$$-\int_{\mathbb{R}} f\left(\frac{f'}{f} - \frac{\omega'}{\omega}\right)\left(\frac{D_{\lambda-1}f}{f} - \frac{D_{\lambda-1}\omega}{\omega}\right)dx = -\bar{I}_\lambda(f(t)) \le 0.$$

- By Cauchy-Schwarz inequality, for any given density f in the domain of attraction of the fractional Fisher information

$$\bar{I}_\lambda(f) \le I(f)^{1/2} I_\lambda(f)^{1/2}.$$

- Since

$$I_\lambda(\delta^{1/\lambda}X_1 + (1-\delta)^{1/\lambda}X_2) \le \delta^{2/\lambda} I_\lambda(X_1) + (1-\delta)^{2/\lambda} I_\lambda(X_2),$$

$$I_\lambda(f(t)) = I_\lambda(X_t) \le \alpha(t)^2 I_\lambda(Y) = \alpha(t)^2 I_\lambda(\varphi)$$

with $\alpha(t) = e^{-t/\lambda}$.

- Consider that

$$\max\{\alpha(t)^{\lambda}, \beta(t)^{\lambda}\} \geq \frac{1}{2}.$$

- Then

$$I(X_t) = I(\alpha(t)Y + \beta(t)Z) \leq \min\{I(\alpha(t)Y),\ I(\beta(t)Z_{\lambda}) = \min\{\alpha(t)^{-2}I(Z_{\lambda}),\ \beta(t)^{-2}I(Z_{\lambda})\} \leq 2^{2/\lambda}\min\{I(Y),\ I(Z_{\lambda})\}.$$

- This implies

$$\bar{I}_{\lambda}(f) \leq e^{-t/\lambda}\, 2^{1/\lambda}\, \min\{I(\varphi),\ I(\omega)\}^{1/2}\, I_{\lambda}(\varphi)^{1/2}.$$

- integrating from zero to infinity, and recalling that the relative entropy converges to zero, we obtain

Theorem

Let X be a random variable with density φ in the domain of normal attraction of the Lévy symmetric random variable Z_λ, $1 < \lambda < 2$. If in addition X has bounded Fisher information, and lies in the domain of attraction of the fractional Fisher information, the Shannon relative entropy $H(X|Z_\lambda)$ is bounded, and the following inequality holds

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

- We proved the analogous of the logarithmic Sobolev inequality, which is obtained when $\lambda = 2$ (Gaussian case).
- In this case, the fractional Fisher information coincides with the classical Fisher information.
- As for the classical logarithmic Sobolev inequality, the inequality is saturated when the laws of X and Z_λ coincide.
- Let us take $\lambda = 2$. The steady state of the Fokker–Planck equation is the Gaussian density and

$$\frac{d}{dt} H(f(t)|\omega) = -I_2(f(t)) = -I(f(t)|\omega_2).$$

- Let us consider the normalized sum

$$T_n = \frac{1}{n^{1/\lambda}} \sum_{j=1}^{n} X_j.$$

- If the density f of X_i has bounded Fisher information, and belongs to the domain of attraction of the relative fractional Fisher information, so that $I_\lambda(f) < +\infty$,

$$H(T_n | Z_\lambda) \le \lambda\, 2^{1/\lambda} \min\{I(T_n), I(Z_\lambda)\}^{1/2} I_\lambda(T_n)^{1/2},$$

- and

$$I_\lambda(T_n)^{1/2} \le \left(\frac{1}{n}\right)^{(2-\lambda)/(2\lambda)} I_\lambda(X)^{1/2}.$$

- Convergence in relative entropy at the rate $n^{-(2-\lambda)/(2\lambda)}$ follows $I(T_n)$ is uniformly bounded.
- We have

Theorem

Let f belong to the domain of normal attraction of the Lévy symmetric random variable Z_λ, $1 < \lambda < 2$ and assume that there exists $M > 0$ such that

$$\int_{\mathbb{R}} |\widehat{f}(\xi)|^M (1 + |\xi|^2)^k \, d\xi = C_M < +\infty.$$

Then, for $n \geq M/2$, $f_n \in H^k(\mathbb{R})$. In addition, this condition holds with $M = 2$ if $f \in H^k(\mathbb{R})$, with $M > (2k+1)/\varepsilon$ if $|\widehat{f}(\xi)||\xi|^\varepsilon$ is bounded for $|\xi| \geq 1$, where $\varepsilon > 0$ is arbitrary, and with $M > 2k+1$ if $I(f)$ is bounded.

- This allows to conclude that, provided $I(f) < +\infty$, for all $n \geq 1$, $I(T_n) \leq C$.
- We have

Theorem

Let the random variable X belong to the domain of normal attraction of the random variable Z_λ with Lévy symmetric stable density ω. If in addition the density f of X has bounded Fisher information, and belongs to the domain of attraction of the relative fractional Fisher information, so that $I_\lambda(f) < +\infty$, the sequence of density functions f_n of the normalized sums T_n, converges to zero in relative entropy and

$$H(T_n | Z_\lambda) \leq C_\lambda(X) \left(\frac{1}{n}\right)^{(2-\lambda)/(2\lambda)} I_\lambda(X)^{1/2}.$$

- Thanks to Csiszar–Kullback inequality, convergence in relative entropy implies convergence in $L^1(\mathbb{R})$ at the sub-optimal rate $n^{-(2-\lambda)/(4\lambda)}$.
- Using the convergence in $L^1(\mathbb{R})$ of f_n to ω we obtain

Corollary

Let f satisfy the conditions of the previous Theorem. Then f_n converges to ω in $H^k(\mathbb{R})$ for all $k \geq 0$. Moreover, there is convergence of f_n to ω in the homogeneous Sobolev space $\dot{H}^k(\mathbb{R})$ at the rate $[n^{-(2-\lambda)/(4\lambda)}]^{2/(2k+3)}$.

Conclusions

- We introduced the definition of relative fractional Fisher information. This nonlocal functional is based on a suitable modification of the linear score function used in theoretical statistics.
- 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 score is linear (i.e. $f'(X)/f(X) = CX$), Lévy symmetric stable laws are identified as the unique random variables for which the new defined fractional score is linear.
- We showed that the fractional Fisher information can be fruitfully used to bound the relative (to the Lévy stable law) Shannon entropy, through an inequality similar to the classical logarithmic Sobolev inequality.
- Analogously to the central limit theorem, where monotonicity of entropy along the sequence provides an explicit rate of convergence to the Gaussian law for some smooth densities, in the case of the central limit theorem for stable laws convergence in $L^1(\mathbb{R})$ at explicit rate is proven, and, for smooth densities, convergence in various Sobolev spaces (still with rate).

References

- G.T. , *A strengthened entropy power inequality for log-concave densities*, IEEE Transactions on Information Theory **61** (12) 6550–6559 (2015).
- G.T., *The fractional Fisher information and the central limit theorem for stable laws*, Ricerche Mat., **65** (1) 71–91 (2016)
- G.T., *Entropy inequalities for stable densities and strengthened central limit theorems*, J. Stat. Phys., **165** 371–389 (2016)
- G.T., *Score functions, generalized relative Fisher information and applications*, Ricerche Mat. (in press) (2017).

- Part of a project on inequalities, developed in last years with Jean Dolbeault
  [J. Dolbeault, and G. T. (2013) Ann. I.H. Poincar AN,]
  [J. Dolbeault, and G. T. (2015) J. Phys. A: Math. Theor.]
  [J. Dolbeault, and G. T. (2015) Int. Math. Res. Notices]
  [J. Dolbeault, and G. T. (2016) Nonlinear Anal. Series A]
  about inequalities and nonlinear diffusion equations
- Same strategy here for inequalities in information theory and the heat equation [G.T. (2014) Milan J. Math.]

- A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter, On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations, *Comm. Partial Differential Equations*, **26** (2001) 43–100.
- T. Aubin, Problémes isopérimétriques et espaces de Sobolev. *J. Differential Geometry* **11**, (1976) 573–598.
- K.I. Babenko, An inequality in the theory of Fourier analysis. *Izv. Akad. Nauk SSSR, Ser. Mat.* **25** (1961), 531–542; English transl. *Amer. Math. Soc. Transl.* (2) **44** (1965), 115–128.
- D. Bakry and M. Emery, *Diffusions hypercontractives, Sém. Proba. XIX*, Lect. Notes in Math. n. 1123, Springer Verlag, Berlin, pp. 177–206, 1985.
- K. Ball, F. Barthe, and A. Naor, Entropy jumps in the presence of a spectral gap. *Duke Math. J.* **119** (2003) 41–63.

- K. Ball, and V.H. Nguyen, Entropy jumps for isotropic log-concave random vectors and spectral gap. *Studia Math.*, **213** (1)(2012) 81–96.
- G. I. Barenblatt, On some unsteady motions of a liquid and gas in a porous medium. *Akad. Nauk SSSR. Prikl. Mat. Meh.*, **16** (1952) 67–78.
- G.I. Barenblatt, *Scaling. Self-Similarity and Intermediate Asymptotics*. Cambridge Univ. Press, Cambridge 1996.
- A.R. Barron, Entropy and the central limit theorem. *Ann. Probab.*, **14** (1986) 336–342.
- F. Barthe, Optimal Young's inequality and its converse: a simple proof. *Geom. Funct. Anal.* **8** (1998) 234–242.
- F. Barthe and D. Cordero-Erausquin, Inverse Brascamp-Lieb inequalities along the heat equation. In: *Geometric Aspects of*

Functional Analysis. Lecture Notes in Math., vol. 1850, 65–71. Springer, Berlin 2004.

F. Barthe and N. Huet, On Gaussian Brunn-Minkowski inequalities. *Stud. Math.* **191** (2009) 283–304.

J-P. Bartier, A. Blanchet, J. Dolbeault and M. Escobedo, Improved intermediate asymptotics for the heat equation. *Appl. Math. Letters*, **24** (2011) 76–81.

W. Bechner, Inequalities in Fourier Analysis on $\mathbb{R}^d$. *Proc. Nat. Acad. Sci. USA* **72** (1975) 638–641.

J. Bennett, Heat-flow monotonicity related to some inequalities in euclidean analysis. *Contemporary Mathematics* **505** (2010) 85–96.

J. Bennett and N. Bez,Closure Properties of Solutions to Heat Inequalities. *J. Geom. Anal.* **19** (2009) 584–600.

J. Bennett, N. Bez and A. Carbery, Heat-flow monotonicity related to the Hausdorff–Young inequality, *Bull. London Math. Soc.* **41** (6) (2009) 971–979.

J. Bennett, A. Carbery, M. Christ and T. Tao, The Brascamp–Lieb Inequalities: Finiteness, Structure and Extremals. *Geometric And Functional Analysis* **17** (2007) 1343–1415.

P.L. Bhatnagar, E.P. Gross and M. Krook, A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems. *Phys. Rev.* **94** (1954) 511–525.

N.M. Blachman, The convolution inequality for entropy powers. *IEEE Trans. Inform. Theory* **2** (1965) 267–271.

A. Blanchet, M. Bonforte, J. Dolbeault, G. Grillo and J. L Vázquez, Asymptotics of the Fast Diffusion Equation via

Entropy Estimates. *Arch. Ration. Mech. Anal.*, **191** (2009) 347–385.

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.

S.G. Bobkov, and M. Madiman, Dimensional behaviour of entropy and information. *C. R. Acad. Sci. Paris Sér. I Math.*, **349** (2011) 201–204.

S.G. Bobkov, and M. Madiman, Reverse Brunn–Minkowski and reverse entropy power inequalities for convex measures, *J. Funct. Anal.*

S.G. Bobkov, and M. Madiman, On the problem of reversibility of the entropy power inequality, in P. Eichelsbacher et al., editors, *Limit Theorems in Probability, Statistics, and Number Theory (in honor of Friedrich Götze)*, vol. 42 of Springer

Proceedings in Mathematics and Statistics. Springer-Verlag, Berlin 2013.

- L. Boltzmann, Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Sitzungsberichte der Akademie der Wissenschaften*, **66** 275–370, in *Lectures on Gas Theory*. Berkeley: University of California Press (1964). Translated by S.G. Brush. Reprint of the 1896–1898 Edition. Reprinted by Dover Publ.
- C. Borell, The Ehrhard inequality. *C.R. Math. Acad. Sci. Paris* **337** (2003) 663–666.
- H.J. Brascamp and E.H. Lieb, Best constants in Young's inequality, its converse and its generalization to more than three functions. *Adx. Math.* **20** (1976) 151–173.

- E.A. Carlen, E.H. Lieb, and M. Loss, A sharp analog of Young's inequality on S^N and related entropy inequalities. *J. Geom. Anal.* **14** (2004) 487–520.
- E.A. Carlen, and A. Soffer, Propogation of localization, optimal entropy production, and convergence rates for the central limit theorem, Preprint, arXiv:1106.2256, (2011).
- J.A. Carrillo, G. Toscani, Exponential convergence toward equilibrium for homogeneous Fokker-Planck-type equations. *Math. Methods Appl. Sci.*, **21** (1998) 1269–1286.
- J.A. Carrillo, G. Toscani,
  Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity. *Indiana Univ. Math. J.*, **49** (2000) 113–141.
- J.A. Carrillo, G. Toscani, Contractive probability metrics and asymptotic behavior of dissipative kinetic equations. *Riv. Mat. Univ. Parma* (7) **6**(2007) 75–198.

- K.M. Case, and P.F. Zweifel *Linear Transport Theory*. Addison-Wesley, Reading 1967.
- S. Chandrasekhar, Stochastic problems in physics and astronomy. *Rev. Modern Phys.* **15**, (1943) 1–110.
- S. Chandrasekhar *Radiative Transfer*. Reprinted by Clarendon Press, Oxford 1960, Dover Publications, New York 1950.
- F. Cheng, and Y. Geng, Higher order derivatives in Costa's entropy power inequality. *IEEE Trans. Inform. Theory* **61** (2015) 5892–5905.
- D. Cordero-Erausquin, B. Nazaret, and C. Villani, A mass–transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities. *Advances in Mathematics*, **182** (2004) 307–332.
- M.H.M. Costa, *A new entropy power inequality*. IEEE Trans. Inform. Theory, **IT-31**, (6) (1985) 751–760.

- J. Costa, A. Hero and C. Vignat, On solutions to multivariate maximum alpha-entropy problems. *Lecture Notes in Computer Science*, **2683**, no. EMMCVPR 2003, Lisbon, 7-9 July 2003,
- T.M. Cover and J.A. Thomas, *Elements of Information Theory*. 2nd ed. John Wiley & Sons, New York, 2006.
- T.M. Cover, and Z. Zhang, On the maximum entropy of the sum of tweo dependent random variables. *IEEE Trans. Inform. Theory*, **40**, (1994) 1244–1246.
- D.R. Cox, and D.V. Hinkley, *Theoretical Statistics*, Chapman & Hall, London 1974
- M. Del Pino, and J. Dolbeault, Best constants for Gagliardo-Nirenberg inequalities and application to nonlinear diffusions. *J. Math. Pures Appl.*, **81** (2002) 847–875.

- A. Dembo, A simple proof of the concavity of the entropy power with respect to the variance of additive normal noise. *IEEE Trans. Inform. Theory* **35**, (1989) 887–888.
- A. Dembo, T.M. Cover, and J.A. Thomas, Information theoretic inequalities. IEEE Trans. Inform. Theory **37** (6) (1991) 1501–1518.
- S. Dharmadhikari, and K. Joag-Dev, *Unimodality, Convexity, and Applications*, Academic Press, Boston (1988).
-
-
- J. Dolbeault, and G. Toscani, Nonlinear diffusions: extremal properties of Barenblatt profiles, best matching and delays, *Nonlinear Anal. Series A* **138** (2016) 31–43.
-

E. Gabetta, P.A. Markowich, and A. Unterreiter, A note on the entropy production of the radiative transfer equation. *Appl. Math. Lett.*, **12** (1999) 111–116.

E. Gabetta, G. Toscani, and B. Wennberg, Metrics for probability distributions and the trend to equilibrium for solutions of the Boltzmann equation. *J. Statist. Phys.* **81** (1995) 901–934.

R.J. Gardner, The Brunn-Minkowski inequality. *Bull. Amer. Math. Soc.* **39** (2002) 355–405.

L. Gross, Logarithmic Sobolev inequalities, *Amer. J. Math.* **97** (1975) 1061–1083.

T. Goudon, S. Junca, and G. Toscani, Fourier-based distances and Berry–Esseen like inequalities for smooth densities. *Monatsh. Math.* **135** (2002) 115–136.

D. Guo, S. Shamai, and S. Verdú, Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inform. Theory*, **51**, (4) (2005) 1261–1282.

D. Guo, S. Shamai, and S. Verdú, A Simple Proof of the Entropy-Power Inequality. *IEEE Trans. Inform. Theory*, **52**, (5) (2006) 2165–2166.

O. Johnson, and A.R. Barron, Fisher information inequalities and the central limit theorem. *Probab. Theory Related Fields*, **129** (3), (2004) 391–409.

S. Kullback, *Information Theory and Statistics*. John Wiley & Sons, New York 1959.

R. G. Laha and V. K. Rohatgi, *Probability theory*. John Wiley & Sons, New York, 1979.

L. Leindler, On a certain converse of Hölder's inequality. II. *Acta. Sci. Math. Szeged* **33** (1972), 217–223.

E.H. Lieb, Proof of an entropy conjecture of Wehrl. *Commun. Math. Phys.*, **62** (1978) 35–41.

E.H. Lieb, Gaussian kernels have only gaussian maximizers. *Invent. Math.*, **102** (1990) 179–208.

Yu.V. Linnik, An information-theoretic proof of the central limit theorem with the Lindeberg condition. *Theory Probab. Appl.* **4** (1959) 288-299.

P.L. Lions, G. Toscani, A strenghened central limit theorem for smooth densities. *J. Funct. Anal.* **129** (1995) 148–167.

E. Lutwak, D. Yang, and G. Zhang, Cramér-Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information. *IEEE Trans. Inform. Theory*, **51** (2005) 473–478.

E. Lutwak, D. Yang and G. Zhang, Moment-entropy inequalities for a random vector. *IEEE Trans. Inform. Theory*, **53** (2007) 1603–1607.

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

A.W. Marshall, and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. Academic Press, Orlando 1979.

H.P. McKean Jr., Speed of approach to equilibrium for Kac caricature of a Maxwellian gas. *Arch. Rat. Mech. Anal.* **21** (1966) 343–367.

J. Nash, Continuity of solutions of parabolic and elliptic equations, *Amer. J. Math.* **80**, (1958) 931–954.

F. Otto, The geometry of dissipative evolution equations: the porous medium equation. *Commun. Part. Diff. Eq.*, **26** (2001) 101–174.

A. Prékopa, Logarithmic concave measures with application to stochastic programming. *Acta. Sci. Math. Szeged* **32** (1971) 301–315.

A. Prékopa, On logarithmic measures and functions. *Acta. Sci. Math. Szeged* **34** (1973) 335–343.

A. Rényi, On measures of entropy and information. *Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1*, 547–561. University of California Press, Berkeley 1961.

O. Rioul, A simple proof of the entropy-power inequality via properties of mutual information, in *Proceedings of the 2007 IEEE International Symposium on Information Theory (ISIT 2007)*, Nice, France, June 24-29-th, 2007.

- G. Savaré, and G. Toscani, The concavity of Renyi entropy power, *IEEE Trans. Inform. Theory*, **60** (5) (2014) 2687–2693.
- C.E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27** Jul.(1948), 379–423; Oct. (1948), 623–656.
- A.J. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Contr.* **2** (1959) 101–112.
- G. Talenti, Best constant in Sobolev inequality. *Ann. Mat. Pura Appl.*, **110** (1976) 353–372.
- G. Toscani, Sur l'iégalité logarithmique de Sobolev. *C. R. Acad. Sci. Paris Sér. I Math.*, **324**, Série I (1997) 689–694.
- G. Toscani, Entropy production and the rate of convergence to equilibrium for the Fokker-Planck equation, *Quarterly of Appl. Math.*, Vol. LVII (1999) 521–541.

G. Toscani, An information-theoretic proof of Nash's inequality. *Rend. Lincei Mat. Appl.*, **24** (2013) 83–93.

G. Toscani, Lyapunov functionals for the heat equation and sharp inequalities. *Atti Acc. Peloritana Pericolanti, Classe Sc. Fis. Mat. e Nat.*, **91** (2013) 1–10.

G. Toscani, Heat equation and convolution inequalities, *Milan J. Math.* **82** (2) (2014) 183–212.

G. Toscani, A strengthened entropy power inequality for log-concave densities, *IEEE Trans. Inform. Theory*, **61** (12) (2015) 6550–6559.

G. Toscani, Rényi entropies and nonlinear diffusion equations, *Acta. Appl. Math.*, **132** (2014) 595–604.

G. Toscani, A concavity property for the reciprocal of Fisher information and its consequences on Costa's EPI, *Physica A* **432** (2015) 35–42.

C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.*, **52** (1988) 479–487.

C. Tsallis, *Introduction to Nonextensive Statistical Mechanics - Approaching a Complex World*. Springer Verlag, New York 2009.

J.L. Vazquez, *The Porous Medium Equation: Mathematical Theory*, Oxford University Press, Oxford 2007.

C. Villani, A Short Proof of the Concavity of Entropy Power. *IEEE Trans. Inform. Theory*, **46** (2000) 1695–1696.

R. Zamir and M. Feder, A generalization of the entropy power inequality with applications. *IEEE Trans. Inf. Theory*, **39** (5) (1993) 1723–1728.

Ya.B. Zel'dovich, and A.S. Kompaneetz, Towards a theory of heat conduction with thermal conductivity depending on the temperature. *Collection of Papers Dedicated to 70th Birthday* *of Academician* A. F. Ioffe, *Izd. Akad. Nauk SSSR*, Moscow (1950)