On generalized Fisher information, generalized Cramér-Rao inequalities, and an extension of the Shannon-Fisher-Gauss setting

Jean-François Bercher
Laboratoire d’Informatique - Institut Gaspard Monge, UMR 8049 Université Paris-Est, Esiee-Paris, France

Séminaire LSS
Gif, le 28 novembre 2013
Averaging

- **Estimation context**: estimate $h(\theta)$ by a function $T(X)$ of the data
- In estimation, the error is $T(X) - h(\theta)$, and
- **Idea**: compute a moment of order $\alpha$ of the error
Averaging

- **Estimation context:** estimate $h(\theta)$ by a function $T(X)$ of the data
- In estimation, the error is $T(X) - h(\theta)$, and
- **Idea:** compute a moment of order $\alpha$ of the error with respect to a distribution $g(x; \theta)$ instead of $f(x; \theta)$:
Averaging

- **Estimation context**: estimate $h(\theta)$ by a function $T(X)$ of the data.
- In estimation, the error is $T(X) - h(\theta)$, and
- **Idea**: compute a moment of order $\alpha$ of the error with respect to a distribution $g(x; \theta)$ instead of $f(x; \theta)$:

  the bias can be evaluated as
  \[
  \int_X (T(x) - h(\theta)) f(x; \theta) \, dx = E_f [T(X) - h(\theta)] = \eta(\theta) - h(\theta),
  \]
  while a general moment of of the error can be computed with respect to another probability distribution $g(x; \theta)$, as in
  \[
  E_g \left[ |T(X) - h(\theta)|^{\beta} \right] = \int_X |T(x) - h(\theta)|^{\alpha} g(x; \theta) \, dx.
  \]
Averaging

- **Estimation context:** estimate $h(\theta)$ by a function $T(X)$ of the data.
- In estimation, the error is $T(X) - h(\theta)$, and
- **Idea:** compute a moment of order $\alpha$ of the error with respect to a distribution $g(x; \theta)$ instead of $f(x; \theta)$:

  \[ \int_X (T(x) - h(\theta)) f(x; \theta) \, dx = E_f [T(X) - h(\theta)] = \eta(\theta) - h(\theta), \]

  while a general moment of the error can be computed with respect to another probability distribution $g(x; \theta)$, as in

  \[ E_g \left[ |T(X) - h(\theta)|^\beta \right] = \int_X |T(x) - h(\theta)|^\alpha g(x; \theta) \, dx. \]

- The two distributions $f(x; \theta)$ and $g(x; \theta)$ can be chosen very arbitrary, e.g. as a pair of escorts

  \[ f(x; \theta) = \frac{g(x; \theta)^q}{\int g(x; \theta)^q \, dx} \quad \text{and} \quad g(x; \theta) = \frac{f(x; \theta)^\bar{q}}{\int f(x; \theta)^\bar{q} \, dx} \]
Escort distributions

“Escort probabilities are naturally induced from the studies of multifractals and non-extensive entropy to play an important but mysterious role.”
Ohara-Matsuzoe-Amari (2010)

A pair of escort distributions is defined by

\[ f(x; \theta) = \frac{g(x; \theta)^q}{\int g(x; \theta)^q dx} \quad \text{and} \quad g(x; \theta) = \frac{f(x; \theta)^{\bar{q}}}{\int f(x; \theta)^{\bar{q}} dx}, \]

where \( q \) is a positive parameter, \( \bar{q} = 1/q \).

- Applications in source coding (a)
- Entropy estimation (Regnault & Girardin)
- A geometric model (b)

(a) jfb - Source Coding with Escort Distributions and Rényi Entropy Bounds, Physics Letters A, 2009
(b) jfb - A simple probabilistic construction yielding generalized entropies and divergences, escort distributions and q-Gaussians, Physica A, 2012
Illustration

On generalized Fisher information, etc, etc...
The Shannon-Fisher-Gauss setting

\[ N[f] = \frac{1}{2} \pi e^{\exp(2nH[f])} \]

the entropy power,

\[ m_2[f] \]

the moment of order 2,

\[ G \]

the standard Gaussian

Maximum entropy: variance-entropy inequality

\[ m_2[f] \geq m_2[G] \]

de Bruijn's identity

If \( X_t = X + \sqrt{t}N, \) \( N \sim G \)

\[ d \frac{d}{dt} H[f_t] = I_2, \]

Stam's inequality

\[ I_2,1[f] \geq I_2,1[G] \]

Minimum Fisher information: Cramér-Rao inequality

\[ I_2,1[f] m_2[f] \geq I_2,1[G] m_2[G] = n \]

The Fisher Information Inequality

\[ I_2,1[f_1 \ast f_2] - 1 \geq I_2,1[f_1] - 1 + I_2,1[f_2] - 1 \]

All the inequalities are saturated by the standard Gaussian.

On generalized Fisher information, etc, etc...
The Shannon-Fisher-Gauss setting

Let $N[f] = \frac{1}{2\pi e} \exp \left( \frac{2}{n} H[f] \right)$ the entropy power, $m_2[f]$ the moment of order 2, $G$ the standard Gaussian
The Shannon-Fisher-Gauss setting

Let $N[f] = \frac{1}{2\pi e} \exp \left( \frac{2}{n} H[f] \right)$ the entropy power, $m_2[f]$ the moment of order 2, $G$ the standard Gaussian

1. **Maximum entropy: variance-entropy inequality**

$$\frac{m_2[f]}{N[f]} \geq \frac{m_2[G]}{N[G]}$$

2. **de Bruijn’s identity** If $X_t = X + \sqrt{t}N$, $N \sim G$

$$\frac{d}{dt} H[f_t] = I_{2,1}[f_t]$$

3. **Stam’s inequality**

$$I_{2,1}[f] N[f] \geq I_{2,1}[G] N[G]$$

4. **Minimum Fisher information: Cramér-Rao inequality**

$$I_{2,1}[f] m_2[f] \geq I_{2,1}[G] m_2[G] = n$$

5. **The Fisher Information Inequality**

$$I_{2,1}[f_1 * f_2]^{-1} \geq I_{2,1}[f_1]^{-1} + I_{2,1}[f_2]^{-1}$$

On generalized Fisher information, etc, etc...
**q-entropies and q-Gaussians**

- Rényi-Tsallis $q$-entropies
  - Renyi entropy (1961)
    \[
    H_q[f] = \frac{1}{1-q} \log \left( \int f(x)^q dx \right).
    \]
    \[
    H_q[f] = \frac{1}{1-q} \left( \int f(x)^q dx - 1 \right).
    \]

- Maximum $q$-entropy $\implies$ $q$-Gaussians

  \[
  f(x) \propto \left( 1 - (q - 1)\gamma |x|^\alpha \right)^{\frac{1}{q-1}}, \quad f(x) \propto \exp \left( -\gamma |x|^\alpha \right) \text{ if } q = 1
  \]

  Compact support $q > 1$; **power like distribution** for $q \leq 1$

- Is it possible to find similar interrelations for $q$-entropies, higher orders and a suitable extension of Fisher information?
Generalized $q$-Gaussians

They have the probability density

$$G_\gamma(x) = \begin{cases} \frac{1}{Z(\gamma)} \exp(-\gamma |x|^\alpha) & \text{for } q = 1 \\ \end{cases}$$

Statistics of wavelet coefficients, non Gaussian prior in image restoration, in communications.
Generalized $q$-Gaussians

They have the probability density

$$G_\gamma(x) = \begin{cases} \frac{1}{Z(\gamma)} \exp (-\gamma|x|^{\alpha}) & \text{for } q = 1 \\ \frac{1}{Z(\gamma)} (1 - \gamma(q - 1)|x|^{\alpha})^{\frac{1}{q-1}} & \text{if } q \neq 1 \end{cases}$$

where $\gamma > 0$ and $(x)_+ = \max\{x, 0\}$.

- Statistics of wavelet coefficients, non Gaussian prior in image restoration, in communications.
Generalized $q$-Gaussians

They have the probability density

$$G_\gamma(x) = \begin{cases} \frac{1}{Z(\gamma)} \exp(-\gamma|x|^\alpha) & \text{for } q = 1 \\ \frac{1}{Z(\gamma)} (1 - \gamma(q - 1)|x|^\alpha)^{\frac{1}{q-1}} & \text{if } q \neq 1 \end{cases}$$

where $\gamma > 0$ and $(x)_+ = \max\{x, 0\}$.

- Statistics of wavelet coefficients, non Gaussian prior in image restoration, in communications.
- Maximum Entropy distribution on nonextensive statistical physics, solutions of nonlinear diffusion equations, saturation of Sobolev inequalities on $\mathbb{R}^n$. 

On generalized Fisher information, etc, etc...
Generalized $q$-Gaussians

They have the probability density

$$G_{\gamma}(x) = \begin{cases} \frac{1}{Z(\gamma)} \exp(-\gamma|x|^\alpha) & \text{for } q = 1 \\ \frac{1}{Z(\gamma)} (1 - \gamma(q - 1)|x|^\alpha)^{\frac{1}{q-1}} & \text{if } q \neq 1 \end{cases}$$

where $\gamma > 0$ and $(x)_+ = \max\{x, 0\}$.

- Statistics of wavelet coefficients, non Gaussian prior in image restoration, in communications.
- Maximum Entropy distribution on nonextensive statistical physics, solutions of nonlinear diffusion equations, saturation of Sobolev inequalities on $\mathbb{R}^n$.
- Compact support $q > 1$; power like distribution for $q < 1$.  

On generalized Fisher information, etc, etc...
Generalized Gaussians

$q > 1$ – Compact support. $\alpha = 2$, $\sigma^2 = 1$
Generalized Gaussians

$q \leq 1$ – non compact support. $\alpha = 2$, $\sigma^2 = 1$
Fisher information

- Measure of the information about a parameter \( \rightarrow \) Cramér-Rao bound on the variance of an estimator
- Inference in statistical physics (Frieden)
- Tool for characterizing complex signals and systems, with applications (information planes).

**Definition**

Let \( f(x; \theta) \) denote a probability density defined over a subset \( X \) of \( \mathbb{R}^n \), and \( \theta \in \Theta \) a real parameter.

\[
l_{2,1}[f, \theta] = \int_X \left( \frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2 f(x; \theta) dx
\]

When \( \theta \) is the location parameter, i.e. \( f(x; \theta) = f(x - \theta) \),

\[
l_{2,1}[f, \theta = 0] = \int_X \left( \frac{d \ln f(x)}{dx} \right)^2 f(x) dx.
\]
The standard Cramér-Rao inequality

- Estimation: find a function $T(x)$ of the data $x$, that approaches the unknown value of a function $h(\theta)$ of the parameter $\theta$ of the pdf of these data.
The standard Cramér-Rao inequality

Estimation: find a function $T(x)$ of the data $x$, that approaches the unknown value of a function $h(\theta)$ of the parameter $\theta$ of the pdf of these data.

Cramér-Rao inequality

Let $f(x; \theta)$ be a probability density function defined over a subset $X$ of $\mathbb{R}^n$, $\theta \in \Theta \subset \mathbb{R}$, $h(\theta)$ a scalar valued function and $\eta(\theta) = E[T(X)]$. Under some regularity conditions,

$$E \left[ \left| T(X) - h(\theta) \right|^2 \right]_{L_2,1} \geq \left| \frac{\partial}{\partial \theta} \eta(\theta) \right|^2.$$

The estimator is said efficient if it is unbiased and saturates the inequality. This can happen if the probability density and the estimator satisfy $\frac{\partial}{\partial \theta} \ln f(x; \theta) = k(\theta)(T(x) - h(\theta)).$
The standard Cramér-Rao inequality

Estimation: find a function $T(x)$ of the data $x$, that approaches the unknown value of a function $h(\theta)$ of the parameter $\theta$ of the pdf of these data.

Cramér-Rao inequality (matrix case)

If $\theta \in \Theta \subset \mathbb{R}^m$, $h(\theta) : \mathbb{R}^m \to \mathbb{R}^k$, denoting $\dot{\eta}(\theta) = \nabla_{\theta} \eta(\theta)^T$ the $k \times m$ matrix of partial derivatives, and under some regularity conditions

$$E \left[ (T(X) - h(\theta))(T(X) - h(\theta))^T \right] \succeq \dot{\eta}(\theta)^T l_{2,1}[f; \theta]^{-1} \dot{\eta}(\theta)$$

where $l_{2,1}[f, \theta] = E_f \left[ \frac{\nabla_{\theta} f(x; \theta)}{f(x; \theta)} \right]^T$ is the Fisher information matrix. Equality holds iff $T(x) - h(\theta) = \dot{\eta}(\theta)^T l_{2,1}[f; \theta]^{-1} \frac{\nabla_{\theta} f(x; \theta)}{f(x; \theta)}$
Agenda

In this presentation...

- Ingredients
- $\chi^\beta$ – divergence
- Extension of de Bruijn’s identity
- Extension of Stam’s inequality
- Extensions of Cramér-Rao inequality, including matrix forms
- Particular case of location parameter
- Fisher Information Inequality
The $\chi^\beta$ divergence and Fisher information

- $\chi^\beta$-divergence, $\beta > 1$, between $f_1$ and $f_2$
  \[
  \chi^\beta(f_1, f_2) = E_{f_2} \left[ \left| 1 - \frac{f_1}{f_2} \right|^\beta \right]
  \]

- the Fisher information of $f_{\theta}$ is nothing but
  \[
  l_2[f_{\theta}; \theta] = \lim_{|t| \to 0} \chi^2(f_{\theta+t}, f_{\theta}) / t^2
  \]

- Vajda’s generalized Fisher information
  \[
  l_\beta[f_{\theta}; \theta] = \lim_{|t| \to 0} \chi^\beta(f_{\theta+t}, f_{\theta}) / |t|^\beta = E_{f_{\theta}} \left[ \left| \frac{f_{\theta}}{f_{\theta}} \right|^\beta \right]
  \]

A modified $\chi^\beta$-divergence and generalized Fisher information

- A $\chi^\beta$-divergence modified in order to average wrt a third distribution $g(x; \theta)$:

$$\chi^\beta_g(f_1, f_2) = E_g \left[ \left| \frac{f_2 - f_1}{g} \right|^\beta \right]$$
A modified $\chi^\beta$-divergence and generalized Fisher information

- A $\chi^\beta$-divergence modified in order to average wrt a third distribution $g(x; \theta)$:

$$\chi^\beta_g(f_1, f_2) = E_g \left[ \frac{f_2 - f_1}{g} \right]^\beta$$

- Generalized Fisher information, taken wrt $g_\theta$

$$l_\beta[f_\theta | g_\theta; \theta] = \lim_{|t| \to 0} \chi^\beta(f_{\theta+t}, f_\theta)/|t|^\beta = E_{g_\theta} \left[ \frac{\dot{f}_\theta}{g_\theta} \right]^\beta$$
A modified $\chi^\beta$-divergence and generalized Fisher information

- A $\chi^\beta$-divergence modified in order to average wrt a third distribution $g(x; \theta)$:

$$\chi^\beta_g(f_1, f_2) = E_g \left[ \left| \frac{f_2 - f_1}{g} \right|^\beta \right]$$

- Generalized Fisher information, taken wrt $g_\theta$

$$l_\beta[f_\theta|g_\theta; \theta] = \lim_{|t| \to 0} \chi^\beta(f_\theta + t, f_\theta)/|t|^\beta = E_{g_\theta} \left[ \left| \frac{f_\theta}{g_\theta} \right|^\beta \right]$$

- In the multivariate case, $l_\beta[f_\theta|g_\theta; \theta] = E_{g_\theta} \left[ \left| \frac{\nabla f_\theta}{g_\theta} \right|^\beta \right]$
A modified $\chi^\beta$-divergence and generalized Fisher information

- A $\chi^\beta$-divergence modified in order to average wrt a third distribution $g(x; \theta)$:

$$
\chi^\beta_g(f_1, f_2) = E_g \left[ \frac{f_2 - f_1}{g} \right]^\beta
$$

- Generalized Fisher information, taken wrt $g_\theta$

$$
I^\beta_{\theta} [f_\theta | g_\theta; \theta] = \lim_{|t| \to 0} \chi^\beta(f_\theta + t, f_\theta)/|t|^\beta = E_{g_\theta} \left[ \left| \frac{\dot{f}_\theta}{g_\theta} \right|^\beta \right]
$$

- In the multivariate case, $I^\beta_{\theta} [f_\theta | g_\theta; \theta] = E_{g_\theta} \left[ \left| \frac{\nabla f_\theta}{g_\theta} \right|^\beta \right]$

This generalized Fisher information is involved in a generalized Cramér-Rao inequality for parameter estimation.
Properties

1 - The modified $\chi^\beta$-divergence has information monotonicity. Coarse-graining the data leads to a loss of information: 
$\chi^\beta_g(f_1, f_2) \geq \chi^\beta_{\tilde{g}}(\tilde{f}_1, \tilde{f}_2)$. Data processing inequality: if $Y = \phi(X)$, then

$$
\begin{align*}
\chi^\beta_g(f_1, f_2) & \geq \chi^\beta_{\phi}(f_1^\phi, f_2^\phi), \\
l^\beta[f_\theta \mid g^\theta; \theta] & \geq l^\beta[f_\theta^\phi \mid g^\phi; \theta].
\end{align*}
$$

2 - Matrix Fisher data processing inequality: $l_{2,g}[\theta] \geq l_{2,g^\phi}[\theta]$

Taylor expansion $f_{\theta+t} = f_{\theta} + \sum_i t_i \partial_i f_{\theta} + \frac{1}{2} \sum_i \sum_j t_i t_j \partial^2_{ij} f_{\theta} + \ldots$

$$
\chi^2_g(f_{\theta}, f_{\theta+t}) = E_g \left[ \left( \frac{f_{\theta+t} - f_{\theta}}{g^\theta} \right)^2 \right] = E_g \left[ \sum_i \sum_j t_i t_j \frac{\partial_i f_{\theta} \partial_j f_{\theta}}{g^2} \right] = t^T l_{2,g}[\theta] t,
$$

where $l_{2,g}[\theta] = E_g \left[ \psi_\theta \psi_\theta^T \right] = E_g \left[ \frac{\nabla f_{\theta}}{g^\theta} \frac{\nabla f_{\theta}^T}{g^\theta} \right]$. $\nabla f_{\theta}$

Since $\chi^2_g(f_{\theta}, f_{\theta+t}) \geq \chi^2_{g^\phi}(f_{\theta}^\phi, f_{\theta+t}^\phi)$, then $l_{2,g}[\theta] \geq l_{2,g^\phi}[\theta]$
3 - If $T(X)$ is a statistic, then with $\alpha^{-1} + \beta^{-1} = 1$, $\alpha \geq 1$, we have

$$|E_{f_2}[T] - E_{f_1}[T]| \leq E_g[|T|^\alpha]^{\frac{1}{\alpha}} \chi_g(f_1, f_2)^{\frac{1}{\beta}}.$$ 

It suffices here to consider

$$E_g[T \left( \frac{f_2 - f_1}{g} \right)] = |E_{f_2}[T] - E_{f_1}[T]|$$

and then apply the Hölder inequality.
3 - If \( T(X) \) is a statistic, then with \( \alpha^{-1} + \beta^{-1} = 1, \alpha \geq 1 \), we have

\[
\left| E_{f_2} [T] - E_{f_1} [T] \right| \leq E_g \left[ |T|^\alpha \right]^{\frac{1}{\alpha}} \chi^\beta_g (f_1, f_2)^{\frac{1}{\beta}}.
\]

It suffices here to consider \( \left| E_g \left[ T \left( \frac{f_2-f_1}{g} \right) \right] \right| = \left| E_{f_2} [T] - E_{f_1} [T] \right| \) and then apply the Hölder inequality

**Generalized Cramér-Rao inequality.** Set \( f_2 = f_{\theta+t}, f_1 = f_{\theta} \), and denote \( \eta = E_f [T(X)] \). Then divide both sides by \( t \), substitute \( T(X) \) by \( T(X) - h(\theta) \), and take the limit \( t \to 0 \). Then

\[
E_g \left[ |T(X) - h(\theta)|^\alpha \right]^{\frac{1}{\alpha}} l_\beta [f_\theta | g_\theta; \theta]^{\frac{1}{\beta}} \geq \left| \frac{d}{d\theta} \eta \right|
\]

with \( l_\beta [f_\theta | g_\theta; \theta] = E_{g_\theta} \left[ \left| \frac{f_\theta}{g_\theta} \right|^\beta \right] \).
De Bruijn’s identity: if $Y_t = X + \sqrt{2t}Z$ where $Z$ is a standard Gaussian vector and $X \in \mathbb{R}^n$, then

$$\frac{d}{dt} H[f_{Y_t}] = l_{2,1}[f_{Y_t}]$$

The classical proof uses the fact that $Y_t$ satisfies the heat equation $\frac{\partial f}{\partial t} = \Delta f$, where $\Delta$ denotes the Laplace operator.
Nonlinear heat equation

Nonlinear versions of the heat equation: the doubly nonlinear equation [J. L. Vazquez]

$$\frac{\partial}{\partial t} f = \Delta_\beta f^m = \text{div} \left( |\nabla f^m|^{\beta - 2} \nabla f^m \right),$$

with $\Delta_\beta f := \text{div} \left( |\nabla f|^{\beta - 2} \nabla f \right)$, $\beta$-Laplacian operator

- the standard heat equation ($\beta = 2, m = 1$),
- the $\beta$-Laplace equation ($\beta \neq 2, m = 1$),
- porous medium equation ($\beta = 2, m > 1$)
- fast diffusion equation ($\beta = 2, m < 1$).

For $m(\beta - 1) + (\beta/n) - 1 > 0$, unique solution: Barenblatt profile. With $q = m + 1 - \frac{\alpha}{\beta}$, it is nothing but a generalized $q$-Gaussian distribution.
An extended de Bruijn identity

[Extended de Bruijn identity] Let $f(x, t), x \in X$ of $\mathbb{R}^n$ satisfying the doubly nonlinear equation. Then, for $\beta > 1$, $q = m + 1 - \frac{\alpha}{\beta}$, $M_q[f] = \int f^q$ and $S_q[f]$ the Tsallis entropy

$$\frac{d}{dt} S_q[f] = \left( \frac{m}{q} \right)^{\beta-1} M_q[f]^\beta I_{\beta,q}[f]$$

with

$$I_{\beta,q}[f] = \left( \frac{q}{M_q[f]} \right)^\beta \int_X f(x)^{\beta(q-1)+1} \left( \frac{\nabla f(x)}{f(x)} \right)^\beta dx$$
An extended de Bruijn identity

[Extended de Bruijn identity] Let $f(x, t), x \in X$ of $\mathbb{R}^n$ satisfying the doubly nonlinear equation. Then, for $\beta > 1$, $q = m + 1 - \frac{\alpha}{\beta}$, $M_q[f] = \int f^q$ and $S_q[f]$ the Tsallis entropy

$$\frac{d}{dt} S_q[f] = \left( \frac{m}{q} \right)^{\beta - 1} M_q[f]^\beta I_{\beta, q}[f]$$

with

$$I_{\beta, q}[f] = \left( \frac{q}{M_q[f]} \right)^\beta \int_X f(x)^{\beta(q-1)+1} \left( \frac{\nabla f(x)}{f(x)} \right)^\beta \, dx$$

- The proof of this result relies on integration by part (actually using Green’s identity) along the solutions of the nonlinear heat equation.
- The standard Fisher information is recovered in the particular case $\alpha = \beta = 2$, and $q = m = 1$, and so is the de Bruijn identity.
Standard Stam inequality

The minimum of the standard Fisher information for probability densities with a given (Shannon) entropy is attained by the standard normal distribution:

\[ I_{2,1}[f] N[f] \geq I_{2,1}[G] N[G]. \]

where \( N[f] \) is the entropy power

\[ N[f] = \frac{1}{2\pi e} \exp \left( \frac{2}{n} H[f] \right), \]

and with equality if and only if \( f \) is any Gaussian distribution.
Generalized Stam Inequality

For \( n \geq 1 \), \( \beta \) and \( \alpha \) Hölder conjugates of each other, \( \alpha > 1 \), and \( q > \max\left\{(n-1)/n, n/(n+\alpha)\right\} \), then for any probability density on \( \mathbb{R}^n \), the following generalized Stam inequality holds

\[
I_{\beta,q}[f]^{\frac{1}{\beta}} N_q[f]^{\frac{1}{2}} \geq I_{\beta,q}[G]^{\frac{1}{\beta}} N_q[G]^{\frac{1}{2}}.
\]

where

\[
N_q[f] = M_q[f]^{\frac{2}{n}} = \frac{1}{1-q} \exp\left(\frac{2}{n} H_q[f]\right) = \left(\int_{\Omega} f(x)^q dx\right)^{\frac{2}{n} \frac{1}{1-q}},
\]

and with equality if and only if \( f \) is any generalized \( q \)-Gaussian.

- The proof relies on manipulations on a general sharp Gagliardo-Nirenberg inequality

---

jfb - Some properties of generalized Fisher information in the context of nonextensive thermostatistics, Physica A, August 2013
[Sketch of Proof] The inequality follows from a sharp Gagliardo-Nirenberg inequality due to Cordero et al (2004): with $a > 1$ and $\|\nabla u\|_\beta = \left( \int \|\nabla u\|_\beta^\beta \, dx \right)^{1/\beta}$, 

$$\|\nabla u\|_\beta \|u\|^{1-1}_{\theta a(\beta - 1) + 1} \geq K \|u\|^{1}_{a\beta}$$

where $K$ is a sharp constant attained iff $u$ is a generalized Gaussian with exponent $1/(1-a)$, and where $\theta$ is given by

$$\theta = n(a-1)/a(n\beta - (a\beta + 1 - a)(n - \beta)).$$

The idea is to take $u = g^t$, $g$ being a probability density function, with $a\beta t = 1$, and to note $q = [a(\beta - 1) + 1] t$. With these notations, we get that $\beta t = \beta (q - 1) + 1$, and the result follows.
Averaging

- **Estimation context**: estimate $h(\theta)$ by a function $T(X)$ of the data
- In estimation, the error is $T(X) - h(\theta)$, and
- **Idea**: compute a moment of order $\alpha$ of the error
Averaging

- **Estimation context:** estimate \( h(\theta) \) by a function \( T(X) \) of the data
- In estimation, the error is \( T(X) - h(\theta) \), and
- **Idea:** compute a moment of order \( \alpha \) of the error with respect to a distribution \( g(x; \theta) \) instead of \( f(x; \theta) \):

\[
\text{bias} = \mathbb{E}_f [ T(X) - h(\theta) ] = \eta(\theta) - h(\theta)
\]

while a general moment of the error can be computed with respect to another probability distribution \( g(x; \theta) \), as in

\[
\mathbb{E}_g [ |T(X) - h(\theta)|^\beta ] = \mathbb{E}_g [ |T(X) - h(\theta)|^\alpha ]
\]

The two distributions \( f(x; \theta) \) and \( g(x; \theta) \) can be chosen very arbitrary, e.g. as a pair of escorts

\[
f(x; \theta) = g(x; \theta)^q \frac{\int g(x; \theta)}{d\gamma} \quad \text{and} \quad g(x; \theta) = f(x; \theta)^{\bar{q}} \frac{\int f(x; \theta)}{d\gamma}
\]
Averaging

- **Estimation context:** estimate $h(\theta)$ by a function $T(X)$ of the data.
- In estimation, the error is $T(X) - h(\theta)$, and
- **Idea:** compute a moment of order $\alpha$ of the error with respect to a distribution $g(x; \theta)$ instead of $f(x; \theta)$:
  
  the bias can be evaluated as
  \[
  \int_X (T(x) - h(\theta)) f(x; \theta) \, dx =Ef[T(X) - h(\theta)] = \eta(\theta) - h(\theta),
  \]
  while a general moment of the error can be computed with respect to another probability distribution $g(x; \theta)$, as in
  \[
  Eg\left[|T(X) - h(\theta)|^\beta\right] = \int_X |T(x) - h(\theta)|^\alpha g(x; \theta) \, dx.
  \]
**Averaging**

- **Estimation context**: estimate \( h(\theta) \) by a function \( T(X) \) of the data

- In estimation, the error is \( T(X) - h(\theta) \), and

- **Idea**: compute a moment of order \( \alpha \) of the error with respect to a distribution \( g(x; \theta) \) instead of \( f(x; \theta) \):

  \[
  \text{the bias can be evaluated as } \int_X (T(x) - h(\theta)) f(x; \theta) \, dx = E_f [T(X) - h(\theta)] = \eta(\theta) - h(\theta),
  \]

- while a general moment of the error can be computed with respect to another probability distribution \( g(x; \theta) \), as in

  \[
  E_g |T(X) - h(\theta)|^\beta = \int_X |T(x) - h(\theta)|^\alpha g(x; \theta) \, dx.
  \]

- The two distributions \( f(x; \theta) \) and \( g(x, \theta) \) can be chosen very arbitrary, e.g. as a pair of escorts

  \[
  f(x; \theta) = \frac{g(x; \theta)^q}{\int g(x; \theta)^q \, dx} \quad \text{and} \quad g(x; \theta) = \frac{f(x; \theta)^\bar{q}}{\int f(x; \theta)^\bar{q} \, dx}
  \]
Generalized CR - ingredients

**Dual norm.** If $\|\cdot\|$ is an arbitrary norm, the dual norm $\|\cdot\|_*$ is defined by

$$\| Y \|_* = \sup_{\| X \| \leq 1} X \cdot Y,$$

where $X \cdot Y$ is the standard scalar product. E.g. if $\|\cdot\|$ is a $L_p$-norm, then $\|\cdot\|_*$ is the $L_q$-norm

Hölder-type inequality for vector-valued functions, with arbitrary norms. If $w(t)$ is a weight function, then

$$\left( \int \| X(t) \|^\alpha w(t) dt \right)^{\frac{1}{\alpha}} \left( \int \| Y(t) \|_*^\beta w(t) dt \right)^{\frac{1}{\beta}} \geq \int X(t) \cdot Y(t) w(t) dt$$

with $\alpha^{-1} + \beta^{-1} = 1$, $\alpha \geq 1$. With $K > 0$, the equality is obtained if

$$Y(t) = K \| X(t) \|^{\alpha^{-1}} \nabla X(t) \cdot X(t)$$
A generalized Cramér-Rao inequality

Under some regularity conditions, then for any estimator \( \hat{\theta}(X) \) of \( \theta \in \mathbb{R}^n \), and arbitrary norms,

\[
E_g \left[ \left\| \hat{\theta}(X) - \theta \right\|^\alpha \right] \frac{1}{\alpha} \geq \left| \nabla_{\theta} . E_f[\hat{\theta}(X)] \right|
\]

where \( \alpha^{-1} + \beta^{-1} = 1 \), \( \alpha \geq 1 \), and where

\[
I_{\beta}[f|g; \theta] = \int_{X} \left\| \frac{\nabla_{\theta} f(x; \theta)}{g(x; \theta)} \right\|_{*}^{\beta} g(x; \theta) \, dx
\]

The equality case is obtained if

\[
\frac{\nabla_{\theta} f(x; \theta)}{g(x; \theta)} = K \left\| \hat{\theta}(x) - \theta \right\|^{\alpha^{-1}} \nabla_{\hat{\theta}(x)-\theta} \left\| \hat{\theta}(x) - \theta \right\|
\]

- Proof: Divergence computation, averaging and Hölder inequality for vector valued functionals

A generalized Cramér-Rao inequality

[Sketch of proof for arbitrary norms on \( \mathbb{R}^n \)]

1. (a) evaluate the divergence of \( \eta(\theta) = E_f[\hat{\theta}(X)] \),
2. (b) introduce an averaging with respect to \( g \) and use the fact that the expectation of the score is zero
   \( \nabla_\theta . \eta(\theta) = \int_X \frac{\nabla_\theta f(x;\theta)}{g(x;\theta)} . (\hat{\theta}(x) - \theta) \ g(x;\theta) \ dx \),
3. (c) apply the version of the Hölder inequality for arbitrary norms, with \( X(x) = \hat{\theta}(x) - \theta \), \( Y(x) = \frac{\nabla_\theta f(x;\theta)}{g(x;\theta)} \), and \( w(x) = g(x;\theta) \).
Generalized Cramér-Rao - matrix version

When \( h(\theta) \) is scalar valued, we have the following variation on the Cramér-Rao bound theme, which involves a Fisher information matrix, but unfortunately in a non-explicit form.

**Theorem**

Let \( T(x) \) be an estimator of a scalar-valued function \( h(\theta) \) and set \( \eta(\theta) = E_f[T(X)] \). Then, under some regularity conditions

\[
E_g \left[ \left| T(X) - h(\theta) \right|^{\alpha} \right]^{\frac{1}{\alpha}} \geq \sup_{A > 0} \frac{\hat{\eta}(\theta)^T A \hat{\eta}(\theta)}{E_g \left[ \left| \hat{\eta}(\theta)^T A \psi_g(X; \theta) \right|^{\beta} \right]^{\frac{1}{\beta}}}
\]

with a certain equality condition, and where \( \psi_g(x; \theta) \) a score function given with respect to \( g(x; \theta) \):

\[
\psi_g(x; \theta) := \frac{\nabla_\theta f(x; \theta)}{g(x; \theta)}.
\]
Generalized Cramér-Rao - matrix version

[Proof] Differentiating $\eta(\theta) = E_f[T(X)]$ with respect to $\theta$ we get

$$\dot{\eta}(\theta) = \nabla_{\theta} \eta(\theta) = \nabla_{\theta} \int_X T(x) f(x; \theta) \, dx$$

$$= \int_X (T(x) - h(\theta)) \frac{\nabla_{\theta} f(x; \theta)}{g(x; \theta)} g(x; \theta) \, dx.$$ 

With $A > 0$, multiplying on the left by $\dot{\eta}(\theta)^T A$ gives

$$\dot{\eta}(\theta)^T A \dot{\eta}(\theta) = \int_X (T(x) - h(\theta)) \dot{\eta}(\theta)^T A \psi_g(x; \theta) g(x; \theta) \, dx,$$ 

and by the Hölder inequality, we obtain

$$E_g \left[ |T(x) - h(\theta)|^\alpha \right]^{\frac{1}{\alpha}} E_g \left[ \left| \dot{\eta}(\theta)^T A \psi_g(x; \theta) \right|^\beta \right]^{\frac{1}{\beta}} \geq \dot{\eta}(\theta)^T A \dot{\eta}(\theta).$$
Corollary

In the scalar case (or the case of a single component of \( \theta \)), the following inequality holds

\[
E_g \left[ |T(X) - h(\theta)|^\alpha \right] \frac{1}{\alpha} \geq \frac{|\dot{\eta}(\theta)|}{E_g \left[ |\psi_g(X; \theta)|^\beta \right]^{\frac{1}{\beta}}},
\]

with equality if and only if

\[
\psi_g(x; \theta) = c(\theta) \text{sign}(T(x) - h(\theta)) |T(x) - h(\theta)|^{\alpha-1}.
\]

This inequality recovers at once the generalized Cramér-Rao inequality we presented in the univariate case and with \( g \) the escort distribution associated with \( f \).

---

Matrix Fisher - quadratic case

A second interesting case is the multivariate case with $\alpha = \beta = 2$. Indeed, in that case, we get an explicit form for the generalized Fisher information matrix and an inequality which looks like the classical one.

**Corollary**

[Multivariate Cramér-Rao inequality with $\alpha = \beta = 2$] For $\alpha = \beta = 2$, we have

$$E_g \left[ |T(X) - h(\theta)|^2 \right] \geq \dot{\eta}(\theta)^T J_g(\theta)^{-1} \dot{\eta}(\theta)$$

with $J_g(\theta) = E_g \left[ \psi_g(X; \theta) \psi_g(X; \theta)^T \right]$, and with equality if and only if $|T(X) - h(\theta)| = k(\theta) |\dot{\eta}(\theta)^T J_g(\theta)^{-1} \psi_g(X; \theta)|$. 

On generalized Fisher information, etc, etc...
The Matrix version - Quadratic case

Analog of a well known result: the covariance of the estimation error is greater than the inverse of the Fisher information matrix (in the Löwner sense). The proof follows the lines in Liese (2008).

**Theorem**

Let $T(x)$ estimator of a vector valued function $h(\theta)$. Set

$\eta(\theta) = E_f[T(X)]$, denote $\psi_g(x; \theta) := \frac{\nabla_\theta f(x; \theta)}{g(x; \theta)}$ and

$J_g(\theta) = E_g \left[ \psi_g(X; \theta) \psi_g(X; \theta)^T \right]$ the Fisher Fisher information matrix $J_g(\theta)$. Then we have that

$$E_g \left[ (T(X) - h(\theta)) (T(X) - h(\theta))^T \right] \succeq \dot{\eta}(\theta)^T J_g(\theta)^{-1} \dot{\eta}(\theta)$$

with $\dot{\eta}(\theta) = \nabla_\theta E_f[T(X)^T]$, and equality iff

$\dot{\eta}(\theta) J_g(\theta)^{-1} \psi_g(X) = (T(X) - h(\theta))$. 

On generalized Fisher information, etc, etc...
Proof.

With \( \eta(\theta) = E_f[T(X)] \), then
\[
\dot{\eta}(\theta) = \nabla_{\theta} E_f[T(X)^T] = E_g[\psi_g(X) T(X)^T].
\]
For every vectors \( u \) and \( v \)
\[
E_g \left[ u^T (T(X) - h(\theta)) \psi_g(X)^T v \right] = u^T \dot{\eta}(\theta)^T v.
\]
By the Schwarz inequality, we have
\[
\left( u^T \dot{\eta}(\theta)^T v \right)^2 \leq \left( u^T C_g u \right) \left( v^T J_g v \right)
\]
with \( C_g := E_g \left[ (T(X) - h(\theta))(T(X) - h(\theta))^T \right] \). If \( v = J_g^{-1} \dot{\eta}(\theta)u \), then this reduces to
\[
\left( u^T \dot{\eta}(\theta)^T J_g^{-1} \dot{\eta}(\theta)u \right) \leq \left( u^T C_g u \right).
\]
The generalized CR inequality provides a bound, but not the way to achieve it...

In the generalized CR inequality, equality occurs if

\[ \nabla_{\theta} f(x; \theta)^{1-\bar{q}} = (1 - \bar{q})K(\theta) \left\| \hat{\theta}(x) - \theta \right\|^{\alpha - 1} \nabla_{\hat{\theta}(x)-\theta} \| \hat{\theta}(x) - \theta \|, \quad \text{with } K(\theta) > 0. \]

Thus, we see that if the bound is attained then if

\[ \hat{\theta}_{MEL}(x) = \arg \max_{\theta} f(x; \theta) = \arg \max_{\theta} \frac{g(x; \theta)^{q}}{M_q[g(x; \theta)]}, \]

Indeed, we have that the derivative in the left of is zero, and that \( \hat{\theta}(x) = \hat{\theta}_{MEL}(x) \). So a possible procedure is “Find the parameters that maximize the escort of the likelihood”

The CR bound is reached only if the likelihood has a special analytical form

A related rule has been proposed and studied in the literature (Ferrari, Ann. of Stats, 2010)
Figure: Estimated Probability density of the effective temperature of the stars in the CYG OB1 cluster [Rousseeuw 2005] – This real data series presents outliers.

Location and standard deviation parameters are estimated from the escort distribution for several values of $q$

<table>
<thead>
<tr>
<th>$q$</th>
<th>1</th>
<th>1.4</th>
<th>2</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>4.31</td>
<td>4.36</td>
<td>4.39</td>
<td>4.42</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.298</td>
<td>0.27</td>
<td>0.22</td>
<td>0.15</td>
</tr>
</tbody>
</table>


**Illustration**

Figure: Estimated Probability density of delays in light speed measurements [Stigler 1977] – This real data series shows outliers.

Location and standard deviation parameters are estimated from the escort distribution for several values of $q$

<table>
<thead>
<tr>
<th>$q$</th>
<th>1</th>
<th>1.4</th>
<th>2</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>26.21</td>
<td>27.18</td>
<td>27.37</td>
<td>27.12</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>10.85</td>
<td>8.04</td>
<td>5.83</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Table: Estimates for different values of $q$. Results are similar to a trimming of data.
Case of a translation family

In the particular case of a translation parameter, the generalized Cramér-Rao inequalities leads to new functional inequalities and to new characterizations of $q$-Gaussians.

- Let $\theta$ be a location parameter, and define by $f(x; \theta)$ the family of density $f(x; \theta) = f(x - \theta)$. In this case, we have that $\nabla_\theta f(x; \theta) = -\nabla_x f(x - \theta)$

- Without loss of generality, assume that $f(x)$ has zero mean.
- The estimator $\hat{\theta}(x) = x$ is unbiased (wrt $f$).
- Assume the $\theta = 0$
Case of a translation family

In the particular case of a translation parameter, the generalized Cramér-Rao inequalities leads to new functional inequalities and to new characterizations of \(q\)-Gaussians.

- Let \(\theta\) be a location parameter, and define by \(f(x; \theta)\) the family of density \(f(x; \theta) = f(x - \theta)\). In this case, we have that \(\nabla_\theta f(x; \theta) = -\nabla_x f(x - \theta)\).

- Without loss of generality, assume that \(f(x)\) has zero mean.

- The estimator \(\hat{\theta}(x) = x\) is unbiased (wrt \(f\)).

- Assume the \(\theta = 0\)

- This leads to a general functional inequality

- In the special case of an escort pair,
General functional inequality

Let $f(x)$ and $g(x)$ be two multivariate probability density functions defined over a subset $X$ of $\mathbb{R}^n$, plus some technical conditions. Then, the following inequality holds

$$\left( \int_X \|x\|^{\alpha} g(x) \, dx \right)^{\frac{1}{\alpha}} \left( \int_X \left\| \frac{\nabla_x f(x)}{g(x)} \right\|^{\beta} \, dx \right)^{\frac{1}{\beta}} \geq n,$$

with equality if (and only if when the dual norm is strictly convex)

$$\nabla_x f(x) = -K g(x) \|x\|^{\alpha^{-1}} \nabla_x \|x\|,$$

with $K > 0$

---

Generalized $q$-Cramér-Rao inequality

**Corollary**

Under some regularity conditions, and for a pair $f, g$ of escort distributions, the following holds

$$E_g \left[ \|X\|^\alpha \right]^{1/\alpha} I_{\beta,q}[g]^{1/\beta} \geq n,$$

with $I_{\beta,q}[g] = (q/M_q[g])^\beta \ E_g \left[ g(x)^{\beta(q-1)} \left\| \nabla_x g(x) \right\|_*^{\beta} \right],$

with equality if and only if $g(x)$ is a generalized $q$-Gaussian, i.e.

$$g(x) \propto \left(1 - \gamma \|x\|^\alpha\right)^{\frac{1}{q-1}}, \text{ with } \gamma > 0.$$

Also: $\text{Cov}(X) \geq l_{2,q}(X)^{-1}$ with equality iff $X$ is a $q$-Gaussian

- This generalizes the well-known fact that the Gaussian with a given variance minimizes Fisher information
- Complements the entropic characterization of generalized $q$-Gaussians
Fisher Information Inequality

Standard proofs of the FII (eg Stam 1959 or Rioul 2011)

\[ I_{2,1}[f_1 * f_2]^{-1} \geq I_{2,1}[f_1]^{-1} + I_{2,1}[f_2]^{-1} \]

use two basic ingredients

1. monotonicity and
2. additivity.
Fisher Information Inequality

Standard proofs of the FII (e.g., Stam 1959 or Rioul 2011)

\[ l_{2,1}[f_1 * f_2]^{-1} \geq l_{2,1}[f_1]^{-1} + l_{2,1}[f_2]^{-1} \]

use two basic ingredients

1. monotonicity and
2. additivity.

However, the additivity of Fisher information does not hold in the extended context. In the case of a pair of escorts, we still have (by Minkovski inequality)

\[
I_{\beta,q}[f_X f_Y]^\frac{1}{\beta} \leq \frac{M_\beta(q-1) + 1}{M_q[f_Y]} I_{\beta,q}[f_X]^{\frac{1}{\beta}} + \frac{M_\beta(q-1) + 1}{M_q[f_X]} I_{\beta,q}[f_Y]^{\frac{1}{\beta}},
\]

and an equality in the quadratic case.

\[
l_{2,1}[f_X f_Y] = M_{2q-1}[f_Y] l_{2,1}[f_Y] + M_{2q-1}[f_X] l_{2,1}[f_X].
\]
Following (Zamir, 1998) and using only the monotonicity property for the Fisher Information Matrix, we have also a matrix form of the FII:

\[
\begin{align*}
    A^t J_g(AX) A & \leq J_g(X) \\
    J_g(AX) & \leq (AJ_g(X)^{-1} A^t)^{-1}
\end{align*}
\]

Equality occurs if \( m = n \) or if \( \psi_g(X) = J_g(X) X \). In the case of a pair of escort distributions, equality occurs if \( g \) is a \( q \)-Gaussian distribution with covariance matrix \( J_g(X)^{-1} \).

**Proof.**

\( \text{Cov}_g(X) \geq J_g(X)^{-1} \) holds with equality iff \( X \) is a \( q \)-Gaussian. If \( X \) is \( q \)-Gaussian distributed, so is \( AX \), respectively with covariances \( \text{Cov}_g(X) = J_g(X)^{-1} \) and \( \text{Cov}_g(AX) = J_g(AX)^{-1} \). Therefore, we have

\[
\text{Cov}_g(AX) = J_g(AX)^{-1} \geq AJ_g(X)^{-1} A^t = A\text{Cov}_g(X)A^t.
\]

Since \( \text{Cov}_g(AX) = A\text{Cov}_g(X)A^t \), we see that equality must hold.
FII and case of equality - If \((f, g)\) is a pair of escorts, for any random vector \(X^T = [U_1 \ldots U_n]\) with generalized Fisher information matrix \(J_g(X)\), we have

\[
    I_{2,q}\left(\sum_{i=1}^{n} U_i\right)^{-1} \geq 1^T J_g(X)^{-1} 1.
\]

The lower bound is attained if \(X\) is a \(q\)-Gaussian vector with covariance matrix \(J_g(X)^{-1}\), i.e.

\[
    g(x) \propto \left(1 - \gamma(q - 1)x^T J_g(X)^{-1} x\right)^{\frac{1}{q-1}}
\]

For independent random variables \(U_i\) with respective distributions \(f_i\), it holds that

\[
    I_{2,q}\left(\sum_{i=1}^{n} U_i\right)^{-1} \geq \sum_{i=1}^{n} \prod_{k \neq i} \frac{M_q[f_i]^2}{M_{2q-1}[f_i]} I_{2,q}(U_i)^{-1}.
\]

On generalized Fisher information, etc, etc...
The \( q \)-Entropy-\( q \)Fisher-\( q \)Gauss setting

The classical Shannon-Fisher-Gauss setting can be extended to \( q \)-entropies, higher moments, and a generalized Fisher info

\[ \text{Maximum entropy: (Lutwak et al 2005)} \]

\[ m_{\alpha}^{\left[ f \right]} \leq m_{\alpha}^{\left[ G \right]} \]

Extended de Bruijn's identity (nonlinear diffusion equation)

\[ \frac{d}{dt} S_{q}^{\left[ f \right]} = \left( m_{q}^{\left[ f \right]} \right)^{\beta - 1} M_{q}^{\left[ f \right]} I_{\beta, q}^{\left[ f \right]} \]

Generalized Stam's inequality

\[ I_{\beta, q}^{\left[ f \right]}^{1/\beta} \leq I_{\beta, q}^{\left[ G \right]}^{1/\beta} \]

(*) jfb - Some properties of generalized Fisher information in the context of nonextensive thermostatistics, Physica A, august 2013
The \( q \)-Entropy-\( q \)Fisher-\( q \)Gauss setting

The classical Shannon-Fisher-Gauss setting can be extended to \( q \)-entropies, higher moments, and a generalized Fisher info

1. Maximum entropy: (Lutwak et al 2005)

\[
\frac{m_\alpha[f]^{\frac{1}{2}}}{N_q[f]} \geq \frac{m_\alpha[G]^\frac{1}{2}}{N_q[G]}
\]

2. Extended de Bruijn’s identity (nonlinear diffusion equation) (*)

\[
\frac{d}{dt} S_q[f] = \left(\frac{m}{q}\right)^{\beta-1} M_q[f]^{\beta} I_{\beta,q}[f]
\]

3. Generalized Stam’s inequality (*)

\[
I_{\beta,q}[f]^{\frac{1}{\beta}} N_q[f]^{\frac{1}{2}} \geq I_{\beta,q}[G]^{\frac{1}{\beta}} N_q[G]^{\frac{1}{2}}
\]

(*) jfb - Some properties of generalized Fisher information in the context of nonextensive thermostatistics, Physica A, August 2013
The $q$-setting (continued)

4. Generalized Cramér-Rao inequality

$$E_g \left[ \|X\|^{\alpha} \right]^{\frac{1}{\alpha}} l_{\beta,q}[g]^{\frac{1}{\beta}} \geq n$$

5. Fisher Information Inequality

$$l_{2,q}(X + Y)^{-1} \geq l_{2,q}(X)^{-1} + l_{2,q}(Y)^{-1}$$
Remarks, further and future work

- We got new Cramér-Rao inequalities in the estimation setting

\[ N_q(X + Y) \geq c_q \left( N_q(X) + N_q(Y) \right) \quad (q \geq 1) \]

has been proved very recently and the subject is very hot (Bobkov, S. G.; Chistyakov – submitted in May 2013, also Wang, L; Madiman, M. – July 2013)

In the usual case de Bruijn + FII → EPI. No such proof here, but there are ongoing efforts on generalized Fisher information, etc, etc...
Remarks, further and future work

- We got new Cramér-Rao inequalities in the estimation setting
- Which also suggests new robust estimation methods
Remarks, further and future work

- We got new Cramér-Rao inequalities in the estimation setting
- Which also suggests new robust estimation methods
- Future: give/study numerical examples, role of escort distributions

Entropic Power Inequality

\[ N_q(X + Y) \geq c N_q(N_q(X)) + N_q(Y) \] (\(q \geq 1\))

has been proved very recently and the subject is very hot (Bobkov, S. G.; Chistyakov – submitted in May 2013, also Wang, L; Madiman, M. – July 2013)
Remarks, further and future work

- We got new Cramér-Rao inequalities in the estimation setting
- Which also suggests new robust estimation methods
- Future: give/study numerical examples, role of escort distributions
- It is possible to get even more general CR inequalities for Young cost functions
Remarks, further and future work

- We got new Cramér-Rao inequalities in the estimation setting
- Which also suggests new robust estimation methods
- Future: give/study numerical examples, role of escort distributions
- It is possible to get even more general CR inequalities for Young cost functions
- Entropy Power Inequality

\[ N_q(X + Y) \geq c_q (N_q(X) + N_q(Y)) \quad (q \geq 1) \]

has been proved very recently and the subject is very hot (Bobkov, S. G.; Chistyakov – submitted in May 2013, also Wang, L; Madiman, M. – July 2013)

In the usual case de Bruijn + FII \(\rightarrow\) EPI. No such proof here, but there are ongoing efforts