Extensions of Fisher information and Stam's inequality

Erwin Lutwak, Songjun Lv, Deane Yang, and Gaoyong Zhang

***Abstract*—We explain how the classical notions of Fisher information of a random variable and Fisher information matrix of a random vector can be extended to a much broader setting. We also show that Stam's inequality for Fisher information and Shannon entropy, as well as the more generalized versions proved earlier by the authors, are all special cases of more general sharp inequalities satisfied by random vectors. The extremal random vectors, which we call generalized Gaussians, contain Gaussians as a limiting case but are noteworthy because they are heavy-tailed.**

I. Introduction

Two fundamental relationships among the entropy, second moment, and Fisher information of a continuous random variable are the following:

- (Moment-entropy inequality) A continuous random variable with given second moment has maximal Shannon entropy if and only if it is Gaussian (see, for example, Theorem 9.6.5 in the book of Cover and Thomas [1]).
- (Stam's inequality) A continuous random variable with given Fisher information has minimal Shannon entropy if and only if it is Gaussian (see Stam [2]).

A consequence of the inequalities underlying these facts is the Cramér-Rao inequality, which states that the second moment is always bounded from below by the reciprocal of the Fisher information.

The authors [3] showed that these results still hold in a broader setting, establishing a moment-entropy inequality for Renyi entropy and arbitrary moments. Moreover, associated with each λ-Renyi entropy and p-th moment is a corresponding notion of Fisher information, which satisfies a corresponding Stam's inequality. Both inequalities have the same family of extremal random variables, which the authors call generalized Gaussians. These contain the standard Gaussian as a limiting case but are notable in that they are heavy-tailed.

The authors [4] introduced different notions of moments for random vectors, established corresponding sharp moment-entropy inequalities, and identified the extremal random vectors, which the authors also call generalized Gaussians. The authors [5] then showed that the moment-entropy inequalities established in [3] are special cases of more general inequalities that hold for random vectors and not just random variables (For discrete random variables, also see Arikan [6]).

In this paper, we show that the concept of Fisher information and the corresponding Stam's inequality for random vectors holds in an even broader setting than above. In one direction, we show that associated with each λ-Renyi entropy and exponent $p \geq 1$, there is a corresponding notion of a Fisher information matrix for a multivariate random vector. We also show that there is a corresponding version of Stam's inequality, extending the one in [3]. These inequalities together with the moment-entropy inequalities in [5] show that the Cramér-Rao inequality in [3] holds for multivariate random vectors. Other extensions of the results in [3] were also obtained in [7], [**?**] and [**?**].

We also show that there is a notion of Fisher information, which we call affine Fisher information, that is invariant under all entropy-preserving (volume-preserving) linear transformations. Such a notion is natural to study when there is no *a priori* best or natural choice (of, say, weights in a weighted sum of squares or other powers) for defining the total error given an error vector. The affine Fisher information is an information measure that is well-defined independent of the weights used. Again, there is a corresponding Stam's inequality, where the extremal random vectors are generalized Gaussians. This one is in fact stronger than and implies the ones cited above, and a consequence is also an affine and stronger version of the Cramér-Rao inequality.

II. Preliminaries

Let X be a random vector in $\mathbb{R}^n$ with probability density function f_X. We also write f_X simply as f.

A. Linear transformation of a random vector

If A is a nonsingular $n \times n$ matrix, then the probability density of a random vector transforms under a linear transformation by

$$f_{AX}(y) = |A|^{-1} f_X(A^{-1}y), \tag{1}$$

where $|A|$ is the absolute value of the determinant of A. A special case of this is the formula for scalar multiplication,

$$f_{aX}(y) = |a|^{-n} f_X\big(\frac{y}{a}\big), \tag{2}$$

for each real $a \neq 0$.

E. Lutwak (elutwak@poly.edu), D. Yang (dyang@poly.edu), and G. Zhang (gzhang@poly.edu) are with the Department of Mathematics, Polytechnic Institute of NYU, Brooklyn, New York, and were supported in part by NSF Grant DMS-0706859. S. Lv (lvsongjun@126.com) is with the College of Mathematics and Computer Science, Chongqing Normal University, Chongqing, China, and was supported in part by Chinese NSF Grant 10801140.

The authors would like to thank Christoph Haberl, Tuo Wang, and Guangxian Zhu for their careful reading of this paper and the many improvements.

B. Entropy power

The Shannon entropy of a random vector X with density f is defined to be

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

The λ-*Renyi entropy power* for $\lambda > 0$ is defined to be

$$N_\lambda(X) = \begin{cases} \left(\int_{\mathbf{R}^n} f(x)^\lambda dx\right)^{\frac{2}{n(1-\lambda)}} & \text{if } \lambda \neq 1, \\ e^{\frac{2}{n}h(X)} & \text{if } \lambda = 1. \end{cases}$$

Note that

$$N_1(X) = \lim_{\lambda\to 1} N_\lambda(X).$$

The λ-Renyi entropy of a random vector X is defined to be

$$h_\lambda(X) = \frac{n}{2}\log N_\lambda(X).$$

In particular, $h_1 = h$. The λ-Renyi entropy $h_\lambda(X)$ is a continuous and, by the Hölder inequality, decreasing function of $\lambda \in (0,\infty)$.

By (1),

$$N_\lambda(AX) = |A|^{\frac{2}{n}} N_\lambda(X), \tag{3}$$

for any invertible matrix A.

C. Fisher information

The Fisher information of the random vector X is

$$\Phi(X) = \int_{\mathbf{R}^n} f^{-1}|\nabla f|^2\,dx. \tag{4}$$

The Fisher information matrix of the random vector X is

$$J(X) = E(X_1 \otimes X_1), \tag{5}$$

where $X_1 = f(X)^{-1}\nabla f(X)$. The Fisher information is the trace of the Fisher information matrix.

By (1),

$$J(AX) = A^{-t}J(X)A^{-1}, \tag{6}$$

for any invertible matrix A, where A^{-t} is the transpose of the inverse A^{-1}.

III. Generalized Gaussian distributions

For $\alpha > 0$ and $s < \frac{\alpha}{n}$, let Z be the random vector in $\mathbf{R}^n$ with density function

$$f_Z(x) = \begin{cases} b_{\alpha,s}\,(1-\frac{s}{\alpha}|x|^\alpha)_+^{\frac{1}{s}-\frac{n}{\alpha}-1} & \text{if } s\neq 0, \\ b_{\alpha,0}\,e^{-\frac{1}{\alpha}|x|^\alpha} & \text{if } s = 0, \end{cases} \tag{7}$$

where $t_+ = \max(t,0)$ and

$$b_{\alpha,s} = \begin{cases} \frac{\frac{\alpha}{n}|\frac{s}{\alpha}|^{\frac{n}{\alpha}}\Gamma(\frac{n}{2}+1)}{\pi^{\frac{n}{2}}B(\frac{n}{\alpha},1-\frac{1}{s})} & \text{if } s<0, \\ \frac{\Gamma(\frac{n}{2}+1)}{\pi^{\frac{n}{2}}\alpha^{\frac{n}{\alpha}}\Gamma(\frac{n}{\alpha}+1)} & \text{if } s=0, \\ \frac{\frac{\alpha}{n}(\frac{s}{\alpha})^{\frac{n}{\alpha}}\Gamma(\frac{n}{2}+1)}{\pi^{\frac{n}{2}}B(\frac{n}{\alpha},\frac{1}{s}-\frac{n}{\alpha})} & \text{if } s>0, \end{cases} \tag{8}$$

$\Gamma(\cdot)$ denotes the gamma function, and $B(\cdot,\cdot)$ denotes the beta function. The random variable Z is called the *standard generalized Gaussian*. It is normalized so that the αth moment is given by

$$E(|Z|^\alpha) = n. \tag{9}$$

The mean of Z is 0 and the covariance matrix of Z is $\frac{\sigma_2}{n}I$, where $\sigma_2 = \sigma_2(Z)$ is the second moment of Z. The p-th moment $\sigma_p(Z)$ is

$$E(|Z|^p) = \begin{cases} \left|\frac{\alpha}{s}\right|^{\frac{p}{\alpha}}\frac{B(\frac{n+p}{\alpha},1-\frac{p}{\alpha}-\frac{1}{s})}{B(\frac{n}{\alpha},1-\frac{1}{s})} & \text{if } s<0, \\ \alpha^{\frac{p}{\alpha}}\frac{\Gamma(\frac{n+p}{\alpha})}{\Gamma(\frac{n}{\alpha})} & \text{if } s=0, \\ \left(\frac{\alpha}{s}\right)^{\frac{p}{\alpha}}\frac{B(\frac{n+p}{\alpha},\frac{1}{s}-\frac{n}{\alpha})}{B(\frac{n}{\alpha},\frac{1}{s}-\frac{n}{\alpha})} & \text{if } s>0. \end{cases} \tag{10}$$

When $\alpha = 2$ and $s = 0$, Z is the usual standard Gaussian random vector. Any random vector of the form $Y = A(Z-\mu)$, where A is a nonsingular matrix, is called a *generalized Gaussian*. Let C be a positive definite symmetric matrix. If we let $C = AA^t$, then the density function of Y is given explicitly by

$$f_Y(x) \tag{11}$$

$$= \begin{cases} \frac{b_{\alpha,s}}{|C|^{\frac{1}{2}}}\left(1-\frac{s}{\alpha}\left((x-\mu)^tC^{-1}(x-\mu)\right)^{\frac{\alpha}{2}}\right)_+^{\frac{1}{s}-\frac{n}{\alpha}-1} & \text{if } s\neq 0, \\ \frac{b_{\alpha,0}}{|C|^{\frac{1}{2}}}e^{-\frac{1}{\alpha}\left((x-\mu)^tC^{-1}(x-\mu)\right)^{\frac{\alpha}{2}}} & \text{if } s=0. \end{cases} \tag{12}$$

The mean of Y is μ and the covariance matrix of Y is $\frac{\sigma_2}{n}C$.

Generalized Gaussians appear naturally as extremal distributions for various moment, entropy, and Fisher information inequalities [3]–[5], [7]–[11]. The usual Gaussians and $t-$student distributions are generalized Gaussians. Special cases of generalized Gaussians were studied by many authors, see for example, [?], [?], **[?]**, **[?]**, **[?]**, [7], [10], [12], [13]. We also note that the term of generalized Gaussians in the literature sometimes may refer to different families of distributions.

IV. Generalized Fisher information

A. $(p,\lambda)-$Fisher information

Let X be a random vector in $\mathbf{R}^n$ with probability density f. Define the λ-score of X to be the random vector

$$X_\lambda = f^{\lambda-2}(X)\nabla f(X), \tag{13}$$

and the $(p,\lambda)-$Fisher information of X to be the p-th moment of the λ-score of X,

$$\Phi_{p,\lambda}(X) = E(|X_\lambda|^p). \tag{14}$$

The classical Fisher information $\Phi(X) = \Phi_{2,1}(X)$. Note that the generalized Fisher information defined in [3] is normalized differently.

Fisher information, as defined above, relies on the standard inner product on $\mathbf{R}^n$. The formula for how the Fisher information behaves under linear transformations of the random vector and an arbitrary inner product is given by the following lemma.

Lemma 4.1: If A is a nonsingular matrix and X_λ is the λ-score of the random vector X in $\mathbf{R}^n$, then

$$(AX)_\lambda = |A|^{1-\lambda}A^{-t}X_\lambda. \tag{15}$$

Proof: Let $Y = AX$. Since

$$f_Y(y) = |A|^{-1} f_X(A^{-1}y),$$
$$\nabla f_Y(y) = |A|^{-1} A^{-t} \nabla f_X(A^{-1}y),$$

it follows that

$$\begin{aligned}(AX)_\lambda &= f_Y^{\lambda-2}(AX)\nabla f_Y(AX)\\ &= |A|^{1-\lambda} A^{-t}(f_X^{\lambda-2}(X)\nabla f_X(X))\\ &= |A|^{1-\lambda} A^{-t} X_\lambda.\end{aligned}$$

■

B. $(p, \lambda)-$Fisher information matrix

The Fisher information matrix $J(X)$ of a random vector X can be characterized as the square of the unique matrix with minimal determinant among all positive definite symmetric matrices A such that $E(|A^{-1}X_1|^2) = n$. This characterization motivates the following definition.

Definition 4.2: For $p, \lambda > 0$, the *(p, λ)-Fisher information matrix* $J_{p,\lambda}(X)$ of a random vector X in $\mathbf{R}^n$ is the pth power of the unique matrix with minimal determinant among all positive definite symmetric matrices A such that $E(|A^{-1}X_\lambda|^p) = n$.

The existence and uniqueness of the (p, λ)-Fisher information matrix is established by the following theorem.

Theorem 4.3: If $p > 0$ and X is a random vector in $\mathbf{R}^n$ with finite (p, λ)-Fisher information, then there exists a unique positive definite symmetric matrix A of minimal determinant so that $E(|A^{-t}X_\lambda|^p) = n$. Moreover, the matrix A is the unique positive definite symmetric matrix satisfying

$$E\big(|A^{-t}X_\lambda|^{p-2}(A^{-t}X_\lambda) \otimes (A^{-t}X_\lambda)\big) = I. \tag{16}$$

Proof: Theorem 4.3 will be proved by showing the existence and uniqueness of a positive definite symmetric matrix B satisfying $E(|BX_\lambda|^p) = n$ with maximum determinant. Then $A = B^{-t}$ is the desired matrix. Existence follows by showing that the set of positive definite symmetric matrices B satisfying the constraint $E(|BX_\lambda|^p) = n$ is compact. Since this set is closed, it suffices to show that this set is bounded. This, in turn, follows from an upper bound on the eigenvalues of any matrix B satisfying the constraint. If e is a unit eigenvector of B with eigenvalue η, then

$$\eta^p E(|e \cdot X_\lambda|^p) \le E(|B \cdot X_\lambda|^p) = n.$$

We claim that if X is a random vector in $\mathbf{R}^n$ with finite $(p, \lambda)-$Fisher information for $p > 0$, then there exists a constant $c > 0$ such that

$$E(|e \cdot X_\lambda|^p) \ge c > 0 \tag{17}$$

for every unit vector e.

Since the left side of (17) is a continuous function of the unit sphere, which is compact, it achieves its minimum. If the minimum is zero, then there exists a unit vector e such that $e \cdot \nabla f(x) = 0$ for almost every x in the support of f. This, however, is impossible for a differentiable probability density function on $\mathbf{R}^n$. See the proof of Lemma 4 in [8] for details. This shows the claim.

Therefore, there is a uniform upper bound for the eigenvalues of B, proving the existence of a minimum. The uniqueness and (16) follow by the same argument used in the proof of Theorem 8 in [5]. ■

The Fisher information matrix is defined implicitly. When $p = 2$, it has an explicit formula. For this case, equation (16) holds if and only if

$$E(X_\lambda \otimes X_\lambda) = A^2.$$

In other words,

$$J_{2,\lambda}(X) = E(X_\lambda \otimes X_\lambda). \tag{18}$$

This definition of the $(2, \lambda)$-Fisher information matrix was given by Johnson and Vignat [7].

Using Lemma 4.1 and Theorem 4.3, we obtain the following formula for the (p, λ)-Fisher information matrix when a linear transformation is applied to the random vector.

Proposition 4.4: If X is a random vector in $\mathbf{R}^n$ with finite (p, λ)-Fisher information and A is a nonsingular $n \times n$ matrix, then

$$J_{p,\lambda}(AX) = |A|^{p(1-\lambda)}(A^{-t} J_{p,\lambda}(X)^{\frac{2}{p}} A^{-1})^{\frac{p}{2}}. \tag{19}$$

Proof: Let

$$B = J_{p,\lambda}(AX)^{\frac{1}{p}}.$$

By Theorem 4.3 and (15),

$$\begin{aligned}I &= E(|B^{-t}(AX)_\lambda|^{p-2}(B^{-t}(AX)_\lambda) \otimes (B^{-t}(AX)_\lambda))\\ &= |A|^{p(1-\lambda)} E(|B^{-t}A^{-t}X_\lambda|^{p-2}(B^{-t}A^{-t}X_\lambda) \otimes (B^{-t}A^{-t}X_\lambda))\\ &= E(|L^{-t}X_\lambda|^{p-2} L^{-t}X_\lambda \otimes L^{-t}X_\lambda),\end{aligned} \tag{20}$$

where

$$L = |A|^{-1+\lambda} BA.$$

Using polar decomposition, there exists an orthogonal matrix T and a positive definite symmetric matrix P such that $L = T^t P$. Therefore, multiplying (20) on the left by T and on the right by T^t,

$$\begin{aligned}I &= TT^t\\ &= E(|TL^{-t}X_\lambda|^{p-2}(TL^{-t}X_\lambda) \otimes (TL^{-t}X_\lambda))\\ &= E(|P^{-t}X_\lambda|^{p-2}(P^{-t}X_\lambda) \otimes (P^{-t}X_\lambda)).\end{aligned}$$

By Theorem 4.3 again, it follows that

$$J_{p,\lambda}(X) = P^p = (TL)^p = (|A|^{(-1+\lambda)} TBA)^p.$$

Solving for B,

$$B = |A|^{1-\lambda} T^t (J_{p,\lambda}(X))^{\frac{1}{p}} A^{-1}.$$

and, since $B^t = B$,

$$\begin{aligned}B^p &= (B^t B)^{\frac{p}{2}}\\ &= |A|^{(1-\lambda)p}(A^{-t} J_{p,\lambda}(X)^{\frac{2}{p}} A^{-1})^{\frac{p}{2}}.\end{aligned}$$

■

Also, a special case of (19) is the following formula,

$$J_{p,\lambda}(aX) = a^{(1-\lambda)np-p} J_{p,\lambda}(X) \tag{21}$$

for any positive constant a and random vector X in $\mathbf{R}^n$ of finite (p, λ)-Fisher information.

C. Notions of affine Fisher information

Let X be a continuous random vector. There is the following fundamental entropy Fisher information inequality,

$$\frac{N_1(X)}{2\pi e} \geq \frac{n}{\Phi(X)}, \tag{22}$$

with equality if X is a standard Gaussian. This is an uncertainty principle of the entropy and the Fisher information.

By (3), the entropy power $N_1(X)$ is linearly invariant. However, the Fisher information $\Phi(X)$ is not linearly invariant. The equality in (22) characterizes only the standard Gaussian. For general Gaussians of fixed entropy, the Fisher information may become very large when the covariance matrices skew away from the identity matrix. Thus, the inequality (22) becomes inacurate. Is there a natural notion of Fisher information that is invariant under entropy-preserving linear transformations? Such a notion is called the affine Fisher information. It would give a stronger inequality than (22) which characterizes the general Gaussian distributions. By (3), we note that entropy-preserving linear transformations are the same as volume-preserving linear transformations.

One possible way for doing this is to minimize the Fisher information over all volume-preserving linear transformations of X. We define

$$\hat{\Phi}_{p,\lambda}(X) = \inf_{A \in \mathrm{SL}(n)} \Phi_{p,\lambda}(AX). \tag{23}$$

We show that this is simply the determinant of the Fisher information matrix.

Theorem 4.5: Let X be a random vector in $\mathbf{R}^n$ with finite $(p,\lambda)-$Fisher information for $p > 0$. Then

$$\hat{\Phi}_{p,\lambda}(X) = n|J_{p,\lambda}(X)|^{\frac{1}{n}}. \tag{24}$$

Proof: Let $\mathcal{S}$ denote the set of positive definite symmetric matrices. Since $E(|A^{-t}X_\lambda|^p)|A|^{\frac{p}{n}}$ is invariant under dilations of A and, for each $A \in \mathrm{GL}(n)$ and $v \in \mathbf{R}^n$,

$$|A^{-t}v| = |P^{-t}v|,$$

where $A = TP$, $T \in O(n)$, $P \in \mathcal{S}$, is the polar decomposition of A, it follows that

$$\begin{aligned}
&\inf_{A \in \mathrm{SL}(n)} \Phi_{p,\lambda}(AX) \\
&= \inf_{A \in \mathrm{SL}(n)} E(|A^{-t}X_\lambda|^p) \\
&= \inf_{A \in \mathrm{GL}(n)} E(|A^{-t}X_\lambda|^p)|A|^{\frac{p}{n}} \\
&= \inf\{n|A|^{\frac{p}{n}} \ : \ E(|A^{-t}X_\lambda|^p) = n, A \in \mathrm{GL}(n)\} \\
&= \inf\{n|A|^{\frac{p}{n}} \ : \ E(|A^{-t}X_\lambda|^p) = n, A \in \mathcal{S}\} \\
&= n|J_{p,\lambda}(X)|^{\frac{1}{n}}.
\end{aligned}$$

■

However, the above definition of linearly invariant Fisher information does not have an explicit formula except for special cases. This is because the $(p,\lambda)-$Fisher information matrix $J_{p,\lambda}$ is defined implicitly. Thus, it is not convenient for computation.

The general approach we will take is the following: Let $\mathscr{F}$ be a class of norms on $\mathbb{R}^n$ that is closed under linear transformations in the sense that if $\|\cdot\| \in \mathscr{F}$, then $\|\cdot\|_A \in \mathscr{F}$ for each $A \in \mathrm{GL}(n)$, where

$$\|x\|_A = \|Ax\|.$$

A form of Fisher information that is invariant under volume-preserving linear transformations can then be defined to be

$$\inf_{\|\cdot\| \in \mathscr{F}} V(B_{\|\cdot\|})^{-p/n} E(\|X_\lambda\|^p), \tag{25}$$

where $B_{\|\cdot\|}$ is the unit ball for the norm $\|\cdot\|$. In this general approach, there is an important and natural class of norms defined by the p-cosine transforms of density functions. We use it to define the affine Fisher information of a random vector.

The p-cosine transform $C(g)$ of the density g of a random vector Y is the function in $\mathbf{R}^n$ defined by

$$C(g)(x) = \int_{\mathbb{R}^n} |x \cdot y|^p g(y)\, dy. \tag{26}$$

It is a variation of the Fourier transform.

The p-cosine transform $C(g)$ gives the following norm on $\mathbb{R}^n$,

$$\|x\|_{Y,p} = (C(g)(x))^{\frac{1}{p}}\,, \ x \in \mathbb{R}^n. \tag{27}$$

If $p > 0$ and X is a random vector, then the *affine (p,λ)-Fisher information of X*, $\Psi_{p,\lambda}(X)$, is defined by

$$\Psi_{p,\lambda}(X) = \inf_{N_\lambda(Y)=c_1} E(\|X_\lambda\|^p_{Y,p}), \tag{28}$$

where each random vector Y is assumed to be independent of X and have λ-Renyi entropy equal to a constant c_1 which is chosen appropriately.

We shall show that the infimum in the definition above is achieved, and an explicit formula of the affine (p,λ)-Fisher information is obtained.

V. Formula of the affine Fisher information

A. Formula of the affine Fisher information

A random vector X in $\mathbb{R}^n$ with density f is said to have finite p-moment for $p > 0$, if

$$\int_{\mathbb{R}^n} |x|^p f(x)\, dx < \infty.$$

The following is the dual Minkowski inequality established in [4].

Lemma 5.1: Let $p > 0$, $\lambda > \frac{n}{n+p}$. If $\|\cdot\|$ is an n-dimensional norm in $\mathbf{R}^n$ and X is a random vector in $\mathbb{R}^n$ with finite p-moment, then

$$\int_{\mathbf{R}^n} \|x\|^p f(x) dx \geq N_\lambda(X)^{\frac{p}{2}} \left(a_1 \int_{S^{n-1}} \|u\|^{-n} dS(u) \right)^{-\frac{p}{n}}, \tag{29}$$

where f is the density of X, S^{n-1} is the unit sphere in $\mathbb{R}^n$, and dS denotes the standard surface area measure on S^{n-1}. The equality in the inequality holds if $\|\cdot\|$ is the Euclidean norm and X is the standard generalized Gaussian Z with parameters $\alpha = p$ and $\frac{1}{\lambda-1} = \frac{1}{s} - \frac{n}{\alpha} - 1$. The constant a_1 is given by

$$a_1 = \frac{\Gamma(\frac{n}{2}) N_\lambda(Z)^{\frac{n}{2}}}{2\pi^{\frac{n}{2}} \sigma_p(Z)^{\frac{n}{p}}} \tag{30}$$

$$= \begin{cases} a_0 B\left(\frac{n}{p}, \frac{1}{1-\lambda} - \frac{n}{p}\right), & \text{if } \lambda < 1, \\ \left(\frac{pe}{n}\right)^{\frac{n}{p}} \Gamma(1 + \frac{n}{p}), & \text{if } \lambda = 1, \\ a_0 B\left(\frac{n}{p}, \frac{\lambda}{\lambda-1}\right), & \text{if } \lambda > 1, \end{cases} \tag{31}$$

where

$$a_0 = \frac{1}{p}\left(1 + \frac{n(\lambda-1)}{p\lambda}\right)^{\frac{1}{\lambda-1}} \left|1 + \frac{p\lambda}{n(\lambda-1)}\right|^{\frac{n}{p}}.$$

Denote the volume of the unit ball in $\mathbb{R}^n$ by

$$\omega_n = \frac{\pi^{\frac{n}{2}}}{\Gamma(1+\frac{n}{2})},$$

and observe that

$$\int_{S^{n-1}} dS(u) = n\omega_n. \tag{32}$$

Let e be a fixed unit vector, and let

$$\omega_{n,p} = \int_{S^{n-1}} |u \cdot e|^p \, dS(u) = \frac{2\pi^{\frac{n-1}{2}} \Gamma(\frac{p+1}{2})}{\Gamma(\frac{n+p}{2})},$$

and

$$c_0 = \frac{1}{n\omega_n}\left(\frac{\omega_{n,p}}{n\omega_n}\right)^{\frac{n}{p}} = \frac{\Gamma(\frac{n}{2})}{2\pi^{\frac{n}{2}}}\left(\frac{\Gamma(\frac{n}{2})\Gamma(\frac{p+1}{2})}{\pi^{\frac{1}{2}}\Gamma(\frac{n+p}{2})}\right)^{\frac{n}{p}}.$$

Choose the constant c_1 in (28) as $c_1 = \left(\frac{a_1}{c_0}\right)^{\frac{2}{n}}$.

The following theorem gives an explicit formula for the affine Fisher information.

Theorem 5.2: If $p > 0$, $\lambda > \frac{n}{n+p}$, and X is a random vector in $\mathbf{R}^n$, then

$$\Psi_{p,\lambda}(X) = \left(c_0 \int_{S^{n-1}} E(|u \cdot X_\lambda|^p)^{-\frac{n}{p}} \, dS(u)\right)^{-\frac{p}{n}}.$$

Proof: Let f be the density of X and Y be a random vector with density g that has finite p-moment and satisfies $N_\lambda(Y) = c_1$. By Fubini's theorem and applying Lemma 5.1 to the norm $\|y\| = E(|y \cdot X_\lambda|^p)^{\frac{1}{p}}$, $y \in \mathbf{R}^n$, we have

$$\begin{aligned} &E(\|X_\lambda\|_{Y,p}^p) \\ &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} |X_\lambda \cdot y|^p g(y)\,dy\right) f(x)\,dx \\ &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} |X_\lambda \cdot y|^p f(x)\,dx\right) g(y)\,dy \\ &= \int_{\mathbb{R}^n} E(|y \cdot X_\lambda|^p) g(y)\,dy \\ &\geq N_\lambda(Y)^{\frac{p}{2}}\left(a_1 \int_{S^{n-1}} E(|u \cdot X_\lambda|^p)^{-\frac{n}{p}}\,dS(u)\right)^{-\frac{p}{n}}. \end{aligned}$$

■

B. Affine versus Euclidean

We show that the affine Fisher information is always less than the Euclidean Fisher information. Thus, when the Fisher information is used as a measure for error of data, the affine Fisher information gives a better measurement.

Lemma 5.3: If $p > 0$ and X is a random vector in $\mathbb{R}^n$, then

$$\Psi_{p,\lambda}(X) \leq \Phi_{p,\lambda}(X). \tag{33}$$

Equality holds if the function $v \mapsto E(|v \cdot X_\lambda|^p)$ is constant for $v \in S^{n-1}$. In particular, equality holds if X is spherically contoured.

Proof: Let e be a fixed unit vector. By Theorem 5.2, (32), Hölder's inequality and Fubini's theorem,

$$\begin{aligned} &(n\omega_n c_0)^{\frac{p}{n}} \Psi_{p,\lambda}(X) \\ &= \left[\frac{1}{n\omega_n}\int_{S^{n-1}} E(|u \cdot X_\lambda|^p)^{-\frac{n}{p}}\,dS(u)\right]^{-\frac{p}{n}} \\ &\leq \frac{1}{n\omega_n}\int_{S^{n-1}} E(|u \cdot X_\lambda|^p)\,dS(u) \\ &= \frac{1}{n\omega_n}\int_{\mathbb{R}^n}\int_{S^{n-1}} |u \cdot X_\lambda|^p f(x)\,dS(u)\,dx \\ &= \frac{1}{n\omega_n}\int_{S^{n-1}} |u \cdot e|^p\,dS(u)\int_{\mathbb{R}^n} |X_\lambda|^p f(x)\,dx \\ &= \frac{\omega_{n,p}}{n\omega_n}\Phi_{p,\lambda}(X). \end{aligned}$$

The equality condition follows by the equality condition of Hölder's inequality. ■

C. Linear invariance of the affine Fisher information

Denote the inverse transpose of $A \in \mathrm{GL}(n)$ by A^{-t}.

Lemma 5.4: If Y is a random vector in $\mathbb{R}^n$ and $A \in \mathrm{GL}(n)$, then for each $x \in \mathbb{R}^n$,

$$\|x\|_{AY,p} = \|A^t x\|_{Y,p}.$$

Proof: By (27), (1), and a change of variable, we have

$$\begin{aligned} \|x\|_{AY,p}^p &= |A|^{-1}\int_{\mathbb{R}^n} |x \cdot y|^p g_Y(A^{-1}y)\,dy \\ &= \int_{\mathbb{R}^n} |x \cdot Az|^p g_Y(z)\,dz \\ &= \int_{\mathbb{R}^n} |A^t x \cdot z|^p g_Y(z)\,dz \\ &= \|A^t x\|_{Y,p}. \end{aligned}$$

■

Proposition 5.5: If $p > 0$ and X is a random vector in $\mathbf{R}^n$, then

$$\Psi_{p,\lambda}(AX) = \Psi_{p,\lambda}(X), \tag{34}$$

for each $A \in \mathrm{SL}(n)$.

Proof: Let $U = AX$. Then for each $u \in \mathbb{R}^n$,

$$f_U(u) = |A|^{-1} f_X(A^{-1}u),$$

and therefore

$$\begin{aligned}U_\lambda &= f_U^{\lambda-2}(U)\nabla f_U(U)\\ &= |A|^{1-\lambda} f_X^{\lambda-2}(A^{-1}U)A^{-t}\nabla f_X(A^{-1}U)\\ &= |A|^{1-\lambda}A^{-t}(f_X^{\lambda-2}(X)\nabla f_X(X))\\ &= |A|^{1-\lambda}A^{-t}X_\lambda.\end{aligned}$$

Since the norm $\|\cdot\|_{Y,p}$ is homogeneous of degree 1,

$$E(\|U_\lambda\|_{Y,p}^p) = |A|^{p(1-\lambda)}E(\|A^{-t}X_\lambda\|_{Y,p}^p).$$

In particular, if $A \in \mathrm{SL}(n)$, then the above, Lemma 5.4, and (3) give

$$\begin{aligned}\Psi_{p,\lambda}(AX) &= \inf_{N_\lambda(Y)=c_1} E(\|U_\lambda\|_{Y,p}^p)\\ &= \inf_{N_\lambda(Y)=c_1} E(\|A^{-t}X_\lambda\|_{Y,p}^p)\\ &= \inf_{N_\lambda(Y)=c_1} E(\|X_\lambda\|_{A^{-1}Y,p}^p)\\ &= \inf_{N_\lambda(A^{-1}Y)=c_1} E(\|X_\lambda\|_{A^{-1}Y,p}^p)\\ &= \inf_{N_\lambda(Y)=c_1} E(\|X_\lambda\|_{Y,p}^p)\\ &= \Psi_{p,\lambda}(X).\end{aligned}$$

That proves the linear invariance of the affine (p,λ)-Fisher information. ■

VI. Contoured random vectors

A random vector X is *spherically contoured* about $x_0 \in \mathbf{R}^n$ if the density function f_X can be written as

$$f_X(x) = F(|x-x_0|^2)$$

for some 1-dimensional function $F : [0,\infty) \to [0,\infty)$. A random vector Y is *elliptically contoured,* if there exists a spherically contoured random vector X and an invertible matrix A such that $Y = AX$. Elliptically contoured random vectors were studied by many authors, see for example, [?], [14]. Many of the explicitly known examples of random vectors are either spherically or elliptically contoured. The contoured property greatly facilitates the computation of information theoretic measures of the random vector [14]. We compute the (p,λ)-Fisher information matrix and the (p,λ)-affine Fisher information for spherically and elliptically contoured distributions.

A. Fisher information matrix of elliptically contoured distributions

Proposition 6.1: If a random vector X in $\mathbf{R}^n$ is spherically contoured, then

$$J_{p,\lambda}(X) = \frac{1}{n}\Phi_{p,\lambda}(X)I. \tag{35}$$

Proof: The λ-score of X is

$$\begin{aligned}X_\lambda &= f^{\lambda-2}(X)\nabla f(X)\\ &= 2(X-x_0)F^{\lambda-2}(|X-x_0|^2)F'(|X-x_0|^2).\end{aligned} \tag{36}$$

Then

$$\begin{aligned}&E(|X_\lambda|^{p-2}X_\lambda\otimes X_\lambda)\\ &= \int_{\mathbf{R}^n} |2F^{\lambda-2}(|x-x_0|^2)F'(|x-x_0|^2)|^p|x-x_0|^{p-2}\\ &\qquad ((x-x_0)\otimes(x-x_0))F(|x-x_0|^2)\,dx\\ &= \int_{\mathbf{R}^n} |2F^{\lambda-2}(|x|^2)F'(|x|^2)|^p|x|^{p-2}(x\otimes x)F(|x|^2)\,dx\\ &= aI,\end{aligned} \tag{37}$$

where by (37) and (36),

$$\begin{aligned}a &= \frac{1}{n}\operatorname{tr}\int_{\mathbf{R}^n} |2F^{\lambda-2}(|x|^2)F'(|x|^2)|^p|x|^{p-2}(x\otimes x)F(|x|^2)\,dx\\ &= \frac{1}{n}\int_{\mathbf{R}^n} |2F^{\lambda-2}(|x|^2)F'(|x|^2)|^p|x|^pF(|x|^2)\,dx\\ &= \frac{1}{n}\int_{\mathbf{R}^n} |f^{\lambda-2}(x)\nabla f(x)|^pf(x)\,dx\\ &= \frac{1}{n}E(|X_\lambda|^p)\\ &= \frac{1}{n}\Phi_{p,\lambda}(X).\end{aligned}$$

■

Equations (35) and (19) imply the following formula for the (p,λ)-Fisher information matrix of an elliptically contoured random vector.

Corollary 6.2: If $Y = AX$, where X is a spherically contoured random vector in $\mathbf{R}^n$ with finite (p,λ)-Fisher information and A is an invertible matrix, then

$$J_{p,\lambda}(Y) = \frac{1}{n}\Phi_{p,\lambda}(X)|A|^{(1-\lambda)p}(AA^t)^{-\frac{p}{2}}. \tag{38}$$

B. Affine Fisher information of elliptically contoured distributions

Proposition 6.3: If a random vector X in $\mathbf{R}^n$ is spherically contoured, then

$$\Psi_{p,\lambda}(AX) = \Phi_{p,\lambda}(X), \quad A \in SL(n). \tag{39}$$

Proof: We first show that if X is spherically contoured, then

$$E(|u\cdot X_\lambda|^p) = \frac{\omega_{n,p}}{n\omega_n}E(|X_\lambda|^p).$$

Indeed, by the using polar coordinates, we have

$$\begin{aligned}E(|u\cdot X_\lambda|^p) &= \int_{\mathbf{R}^n} |u\cdot X_\lambda|^pf(x)dx\\ &= \int_{\mathbf{R}^n} |u\cdot 2(x-x_0)F^{\lambda-2}(|x-x_0|^2)\\ &\qquad F'(|x-x_0|^2)|^pF(|x-x_0|^2)\,dx\\ &= E(|X_\lambda|^p)\frac{1}{n\omega_n}\int_{S^{n-1}} |u\cdot v|^p\,dv\\ &= \frac{\omega_{n,p}}{n\omega_n}E(|X_\lambda|^p).\end{aligned}$$

Then the desired equation follows from Theorem 5.2 and (34). ■

C. The entropy and Fisher information of a generalized Gaussian

A straightforward calculation shows that if

$$\frac{1}{\lambda-1} = \frac{1}{s} - \frac{n}{\alpha} - 1, \tag{40}$$

then the λ-Renyi entropy power of the standard generalized Gaussian Z is given by

$$N_\lambda(Z) = \begin{cases} b_{\alpha,s}^{-\frac{2}{n}} \, (1 - \frac{sn}{\alpha})^{\frac{2}{n(1-\lambda)}} & \text{if } \lambda \neq 1, \\ b_{\alpha,0}^{-\frac{2}{n}} \, e^{\frac{2}{\alpha}} & \text{if } \lambda = 1. \end{cases} \tag{41}$$

If, in addition to (40),

$$\frac{1}{p} + \frac{1}{\alpha} = 1, \tag{42}$$

then the (p, λ)-Fisher information of the standard generalized Gaussian Z is equal to

$$\Phi_{p,\lambda}(Z) = \begin{cases} n b_{\alpha,s}^{(\lambda-1)p} \left|1 - \frac{s(n+\alpha)}{\alpha}\right|^p & \text{if } \lambda \neq 1, \\ n & \text{if } \lambda = 1. \end{cases} \tag{43}$$

VII. Inequalities for entropy and Fisher information

Define the constant $c_{n,p,\lambda}$ by

$$c_{n,p,\lambda} = \Phi_{p,\lambda}(Z) N_\lambda(Z)^{\frac{p}{2}((\lambda-1)n+1)}, \tag{44}$$

where the parameters α and s of the standard generalized Gaussian Z satisfy (40) and (42). The necessary condition $s < \frac{\alpha}{n}$ is equivalent to

$$\lambda \in (-\infty, 0) \cup \left(\frac{n}{n+\alpha}, \infty\right). \tag{45}$$

Theorem 7.1: If $n \geq 2$, X is a random vector in $\mathbf{R}^n$, $1 \leq p < n$, and $\lambda \geq (n-1)/n$. Then

$$\Phi_{p,\lambda}(X) N_\lambda(X)^{\frac{p}{2}((\lambda-1)n+1)} \geq c_{n,p,\lambda} \tag{46}$$

with equality if X is the standard generalized Gaussian Z with parameters given by (40) and (42).

Proof: We use the following sharp Gagliardo-Nirenberg inequality established by Del Pino and Dolbeault [9] (also, see [11] and [8]): If $n \geq 2$, w is a function of $\mathbf{R}^n$, $1 \leq p < n$ and $0 < r \leq \frac{np}{n-p}$, then

$$\|\nabla w\|_p \geq c \, \|w\|_q^{1-\gamma} \|w\|_r^\gamma, \tag{47}$$

where

$$q = r\big(1 - \frac{1}{p}\big) + 1, \quad \frac{1}{p} - \frac{1}{n} = \frac{1-\gamma}{q} + \frac{\gamma}{r}, \tag{48}$$

and the constant c is such that equality holds when

$$w(x) = \begin{cases} b(1 - a|x - x_0|^{\frac{p}{p-1}})_+^{\frac{p}{p-r}}, & p \neq r \\ b \exp(-a|x - x_0|^{\frac{p}{p-1}}), & p = r. \end{cases} \tag{49}$$

where $a, b > 0$ are constant and $x_0 \in \mathbf{R}^n$ is a constant vector.

If p and λ satisfy the assumptions of the theorem, α satisfies (42), and q and r are given by

$$\left(\lambda - \frac{1}{\alpha}\right) r = 1 \text{ and } \left(\lambda - \frac{1}{\alpha}\right) q = \lambda,$$

then p, q, r satisfy (48). If

$$w = f^{\lambda - \frac{1}{\alpha}},$$

where f is the density function of X, then

$$\begin{aligned} \|\nabla w\|_p^p &= \left(\lambda - \frac{1}{\alpha}\right)^p \Phi_{p,\lambda}(X), \\ \|w\|_q^{1-\gamma} &= N_\lambda(X)^{-\frac{1}{2}((\lambda-1)n+1)}, \\ \|w\|_r &= 1. \end{aligned}$$

These equations, the inequality (47), and (49) imply the desired inequality (46). ■

The 1-dimensional analogue of Theorem 7.1 was proved in [3]. In n-dimension, Theorem 7.1 was proved in [?] for the case of $\lambda = 1$.

Theorem 7.2: If X is a random vector in $\mathbf{R}^n$, A a nonsingular $n \times n$ matrix, $1 \leq p < n$, and $\lambda \geq 1 - \frac{1}{n}$, then

$$\Phi_{p,\lambda}(AX) N_\lambda(X)^{\frac{p}{2}((\lambda-1)n+1)} \geq c_{n,p,\lambda} |A|^{(1-\lambda-\frac{1}{n})p}. \tag{50}$$

Equality holds if X is the standard generalized Gaussian Z with parameters given by (40) and (42).

Proof: By equation (3) and inequality (46),

$$\begin{aligned} &\Phi_{p,\lambda}(AX) N_\lambda(X)^{\frac{p}{2}((\lambda-1)n+1)} |A|^{(\lambda+\frac{1}{n}-1)p} \\ &= \Phi_{p,\lambda}(AX) N_\lambda(AX)^{\frac{p}{2}((\lambda-1)n+1)} \\ &\geq \Phi_{p,\lambda}(Z) N_\lambda(Z)^{\frac{p}{2}((\lambda-1)n+1)}. \end{aligned}$$

■

Theorem 7.3: If X is a random vector in $\mathbf{R}^n$, $1 \leq p < n$, and $\lambda \geq 1 - \frac{1}{n}$, then

$$|J_{p,\lambda}(X)|^{\frac{1}{n}} N_\lambda(X)^{\frac{p}{2}((\lambda-1)n+1)} \geq \frac{c_{n,p,\lambda}}{n}. \tag{51}$$

Equality holds if X is a generalized Gaussian.

Proof: Let $A = J_{p,\lambda}(X)^{\frac{1}{p}}$. Taking the trace of both sides of (16), we get

$$E(|A^{-t} X_\lambda|^p) = n. \tag{52}$$

By equation (14), (15), and the definition of the Fisher information matrix and (52),

$$\begin{aligned} \Phi_{p,\lambda}(AX) &= E(|(AX)_\lambda|^p) \\ &= |A|^{(1-\lambda)p} E(|A^{-t} X_\lambda|^p) \\ &= n |J_{p,\lambda}(X)|^{1-\lambda}. \end{aligned}$$

By this and Theorem 7.2,

$$n |J_{p,\lambda}(X)|^{1-\lambda} N_\lambda(X)^{\frac{p}{2}((\lambda-1)n+1)} \geq c_{n,p,\lambda} |A|^{(1-\lambda-\frac{1}{n})p}.$$

This gives the inequality (51). ■

By (24) and Theorem 7.3, we have

Theorem 7.4: If $1 \leq p < n$, $\lambda \geq 1 - 1/n$, and X is a random vector in $\mathbb{R}^n$, then

$$\hat{\Phi}_{p,\lambda}(X) N_\lambda(X)^{\frac{p}{2}((\lambda-1)n+1)} \geq c_{n,p,\lambda}. \tag{53}$$

Equality holds if X is a generalized Gaussian.

VIII. Inequalities for entropy and affine Fisher information

A. A Sobolev inequality

The following is taken from [8].

Given a function w on $\mathbb{R}^n$, define $H : \mathbb{R}^n \to (0, \infty)$ by

$$H(v) = \left(\int_{\mathbb{R}^n} |v \cdot \nabla w(x)|^p \, dx \right)^{\frac{1}{p}}, \tag{54}$$

for each $v \in \mathbb{R}^n \backslash \{0\}$, and the set $B_p w \subset \mathbb{R}^n$ by

$$B_p w = \{ v \in \mathbb{R}^n \ : \ H(v) \leq 1 \}.$$

Using polar co-ordinates, note that the volume of the convex body $B_p w$ is given by

$$V(B_p w) = \frac{1}{n} \int_{S^{n-1}} H^{-n}(u) \, dS(u). \tag{55}$$

The proof of Theorem 8.2 below requires the following:

Theorem 8.1 (Theorem 7.2, [8]): If $1 \leq p < n$ and $0 < r \leq np/(n-p)$, then there exists a constant $c(p, r, n)$ such that for each function w on $\mathbb{R}^n$,

$$V(B_p w)^{-\frac{1}{n}} \geq c(p, r, n) |w|_q^{1-\gamma} |w|_r^{\gamma},$$

where $|w|_q$ and $|w|_r$ are the L_q and L_r norms of w respectively, q and γ are given by (48), and equality holds if and only if there exist $b > 0$, $A \in \mathrm{GL}(n)$ and $x_0 \in \mathbb{R}^n$ such that

$$w(x) = \begin{cases} b\left(1 + (r-p)|A(x-x_0)|^{p/(p-1)}\right)_+^{p/(p-r)} & \text{if } r \neq p, \\ b \exp(-p|A(x-x_0)|^{p/(p-1)}) & \text{if } r = p. \end{cases}$$

B. Affine Fisher information inequalities

Theorem 8.2: Let X be a random vector in $\mathbb{R}^n$, $1 \leq p < n$, and $\lambda \geq 1 - 1/n$. Then

$$\Psi_{p,\lambda}(X) N_\lambda(X)^{\frac{p}{2}((\lambda-1)n+1)} \geq c_{n,p,\lambda}. \tag{56}$$

Equality holds if and only if X is a generalized Gaussian.

Proof: Let r be given by

$$\left(\lambda - 1 + \frac{1}{p} \right) r = 1, \tag{57}$$

q and γ by (48), and $w = f_X^{\lambda - 1 + 1/p}$. By (54), (55), and Theorem 5.2,

$$\begin{aligned} \Psi_{p,\lambda}(X) &= \left(\lambda - 1 + \frac{1}{p} \right)^{-p} (nc_0 V(B_p w))^{-\frac{p}{n}} \\ |w|_q^{1-\gamma} &= N_\lambda(X)^{-\frac{1}{2}((\lambda-1)n+1)}, \\ |w|_r &= 1. \end{aligned} \tag{58}$$

Inequality (56) now follows by Theorem 8.1 and (58). The equality conditions follow from the equality conditions given by Theorem 8.1. ∎

By Lemma 5.3, inequality (56) is stronger than inequality (46).

C. Strengthened Cramér-Rao inequality

The Cramér-Rao inequality is

$$\frac{1}{n} \sigma_2(X) \geq \frac{n}{\Phi(X)}.$$

The reciprocal of the Fisher information $\Phi(X)$ gives a lower bound of the second moment $\sigma_2(X)$. This inequality is generalized to p-moment and (p, λ)-Fisher information. We need the following moment-entropy inequality proved in [4], see also [5]. If $p > 0$ and $\lambda > \frac{n}{n+p}$, then

$$\frac{\sigma_p(X)}{\sigma_p(Z)} \geq \left(\frac{N_\lambda(X)}{N_\lambda(Z)} \right)^{\frac{p}{2}} \tag{59}$$

with equality if X is the generalized Gaussian Z with parameters $\alpha = p$ and (40).

Let p^* be the conjugate of p. Note that $p^* < n$ implies $\frac{n-1}{n} > \frac{n}{n+p}$. By Theorem 7.1 and (59), we have

Theorem 8.3: If X is a random vector in $\mathbf{R}^n$, then for $1 < p^* < n$ and $\lambda > \frac{n-1}{n}$,

$$\left(\frac{\sigma_p(X)}{\sigma_p(Z)} \right)^{(\lambda-1)n+1} \geq \frac{\Phi_{p^*,\lambda}(Z)}{\Phi_{p^*,\lambda}(X)} \tag{60}$$

with equality if $X = aZ$, $a > 0$, where the standard generalized Gaussian Z has parameters $\alpha = p$ and (40).

The 1-dimensional analogue of Theorem 8.3 was proved in [3]. In n-dimension, Theorem 8.3 was proved in [7] for the case of $p = 2$.

The Cramér-Rao inequality can be strengthened by using the affine Fisher information. By Theorem 8.2 and (59), we have

Theorem 8.4: If X is a random vector in $\mathbf{R}^n$, then for $1 < p^* < n$ and $\lambda > \frac{n-1}{n}$,

$$\left(\frac{\sigma_p(X)}{\sigma_p(Z)} \right)^{(\lambda-1)n+1} \geq \frac{\Psi_{p^*,\lambda}(Z)}{\Psi_{p^*,\lambda}(X)} \tag{61}$$

with equality if $X = aZ$, $a > 0$, where the standard generalized Gaussian Z has parameters $\alpha = p$ and (40).

By Lemma 5.3 and Proposition 6.3, inequality (61) is stronger than inequality (60).

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, "Cramer-Rao and moment-entropy inequalities for Renyi entropy and generalized Fisher information," *IEEE Trans. Inform. Theory*, vol. 51, pp. 473–478, 2005.

[4] ——, "Moment–entropy inequalities," *Annals of Probability*, vol. 32, pp. 757–774, 2004.

[5] ——, "Moment-entropy inequalities for a random vector," *IEEE Trans. Inform. Theory*, vol. 53, pp. 1603–1607, 2007.

[6] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inform. Theory*, vol. 42, pp. 99–105, 1996.

[7] O. Johnson and C. Vignat, "Some results concerning maximum Rényi entropy distributions," *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 43, pp. 339–351, 2007.

[8] E. Lutwak, D. Yang, and G. Zhang, "Optimal Sobolev norms and the L^p Minkowski problem," *Int. Math. Res. Not.*, 2006, Art. ID 62987, 21 pp.

[9] M. Del Pino and J. Dolbeault, "Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions," *J. Math. Pures Appl. (9)*, vol. 81, pp. 847–875, 2002.

[10] 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.

[11] 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.

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

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

[14] O. G. Guleryuz, E. Lutwak, D. Yang, and G. Zhang, "Information-theoretic inequalities for contoured probability distributions," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2377–2383, 2002.