Information-theoretical inequalities for log-concave and stable densities

Giuseppe Toscani

Department of Mathematics
University of Pavia, Italy

Nonlocal Nonlinear Partial Differential Equations and Applications
Anacapri, September 14-18, 2015
Outline

1. Information inequalities for entropies
   - Entropy power and Stam’s Fisher information inequalities
   - Further inequalities for log-concave densities
   - A strengthened entropy power inequality

2. The central limit theorem for stable laws
   - The fractional Fisher information
   - Monotonicity of the fractional Fisher information
Outline

1. Information inequalities for entropies
   - Entropy power and Stam’s Fisher information inequalities
   - Further inequalities for log-concave densities
   - A strengthened entropy power inequality

2. The central limit theorem for stable laws
   - The fractional Fisher information
   - Monotonicity of the fractional Fisher information


References


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

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

$$e^{2n H(X+Y)} \geq e^{2n H(X)} + e^{2n H(Y)}.$$
For a Gaussian random vector $N_\sigma$ with covariance $\sigma I$.

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

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:

\[ I(X) = I(f) = \int_{\{f > 0\}} \frac{\left| \nabla f(x) \right|^2}{f(x)} \, dx. \]
For a Gaussian random vector \( N_\sigma \) with covariance \( \sigma I \).

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

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:

\[
l(X) = l(f) = \int_{\{f>0\}} \frac{|\nabla f(x)|^2}{f(x)} \, dx.
\]
For a Gaussian random vector $N_\sigma$ with covariance $\sigma I$.

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

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

$$I(X) = I(f) = \int_{\{f>0\}} \frac{\left|\nabla f(x)\right|^2}{f(x)} \, dx.$$
The heat equation in the whole space $\mathbb{R}^n$

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

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

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

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 $\kappa = 1$

$$I(f) = \left. \frac{d}{dt} \right|_{t=0} H(u(t)); \quad J(f) = -\frac{1}{2} \left. \frac{d}{dt} \right|_{t=0} I(u(t)).$$
Fisher information satisfies the inequality \((a, b > 0)\)

\[
I(X + Y) \leq \frac{a^2}{(a + b)^2} I(X) + \frac{b^2}{(a + b)^2} I(Y)
\]

Optimizing over \(a\) and \(b\) one obtains Stam’s Fisher information inequality

\[
\frac{1}{I(X + Y)} \geq \frac{1}{I(X)} + \frac{1}{I(Y)}.
\]

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.
Fisher information satisfies the inequality \((a, b > 0)\)

\[
I(X + Y) \leq \frac{a^2}{(a + b)^2} I(X) + \frac{b^2}{(a + b)^2} I(Y)
\]

Optimizing over \(a\) and \(b\) one obtains Stam’s Fisher information inequality

\[
\frac{1}{I(X + Y)} \geq \frac{1}{I(X)} + \frac{1}{I(Y)}.
\]

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.
Fisher information satisfies the inequality \((a, b > 0)\)

\[
I(X + Y) \leq \frac{a^2}{(a + b)^2} I(X) + \frac{b^2}{(a + b)^2} I(Y)
\]

Optimizing over \(a\) and \(b\) one obtains Stam’s Fisher information inequality

\[
\frac{1}{I(X + Y)} \geq \frac{1}{I(X)} + \frac{1}{I(Y)}.
\]

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 $t \to 0$

$$I(X)e^{\frac{2}{n}H(X)} \geq 2\pi en.$$
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 $t \rightarrow 0$

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

Likewise Dembo(1989), (cf. Villani(2000)) if $N$ is a Gaussian random vector with covariance $I$, for $t > 0$

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

This implies

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

Letting $t \rightarrow 0$ gives the inequality

$$\frac{1}{I^2(X)} J(X) \geq \frac{1}{n}.$$
The isoperimetric inequality for entropies implies logarithmic Sobolev inequality.

Likewise Dembo(1989), (cf. Villani(2000)) if $N$ is a Gaussian random vector with covariance $I$, for $t > 0$

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

This implies

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

Letting $t \to 0$ gives the inequality

\[
\frac{1}{I^2(X)} J(X) \geq \frac{1}{n}.
\]
The isoperimetric inequality for entropies implies logarithmic Sobolev inequality.
Likewise Dembo(1989), (cf. Villani(2000)) if $N$ is a Gaussian random vector with covariance $I$, for $t > 0$

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

This implies

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

Letting $t \to 0$ gives the inequality

$$\frac{1}{I^2(X)} J(X) \geq \frac{1}{n}.$$
The isoperimetric inequality for entropies implies logarithmic Sobolev inequality.

Likewise Dembo(1989), (cf. Villani(2000)) if $N$ is a Gaussian random vector with covariance $I$, for $t > 0$

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

This implies

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

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. If $N$ is a Gaussian random vector with covariance $I$, the entropy power is concave in $t$.

$$d^2 dt^2 e_n^2 H(X+tN) \leq 0.$$ 

Concavity of entropy power generalized to Renyi entropies GT and Savaré (2014).
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

$$e_n^2 H(X + tN)$$

is concave in $t$.

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

Concavity of entropy power generalized to Renyi entropies GT
and Savaré (2014).
The functional $J(X)$ is given by

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

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

Let $f, g$ be log-concave densities. Our goal will be to prove the inequality

$$J(f \ast g) \leq \frac{a^4}{(a + b)^4} J(f) + \frac{b^4}{(a + b)^4} J(g) + \frac{2a^2 b^2}{(a + b)^4} H(f, g),$$

$$H(f, g) = \sum_{i,j=1}^{n} \int\{f>0\} \frac{\partial_i f \partial_j f}{f} \, dx \int\{g>0\} \frac{\partial_i g \partial_j g}{g} \, dx.$$
Further inequalities for log-concave densities

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

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

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

- Let $f, g$ be log-concave densities. Our goal will be to prove the inequality

$$J(f \ast g) \leq \frac{a^4}{(a + b)^4} J(f) + \frac{b^4}{(a + b)^4} J(g) + \frac{2a^2 b^2}{(a + b)^4} H(f, g),$$

$$H(f, g) = \sum_{i,j=1}^{n} \int \{f>0\} \frac{\partial_i f \partial_j f}{f} \, dx \int \{g>0\} \frac{\partial_i g \partial_j g}{g} \, dx.$$
Equality holds if and only $X$ and $Y$ are Gaussian random vectors with proportional covariance matrices.

Optimizing over $a, b$ one obtains

$$\frac{1}{\sqrt{J(X + Y)}} \geq \frac{1}{\sqrt{J(X)}} + \frac{1}{\sqrt{J(Y)}},$$

with equality if and only $X$ and $Y$ are Gaussian random vectors with proportional covariance matrices.

Indeed, $f N_{\sigma}$ is a Gaussian random vector with covariance $\sigma I$

$$J(M_{\sigma}) = \frac{n}{\sigma^2}.$$
Outlines

Information inequalities for entropies

The central limit theorem for stable laws

Further inequalities for log-concave densities

- Equality holds if and only if $X$ and $Y$ are Gaussian random vectors with proportional covariance matrices.
- Optimizing over $a$, $b$ one obtains

\[
\frac{1}{\sqrt{J(X+Y)}} \geq \frac{1}{\sqrt{J(X)}} + \frac{1}{\sqrt{J(Y)}},
\]

with equality if and only if $X$ and $Y$ are Gaussian random vectors with proportional covariance matrices.

- Indeed, $f \mathcal{N}_\sigma$ is a Gaussian random vector with covariance $\sigma I$

\[
J(M_\sigma) = \frac{n}{\sigma^2}.
\]
Equality holds if and only $X$ and $Y$ are Gaussian random vectors with proportional covariance matrices.

Optimizing over $a, b$ one obtains

$$\frac{1}{\sqrt{J(X + Y)}} \geq \frac{1}{\sqrt{J(X)}} + \frac{1}{\sqrt{J(Y)}},$$

with equality if and only $X$ and $Y$ are Gaussian random vectors with proportional covariance matrices.

Indeed, $f N_\sigma$ is a Gaussian random vector with covariance $\sigma I$

$$J(M_\sigma) = \frac{n}{\sigma^2}.$$
In particular, inequality for $J$ is stronger

\[
\frac{1}{\sqrt{J(X + Y)}} \geq \left( \frac{1}{\sqrt{J(X)}} + \frac{1}{\sqrt{J(Y)}} \right) R(X, Y),
\]

with $R(X, Y) \geq 1$.

In one dimension one has the explicit expression

\[
1 \leq R(X, Y) = \left( 1 - 2 \frac{\sqrt{J(X)} \sqrt{J(Y)} - I(X)I(Y)}{\left( \sqrt{J(X)} + \sqrt{J(Y)} \right)^2} \right)^{-1/2}.
\]

$R(X, Y) = 1$ if and only if $X$ and $Y$ are Gaussian random variables.
In particular, inequality for $J$ is stronger

\[
\frac{1}{\sqrt{J(X+Y)}} \geq \left( \frac{1}{\sqrt{J(X)}} + \frac{1}{\sqrt{J(Y)}} \right) R(X, Y),
\]

with $R(X, Y) \geq 1$.

In one dimension one has the explicit expression

\[
1 \leq R(X, Y) = \left( 1 - 2 \frac{\sqrt{J(X)}\sqrt{J(Y)} - I(X)I(Y)}{\left(\sqrt{J(X)} + \sqrt{J(Y)}\right)^2} \right)^{-1/2}.
\]

$R(X, Y) = 1$ if and only if $X$ and $Y$ are Gaussian random variables.
In particular, inequality for $J$ is stronger

$$\frac{1}{\sqrt{J(X + Y)}} \geq \left( \frac{1}{\sqrt{J(X)}} + \frac{1}{\sqrt{J(Y)}} \right) \mathcal{R}(X, Y),$$

with $\mathcal{R}(X, Y) \geq 1$.

In one dimension one has the explicit expression

$$1 \leq \mathcal{R}(X, Y) = \left( 1 - 2 \frac{\sqrt{J(X)} \sqrt{J(Y)} - I(X)I(Y)}{(\sqrt{J(X)} + \sqrt{J(Y)})^2} \right)^{-1/2}.$$

$\mathcal{R}(X, Y) = 1$ if and only if $X$ and $Y$ are Gaussian random variables.
A consequence of the inequality is linked to the functional

$$
\Lambda(t) = H(f(t) * g(t)) - \kappa H(f(t)) - (1 - \kappa) H(g(t))
$$

- Here $0 < \kappa < 1$, $f(x, t)$ (respectively $g(x, t)$) are the solutions to the heat equation with diffusion constant $\kappa$ (respectively $1 - \kappa$), and initial data $f$ and $g$ log-concave probability densities in $\mathbb{R}^n$.

- $\Lambda(t)$ is a convex function of time, and this implies, optimizing over $\kappa$, a strengthened entropy power inequality.
A consequence of the inequality is linked to the functional

\[ \Lambda(t) = H(f(t) \ast g(t)) - \kappa H(f(t)) - (1 - \kappa) H(g(t)) \]

Here \(0 < \kappa < 1\), \(f(x, t)\) (respectively \(g(x, t)\)) are the solutions to the heat equation with diffusion constant \(\kappa\) (respectively \(1 - \kappa\)), and initial data \(f\) and \(g\) log-concave probability densities in \(\mathbb{R}^n\).

\(\Lambda(t)\) is a convex function of time, and this implies, optimizing over \(\kappa\), a strengthened entropy power inequality.
A consequence of the inequality is linked to the functional

\[ \Lambda(t) = H(f(t) \ast g(t)) - \kappa H(f(t)) - (1 - \kappa) H(g(t)) \]

Here \( 0 < \kappa < 1 \), \( f(x, t) \) (respectively \( g(x, t) \)) are the solutions to the heat equation with diffusion constant \( \kappa \) (respectively \( 1 - \kappa \)), and initial data \( f \) and \( g \) log-concave probability densities in \( \mathbb{R}^n \).

\( \Lambda(t) \) is a convex function of time, and this implies, optimizing over \( \kappa \), a strengthened entropy power inequality.
Theorem

Let \( N(X) \) denote the entropy power of the random vector \( X \). Then, if the densities of the independent pair \( X \) and \( Y \) of random vectors are log-concave

\[
N(X + Y) \geq [N(X) + N(Y)] R(X, Y)
\]

Here the quantity \( R(X, Y) \geq 1 \) can be interpreted as a measure of the non-Gaussianity of the two random vectors \( X, Y \). Indeed, \( R(X, Y) = 1 \) if and only if both \( X \) and \( Y \) are Gaussian random vectors.

How to quantify \( R(X, Y) \) in terms of a distance from the manifold of Gaussian random vectors?
Theorem

Let $N(X)$ denote the entropy power of the random vector $X$. Then, if the densities of the independent pair $X$ and $Y$ of random vectors are log-concave

$$N(X + Y) \geq [N(X) + N(Y)] R(X, Y)$$

Here the quantity $R(X, Y) \geq 1$ can be interpreted as a measure of the non-Gaussianity of the two random vectors $X$, $Y$. Indeed, $R(X, Y) = 1$ if and only if both $X$ and $Y$ are Gaussian random vectors.

How to quantify $R(X, Y)$ in terms of a distance from the manifold of Gaussian random vectors?
Further inequalities for log-concave densities

- **Interesting connections of entropy power inequality with the central limit theorem**
- Consider the law of \((X_i \text{ i.i.d.})\)

\[
S_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}, \quad n \geq 1.
\]

- Then \(H(S_n)\) is non-decreasing with respect to \(n\) Artstein, Ball, Barthe, Naor (2002), Madiman, Barron (2007).
- How to quantify the entropy jump?

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

Further inequalities for log-concave densities

- Interesting connections of entropy power inequality with the central limit theorem
- Consider the law of \((X_i \text{ i.i.d.})\)

\[ S_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}, \quad n \geq 1. \]

- Then \(H(S_n)\) is non-decreasing with respect to \(n\) Artstein, Ball, Barthe, Naor (2002), Madiman, Barron (2007).
- How to quantify the entropy jump?

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

Further inequalities for log-concave densities

- Interesting connections of entropy power inequality with the central limit theorem
- Consider the law of \((X_i \text{ i.i.d.})\)

\[ S_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}, \quad n \geq 1. \]

- Then \(H(S_n)\) is non-decreasing with respect to \(n\) Artstein, Ball, Barthe, Naor (2002), Madiman, Barron (2007).

- How to quantify the entropy jump?

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

Interesting connections of entropy power inequality with the central limit theorem

Consider the law of \((X_i \text{ i.i.d.})\)

\[ S_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}, \quad n \geq 1. \]

Then \(H(S_n)\) is non-decreasing with respect to \(n\) Artstein, Ball, Barthe, Naor (2002), Madiman, Barron (2007).

How 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.
Outlines

Information inequalities for entropies

The central limit theorem for stable laws

Further inequalities for log-concave densities

• Interesting connections of entropy power inequality with the central limit theorem
• Consider the law of \((X_i \text{ i.i.d.})\)

\[
S_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}, \quad n \geq 1.
\]

• Then \(H(S_n)\) is non-decreasing with respect to \(n\) Artstein, Ball, Barthe, Naor (2002), Madiman, Barron (2007).
• How 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 central limit theorem for stable laws studies convergence of the law of $(X_i \text{ i.i.d.})$

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

- If the random variable $X_i$ lies in the domain of attraction of the Lévy symmetric stable variable $Z_{\lambda}$, the law of $T_n$ converges weakly to the law of $Z_{\lambda}$.
- A Lévy symmetric stable law $L_{\lambda}$ defined in Fourier by

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

- While the Gaussian density is related to the linear diffusion equation, Lévy distributions are related to linear fractional diffusion equations.
The central limit theorem for stable laws studies convergence of the law of \((X_i \text{ i.i.d.})\)

\[ T_n = \frac{X_1 + X_2 + \cdots + X_n}{n^{1/\lambda}}, \quad n \geq 1. \]

If the random variable \(X_i\) lies in the domain of attraction of the Lévy symmetric stable variable \(Z_\lambda\), the law of \(T_n\) converges weakly to the law of \(Z_\lambda\).

A Lévy symmetric stable law \(L_\lambda\) defined in Fourier by

\[ \hat{L}_\lambda(\xi) = e^{-|\xi|^\lambda}. \]

While the Gaussian density is related to the linear diffusion equation, Lévy distributions are related to linear fractional diffusion equations.
The central limit theorem for stable laws studies convergence of the law of \((X_i \text{ i.i.d.})\)

\[ T_n = \frac{X_1 + X_2 + \cdots + X_n}{n^{1/\lambda}}, \quad n \geq 1. \]

If the random variable \(X_i\) lies in the domain of attraction of the Lévy symmetric stable variable \(Z_\lambda\), the law of \(T_n\) converges weakly to the law of \(Z_\lambda\).

A Lévy symmetric stable law \(L_\lambda\) defined in Fourier by

\[ \hat{L}_\lambda(\xi) = e^{-|\xi|^\lambda}. \]

While the Gaussian density is related to the linear diffusion equation, Lévy distributions are related to linear fractional diffusion equations.
The central limit theorem for stable laws studies convergence of the law of \((X_i \text{ i.i.d.})\)

\[
T_n = \frac{X_1 + X_2 + \cdots + X_n}{n^{1/\lambda}}, \quad n \geq 1.
\]

If the random variable \(X_i\) lies in the domain of attraction of the Lévy symmetric stable variable \(Z_\lambda\), the law of \(T_n\) converges weakly to the law of \(Z_\lambda\).

A Lévy symmetric stable law \(L_\lambda\) defined in Fourier by

\[
\hat{L}_\lambda(\xi) = e^{-|\xi|^\lambda}.
\]

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 $S_n$, 
\[
S_n = \frac{X_1 + X_2 + \cdots + X_n}{n^{1/2}}, \quad n \geq 1.
\]
is a consequence of the monotonicity of Fisher information of $S_n$,
Artstein, Ball, Barthe, Naor (2002), simplified in 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.
In the classical **central limit theorem** the monotonicity of Shannon’s entropy of \( S_n \),

\[
S_n = \frac{X_1 + X_2 + \cdots + X_n}{n^{1/2}}, \quad n \geq 1.
\]

is a consequence of the monotonicity of Fisher information of \( S_n \) \[Artstein, Ball, Barthe, Naor (2002), simplified in Madiman, Barron (2007).\]

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.
The fractional Fisher information

- In the classical **central limit theorem** the monotonicity of Shannon’s entropy of $S_n$,

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

is a consequence of the monotonicity of Fisher information of $S_n$ Artstein, Ball, Barthe, Naor (2002), simplified in 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 $Z_\sigma$ of variance $\sigma$ is uniquely defined by the score function

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

The relative (to $X$) score function of $X$ and $Z_\sigma$

$$\tilde{\rho}(X) = \frac{f'(X)}{f(X)} + \frac{X}{\sigma}.$$
The fractional 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 $Z_\sigma$ of variance $\sigma$ is uniquely defined by the score function

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

- The relative (to $X$) score function of $X$ and $Z_\sigma$

$$\tilde{\rho}(X) = \frac{f'(X)}{f(X)} + \frac{X}{\sigma}.$$
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 $Z_\sigma$ of variance $\sigma$ is uniquely defined by the score function

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

The relative (to $X$) score function of $X$ and $Z_\sigma$

$$\tilde{\rho}(X) = \frac{f'(X)}{f(X)} + \frac{X}{\sigma}.$$
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 $Z_\sigma$ of variance $\sigma$ is uniquely defined by the score function

$$\rho(Z_\sigma) = -\frac{Z_\sigma}{\sigma}.$$

The relative (to $X$) score function of $X$ and $Z_\sigma$

$$\tilde{\rho}(X) = \frac{f'(X)}{f(X)} + \frac{X}{\sigma}.$$
The fractional 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 \( \sigma \).
- 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 \( \alpha \), with \( 0 < \alpha < 1 \), the linear fractional score

\[ \rho_{\alpha+1}(X) = \frac{D_\alpha f(X)}{f(X)}. \]
The (relative to the Gaussian) Fisher information

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

\[ \tilde{I}(X) \geq 0, \text{ while } \tilde{I}(X) = 0 \text{ if (and only if) } X \text{ is a centered Gaussian variable of variance } \sigma. \]

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 \( \alpha \), with \( 0 < \alpha < 1 \), the linear fractional score

\[ \rho_{\alpha+1}(X) = \frac{D_\alpha f(X)}{f(X)}. \]
The (relative to the Gaussian) Fisher information

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

\[ \tilde{I}(X) \geq 0, \text{ while } \tilde{I}(X) = 0 \text{ if (and only if) } X \text{ is a centered Gaussian variable of variance } \sigma \]

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 \( \alpha \), with \( 0 < \alpha < 1 \), the linear fractional score

\[ \rho_{\alpha+1}(X) = \frac{\mathcal{D}_\alpha f(X)}{f(X)}. \]
To fix notations, for $0 < \alpha < 1$, we let $R_\alpha$ 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

$$\hat{R}_\alpha(f)(\xi) = |\xi|^{\alpha} \hat{f}(\xi).$$

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

$$\frac{d^\alpha f(x)}{d x^\alpha} = D_\alpha f(x) = \frac{d}{d x} R_{1-\alpha}(f)(x).$$

In Fourier variables

$$\hat{D}_\alpha f(\xi) = i \frac{\xi}{|\xi|} |\xi|^{\alpha} \hat{f}(\xi).$$
To fix notations, for $0 < \alpha < 1$, we let $R_\alpha$ 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 then define the fractional derivative of order $\alpha$ of a real function $f$ as $(0 < \alpha < 1)$

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

In Fourier variables

$$\widehat{D_\alpha f}(\xi) = i \frac{\xi}{|\xi|} |\xi|^\alpha \widehat{f}(\xi).$$
To fix notations, for $0 < \alpha < 1$, we let $R_\alpha$ 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

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

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

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

In Fourier variables

$$\widehat{D_\alpha f}(\xi) = i \frac{\xi}{|\xi|} |\xi|^{\alpha} \hat{f}(\xi).$$
To fix notations, for $0 < \alpha < 1$, we let $R_\alpha$ 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

$$\hat{R}_\alpha(f)(\xi) = |\xi|^{\alpha} \hat{f}(\xi).$$

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

$$\frac{d^\alpha f(x)}{d x^\alpha} = D_\alpha f(x) = \frac{d}{d x} R_{1-\alpha}(f)(x).$$

In Fourier variables

$$\hat{D}_\alpha f(\xi) = i \frac{\xi}{|\xi|} |\xi|^{\alpha} \hat{f}(\xi).$$
The fractional Fisher information

- 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 $\{f > 0\}$

$$D_\alpha f(x) = -Cf(x)$$

- Passing to Fourier transform, this identity yields

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

Consequently

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

$$D_\alpha f(x) = -Cxf(x)$$

Passing to Fourier transform, this identity yields

$$i\xi|\xi|^{\alpha-1} \hat{f}(\xi) = -iCe^\frac{|\xi|^{\alpha+1}}{C(\alpha+1)}.$$
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 $\{f > 0\}$

$$D_{\alpha}f(x) = -Cxf(x)$$

Passing to Fourier transform, this identity yields

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

Consequently

$$\hat{f}(\xi) = \hat{f}(0)e^{-|\xi|^{\alpha+1}/C(\alpha + 1)}.$$
### The fractional Fisher information

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

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

- The relative (to $X$) fractional score function of $X$ and $Z_\lambda$ assumes the simple expression

\[ \tilde{\rho}_\lambda(X) = \frac{D_{\lambda-1}f(X)}{f(X)} + \frac{X}{\lambda}. \]

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

\[ I_\lambda(X) = I_\lambda(f) = \int_{\{f>0\}} \left( \frac{D_{\lambda-1}f(x)}{f(x)} + \frac{x}{\lambda} \right)^2 f(x) \, dx. \]
The fractional Fisher information

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

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

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

\[
\tilde{\rho}_\lambda(X) = \frac{D_{\lambda-1}f(X)}{f(X)} + \frac{X}{\lambda}.
\]

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

\[
l_\lambda(X) = l_\lambda(f) = \int_{\{f > 0\}} \left( \frac{D_{\lambda-1}f(x)}{f(x)} + \frac{x}{\lambda} \right)^2 f(x) \, dx.
\]
Arranging constants, we show that, if $Z_\lambda$ is a Lévy distribution of density $L_\lambda$ ($1 < \lambda < 2$)

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

The relative (to $X$) fractional score function of $X$ and $Z_\lambda$ assumes the simple expression

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

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

$$I_\lambda(X) = I_\lambda(f) = \int_{\{f>0\}} \left(\frac{D_{\lambda-1}f(x)}{f(x)} + \frac{x}{\lambda}\right)^2 f(x) \, dx.$$
The fractional Fisher information

- 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 $\lambda$.
- At difference with the relative standard relative Fisher information, $I\lambda$ is well-defined any time that 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 $\nu > 0$

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

This leads to the relative fractional Fisher information $I\lambda,\nu(X)$. 
The fractional Fisher information

- 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 $\lambda$.
- At difference with the relative standard relative Fisher information, $I_\lambda$ is well-defined any time that 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 $\nu > 0$

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

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

- 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 $\lambda$.
- At difference with the relative standard relative Fisher information, $I_{\lambda}$ is well-defined any time that 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 $P_{\lambda}$ the set of probability density functions such that $I_{\lambda}(f) < +\infty$
- The concept of fractional score can be generalized. For $\nu > 0$

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

This leads to the relative fractional Fisher information $I_{\lambda,\nu}(X)$.
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 $\lambda$, with $1 < \lambda < 2$, and each positive constant $\delta$, with $0 < \delta < 1$, the relative fractional score function of the sum $X_1 + X_2$ can be expressed as

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

This Lemma has several interesting consequences.
• 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 $\lambda$, with $1 < \lambda < 2$, and each positive constant $\delta$, with $0 < \delta < 1$, the relative fractional score function of the sum $X_1 + X_2$ can be expressed as

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

• 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 \( \nu \), the following identity holds

\[ I_{\lambda,\nu}(\nu^{1/\lambda}X) = \nu^{-2(1-1/\lambda)}I_\lambda(X). \]
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 \( \nu \), the following identity holds

\[ I_{\lambda,\nu}(\nu^{1/\lambda}X) = \nu^{-2(1-1/\lambda)} I_\lambda(X). \]
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 $\lambda$, with $1 < \lambda < 2$. Then, for each constant $\delta$ with $0 < \delta < 1$, $I_\lambda(\delta^{1/\lambda}X_1 + (1 - \delta)^{1/\lambda}X_2)$ is bounded, and

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

Moreover, there is equality if and only if, up to translation, both $X_j$, $j = 1, 2$ are Lévy variables of exponent $\lambda$. 
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 \(s \subset [n]\), let \(X_s\) stand for the collection of random variables \((X_i : i \in s)\), with the indices taken in their natural increasing order. Then

**Theorem**

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

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

Then

\[
E[U^2] \leq \frac{m}{n} E[\Phi^2].
\]

This quantifies the reduction.
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 \(s \subset [n]\), let \(X_s\) stand for the collection of random variables \((X_i : i \in s)\), with the indices taken in their natural increasing order. Then

**Theorem**

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

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

Then

\[
E[U^2] \leq \frac{m}{n} E[\Phi^2].
\]

This quantifies the reduction.
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 \(s \subset [n]\), let \(X_s\) stand for the collection of random variables \((X_i : i \in s)\), with the indices taken in their natural increasing order. Then

**Theorem**

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

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

Then

\[
E[U^2] \leq \frac{m}{n} E[\Phi^2].
\]

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 $\lambda$-Fisher information, $1 < \lambda < 2$. Then, for each $n > 1$, the relative $\lambda$-Fisher information of $T_n$ is decreasing in $n$, and the following bound holds

$$I_\lambda(T_n) \leq \left(\frac{n-1}{n}\right)^{(2-\lambda)/\lambda} I_\lambda(T_{n-1}).$$
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 $\lambda$-Fisher information, $1 < \lambda < 2$. Then, for each $n > 1$, the relative $\lambda$-Fisher information of $T_n$ is decreasing in $n$, and the following bound holds

$$I_{\lambda}(T_n) \leq \left(\frac{n - 1}{n}\right)^{(2-\lambda)/\lambda} I_{\lambda}(T_{n-1}).$$
At difference with the classical entropic central limit theorem, this quantifies the decay.

\[
I_\lambda(T_n) \leq \left( \frac{1}{n} \right)^{(2-\lambda)/\lambda} I_\lambda(X).
\]

There is convergence in relative $\lambda$-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.
At difference with the classical entropic central limit theorem, this quantifies the decay.

\[ I_\lambda(T_n) \leq \left( \frac{1}{n} \right)^{(2-\lambda)/\lambda} I_\lambda(X). \]

There is convergence in relative \( \lambda \)-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.
At difference with the classical entropic central limit theorem, this quantifies the **decay**.

\[
I_\lambda(T_n) \leq \left(\frac{1}{n}\right)^{(2-\lambda)/\lambda} I_\lambda(X).
\]

- There is convergence in relative \(\lambda\)-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.
At difference with the classical entropic central limit theorem, this quantifies the decay.

\[ I_\lambda(T_n) \leq \left( \frac{1}{n} \right)^{(2-\lambda)/\lambda} I_\lambda(X). \]

There is convergence in relative $\lambda$-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.
On the domain of attraction

- The leading example of a function which belongs to the domain of attraction of the $\lambda$-stable law is the so-called Linnik distribution:

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

- For all $0 < \lambda \leq 2$, this function is the characteristic function of a symmetric probability distribution. In addition, when $\lambda > 1$, $\hat{p}_\lambda \in L^1(\mathbb{R})$, which, by applying the inversion formula, shows that $p_\lambda$ 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 $\lambda$-stable law)?

- Main open question: Does (as in the classical case) convergence in relative fractional Fisher information imply convergence in $L^1(\mathbb{R})$?
On the domain of attraction

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

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

- For all $0 < \lambda \leq 2$, this function is the characteristic function of a symmetric probability distribution. In addition, when $\lambda > 1$, $\hat{p}_\lambda \in L^1(\mathbb{R})$, which, by applying the inversion formula, shows that $p_\lambda$ 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 $\lambda$-stable law)?

- Main open question: Does (as in the classical case) convergence in relative fractional Fisher information imply Convergence in $L^1(\mathbb{R})$?
On the domain of attraction

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

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

- For all $0 < \lambda \leq 2$, this function is the characteristic function of a symmetric probability distribution. In addition, when $\lambda > 1$, $\hat{p}_\lambda \in L^1(\mathbb{R})$, which, by applying the inversion formula, shows that $p_\lambda$ 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 $\lambda$-stable law)?

- Main open question: Does (as in the classical case) convergence in relative fractional Fisher information imply Convergence in $L^1(\mathbb{R})$?
On the domain of attraction

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

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

- For all $0 < \lambda \leq 2$, this function is the characteristic function of a symmetric probability distribution. In addition, when $\lambda > 1$, $\hat{p}_\lambda \in L^1(\mathbb{R})$, which, by applying the inversion formula, shows that $p_\lambda$ 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 $\lambda$-stable law)?

- Main open question: Does (as in the classical case) convergence in relative fractional Fisher information imply convergence in $L^1(\mathbb{R})$?