A Unified Approach to Cramér-Rao Inequalities

Andrea Cianchi, Erwin Lutwak, Deane Yang, and Gaoyong Zhang

***Abstract*—A unified approach is presented for establishing a broad class of Cramér-Rao inequalities for the location parameter, including, as special cases, the original inequality of Cramér and Rao, as well as an L^p version recently established by the authors. The new approach allows for generalized moments and Fisher information measures to be defined by convex functions that are not necessarily homogeneous.**

In particular, it is shown that associated with any log-concave random variable whose density satisfies certain boundary conditions is a Cramér-Rao inequality for which the given log-concave random variable is the extremal. Applications to specific instances are also provided.

***Index Terms*—moment, Fisher information, entropy, Shannon entropy, Rényi entropy, Cramér-Rao inequality, information measure**

I. Introduction

A fundamental result of information theory and a real analogue of the uncertainty principle in quantum mechanics is the classical non-Bayesian Cramér-Rao inequality [2], [3] (also, see [4]) for a smooth 1-parameter family of continuous random variables X with corresponding distribution function f_θ. The inequality states that an unbiased estimator $g(x)$ for the parameter θ satisfies

$$E[(g(X)-\theta)^2]E\left[\left(\frac{\partial \log f_\theta}{d\theta}(X)\right)^2\right] \ge 1. \tag{1}$$

We will focus only on an important special case, where the parameter θ is the location parameter. In that case, $f_\theta(x) = f(x-\theta)$ and $g(x) = x$. Since the estimator is unbiased, it follows that $E[X] = \theta$ and therefore (1) becomes

$$E[X^2]E\left[\left(\frac{f'(X)}{f(X)}\right)^2\right] \ge 1, \tag{2}$$

where $E[(f'(X)/f(X))^2]$ is known as the Fisher information of X. Moreover, equality holds if and only if X is a Gaussian.

Vajda [5] introduced a generalized Fisher information, $E[|f'(X)/f(X)|^q]$, sometimes known as the Vajda information measure (see, for example, [6]) and showed that it and the p-moment, where $p^{-1}+q^{-1} = 1$, satisfy a generalized Cramér-Rao inequality (also, see [7], [1]). Here, equality holds if and only if X is an L^p Gaussian. In [8] it was shown that the L^p Fisher information defined is linked to not only the p-th moment but also Shannon entropy and extended the definition of L^p Fisher information introduced in [1] to an even more general Fisher information linked to the p-th moment and Rényi entropy. It was also shown that a Cramér-Rao inequality exists in this setting and equality holds if and only if the random variable X is a generalized Gaussian, where if the entropy used is not Shannon but λ-Rényi for $\lambda \ne 1$, the density function does not decay exponentially and is fat-tailed.

More recently, Bercher [9], [10], [11], [12] has studied generalized Fisher information and established generalized Cramér-Rao inequalities for parameterized distributions, which, when the parameter is the location parameter, imply the generalized Cramér-Rao inequalities stated above. Bercher and Vignat [13], extending earlier work of Uhrmann-Klingen [14], have also studied the extremal distributions for the classical Cramér-Rao inequality restricted to random variables that lie within a fixed interval.

In 2002 Vajda [15] introduced and studied an even more generalized Fisher information, where the power function, $x \mapsto |x|^p$, is replaced by an arbitrary convex function.

Here, we show how Vajda's generalized Fisher information can be used to obtain a single unified proof for a broad class of Cramér-Rao inequalities, including the classical one as well as a subset of the inequalities established in [8]. Such a proof extends the original proof of the classical inequality to cover the case of a generalized moment and Fisher information that are defined by convex functions that are not necessarily homogeneous.

The novel aspect of our approach to generalized Fisher information, in contrast to those of Vajda and others, is that, given a convex function ϕ^*, we do not define generalized Fisher information directly as $E[\phi^*(f'(X)/f(X))]$. Instead, we define generalized Fisher information to be the normalization factor m^* such that $E[\phi^*(f'(m^*X)/f(m^*X))]$ is equal to a pre-chosen fixed constant. This indirect approach gives a definition of generalized Fisher information that scales nicely when the random variable is rescaled. We believe that this approach would also be effective for studying other information measures that are defined in terms of a convex function ϕ^*.

The observation that led to the unified proof presented here is that neither the definitions of the moment and Fisher information nor the proof of the classical Cramér-Rao inequality, as well as the ones proved in [8], require the norms used in the definitions of moment and Fisher information to be defined in terms of the power function $x \mapsto |x|^p$. The proof of the classical Cramér-Rao inequality uses only integration by parts and the Cauchy-Schwarz inequality for the L^2 norm

Andrea Cianchi (<cianchi@unifi.it>) is with Università degli Studi di Firenze, Firenze, Italia; 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 Grants DMS-1007347 and DMS-1312181.

The authors would like to thank Monika Ludwig, Christoph Haberl, Jie Xiao, Songjun Lv, Tuo Wang, Jean-François Bercher, and especially Yiming Zhao and Guangxian Zhu for their comments on various versions of this paper. They would also like to thank Almut Burchard, the referees, and the associate editors George Moustakides, who handled this paper, and Marco Lops, who handled an earlier version, for their cogent remarks. They would like to thank one of the referees for calling attention to the paper [1] by Batalama and Kazakos.

of a function. This can be extended to a family of L^p Cramér-Rao inequalities by using the Hölder inequality instead of the Cauchy-Schwarz inequality. In this paper we show that if the moment is defined using a strictly convex function, which we call a gauge, and a corresponding Fisher information is defined by the Legendre transform of the gauge, then the same proof of the Cramér-Rao inequality follows by Young's inequality (5), of which the Cauchy-Schwarz and Hölder inequalities are special cases.

In particular, given an appropriate convex function ϕ (see Section II-A for details) and a random variable Y, we define the ϕ-moment $m_\phi[Y]$ of Y in Section II-E and the ϕ-Fisher information $m^*_\phi[Y]$ in Section II-F, which generalize the standard definitions of moment and Fisher information. Using these definitions we state in §III the main theorem of this paper, Theorem 1.

Theorem 1 is, in turn, a direct consequence of a more general theorem stated in §V, Theorem 2, which shows that associated with any suitable log-concave random variable is a Cramér-Rao inequality for which the given log-concave random variable is the extremal.

Explicit examples of Theorem 1 are given in §VI. In particular, we show in §VI-A that if we set $\phi(x) = x^2/2$, then a corollary of Theorem 1 is the classical Cramér-Rao inequality for the location parameter. The last two examples are, as far as we know, new and give unique characterizations of the logistic and power function distributions.

In the last section, §VII, we show how the equality case of a generalized Cramér-Rao inequality leads to a way to identify whether a random variable has a power function distribution using appropriately defined information measures.

Other previous work on generalizations of Fisher information, the Cramér-Rao inequality, and its proof include [16], [7], [14], [1], [15], [17], [13], [18], [19], [6], [20], [11], [12], [21], [22].

II. Definitions

A. Gaussian Gauge

We define a *gauge* to be a continuously differentiable function $\phi : (x_-, x_+) \to \mathbb{R}$, where $-\infty \le x_- < x_+ \le \infty$, such that $\phi' : (x_-, x_+) \to \mathbb{R}$ is a strictly increasing function. In particular, ϕ is strictly convex. We say a gauge is *Gaussian,* if in addition

$$\int_{x_-}^{x_+} e^{-\phi(x)}\, dx < \infty.$$

B. ϕ-Gaussian random variable

Given a Gaussian gauge $\phi : (x_-, x_+) \to \mathbb{R}$, we define the *standard ϕ-Gaussian random variable* X_ϕ to be the random variable whose density function is positive only on the interval (x_-, x_+) and is given by

$$f_\phi(x) = e^{-\phi(x)} \Big/ \left(\int_{x_-}^{x_+} e^{-\phi(y)}\, dy \right) \quad \text{for } x \in (x_-, x_+).$$

A real random variable X is a *ϕ-Gaussian,* if there exists $\lambda > 0$ such that λX is the standard ϕ-Gaussian.

Note that any strictly log-concave random variable with density f that is continuously differentiable on its support is the standard ϕ-Gaussian, where $\phi = -\log f$.

C. Dual gauge

Denote the image $\phi'((x_-, x_+)) = (\xi_-, \xi_+)$ and define $\phi^* : (\xi_-, \xi_+) \to \mathbb{R}$ as the function obeying

$$\phi^*(\phi'(x)) = x\phi'(x) - \phi(x), \tag{3}$$

for each $x \in (x_-, x_+)$.

Lemma 1: The function $\phi^* : (\xi_-, \xi_+) \to \mathbb{R}$ is a continuously differentiable function whose derivative $(\phi^*)'$ is the inverse function of ϕ' and therefore a strictly increasing function from (ξ_-, ξ_+) onto (x_-, x_+).

Proof: Given $\xi, \eta \in (\xi_-, \xi_+)$ such that $\xi \ne \eta$, let $x, y \in (x_-, x_+)$ satisfy $\xi = \phi'(x)$ and $\eta = \phi'(y)$. The strict convexity of ϕ implies that

$$0 < \phi(x) - \phi(y) - \phi'(y)(x-y) < (y-x)(\phi'(y) - \phi'(x)). \tag{4}$$

By (3) and (4),

$$\begin{aligned}
&\left| \frac{\phi^*(\eta) - \phi^*(\xi)}{\eta - \xi} - (\phi')^{-1}(\xi) \right| \\
&= \left| \frac{\phi^*(\phi'(y)) - \phi^*(\phi'(x))}{\phi'(y) - \phi'(x)} - x \right| \\
&= \left| \frac{y\phi'(y) - \phi(y) - (x\phi'(x) - \phi(x))}{\phi'(y) - \phi'(x)} - x \right| \\
&= |y - x| \left[\frac{\phi(x) - \phi(y) - \phi'(y)(x-y)}{(y-x)(\phi'(y) - \phi'(x))} \right] \\
&< |(\phi')^{-1}(\eta) - (\phi')^{-1}(\xi)|.
\end{aligned}$$

The lemma now follows by taking the limit $\eta \to \xi$. ■

We call ϕ^* the *polar gauge* or *dual gauge* of ϕ. Observe, however, that ϕ^* is not necessarily a Gaussian gauge, because the integral of $e^{-\phi^*}$ over the interval (ξ_-, ξ_+) is not necessarily finite.

The following is known as Young's inequality (see, for example, Arnol'd [23]) or Fenchel's inequality (see, for example, Rockafellar [24]).

Lemma 2: If ϕ is a gauge and ϕ^* its dual gauge, then

$$x\xi \le \phi(x) + \phi^*(\xi), \tag{5}$$

for each $x \in (x_-, x_+)$ and $\xi \in (\xi_-, \xi_+)$. Moreover, the following are equivalent:

$$\begin{aligned}
x\xi &= \phi(x) + \phi^*(\xi) \\
\xi &= \phi'(x) \\
x &= (\phi^*)'(\xi).
\end{aligned}$$

Proof: Since ϕ' is strictly increasing, if $\xi \in (\xi_-, \xi_+)$, then there exists a unique $\hat{x} \in (x_-, x_+)$ such that $\xi = \phi'(\hat{x})$. Since the function $x \mapsto \xi x - \phi(x)$ is strictly concave and $\hat{x}$ is a critical point, it follows that $\hat{x}$ is in fact the unique maximum. Therefore, if $x \in (x_-, x_+)$, then, by (3),

$$\begin{aligned}
\xi x - \phi(x) &\le \xi\hat{x} - \phi(\hat{x}) \\
&= \xi(\phi')^{-1}(\xi) - \phi\big((\phi')^{-1}(\xi)\big) = \phi^*(\xi)\,.
\end{aligned}$$

This proves inequality (5). Moreover, equality holds in (5) if and only if $\xi = \phi'(x)$. The same argument applied to the function $\xi \mapsto \xi x - \phi^*(\xi)$ shows that equality holds in (5) if and only if $x = (\phi^*)'(\xi)$. ■

D. Examples of Gaussian gauges

The most commonly used gauge is the L^2 gauge given by $\phi(x) = x^2/2$; its polar gauge is $\phi^*(\xi) = \xi^2/2$. The corresponding ϕ-Gaussian is the standard Gaussian. More general is the L^p gauge, for each $p \in (1,\infty)$, where $\phi(x) = |x|^p/p$ and $\phi^*(\xi) = |\xi|^q/q$, where $p^{-1} + q^{-1} = 1$ (see, for example, Arnol'd [23] or Rockafellar [24]). The corresponding ϕ-Gaussian is the L^p Gaussian. Other examples of Gaussian gauges are given in §VI.

E. The ϕ-moment

Given a Gaussian gauge ϕ and the standard ϕ-Gaussian random variable X_ϕ, as defined in §II-B, assume that

$$-\infty < E[\phi(X_\phi)] < \infty$$

and denote $\hat{\phi} = E[\phi(X_\phi)]$. We then say that a continuous random variable X has *finite ϕ-moment,* if there exists $m > 0$ such that

$$P\left[x_- < \frac{X}{m} < x_+\right] = 1 \tag{6}$$

and

$$E\left[\phi\left(\frac{X}{m}\right)\right] = \hat{\phi}. \tag{7}$$

In that case, we define the *ϕ-moment* of X to be

$$m_\phi[X] = m. \tag{8}$$

If $\phi(x) = x^2/2$, then $m_\phi[X]$ is the second moment of X. If $\phi(x) = |x|^p/p$, then $m_\phi[X]$ is the usual p-th moment of X.

If c is a positive constant, then

$$m_\phi[cX] = cm_\phi[X].$$

By definition, if $-\infty < \hat{\phi} < \infty$, then $m_\phi[X_\phi] = 1$, and a random variable X is a ϕ-Gaussian if and only if $X/m_\phi[X]$ is the standard ϕ-Gaussian.

F. The ϕ-Fisher information

Given a Gaussian gauge ϕ, define the random variable Ξ_ϕ by

$$\Xi_\phi = \phi'(X_\phi)$$

and denote

$$\hat{\phi}^* = E[\phi^*(\Xi_\phi)], \tag{9}$$

if it exists. Observe that

$$\hat{\phi}^* = E[X_\phi \phi'(X_\phi)] - \hat{\phi}. \tag{10}$$

Given a continuous random variable X, let Ξ be the continuous random variable given by

$$\Xi = \begin{cases} (-\log f)'(X), & \text{if } f(X) > 0 \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

where f is the density function of X. We say that X has *finite ϕ-Fisher information,* if there exists $m^* > 0$ such that

$$P\left[\xi_- < \frac{\Xi}{m^*} < \xi_+\right] = 1 \tag{12}$$

and

$$E\left[\phi^*\left(\frac{\Xi}{m^*}\right)\right] = \hat{\phi}^*. \tag{13}$$

The *ϕ-Fisher information* of X is defined to be

$$m^*_\phi[X] = m^*.$$

If $\phi(x) = x^2/2$, then $\phi^*(\xi) = \xi^2/2$ and the ϕ-Fisher information is the classical Fisher information. If $\phi(x) = |x|^p/p$ for $p > 1$, then $\phi^*(\xi) = |\xi|^q/q$, where $p^{-1}+q^{-1} = 1$. In this case, the ϕ-Fisher information is the (p,λ)-th Fisher information with $\lambda = 1$, as introduced in §II-D of [8],

$$m^*_\phi[X] = \left(\int_{\mathbb{R}} |(\log f)'|^q f\right)^{1/q}.$$

If c is a positive constant, then

$$m^*_\phi[cX] = c^{-1} m^*_\phi[X].$$

Also, if $-\infty < \hat{\phi}^* < \infty$, then the ϕ-Fisher information of the ϕ-Gaussian X_ϕ is equal to 1.

Note that given a Gaussian gauge ϕ, the ϕ-moment and ϕ-Fisher information do not necessarily exist for a random variable. Assumptions on the probability density and its derivative must hold in order for these information measures to exist. Throughout the rest of this paper we restrict our attention to random variables for which these conditions hold.

G. Duality lemma

Lemma 3: Let ϕ be a Gaussian gauge such that $-\infty < \hat{\phi} < \infty$ and f_ϕ be the density function of the standard ϕ-Gaussian. Assume that

$$\begin{aligned} \ell_- &= \lim_{x \to x_-} x f_\phi(x) \\ \ell_+ &= \lim_{x \to x_+} x f_\phi(x) \end{aligned} \tag{14}$$

exist. Then $\hat{\phi}^*$ exists, and

$$\hat{\phi} + \hat{\phi}^* = 1 - \ell_+ + \ell_-.$$

Proof: Denote

$$M = \int_{x_-}^{x_+} e^{-\phi(x)}\,dx.$$

By (10), integration by parts, and (14),

$$\begin{aligned} \hat{\phi} + \hat{\phi}^* &= E[X_\phi \phi'(X_\phi)] \\ &= \frac{1}{M}\int_{x_-}^{x_+} x\phi'(x) e^{-\phi(x)}\,dx \\ &= -\frac{1}{M}\int_{x_-}^{x_+} x(e^{-\phi(x)})'\,dx \\ &= -\ell_+ + \ell_- + \frac{1}{M}\int_{x_-}^{x_+} e^{-\phi(x)}\,dx \\ &= 1 - \ell_+ + \ell_-. \end{aligned}$$

■

III. Main theorem

Using the definitions from the previous section, we can now state the main theorem of this paper.

Theorem 1: Let $\phi : \mathbb{R} \to \mathbb{R}$ be a Gaussian gauge and f_ϕ the density function of the standard ϕ-Gaussian. If Y is a continuous random variable with continuously differentiable density function $f : \mathbb{R} \to [0, \infty)$ such that

$$m_\phi[Y], m_\phi^*[Y] < \infty, \tag{15}$$

then

$$m_\phi[Y] m_\phi^*[Y] \geq 1. \tag{16}$$

Equality holds if and only if Y is a ϕ-Gaussian.

The rest of this paper is devoted to the proof of this theorem.

IV. The generalized Hölder inequality

The following is a version of the Hölder inequality due to Zygmund [25].

Lemma 4: Let ϕ be a gauge and ϕ^* its dual gauge such that $-\infty < \hat{\phi}, \hat{\phi}^* < \infty$. If m and m^* are positive reals and X and Y random variables such that

$$\begin{aligned} E\left[\phi\left(\frac{X}{m}\right)\right] &= \hat{\phi} \\ E\left[\phi^*\left(\frac{Y}{m^*}\right)\right] &= \hat{\phi}^*, \end{aligned} \tag{17}$$

then

$$E[XY] \leq (\hat{\phi} + \hat{\phi}^*) m m^*, \tag{18}$$

with equality if and only if

$$\frac{Y}{m^*} = \phi'\left(\frac{X}{m}\right) \tag{19}$$

with probability one.

Proof: By (5) and (17),

$$\begin{aligned} \frac{E[XY]}{mm^*} &= E\left[\frac{X}{m}\frac{Y}{m^*}\right] \\ &\leq E\left[\phi\left(\frac{X}{m}\right) + \phi^*\left(\frac{Y}{m^*}\right)\right] \\ &= E\left[\phi\left(\frac{X}{m}\right)\right] + E\left[\phi^*\left(\frac{Y}{m^*}\right)\right] \\ &= \hat{\phi} + \hat{\phi}^*, \end{aligned} \tag{20}$$

proving (18). Equality holds in (18) if and only if it holds in (20). This in turn holds if and only if

$$\frac{X}{m}\frac{Y}{m^*} = \phi\left(\frac{X}{m}\right) + \phi^*\left(\frac{Y}{m^*}\right)$$

with probability one. By Lemma 2, this holds if and only if (19) holds. ∎

V. Cramér-Rao inequalities

We now prove the following general theorem.

Theorem 2: Let $\phi : (x_-, x_+) \to \mathbb{R}$ be a Gaussian gauge with ϕ-Gaussian X_ϕ, and Y a continuous random variable with continuously differentiable density function $f : (y_-, y_+) \to [0, \infty)$. Assume that

$$m_\phi[Y], m_\phi^*[Y] < \infty,$$

and that $\lim_{y \to y_\pm} y f(y)$ exist and are both finite. Define

$$\epsilon_\pm = \lim_{y \to y_\pm} y f(y). \tag{21}$$

Then

$$(\hat{\phi} + \hat{\phi}^*) m_\phi[Y] m_\phi^*[Y] \geq 1 - \epsilon_+ + \epsilon_-. \tag{22}$$

Equality holds in (22) if and only if Y is a $\tilde{\phi}$-Gaussian, where $\tilde{\phi} = m_\phi[Y] m_\phi^*[Y] \phi$. If, in addition, $\epsilon_+ - \epsilon_- = \ell_+ - \ell_-$, then equality holds in (22) if and only if Y is a ϕ-Gaussian.

Proof: Let Ξ be the random variable given by

$$\Xi = \begin{cases} -f'(Y)/f(Y) & \text{if } f(Y) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

By integration by parts, (21), Lemma 4, and equations (7) and (13),

$$\begin{aligned} 1 &= \int_{y_-}^{y_+} f(y)\, dy \\ &= \epsilon_+ - \epsilon_- + \int_{y_-}^{y_+} -y f'(y)\, dy \\ &= \epsilon_+ - \epsilon_- + \int_{y_-}^{y_+} y(-f'(y)/f(y)) f(y)\, dy \\ &= \epsilon_+ - \epsilon_- + E[Y\Xi] \\ &\leq \epsilon_+ - \epsilon_- + (\hat{\phi} + \hat{\phi}^*) m_\phi[Y] m_\phi^*[Y], \end{aligned}$$

proving the inequality.

By the equality condition of Lemma 4 and the continuity of f and f', equality holds if and only if, for each $y \in (y_-, y_+)$ such that $f(y) > 0$,

$$(-\log f)'(y) = \frac{d}{dy}[mm^* \phi(y/m)],$$

where $m = m_\phi[Y]$ and $m^* = m_\phi^*[Y]$. This in turn implies that on each interval in (y_-, y_+) where $f > 0$ there exists $\tilde{a} > 0$ such that

$$f(y) = (\tilde{a}/m) e^{-\tilde{\phi}(y/m)}, \tag{23}$$

where $\tilde{\phi}$ is as defined in the statement of this theorem. Since the right side of (23) is positive for any $y \in (y_-, y_+)$, it follows that each interval on which $f > 0$ is relatively open and closed. Since f is assumed to be a probability density, this set must also be nonempty. We conclude that f is strictly positive and (23) holds on the entire interval (y_-, y_+). Therefore, Y is a $\tilde{\phi}$-Gaussian. If $\epsilon_+ - \epsilon_- = \ell_+ - \ell_-$ and equality holds in (22), then by Lemma 3, $mm^* = 1$ and $\tilde{\phi} = \phi$. ∎

To prove Theorem 1, first observe that if ϕ is a Gaussian gauge on $\mathbb{R}$, then there exists $c > 0$ and $R > 0$ such that $\phi(y) \geq c|y|$ if $|y| > R$. Therefore, since $m = m_\phi[Y] < \infty$,

$$\int_R^\infty yf(y)\,dy < \frac{m}{c}\int_R^\infty \phi\left(\frac{y}{m}\right) f(y)\,dy < \infty.$$

It follows that

$$\liminf_{y\to+\infty} yf(y) = 0.$$

In particular, given any $\epsilon > 0$ and $R' > 0$, there exists $R > R'$ such that

$$Rf(R) < \frac{\epsilon}{2}. \tag{24}$$

On the other hand, since f is a probability density on $\mathbb{R}$ and (15) holds, it follows that for any $\epsilon > 0$, there exists $R' > 0$ such that for any $R > R'$,

$$\left|\int_R^y \left[1 + mm^*\left(\phi\left(\frac{x}{m}\right) + \phi^*\left(\frac{f'(x)/f(x)}{m^*}\right)\right)\right] f(x)\,dx\right| < \frac{\epsilon}{2}. \tag{25}$$

Therefore, given any $\epsilon > 0$, there exists $R > 0$ such that both (24) and (25) hold and therefore for any $y > R$,

$$\begin{aligned}
&yf(y)\\
&= Rf(R) + \int_R^y f(x) + xf'(x)\,dx\\
&= Rf(R) + \int_R^y f(x) + x(f'(x)/f(x))f(x)\,dx\\
&\leq Rf(R)\\
&\quad + \int_R^\infty \left[1 + mm^*\left(\phi\left(\frac{x}{m}\right) + \phi^*\left(\frac{f'(x)/f(x)}{m^*}\right)\right)\right] f(x)\,dx\\
&< \epsilon.
\end{aligned}$$

This and a similar argument for R and y negative applied to both f and f_ϕ gives

$$\lim_{y\to\pm\infty} yf(y) = \lim_{y\to\pm\infty} yf_\phi(y) = 0.$$

Theorem 1 now follows from Theorem 2, and Lemma 3.

VI. Examples

In this section we present specific examples of the Cramér-Rao inequality implied by the unified Cramér-Rao inequality given by Theorem 1. In §VI-A and §VI-B we show how the classical (non-Bayesian) Cramér-Rao inequality for the location parameter and the L^p generalization by Batalama and Kazakos [1] (also, see [8]) are special cases of Theorem 1.

In §VI-C and §VI-D we present two Cramér-Rao inequalities that, as far as we know, are new. In §VI-C we define a Gaussian gauge ϕ naturally associated with the logistic random variable and use it to define two new information measures of a random variable, namely the "logistic moment" and "logistic Fisher information"'. Theorem 1 then implies that these information measures satisfy a sharp Cramér-Rao inequality with equality holding if and only if the random variable is logistic. In §VI-D we do the same for the power function distribution.

A. The classical Cramér-Rao inequality

Theorem 2 with $\phi(x) = x^2/2$ yields the following.

Corollary 1: ([2], [3]) If X is a real random variable with density function f that is continuously differentiable on $\mathbb{R}$ and such that

$$E[X^2], E[(-f'(X)/f(X))^2] < \infty,$$

then

$$E[X^2]E[(-f'(X)/f(X))^2] \geq 1,$$

with equality holding if and only if X is Gaussian.

B. The L^p Cramér-Rao inequality

Theorem 2 with $\phi(x) = |x|^p/p$, where $1 < p < \infty$, yields the following.

Corollary 2: (the case $p > 1$ and $\lambda = 1$ of Theorem 5 in [8]) If $1 < p, q < \infty$ satisfy

$$\frac{1}{p} + \frac{1}{q} = 1$$

and X is a real random variable with density function f that is continuously differentiable on $\mathbb{R}$ and such that

$$E[|X|^p], E[|-f'(X)/f(X)|^q] < \infty,$$

then

$$E[|X|^p]^{1/p}E[|-f'(X)/f(X)|^q]^{1/q} \geq 1,$$

with equality holding if and only if the density function of X is of the form

$$f(x) = ae^{-b|x|^p},$$

where a and b are positive constants.

C. The Cramér-Rao inequality for the logistic random variable

If $\phi : \mathbb{R} \to \mathbb{R}$ is given by

$$\phi(x) = x + 2\log(1 + e^{-x}),$$

then the probability distribution for any constant multiple of the ϕ-Gaussian X_ϕ is known as a logistic distribution [26]. The dual gauge is the function $\phi^* : (-1, 1) \to \mathbb{R}$ given by

$$\phi^*(\xi) = (1 + \xi)\log\left(\frac{1+\xi}{2}\right) + (1 - \xi)\log\left(\frac{1-\xi}{2}\right).$$

A straightforward calculation shows that $\hat{\phi} = 2$ and $\hat{\phi}^* = -1$.

Let X be a continuous random variable with density function f that is continuously differentiable on $\mathbb{R}$. Define the *logistic moment* of X to be $m > 0$ be such that

$$E\left[\frac{X}{m} + 2\log(1 + e^{-X/m})\right] = 2.$$

Denoting

$$\Xi = \begin{cases}(-\log f)'(X) & \text{if } f(X) > 0\\ 0 & \text{otherwise,}\end{cases}$$

suppose there exists $m^* > 0$ such that

$$P[-m^* < \Xi < m^*] = 1.$$

and

$$E\left[\log\left(\frac{1+\Xi/m^*}{2}\right)^{1+\Xi/m^*}\left(\frac{1-\Xi/m^*}{2}\right)^{1-\Xi/m^*}\right] = -1.$$

Call m^* the *logistic Fisher information* of X.

Theorem 2 then implies the following.

Corollary 3: The logistic moment m and Fisher information m^* of X satisfy

$$mm^* \geq 1,$$

with equality holding if and only if X is logistic.

D. A Cramér-Rao inequality for the power function distribution

Here, we state and prove a Cramér-Rao inequality, where the extremal distribution is a power function.

Corollary 4: Let X be a continuous random variable with density f satisfying the following:

$$P[0 < X < 1] = 1;$$

$$f : [0,1] \to [0,\infty) \text{ is continuous;}$$

$$f \text{ is continuously differentiable on } (0,1); \tag{26}$$

$$f' > 0 \text{ on } (0,1);$$

$$E[\log X] > -1; \tag{27}$$

$$-\infty < E[\log(f'(X)/f(X))] < \infty; \tag{28}$$

$$\log\left(-1 - \frac{1}{E[\log X]}\right) \leq E[\log X] + E[\log(f'(X)/f(X))]; \tag{29}$$

$$\inf_{0<x<1} \log(f'(x)/f(x)) \geq E[\log X] + E[\log(f'(X)/f(X))]. \tag{30}$$

Then

$$E[\log X] + E[\log(f'(X)/f(X))] \leq \log(f(1) - 1). \tag{31}$$

Equality holds if and only if $f(x) = (p+1)x^p$, where $p > 0$ satisfies

$$p + 1 = -\frac{1}{E[\log X]}. \tag{32}$$

Proof: Given a random variable X satisfying the assumptions of the theorem, let $p > 0$ be given by (32). Define the gauge $\phi : (0,1) \to \mathbb{R}$ by

$$\phi(x) = -p\log x.$$

The ϕ-Gaussian X_ϕ has density

$$f_\phi(x) = (p+1)x^p,$$

and the dual gauge is $\phi^* : (-\infty, -p) \to \mathbb{R}$, where

$$\phi^*(\xi) = p(-1 + \log p - \log(-\xi)).$$

A straightforward calculation shows that

$$\begin{aligned}\hat{\phi} &= \frac{p}{p+1};\\ \hat{\phi}^* &= -\frac{p(p+2)}{p+1};\\ \hat{\phi} + \hat{\phi}^* &= -p.\end{aligned}$$

Given a random variable X on $(0,1)$, another straightforward calculation shows that if (27) holds, then X has a finite ϕ-moment and

$$m_\phi[X] = 1. \tag{33}$$

Similarly, if (27), (28), (29), and (30) hold, then X has a finite ϕ-Fisher information and

$$\log m_\phi^*[X] = E[\log X] + E[\log(f'(X)/f(X))] - \log p. \tag{34}$$

We also claim that $f(1) > 1$. If not, then since f is strictly increasing, it follows that $f < 1$ on the interval $(0,1)$ and therefore $\int_0^1 f(x)\,dx < 1$, which contradicts the assumption that f is a probability density on $(0,1)$.

Theorem 2 now implies that

$$\log m_\phi[X] + \log m_\phi^*[X] + \log p \leq \log(f(1) - 1).$$

Substituting equations (33) and (34) into this inequality gives (31). If equality holds, then by Theorem 2, $f(x) = (p+1)x^p$. ■

It is worth noting that since $\hat{\phi} + \hat{\phi}^* < 0$, this version of the Cramér-Rao inequality gives an upper instead of the usual lower bound for product of the generalized moment and Fisher information.

VII. Identification of a Power Function Distribution and Its Power

The classical Cramér-Rao inequality (2) shows how to identify a Gaussian random variable from its variance and Fisher information. Similarly, Corollary 4 shows how to identify a power function distribution using generalized information measures.

In particular, let Y be a random variable with density g such that the following hold:

$$P[0 < Y < L] = 1;$$

$$g : [0, L] \to [0, \infty) \text{ is continuous};$$

$$g \text{ is continuously differentiable on } (0, L);$$

$$g' < 0 \text{ on } (0, L);$$

$$E[\log(1 - Y/L)] > -1;$$

$$-\infty < E[\log(-g'(Y)/g(Y))] < \infty;$$

$$\log\left(-1 - \frac{1}{E[\log(1 - Y/L)]}\right) \leq E[\log(1 - Y/L)] + E[\log(-g'(Y)/g(Y))];$$

$$\inf_{0<y<L} \log(-g'(y)/g(y)) \geq E[\log(1 - Y/L)] + E[\log(-g'(Y)/g(Y))].$$

By setting $Y = L(1-X)$, the assumptions above for Y imply the assumptions for X in Corollary 4. The fact that

$$f(1) > 1$$

implies that

$$g(0) > 1/L.$$

The corollary therefore implies that any random variable Y satisfying the assumptions above, the inequality

$$E[\log(L - Y)] + E[\log(-g'(Y)/g(Y))] \leq \log(Lg(0) - 1) \tag{35}$$

holds. Moreover, equality holds if and only if Y has a decreasing power function density function

$$g(y) = \frac{p+1}{L^{p+1}}(L - y)^p,$$

where the power p is given by

$$p + 1 = \frac{1}{\log L - E[\log(L - Y)]}.$$

Fig 1 shows the graph of several possible power function distribution functions with $L = 1$.

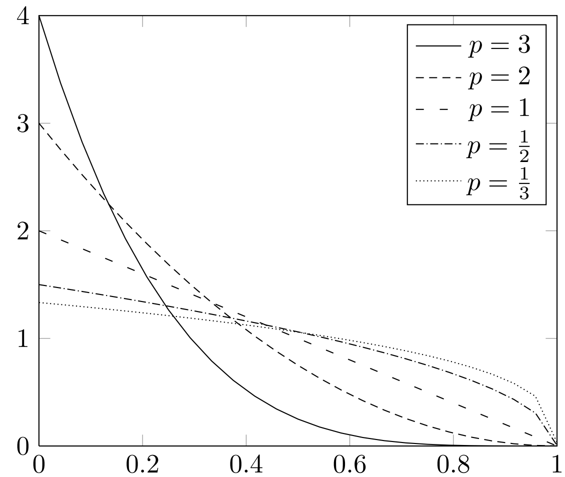

Fig. 1. Decreasing p-th power density functions on $[0, 1]$

References

[1] S. Batalama and D. Kazakos, "On the generalized Cramer-Rao bound for the estimation of the location," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 487–492, 1997.

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

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

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

[5] I. Vajda, "χ^α-divergence and generalized Fisher's information," in *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Tech. Univ. Prague, Prague, 1971; dedicated to the memory of Antonín Špaček)*. Prague: Academia, 1973, pp. 873–886.

[6] C. P. Kitsos and N. K. Tavoularis, "Logarithmic Sobolev inequalities for information measures," *IEEE Trans. Inform. Theory*, vol. 55, no. 6, pp. 2554–2561, 2009.

[7] D. E. Boekee, "Generalized Fisher information with application to estimation problems," in *Information and systems (Proc. IFAC Workshop, Compiègne, 1977)*. Oxford: Pergamon, 1978, pp. 75–82.

[8] 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.

[9] J.-F. Bercher, "On escort distributions, q-gaussians and Fisher information," in *Bayesian inference and maximum entropy methods in science and engineering*, ser. AIP Conf. Proc. Melville, NY: Amer. Inst. Phys., 2010, vol. 1305, pp. 208–215.

[10] ——, "Escort entropies and divergences and related canonical distribution," *Phys. Lett. A*, vol. 375, no. 33, pp. 2969–2973, 2011.

[11] ——, "On a (β, q)-generalized Fisher information and inequalities involving q-Gaussian distributions," *J. Math. Phys.*, vol. 53, no. 6, pp. 063 303, 12, 2012.

[12] ——, "On generalized Cramér-Rao inequalities, generalized Fisher information and characterizations of generalized q-Gaussian distributions," *J. Phys. A*, vol. 45, no. 25, pp. 255 303, 15, 2012.

[13] J.-F. Bercher and C. Vignat, "On minimum Fisher information distributions with restricted support and fixed variance," *Inform. Sci.*, vol. 179, no. 22, pp. 3832–3842, 2009.

[14] E. Uhrmann-Klingen, "Minimal Fisher information distributions with compact-supports," *Sankhyā Ser. A*, vol. 57, no. 3, pp. 360–374, 1995.

[15] I. Vajda, "On convergence of information contained in quantized observations," *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 2163–2172, 2002.

[16] V. Fabian and J. Hannan, "On the Cramér-Rao inequality," *Ann. Statist.*, vol. 5, no. 1, pp. 197–205, 1977.

[17] A. J. Lenstra, "Cramér-Rao revisited," *Bernoulli*, vol. 11, no. 2, pp. 263–282, 2005.

[18] S. Furuichi, "On the maximum entropy principle and the minimization of the Fisher information in Tsallis statistics," *J. Math. Phys.*, vol. 50, no. 1, pp. 013 303, 12, 2009.

[19] B. Grechuk, A. Molyboha, and M. Zabarankin, "Maximum entropy principle with general deviation measures," *Mathematics of Operations Research*, vol. 34, no. 2, pp. 445–467, 2009.

[20] G. Gilardoni, "On Pinsker's and Vajda's type inequalities for Csiszr's f-divergences," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5377–5386, 2010.

[21] J. Bercher, "On multidimensional generalized cramr-rao inequalities, uncertainty relations and characterizations of generalized q-gaussian distributions," *Journal of Physics A: Mathematical and Theoretical*, vol. 46, no. 9, 2013.

[22] J.-F. Bercher, "Some properties of generalized fisher information in the context of nonextensive thermostatistics," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 15, pp. 3140–3154, 2013.

[23] V. I. Arnol′d, *Mathematical methods of classical mechanics*, ser. Graduate Texts in Mathematics. New York: Springer-Verlag, 1989, vol. 60.

[24] R. T. Rockafellar, *Convex analysis*, ser. Princeton Mathematical Series, No. 28. Princeton, N.J.: Princeton University Press, 1970.

[25] A. Zygmund, *Trigonometric series. 2nd ed. Vols. I, II*. New York: Cambridge University Press, 1959.

[26] Wikipedia. (2013) Logistic distribution. [Online]. Available: http://en.wikipedia.org/wiki/Logistic_distribution

Andrea Cianchi Andrea Cianchi received his *Laurea* and Ph.D. degrees in Mathematics from Università degli Studi di Firenze (Firenze, Italia), where he is now Professor of Mathematical Analysis.

Erwin Lutwak Erwin Lutwak is Professor of Mathematics at Polytechnic Institute of NYU. He received his B.S. and M.S. in Mathematics from this institution when it was named Polytechnic Institute of Brooklyn and his Ph.D. in Mathematics when it was named Polytechnic Institute of New York.

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 is Professor of Mathematics at Polytechnic Institute of NYU (Brooklyn, New York).

Gaoyong Zhang Gaoyong Zhang received his B.S. in Mathematics from Wuhan University of Science and Technology, M.S. in Mathematics from Wuhan University, Wuhan, China, and Ph.D. in Mathematics from Temple University, Philadelphia. He is Professor of Mathematics at Polytechnic Institute of NYU (Brooklyn, New York).