It is understood that in the case where ϕ is multiform, then so is $\diamondsuit(f)$. We assume that $f\diamondsuit(f)$ decreases to zero on the boundaries $\mathcal{B}$ of the domain $\mathcal{D}$ of f. In such case, by integration by part, we have

$$\int_{\mathcal{D}} \diamondsuit(f)\mathrm{d}x = [x\diamondsuit(f)]_{\mathcal{B}} - \int_{\mathcal{D}} x\dot{\diamondsuit}(f)\mathrm{d}x,$$
$$= -\int_{\mathcal{D}} x\dot{\diamondsuit}(f)\mathrm{d}x = -\int_{\mathcal{D}} x\frac{\dot{\diamondsuit}(f)}{f(x)}f(x)\,\mathrm{d}x.$$

By the Hölder inequality applied to the last integral, we obtain

$$\int_{\mathcal{D}} \diamondsuit(f)\mathrm{d}x \leq \left(\int_{\mathcal{D}} |x|^{\alpha} f(x)\,\mathrm{d}x\right)^{\frac{1}{\alpha}} \left(\int_{\mathcal{D}} \left|\frac{\dot{\diamondsuit}(f)}{f(x)}\right|^{\beta} f(x)\,\mathrm{d}x\right)^{\frac{1}{\beta}}$$

where $1/\alpha + 1/\beta = 1$. The inequality can also be rewritten as

$$\left(\int_{\mathcal{D}} |x|^{\alpha} f(x)\,\mathrm{d}x\right)^{\frac{1}{\alpha}} \frac{\left(\int_{\mathcal{D}} \left|\frac{\dot{f}(x)}{f(x)}\dot{\diamondsuit}(f)\right|^{\beta} f(x)\,\mathrm{d}x\right)^{\frac{1}{\beta}}}{\int_{\mathcal{D}} \diamondsuit(f)\mathrm{d}x} \geq 1.$$

This has the form of a generalized Cramér-Rao inequality, where the first term is the moment of order α and the second one is a generalized ϕ-Fisher information of order β.

A Cramér-Rao inequality for the estimation of a parameter

The problem of estimation is to determine a function $\hat{\theta}(x)$ in order to estimate an unknown parameter θ. Let $f(x;\theta)$ and $g(x;\theta)$ be two probability density functions, with $x \in X \subseteq \mathbb{R}^k$ and θ a parameter of these densities, $\theta \in \mathbb{R}^n$. An underlying idea in the statement of the new Cramér-Rao inequality is that it is possible to evaluate the moments of the error with respect to different probability distributions. For instance, in the estimation setting the estimation error is $\hat{\theta}(x) - \theta$. The bias can be evaluated with respect to f according to

$$B_f(\theta) = \int_X \left(\hat{\theta}(x) - \theta\right) f(x;\theta)\,\mathrm{d}x = \mathrm{E}_f\left[\hat{\theta}(x) - \theta\right] \tag{18}$$

Theorem 1. *Let $f(x;\theta)$ be a multivariate probability density function defined over a subset $X \subseteq \mathbb{R}^n$, and $\theta \in \Theta \subseteq \mathbb{R}^k$ a parameter of the density. The set Θ is equipped with a norm $\|.\|$, and the corresponding dual norm is denoted $\|.\|_*$. Let $g(x;\theta)$ denote another probability density function also defined on $(X;\Theta)$. Assume that $f(x;\theta)$ is a jointly measurable function of x and θ, is integrable with respect to x, is absolutely continuous with respect to θ, and that the derivatives with respect to each component of θ are locally integrable. For any estimator $\hat{\theta}(x)$ of θ, we have*

$$E\left[\left\|\hat{\theta}(x) - \theta\right\|^{\alpha}\right]^{\frac{1}{\alpha}} I_{\beta}[f|g;\theta]^{\frac{1}{\beta}} \geq |n + \nabla_{\theta}.\,B_f(\theta)| \tag{19}$$

with α and β Hölder conjugates of each other, i.e. $\alpha^{-1} + \beta^{-1} = 1$, $\alpha \geq 1$, and where the (β, g)-Fisher information

$$I_{\beta}[f|g;\theta] = \int_X \left\|\frac{\nabla_{\theta} f(x;\theta)}{g(x;\theta)}\right\|_*^{\beta} g(x;\theta)\ dx \tag{20}$$

is the generalized Fisher information of order β on the parameter θ contained in the distribution f and taken with respect to g. The equality case is obtained if

$$\frac{\nabla_{\theta} f(x;\theta)}{g(x;\theta)} = K\left\|\hat{\theta}(x) - \theta\right\|^{\alpha-1} \nabla_{\hat{\theta}(x)-\theta}\|\hat{\theta}(x) - \theta\|, \tag{21}$$

with $K > 0$.

Proof. The bias in (18) is a n-dimensional vector. Let us consider its divergence with respect to variations of θ:

$$\operatorname{div} B_f(\theta) = \nabla_\theta .\, B_f(\theta). \tag{22}$$

The regularity conditions in the statement of the theorem enable to interchange integration with respect to x and differentiation with respect to θ, and

$$\nabla_\theta .\, B_f(\theta) = \int_X \nabla_\theta .\left(\hat{\theta}(x) - \theta\right) f(x;\theta)\,\mathrm{d}x + \int_X \nabla_\theta f(x;\theta).\left(\hat{\theta}(x) - \theta\right) \mathrm{d}x. \tag{23}$$

In the first term on the right, we have $\nabla_\theta.\theta = n$, and the integral reduces to $-n\int_X f(x;\theta)\,\mathrm{d}x = -n$, since $f(x;\theta)$ is a probability density on X. The second term can be rearranged so as to obtain an integration with respect to the density $g(x;\theta)$, assuming that the derivatives with respect to each component of θ are absolutely continuous with respect to $g(x;\theta)$, i.e. $g(x;\theta) \gg \nabla_\theta f(x;\theta)$. This gives

$$n + \nabla_\theta .\, B_f(\theta) = \int_X \frac{\nabla_\theta f(x;\theta)}{g(x;\theta)}.\left(\hat{\theta}(x) - \theta\right) g(x;\theta)\,\mathrm{d}x. \tag{24}$$

Now, it only remains to apply the generalized Hölder-type inequality (**??**) in Lemma **??** to the integral on the right side, with $X(x) = \hat{\theta}(x) - \theta$, $Y(x) = \frac{\nabla_\theta f(x;\theta)}{g(x;\theta)}$, and $w(x) = g(x;\theta)$. This yields in all generality

$$\left(\int_X \left\|\hat{\theta}(x) - \theta\right\|^\alpha g(x;\theta)\ \mathrm{d}x\right)^{\frac{1}{\alpha}} \left(\int_X \left\|\frac{\nabla_\theta f(x;\theta)}{g(x;\theta)}\right\|_*^\beta g(x;\theta)\ \mathrm{d}x\right)^{\frac{1}{\beta}} \geq |n + \nabla_\theta .\, B_f(\theta)| \tag{25}$$

which is (19). By Lemma **??** again, we know that the case of equality occurs if $Y(t) = K\|X(t)\|^{\alpha-1}\nabla_{X(t)}\|X(t)\|$, $K > 0$, which gives (21). □

5. Some examples

5.1. Normal distribution and second-order moment

For a normal distribution, and second order moment constraint

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{x^2}{2\,\sigma^2}\right) \qquad \text{and} \qquad T_1(x) = x^2 \qquad \text{on} \qquad \mathcal{X} = \mathbb{R}.$$

We begin by computing the inverse of $y = f_X(x)$ where $x \in \mathbb{R}_+$ for instance, which gives

$$\phi'(y) = \left(\lambda_0 - \sigma^2\log(2\pi\sigma^2)\,\lambda_1\right) - 2\sigma^2\lambda_1\log y.$$

The judicious choice

$$\lambda_0 = 1 - \log(\sqrt{2\pi}\sigma) \qquad \text{and} \qquad \lambda_1 = -\frac{1}{2\,\sigma^2}$$

leads to function

$$\phi(y) = y\log y$$

that gives nothing more than the Shannon entropy as expected.