Cramér-Rao and moment-entropy inequalities for Renyi entropy and generalized Fisher information

Erwin Lutwak, Deane Yang, and Gaoyong Zhang

Abstract— **The moment-entropy inequality shows that a continuous random variable with given second moment and maximal Shannon entropy must be Gaussian. Stam's inequality shows that a continuous random variable with given Fisher information and minimal Shannon entropy must also be Gaussian. The Cramér-Rao inequality is a direct consequence of these two inequalities.**

In this paper the inequalities above are extended to Renyi entropy, p-th moment, and generalized Fisher information. Generalized Gaussian random densities are introduced and shown to be the extremal densities for the new inequalities. An extension of the Cramér–Rao inequality is derived as a consequence of these moment and Fisher information inequalities.

Index Terms— **entropy, Renyi entropy, moment, Fisher information, information theory, information measure**

I. Introduction

THE moment-entropy inequality shows that a continuous random variable with given second moment and maximal Shannon entropy must be Gaussian (see, for example, Theorem 9.6.5 in [1]). This follows from the nonnegativity of the relative entropy of two continuous random variables. In this paper we introduce the notion of relative Renyi entropy for two random variables and show that it is always nonnegative. We identify the probability distributions that have maximal Renyi entropy with given p–th moment and call them generalized Gaussians.

In his proof of the Shannon entropy power inequality Stam [2] shows that a continuous random variable with given Fisher information and minimal Shannon entropy must be Gaussian. We introduce below a generalized form of Fisher information associated with Renyi entropy and that is, in some sense, dual to the p–th moment. A generalization of Stam's inequality is established. The probability distributions that have maximal Renyi entropy with given generalized Fisher information are the generalized Gaussians.

The Cramér-Rao inequality (see, for example, Theorem 12.11.1 in [1]) states that the second moment of a continuous random variable is bounded from below by the reciprocal of its Fisher information. We use the moment and Fisher information inequalities to establish a generalization of the Cramér–Rao inequality, where a lower bound is obtained for the p-th moment of a continuous random variable in terms of its generalized Fisher information. Again, the generalized Gaussians are the extremal distributions.

E. Lutwak (elutwak@poly.edu), D. Yang (dyang@poly.edu), and G. Zhang (gzhang@poly.edu) are with the Department of Mathematics, Polytechnic University, Brooklyn, New York and were supported in part by NSF Grant DMS-0104363.

Analogues for convex and star bodies of the moment-entropy, Fisher information-entropy, and Cramér-Rao inequalities had been established earlier by the authors [3], [4], [5], [6], [7]

II. Definitions

Throughout this paper, unless otherwise indicated, all integrals are with respect to Lebesgue measure over the real line $\mathbb{R}$. All densities are probability densities on $\mathbb{R}$.

A. Entropy

The *Shannon entropy* of a density f is defined to be

$$h[f] = -\int_{\mathbb{R}} f \log f, \tag{1}$$

provided that the integral above exists. For $\lambda > 0$ the λ-*Renyi entropy power* of a density is defined to be

$$N_\lambda[f] = \begin{cases} \left(\int_{\mathbb{R}} f^\lambda\right)^{\frac{1}{1-\lambda}} & \text{if } \lambda \neq 1, \\ e^{h[f]} & \text{if } \lambda = 1, \end{cases} \tag{2}$$

provided that the integral above exists. Observe that

$$\lim_{\lambda \to 1} N_\lambda[f] = N_1[f].$$

The λ–*Renyi entropy* of a density f is defined to be

$$h_\lambda[f] = \log N_\lambda[f].$$

The entropy $h_\lambda[f]$ is continuous in λ and, by the Hölder inequality, decreasing in λ. It is strictly decreasing, unless f is a uniform density.

B. Relative entropy

Given two densities $f, g : \mathbb{R} \to \mathbb{R}$, their *relative Shannon entropy* or *Kullback–Leibler distance* [11], [12], [13] (also, see page 231 in [1]) is defined by

$$h_1[f, g] = \int_{\mathbb{R}} f \log\left(\frac{f}{g}\right), \tag{3}$$

provided that the integral above exists. Given $\lambda > 0$ and two densities f and g, we define the *relative λ–Renyi entropy power of f and g* as follows. If $\lambda \neq 1$, then

$$N_\lambda[f, g] = \frac{\left(\int_{\mathbb{R}} g^{\lambda-1} f\right)^{\frac{1}{1-\lambda}} \left(\int_{\mathbb{R}} g^\lambda\right)^{\frac{1}{\lambda}}}{\left(\int_{\mathbb{R}} f^\lambda\right)^{\frac{1}{\lambda(1-\lambda)}}}, \tag{4}$$

and, if $\lambda = 1$, then

$$N_1[f,g] = e^{h_1[f,g]},$$

provided in both cases that the righthand side exists. Define the λ–*Renyi relative entropy of f and g* by

$$h_\lambda[f,g] = \log N_\lambda[f,g].$$

Observe that $h_\lambda[f,g]$ is continuous in λ.

Lemma 1: If f and g are densities such that $h_\lambda[f]$, $h_\lambda[g]$, and $h_\lambda[f,g]$ are finite, then

$$h_\lambda[f,g] \geq 0.$$

Equality holds if and only if $f = g$.

Proof: The case $\lambda = 1$ is well–known (see, for example, page 234 in [1]). The remaining cases are a direct consequence of the Hölder inequality. If $\lambda > 1$, then we have

$$\int_{\mathbb{R}} g^{\lambda-1} f \leq \left(\int_{\mathbb{R}} g^\lambda\right)^{\frac{\lambda-1}{\lambda}} \left(\int_{\mathbb{R}} f^\lambda\right)^{\frac{1}{\lambda}},$$

and if $\lambda < 1$, then we have

$$\begin{aligned}\int_{\mathbb{R}} f^\lambda &= \int_{\mathbb{R}} (g^{\lambda-1} f)^\lambda g^{\lambda(1-\lambda)} \\ &\leq \left(\int_{\mathbb{R}} g^{\lambda-1} f\right)^\lambda \left(\int_{\mathbb{R}} g^\lambda\right)^{1-\lambda}.\end{aligned}$$

The equality conditions follow from the equality conditions of the Hölder inequality. ■

C. The p-th moment

For $p \in (0,\infty)$ define *p-th moment* of a density f to be

$$\mu_p[f] = \int_{\mathbb{R}} |x|^p f(x)\, dx, \tag{5}$$

provided that the integral above exists. For $p \in [0,\infty]$ define the *p-th deviation* by

$$\sigma_p[f] = \begin{cases} \exp\left(\int_{\mathbb{R}} f(x) \log|x|\, dx\right) & \text{if } p = 0, \\ (\mu_p[f])^{\frac{1}{p}} & \text{if } 0 < p < \infty, \\ \operatorname{ess\,sup}\{|x| \;:\; f(x) > 0\} & \text{if } p = \infty, \end{cases} \tag{6}$$

provided in each case that the right side is finite. The deviation $\sigma_p[f]$ is continuous in p and, by the Hölder inequality, strictly increasing in p.

D. The (p,λ)-th Fisher information

Recall that the classical Fisher information [14], [15], [16] of a density $f : \mathbb{R} \to \mathbb{R}$ is given by

$$\phi_{2,1}[f]^2 = \int_{\mathbb{R}} f^{-1} |f'|^2,$$

provided f is absolutely continuous, and the integral exists. If $p \in [1,\infty]$ and $\lambda \in \mathbb{R}$, we denote the *(p,λ)-th Fisher information* of a density f by $\phi_{p,\lambda}[f]$ and define it as follows. If $p \in (1,\infty)$, let $q \in [1,\infty]$ satisfy $p^{-1} + q^{-1} = 1$, and define

$$\phi_{p,\lambda}[f]^{q\lambda} = \int_{\mathbb{R}} |f^{\lambda-2} f'|^q f, \tag{7}$$

provided that f is absolutely continuous, and the norm above is finite. If $p = 1$, then $\phi_{p,\lambda}[f]^\lambda$ is defined to be the essential supremum of $|f^{\lambda-2} f'|$ on the support of f, provided f is absolutely continuous, and the essential supremum is finite. If $p = \infty$, then $\phi_{p,\lambda}[f]^\lambda$ is defined to be the total variation of f^λ/λ, provided that f^λ has bounded variation. (see, for example, [17] for a definition of "bounded variation").

Note that our definition of generalized Fisher information has a different normalization than the standard definition. In particular, the classical Fisher information corresponds to the square of $(2,1)$-th Fisher information, as defined above.

The Fisher information $\phi_{p,\lambda}[f]$ is continuous in (p,λ). For a given λ it is, by the Hölder inequality, decreasing in p.

E. Generalized Gaussian densities

Given $t \in \mathbb{R}$, let

$$t_+ = \max\{t, 0\}.$$

Let

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\, dx$$

denote the Gamma function, and let

$$\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

denote the Beta function.

For each $p \in [0,\infty]$ and $\lambda > 1 - p$, we define the corresponding *generalized Gaussian density* $G : \mathbb{R} \to [0,\infty)$ as follows. If $p \in (0,\infty)$, then G is defined by

$$G(x) = \begin{cases} a_{p,\lambda}(1 + (1-\lambda)|x|^p)_+^{\frac{1}{\lambda-1}} & \text{if } \lambda \neq 1, \\ a_{p,1} e^{-|x|^p} & \text{if } \lambda = 1, \end{cases} \tag{8}$$

where

$$a_{p,\lambda} = \begin{cases} \dfrac{p(1-\lambda)^{\frac{1}{p}}}{2\beta(\frac{1}{p}, \frac{1}{1-\lambda} - \frac{1}{p})} & \text{if } \lambda < 1, \\ \dfrac{p}{2\Gamma(\frac{1}{p})} & \text{if } \lambda = 1, \\ \dfrac{p(\lambda-1)^{\frac{1}{p}}}{2\beta(\frac{1}{p}, \frac{\lambda}{\lambda-1})} & \text{if } \lambda > 1. \end{cases}$$

If $p = 0$ and $\lambda > 1$, then G is defined for almost every $x \in \mathbb{R}$ by

$$G(x) = a_{0,\lambda}(-\log|x|)_+^{\frac{1}{\lambda-1}},$$

where

$$a_{0,\lambda} = \frac{1}{2\Gamma(\frac{\lambda}{\lambda-1})}.$$

If $p = \infty$ and $\lambda > 0$, then G is defined by

$$G(x) = \begin{cases} \dfrac{1}{2} & \text{if } |x| \leq 1 \\ 0 & \text{if } |x| > 1. \end{cases}$$

For consistency we shall also denote $a_{\infty,\lambda} = \frac{1}{2}$.

For $t > 0$, define $G_t : \mathbb{R} \to [0, \infty)$ by

$$G_t(x) = G(x/t)/t. \tag{9}$$

Sz. Nagy [8] established a family of sharp Gagliardo-Nirenberg inequalities on $\mathbb{R}$ and their equality conditions. His results can be used to prove Theorem 3 and identify the generalized Gaussians as the extremal densities for the inequalities proved in this paper. Later, Barenblatt [9] showed that the generalized Gaussians are also the self–similar solutions of the L^p porous media and fast diffusion equations. Generalized Gaussians are also the 1-dimensional versions of the extremal functions for sharp Sobolev, log-Sobolev, and Gagliardo-Nirenberg inequalities (see, for example, [10]).

F. *Information measures of generalized Gaussians*

If $0 < p < \infty$ and $\lambda > 1/(1+p)$, the λ–Renyi entropy power of the generalized Gaussian G as given by (8) is given by

$$N_\lambda[G] = \begin{cases} \left(\dfrac{p\lambda}{p\lambda + \lambda - 1}\right)^{\frac{1}{1-\lambda}} a_{p,\lambda}^{-1} & \text{if } \lambda \neq 1 \\ e^{\frac{1}{p}} a_{p,1}^{-1} & \text{if } \lambda = 1 \end{cases}$$

If $p = 0$ and $\lambda > 1$, then

$$N_\lambda[G] = \left(\frac{\lambda}{\lambda - 1}\right)^{\frac{1}{1-\lambda}} a_{0,\lambda}^{-1}.$$

If $p = \infty$ and $\lambda > 0$, then

$$N_\lambda[G] = 2. \tag{10}$$

If $0 < p < \infty$ and $\lambda > 1/(1+p)$, then the p–th deviation of G is given by

$$\sigma_p[G] = (p\lambda + \lambda - 1)^{-\frac{1}{p}}.$$

If $p = 0$ and $\lambda > 1$, then

$$\sigma_0[G] = e^{-\frac{\lambda}{\lambda-1}}.$$

If $p = \infty$, then

$$\sigma_\infty[G] = 1.$$

If $1 \le p \le \infty$ and $\lambda > 1/(1+p)$, then the (p, λ)-th Fisher information of the generalized Gaussian G is given by

$$\phi_{p,\lambda}[G] = \begin{cases} p^{1/\lambda} a_{p,\lambda}^{(\lambda-1)/\lambda} (p\lambda + \lambda - 1)^{-(1-\frac{1}{p})/\lambda} & \text{if } p < \infty, \\ 2^{(1-\lambda)/\lambda}/\lambda^{1/\lambda} & \text{if } p = \infty. \end{cases}$$

In particular, observe that if $1 \le p \le \infty$ and $\lambda > 1/(1+p)$, then

$$N_\lambda[G]^{1-\lambda} = \lambda \sigma_p[G] \phi_{p,\lambda}[G]^\lambda. \tag{11}$$

Observe that if $\lambda \neq 1$, then

$$\int_{\mathbb{R}} G^\lambda = a_{p,\lambda}^{\lambda-1}(1 + (1-\lambda)\mu_p[G]), \tag{12}$$

and if $\lambda = 1$, then

$$h[G] = -\log a_{p,1} + \mu_p[G]. \tag{13}$$

We will also need the following simple scaling identities:

$$\int_{\mathbb{R}} G_t^\lambda = t^{1-\lambda} \int_{\mathbb{R}} G^\lambda, \tag{14}$$

and

$$\sigma_p[G_t] = t\sigma_p[G]. \tag{15}$$

III. The moment inequality

It is well known that among all probability distributions with given second moment, the Gaussian is the unique distribution that maximizes the Shannon entropy. This follows from the positivity of the relative entropy of a given distribution and a Gaussian distribution of the same variance. This result is generalized to p-th moments in Chapter 11 of [1].

We show that a similar inequality for the p-th moment and λ–Renyi entropy follows from the positivity of the λ–Renyi relative entropy of a given distribution and the appropriate extremal distribution with the same p-th moment.

Theorem 2: Let $f : \mathbb{R} \to \mathbb{R}$ be a density. If $p \in [0, \infty]$, $\lambda > 1/(1+p)$, and $N_\lambda[f], \sigma_p[f] < \infty$, then

$$\frac{\sigma_p[f]}{N_\lambda[f]} \ge \frac{\sigma_p[G]}{N_\lambda[G]}, \tag{16}$$

where G is given by (8). Equality holds if and only if $f = G_t$ for some $t \in (0, \infty)$.

Proof: For convenience let $a = a_{p,\lambda}$. Let

$$t = \frac{\sigma_p[f]}{\sigma_p[G]}. \tag{17}$$

First, consider the case $\lambda \neq 1$. If $p \in (0, \infty)$, then by (8) and (9), (5), (17), and (12),

$$\begin{aligned} &\int_{\mathbb{R}} G_t^{\lambda-1} f \\ &\ge a^{\lambda-1} t^{1-\lambda} + (1-\lambda) a^{\lambda-1} t^{1-\lambda-p} \int_{\mathbb{R}} |x|^p f(x)\, dx \\ &= a^{\lambda-1} t^{1-\lambda} (1 + (1-\lambda) t^{-p} \mu_p[f]) \\ &= a^{\lambda-1} t^{1-\lambda} (1 + (1-\lambda) \mu_p[G]) \\ &= t^{1-\lambda} \int_{\mathbb{R}} G^\lambda, \end{aligned} \tag{18}$$

where equality holds if $\lambda < 1$. For $p = \infty$ observe that f vanishes outside the interval $[-t, t]$ and therefore by (8) and (9) and (6),

$$\begin{aligned} \int_{\mathbb{R}} G_t^{\lambda-1} f &= a^{\lambda-1} t^{-\lambda+1} \int_{-t}^{t} f(x)\, dx \\ &= a^{\lambda-1} t^{-\lambda+1} \\ &= t^{-\lambda+1} \int_{\mathbb{R}} G^\lambda. \end{aligned} \tag{19}$$

It follows that if $p \in (0, \infty]$ and $\lambda \neq 1$, then by Lemma 1, (4), (18), (19), and (14), and (17), we have

$$\begin{aligned}
1 &\le N_\lambda[f, G_t]^\lambda \\
&= \left(\int_{\mathbb{R}} G_t^\lambda\right)\left(\int_{\mathbb{R}} f^\lambda\right)^{-\frac{1}{1-\lambda}}\left(\int_{\mathbb{R}} G_t^{\lambda-1} f\right)^{\frac{\lambda}{1-\lambda}} \\
&\le t\frac{N_\lambda[G]}{N_\lambda[f]} \\
&= \frac{\sigma_p[f]}{N_\lambda[f]}\frac{N_\lambda[G]}{\sigma_p[G]}.
\end{aligned} \tag{20}$$

If $\lambda = 1$ and $p \in (0, \infty)$, then by Lemma 1, (3), (8), and (9), and (17) and (13), we have

$$\begin{aligned}
0 &\le h_1[f, G_t] \\
&= -h[f] - \log a + \log t + t^{-p}\mu_p[f] \\
&= h[G] - h[f] + \log \sigma_p[f] - \log \sigma_p[G].
\end{aligned}$$

If $\lambda = 1$ and $p = \infty$, then by Lemma 1, (3), (1), (8) and (9), and (17), (1), and (6),

$$\begin{aligned}
0 &\le h_1[f, G_t] \\
&= \int_{\mathbb{R}} f \log f - \int_{\mathbb{R}} f \log G_t \\
&= -h[f] - \log a + \log t \\
&= -h[f] + h[G] + \log \sigma_\infty[f] - \log \sigma_\infty[G].
\end{aligned}$$

This gives inequality (16) for $p = \infty$.

If $p = 0$ and $\lambda > 1$, then from (8) and (6), we have

$$\int_{\mathbb{R}} G^\lambda = -a^{\lambda-1} \log \sigma_0[G]. \tag{21}$$

Therefore, by (8) and (9), (6), (17), and (21),

$$\begin{aligned}
&\int_{\mathbb{R}} G_t^{\lambda-1} f \\
&\ge a^{\lambda-1} t^{-\lambda+1} \int_{\mathbb{R}} (\log t - \log |x|) f(x)\, dx \\
&= a^{\lambda-1} t^{-\lambda+1} (\log t - \log \sigma_0[f]) \\
&= -t^{-\lambda+1} a^{\lambda-1} \log \sigma_0[G] \\
&= t^{-\lambda+1} \int_{\mathbb{R}} G^\lambda.
\end{aligned}$$

The inequality for $p = 0$ and $\lambda > 1$ now follows from (20).

In all cases, Lemma 1 shows that equality holds if and only if $f = G_t$. ∎

A higher dimensional version of Theorem 2 was established by the authors in [7]. The case $p = 2$ of Theorem 2 was also established independently by Costa, Hero, and Vignat [18].

It is also worth noting that Arikan [19] obtains a moment-entropy inequality for discrete random variables analogous to Theorem 2. His inequality, however, is for the limiting case $\lambda = 1/(1 + p)$, where Theorem 2 does not apply.

IV. The Fisher information inequality

Stam's inequality [2] shows that among all probability distributions with given Fisher information, the unique distribution that minimizes Shannon entropy is Gaussian. The following theorem extends this fact to λ–Renyi entropy and (p, λ)-th Fisher information.

Theorem 3: Let $p \in [1, \infty]$, $\lambda \in (1/(1 + p), \infty)$, and $f : \mathbb{R} \to [0, \infty)$ be a density. If $p < \infty$, then f is assumed to be absolutely continuous; if $p = \infty$, then f^λ is assumed to have bounded variation. If $N_\lambda[f], \phi_{p,\lambda}[f] < \infty$, then

$$\phi_{p,\lambda}[f] N_\lambda[f] \ge \phi_{p,\lambda}[G] N_\lambda[G], \tag{22}$$

where G is the generalized Gaussian. Equality holds if and only if there exist $t > 0$ and $x_0 \in \mathbb{R}$ such that $f(x) = G_t(x - x_0)$, for all $x \in \mathbb{R}$.

As mentioned earlier, Theorem 3, including its equality conditions, follow from sharp analytic inequalities established by Sz. Nagy [8]. Inequality (22) complements the sharp Gagliardo-Nirenberg inequalities on $\mathbb{R}^n$, with $n \ge 2$ and $n/(n - 1) < p \le \infty$, established by Del Pino and Dolbeault [10] and generalized by Cordero, Nazaret, and Villani [20]. The proof presented here is inspired by the beautiful mass transportation proof of Cordero et al. Observe, however, that there is no overlap between their inequalities and ours.

Before giving the proof of this theorem, we need a change of random variable formula and a lemma on integration by parts.

A. Change of random variable

Let X be a random variable with density f. Let the support of f be contained in an interval (S, T). Given an increasing absolutely continuous function $y : (S, T) \to \mathbb{R}$, the random variable $Y = y(X)$ has density g, where

$$f(x) = g(y(x))y'(x),$$

for almost every x, and $g(z) = 0$, for each $z \in \mathbb{R}\backslash y((S, T))$. Therefore, if $N_\lambda[g] < \infty$, then

$$N_\lambda[g] = \begin{cases} \left(\displaystyle\int_S^T f^\lambda (y')^{1-\lambda}\right)^{\frac{1}{1-\lambda}} & \text{if } \lambda \ne 1, \\ e^{h[g]} & \text{if } \lambda = 1, \end{cases} \tag{23}$$

where

$$h[g] = h[f] + \int_S^T f(x) \log y'(x)\, dx. \tag{24}$$

Similarly, if the p-th moment of g is finite, then it is given by

$$\mu_p[g] = \int_S^T |y(x)|^p f(x)\, dx. \tag{25}$$

B. Integration by parts

Lemma 4: Let $S, T \in [-\infty, \infty]$ and $f : (S, T) \to \mathbb{R}$ be an absolutely continuous function such that

$$\lim_{x\to S} f(x) = \lim_{x\to T} f(x) = 0. \tag{26}$$

Let $g : (S, T) \to \mathbb{R}$ be an increasing absolutely continuous function such that

$$\lim_{t\to T} g(t) > 0,$$

and the integral

$$\int_S^T f' g$$

is absolutely convergent. Then

$$\int_S^T fg' = -\int_S^T f'g.$$

Proof: It suffices to prove

$$\lim_{s\to S} f(s)g(s) = \lim_{t\to T} f(t)g(t) = 0.$$

The same proof works for both limits, so we will show only that the right limit vanishes.

$$\begin{aligned}
0 &= \lim_{t\to T}\int_t^T |f'(x)g(x)|\,dx \\
&\geq \lim_{t\to T} |g(t)| \int_t^T |f'(x)|\,dx \\
&\geq \lim_{t\to T} |g(t)| \left|\int_t^T f'(x)\,dx\right| \\
&= \lim_{t\to T} |g(t)f(t)|.
\end{aligned}$$

■

C. Proof of Theorem 3

Let g be a density that is supported on an open interval $(-R,R)$ for some $R\in(0,\infty]$. Let $S,T\in[-\infty,\infty]$ be such that (S,T) is the smallest interval containing the support of f. Define $y:(S,T)\to(-R,R)$ so that for each $x\in(S,T)$,

$$\int_S^x f(s)\,ds = \int_{-R}^{y(x)} g(t)\,dt.$$

Observe that if X is a random variable with density f, then the random variable $Y=y(X)$ has density g.

If $\lambda\neq 1$ and $p<\infty$, then by (2) and (23), Hölder's inequality, Lemma 4, Hölder's inequality again, and (6) and (7), we have

$$\begin{aligned}
&N_\lambda[f]^{-\lambda}N_\lambda[g] \\
&= \left(\int_S^T f^\lambda\right)^{-\frac{\lambda}{1-\lambda}} \left(\int_S^T f^\lambda (y')^{1-\lambda}\right)^{\frac{1}{1-\lambda}} \\
&\leq \int_S^T f^\lambda y' \\
&= -\int_S^T (f^\lambda)' y \\
&= -\lambda\int_S^T (yf^{1/p})(f^{\lambda-1-1/p}f') \\
&\leq \lambda\left(\int_{\mathbb{R}} |y|^p f\right)^{1/p}\left(\int_{\mathbb{R}} |f^{\lambda-1-1/p}f'|^q\right)^{1/q} \\
&\leq \lambda\sigma_p[g]\phi_{p,\lambda}[f]^\lambda,
\end{aligned} \tag{27}$$

where q is the Hölder conjugate of p.

If $\lambda=1$ and $p<\infty$, then by (24), Jensen's inequality, Lemma 4, Hölder's inequality, and (6) and (7), we have

$$\begin{aligned}
h[g] &= h[f] + \int_S^T f\log y' \\
&\leq h[f] + \log\int_S^T fy' \\
&= h[f] + \log\int_S^T -f'y \\
&\leq h[f] + \log\left(\int_{\mathbb{R}} |(\log f)'|^q f\right)^{\frac{1}{q}}\left(\int_{\mathbb{R}} |y|^p f\right)^{\frac{1}{p}} \\
&= h[f] + \log\phi_{p,1}[f]\sigma_p[g],
\end{aligned} \tag{28}$$

where q is the Hölder conjugate of p.

By the equality conditions of the Hölder inequality, equality holds for (27) and (28), only if there exist for $c_1,c_2,x_0\in\mathbb{R}$ such that $y=c_1(x-x_0)$, and f satisfies the differential equation

$$(f(x)^\lambda)' = c_2|x-x_0|^{p-2}(x-x_0)f(x).$$

This, in turn, implies that there exist $\tau,t>0$ and $x_0\in\mathbb{R}$ such that $g=G_\tau$ and $f(x)=G_t(x-x_0)$, for all $x\in\mathbb{R}$. On the other hand, by (11) equality always holds for (22) if $f=G_t$.

If $p=\infty$, let g be compactly supported on the interval $(-R,R)$ with $R<\infty$, and extend the domain of y to the entire real line by setting $y(x)=-R$ for all $x\in(-\infty,S]$ and $y(x)=R$ for all $x\in[T,\infty)$. Following the same line of reasoning as (27), we get

$$\begin{aligned}
N_\lambda[f]^{-\lambda}N_\lambda[g] &\leq -\int_S^T (f^\lambda)'y \\
&\leq R\int_{\mathbb{R}} |(f^\lambda)'| \\
&= \lambda\sigma_\infty[g]\phi_{\infty,\lambda}[f]^\lambda.
\end{aligned}$$

Equality holds if and only if there exist $c_1,c_2,x_0\in\mathbb{R}$ such that $y=c_1(x-x_0)$, and $|y|$ is constant on the support of $(f^\lambda)'$. This is possible only if $S,T<\infty$, and f is a uniform density for the interval $[S,T]$. In other words, $f=G_t$, for some $t\in(0,\infty)$.

V. The Cramér–Rao inequality

The following theorem generalizes the classical Cramér–Rao inequality [21], [22] (also, see Theorem 12.11.1 in [1]).

Theorem 5: Let $p\in[1,\infty]$, $\lambda\in(1/(1+p),\infty)$, and f be a density. If $p<\infty$, then f is assumed to be absolutely continuous; if $p=\infty$, then f^λ is assumed to have bounded variation. If $\sigma_p[f],\phi_{p,\lambda}[f]<\infty$, then

$$\sigma_p[f]\phi_{p,\lambda}[f] \geq \sigma_p[G]\phi_{p,\lambda}[G].$$

Equality holds if and only if $f=G_t$, for some $t>0$.

The inequality is a direct consequence of (16) and (22).

VI. Inequalities for Shannon and quadratic entropy

The case $p=1$ and $\lambda=1$ of these theorems give the following.

Corollary 6: If $f : \mathbb{R} \to \mathbb{R}$ is an absolutely continuous density with finite Shannon entropy, first moment, and $(1, 1)$-th Fisher information, then

$$(\sup|(\log f)'|)^{-1} \le \frac{N_1[f]}{2e} \le \int_{\mathbb{R}} |x| f(x)\, dx.$$

Equality holds for the first inequality if and only if there exist $t > 0$ and $x_0 \in \mathbb{R}$ such that

$$f(x) = \frac{1}{2t} e^{-|x-x_0|/t}, \tag{29}$$

for all $x \in \mathbb{R}$. Equality holds for the second inequality if and only if there exists $t > 0$ such that (29) holds with $x_0 = 0$.

The cases $p = 1, 2$ and $\lambda = 2$ give the following inequalities for quadratic entropy.

Corollary 7: If $f : \mathbb{R} \to \mathbb{R}$ is an absolutely continuous density with finite 2-Renyi entropy, first moment, and $(1, 2)$-th Fisher information, then

$$\frac{2}{9}\left(\int_{\mathbb{R}} |x| f(x)\, dx\right)^{-1} \le \int_{\mathbb{R}} f^2 \le \frac{2}{3}\left(\sup|f'|\right)^{1/2}.$$

Equality holds for the left inequality if and only if there exist $t > 0$ and $x_0 \in \mathbb{R}$ such that

$$f(x) = (1 - |x - x_0|/t)_+/t, \tag{30}$$

for all $x \in \mathbb{R}$. Equality holds for the right inequality if and only if there exists $t > 0$ such that (30) holds with $x_0 = 0$.

Corollary 8: If $f : \mathbb{R} \to \mathbb{R}$ is an absolutely continuous density with finite 2-Renyi entropy, second moment, and $(2, 2)$-th Fisher information, then

$$\frac{3}{5^{3/2}}\left(\int_{\mathbb{R}} x^2 f(x)\, dx\right)^{-1/2} \le \int_{\mathbb{R}} f^2 \le \frac{6^{1/2}}{5^{3/4}}\left(\int_{\mathbb{R}} (f')^2 f\right)^{1/4}$$

Equality holds for the left inequality if and only if there exist $t > 0$ and $x_0 \in \mathbb{R}$ such that

$$f(x) = (1 - |x - x_0|^2/t^2)_+/t, \tag{31}$$

for all $x \in \mathbb{R}$. Equality holds for the right inequality if and only if there exists $t > 0$ such that (31) holds with $x_0 = 0$.

Acknowledgments

The authors would like to express their gratitude to Professor Wenxiong Chen, the referees, and the Editor for their careful reading of an earlier draft of this paper and valuable suggestions for improving the paper.

References

[1] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: John Wiley & Sons Inc., 1991, a Wiley-Interscience Publication.

[2] A. J. Stam, “Some inequalities satisfied by the quantities of information of Fisher and Shannon,” *Information and Control*, vol. 2, pp. 101–112, 1959.

[3] E. Lutwak, D. Yang, and G. Zhang, “L_p affine isoperimetric inequalities,” *J. Differential Geom.*, vol. 56, no. 1, pp. 111–132, 2000.

[4] O. G. Guleryuz, E. Lutwak, D. Yang, and G. Zhang, “Information-theoretic inequalities for contoured probability distributions,” *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 2377–2383, 2002.

[5] E. Lutwak, D. Yang, and G. Zhang, “The Cramer-Rao inequality for star bodies,” *Duke Math. J.*, vol. 112, no. 1, pp. 59–81, 2002.

[6] ——, “Sharp affine L_p Sobolev inequalities,” *J. Differential Geom.*, vol. 62, no. 1, pp. 17–38, 2002.

[7] ——, “Moment–entropy inequalities,” *Annals of Probability*, vol. 32, pp. 757–774, 2004.

[8] B. v. Sz. Nagy, “Über Integralungleichungen zwischen einer Funktion und ihrer Ableitung,” *Acta Univ. Szeged. Sect. Sci. Math.*, vol. 10, pp. 64–74, 1941.

[9] G. I. Barenblatt, “On self-similar motions of a compressible fluid in a porous medium,” *Akad. Nauk SSSR. Prikl. Mat. Meh.*, vol. 16, pp. 679–698, 1952.

[10] M. Del Pino and J. Dolbeault, “Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions,” *J. Math. Pures Appl. (9)*, vol. 81, no. 9, pp. 847–875, 2002.

[11] S. Kullback and R. A. Leibler, “On information and sufficiency,” *Ann. Math. Statistics*, vol. 22, pp. 79–86, 1951.

[12] I. Csiszár, “Information-type measures of difference of probability distributions and indirect observations,” *Studia Sci. Math. Hungar.*, vol. 2, pp. 299–318, 1967.

[13] S.-i. Amari, *Differential-geometrical methods in statistics*, ser. Lecture Notes in Statistics. New York: Springer-Verlag, 1985, vol. 28.

[14] R. A. Fisher, “Theory of statistical estimation,” *Philos. Trans. Roy. Soc. London Ser A*, vol. 222, pp. 309–368, 1930.

[15] J. L. Doob, “Probability and statistics,” *Trans. Amer. Math. Soc.*, vol. 36, no. 4, pp. 759–775, 1934.

[16] J. M. Borwein, A. S. Lewis, and D. Noll, “Maximum entropy reconstruction using derivative information. I. Fisher information and convex duality,” *Math. Oper. Res.*, vol. 21, no. 2, pp. 442–468, 1996.

[17] W. Rudin, *Real and complex analysis*, 3rd ed. New York: McGraw-Hill Book Co., 1987.

[18] J. A. Costa, A. O. Hero, and C. Vignat, “A characterization of the multivariate distributions maximizing renyi entropy,” in *Proceedings of 2002 IEEE International Symposium on Information Theory*, 2002, p. 263.

[19] E. Arikan, “An inequality on guessing and its application to sequential decoding,” *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 99–105, 1996.

[20] D. Cordero-Erausquin, B. Nazaret, and C. Villani, “A mass-transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities,” *Adv. Math.*, vol. 182, no. 2, pp. 307–332, 2004.

[21] H. Cramér, *Mathematical methods of statistics*, ser. Princeton Landmarks in Mathematics. Princeton, NJ: Princeton University Press, 1999, reprint of the 1946 original.

[22] C. Radhakrishna Rao, “Information and the accuracy attainable in the estimation of statistical parameters,” *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–91, 1945.

Erwin Lutwak Erwin Lutwak received his B.S., M.S., and Ph.D. degrees in Mathematics from Polytechnic University, where he is now Professor of Mathematics.

Deane Yang Deane Yang received his B.A. in mathematics and physics from University of Pennsylvania and Ph.D. in mathematics from Harvard University. He has been an NSF Postdoctoral Fellow at the Courant Institute, a visiting faculty member of Columbia University and Université de Paris-Sud at Orsay. He has been a visiting member of l’Institut des Hautes Etudes Scientiques on numerous occasions. He was an associate professor at Rice University and is now a full professor at Polytechnic University.

Gaoyong Zhang Gaoyong Zhang received his B.S. degree in mathematics from Wuhan University of Science and Technology, M.S. degree in mathematics from Wuhan University, Wuhan, China, and Ph.D. degree in mathematics from Temple University, Philadelphia. He was a Rademacher Lecturer at the University of Pennsylvania, a member of the Institute for Advanced Study at Princeton, and a member of the Mathematical Sciences Research Institute at Berkeley. He was an assistant professor and is now a full professor at Polytechnic University.