Entropy **2015**, *17*, 4918-4939; doi:10.3390/e17074918

OPEN ACCESS
entropy
ISSN 1099-4300
www.mdpi.com/journal/entropy

Article

Fisher Information Properties

Pablo Zegers

Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Monseñor Álvaro del Portillo 12.455, Las Condes, Santiago, Chile; E-Mail: pzegers@miuandes.cl

Academic Editor: Raúl Alcaraz Martínez

Received: 18 June 2015 / Accepted: 10 July 2015 / Published: 13 July 2015

Abstract: A set of Fisher information properties are presented in order to draw a parallel with similar properties of Shannon differential entropy. Already known properties are presented together with new ones, which include: (i) a generalization of mutual information for Fisher information; (ii) a new proof that Fisher information increases under conditioning; (iii) showing that Fisher information decreases in Markov chains; and (iv) bound estimation error using Fisher information. This last result is especially important, because it completes Fano's inequality, *i.e.*, a lower bound for estimation error, showing that Fisher information can be used to define an upper bound for this error. In this way, it is shown that Shannon's differential entropy, which quantifies the behavior of the random variable, and the Fisher information, which quantifies the internal structure of the density function that defines the random variable, can be used to characterize the estimation error.

Keywords: Fisher information; Cramer–Rao bound; Shannon differential entropy; Markov chains

1. Introduction

The birth of information theory was signaled by the publication of Claude Shannon's work [1], which is based on studying the behavior of systems described by density functions. However, much before that work was published, Ronald Fisher had already published the definition of a quantity called Fisher information [2], a hard bound on the capacity to estimate the parameters that define a system [3,4]. Hence, this quantity regulates how well it is possible to determine the internal structure of a system and provides another point of view that can be used to study systems: how they are composed, what they are made of. This work springs from the belief that the combination of these approaches is what completely defines systems: their behavior (Shannon) and their architecture (Fisher). In the following, a series of published results is summarized, together with new results, in order to present a coherent set of Fisher information properties that will hopefully be useful for those that work with this quantity.

1.1. Fisher Information and Other Fields

One connection between Fisher information and the Shannon differential entropy was stated by Kullback [5] (p. 26), who proved that the second derivatives of the Kullback–Leibler divergence with respect to the density functions parameters produce the Fisher information matrix terms. Related results were presented by Blahut [6] (p. 300), and Frieden [7] (p. 37). Another important result that also relates these two frameworks is Bruijn's identity ([8,9] and [10] (p. 672)), which establishes a relation between the derivative of Shannon differential entropy and Fisher information when the underlying random variable is the subject of Gaussian perturbations. This result was recently generalized to non-Gaussian perturbations [11,12]. A consequence of these results is the convolution inequality for Fisher information ([8,9,13–16]; [10] (p. 674)).

Others have been studying the relation between Fisher information and physics. Here, it is important to point out the extreme physical information principle derived by Frieden and others in order to establish a general framework that explains physics [7,17–20]. Of special interest has been the role of Fisher information to generate thermodynamical theory [7,17–22]. It is very common in these approaches to use a special case of Fisher information where the estimated parameter is a location parameter. In this work, the original and general Fisher information definitions, and not the later special case, are addressed only.

Even thought Shannon's ideas have been part of the the machine learning tool set for a long time, Fisher information has not followed the same track. Even though Fisher information is intimately connected to estimation theory [23], its use in the development of learning systems has not been well developed yet. Nevertheless, Amari discovered that natural gradient descent, *i.e.*, common gradient descent corrected with the Fisher information matrix terms, takes into account the topology in a more precise manner, allowing for more efficient training procedures [24,25]. The use of Fisher information has also been taken into account in order to design objective functions to lead the estimation procedure. One of them is mixing maximum entropy with minimum Fisher information [26,27]. On the other hand, mixing Shannon's differential entropy, Fisher information and the central limit theorem has allowed proving that in the presence of large datasets, it is natural to search for minimum Kullback–Leibler, or equivalent, solutions [28].

1.2. Contribution of This Work

This work is focused on presenting already known properties of Fisher information [3,4,7,8,10,29–32] and introducing new ones, such that the reader can have a better grasp of Fisher information and its usefulness. The main results presented in this work are: (i) the generalization of the mutual information concept using Fisher information expressions; (ii) a new proof that conditioning under certain assumptions increases Fisher information; (iii) proving that in Markov chains, the Fisher information increases as the random variables become further away from the estimated parameter; and (iv) an upper bound on estimation error, which is regulated by the Fisher information.

This work is structured roughly in the same way in which is organized the first chapter of the well-known book of Cover and Thomas [30], in order to help the reader to draw a parallel between Shannon and Fisher information.

2. Notation

In the following sections, vectors and matrices are denoted with a bold font [7,31]. Furthermore, density functions are denoted by $f_{\mathbf{X};\boldsymbol{\theta}} \equiv f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})$, where the f is reserved for density functions, the lowercase $\mathbf{X}$ corresponds to the name of the random variable, $\boldsymbol{\theta}$ represents the parameters that define the density function and the symbol within the $(\cdot)$ stands for the instance of the random variable that is used to evaluate the density function. In this way, as an example, a different random variable could be denoted by $f_{\mathbf{Y};\boldsymbol{\theta}} \equiv f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})$. A similar notation is used in [33].

3. Fisher Information

Let there be a random variable $\mathbf{X}$ and its associated density function $f_{\mathbf{X};\boldsymbol{\theta}} \equiv f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})$, which has a support $\mathcal{S}$, and it depends on a set of parameters that is represented by the vector $\boldsymbol{\theta} \in \Theta$. The value θ_k is the k-th component of $\boldsymbol{\theta}$. According to the original definition designed by Fisher to characterize maximum likelihood estimation [2]:

Definition 1 (Fisher Information)**.** *Given a random variable* $\mathbf{X}$ *and its associated density function* $f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})$*, which depends on the parameter vector* $\boldsymbol{\theta} \in \Theta$*, and* θ_k *is the* k*-th component of* $\boldsymbol{\theta}$*, then the Fisher information associated with* θ_k *is defined by:*

$$i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} \equiv \int f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) \left(\frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k}\right)^2 d\mathbf{x} \tag{1}$$

From the definition, it is clear that $i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} \geq 0$. Furthermore, if $f_{\mathbf{X}}$ does not depend on θ_k, then $i_F\left(f_{\mathbf{X}}\right)_{\theta_k} = 0$.

Example 1. *In a Gaussian case with mean* μ *and standard deviation* η*, the density function is given by:*

$$f_{X;\mu,\eta}(x) = \frac{1}{\sqrt{2\pi}\eta} \exp\left(-\frac{(x-\mu)^2}{\eta^2}\right) \tag{2}$$

In this case:

$$\begin{aligned} \ln f_{X;\mu,\eta}(x) &= \ln \frac{1}{\sqrt{2\pi}\eta} - \frac{(x-\mu)^2}{2\eta^2} && (3) \\ &= \ln \frac{1}{\sqrt{2\pi}\eta} - \frac{x^2 - 2\mu x + \mu^2}{2\eta^2} && (4) \end{aligned}$$

If the parameter to be estimated is the mean μ, the previous expression needs to be derived with respect to μ:

$$\begin{aligned} \frac{d \ln f_{X;\mu,\eta}(x)}{d\mu} &= \frac{2x}{2\eta^2} - \frac{2\mu}{2\eta^2} && (5) \\ &= \frac{x-\mu}{\eta^2} && (6) \end{aligned}$$

Replacing into the definition of Fisher information definition:

$$\begin{aligned} i_F\left(f_{X;\mu,\eta}\right)_\mu &= \int f_{X;\mu,\eta}(x)\left(\frac{x-\mu}{\eta^2}\right)^2 dx && (7) \\ &= \int f_{X;\mu,\eta}(x)\frac{x^2-2\mu x+\mu^2}{\eta^4}dx && (8) \\ &= \frac{1}{\eta^4}\int f_{X;\mu,\eta}(x)x^2dx - \frac{2\mu}{\eta^4}\int f_{X;\mu,\eta}(x)xdx + \frac{\mu^2}{\eta^4}\int f_{X;\mu,\eta}(x)dx && (9) \\ &= \frac{1}{\eta^4}\left\{(\eta^2+\mu^2) - 2\mu^2 + \mu^2\right\} && (10) \\ &= \frac{1}{\eta^2} && (11) \end{aligned}$$

This shows that for Gaussian functions, the variance of any estimator of the mean is directly proportional to the variance of the density function.

There is another expression that can be used to represent the Fisher information.

Theorem 1. *Given a random variable $\mathbf{X}$ and its associated density function $f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})$, which depends on the parameter vector $\boldsymbol{\theta} \in \Theta$ and complies with the boundary condition for θ_k (see Appendix A), where θ_k is the k-th component of $\boldsymbol{\theta}$, then the Fisher information associated with θ_k is equal to:*

$$i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} = -\int f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})\frac{\partial^2 \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k^2}d\mathbf{x} \qquad (12)$$

A proof of this theorem can be found in [34] (p. 373).

Example 2. *Continuing the Gaussian example, and using the alternative definition of the Fisher information, the required second derivative is first calculated:*

$$\frac{d^2 \ln f_{X;\mu,\eta}(x)}{d\mu^2} = \frac{d}{d\mu}\left(\frac{x-\mu}{\eta^2}\right) \quad = \quad -\frac{1}{\eta^2} \qquad (13)$$

Replacing into Equation (12), the same result is obtained:

$$\begin{aligned} i_F\left(f_{X;\mu,\eta}\right)_\mu &= -\int f_{X;\mu,\eta}(x)\frac{d^2 \ln f_{X;\mu,\eta}(x)}{d\mu^2}dx && (14) \\ &= -\int f_{X;\mu,\eta}(x)\left(-\frac{1}{\eta^2}\right)dx && (15) \\ &= \frac{1}{\eta^2}\int dx f_{X;\mu,\eta}(x) && (16) \\ &= \frac{1}{\eta^2} && (17) \end{aligned}$$

The importance of the Fisher information quantity stems from the Cramer–Rao bound [3,4,23,35]:

Theorem 2 (Cramer–Rao Bound)**.** *Given a random variable* $\mathbf{X}$ *and its associated density function* $f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})$*, which depends on the parameter vector* $\boldsymbol{\theta} \in \Theta$ *and complies with the boundary condition for* θ_k *(see Appendix A), where* θ_k *is the* k*-th component of* $\boldsymbol{\theta}$*, also given that there is an unbiased estimator* $\hat{\theta}_k(\mathbf{x})$ *of the scalar parameter* θ_k*, then:*

$$\frac{1}{i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k}} \leq \sigma^2_{\hat{\theta}_k} \tag{18}$$

where:

$$\sigma^2_{\hat{\theta}_k} \equiv \int f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) \left(\hat{\theta}_k(\mathbf{x}) - \theta_k\right)^2 d\mathbf{x} \tag{19}$$

is the variance of the estimator. Proofs of this theorem can be found in [7] (p. 29) and [23] (p. 66).

The Cramer–Rao bound establishes that the reciprocal of the Fisher information is a lower bound of the variance of an estimator. Any estimator that reaches the bound imposed by the Cramer–Rao theorem is called efficient [34]. It is important to notice that the bound does not depend on the estimator itself; it only depends on $i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k}$. In this work, the case of biased estimators will not be analyzed, nor when the parameters themselves are random variables.

The following theorem states that the topology of the Fisher information in the density function space is very simple:

Theorem 3. *The Fisher information* $i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k}$ *is convex in* $f_{\mathbf{X};\boldsymbol{\theta}}$*. Proofs of this theorem can be found in [7] (p. 69) and [29].*

4. Several Random Variables Depending on θ_k

4.1. Joint Fisher Information Definition

Definition 2. *Given two random variables* $\mathbf{X}$ *and* $\mathbf{Y}$ *and the associated joint density function* $f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})$*, which depends on the parameter vector* $\boldsymbol{\theta} \in \Theta$*, and* θ_k *is the* k*-th component of* $\boldsymbol{\theta}$*, then the joint Fisher information associated with* θ_k *is defined by:*

$$i_F\left(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} \equiv \iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}) \left(\frac{\partial \ln f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{\partial \theta_k}\right)^2 d\mathbf{x}d\mathbf{y} \tag{20}$$

4.2. An Equivalent Joint Fisher Information Definition

Theorem 4. *Given two random variables* $\mathbf{X}$ *and* $\mathbf{Y}$ *and the associated joint density function* $f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})$*, which depends on the parameter vector* $\boldsymbol{\theta} \in \Theta$ *and complies with the boundary condition for* θ_k *(see Appendix A), where* θ_k *is the* k*-th component of* $\boldsymbol{\theta}$*, then the joint Fisher information associated with* θ_k *is equal to:*

$$i_F\left(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} = -\iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}) \frac{\partial^2 \ln f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{\partial \theta_k^2} d\mathbf{x}d\mathbf{y} \tag{21}$$

Proof. This follows trivially from the alternative definition of the Fisher information. □

4.3. Conditional Fisher Information Definition

Definition 3.

$$i_F\left(f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} \equiv \iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})\left(\frac{\partial \ln f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x})}{\partial\theta_k}\right)^2 d\mathbf{x}d\mathbf{y} \tag{22}$$

4.4. Chain Rule for Two Random Variables

The following result was first published by Zamir [32], who used it to produce an alternative proof of the Fisher information inequality. In the following lines, the same chain rule is proven using the results presented in the previous sections.

Theorem 5 (Chain Rule for Two Random Variables)**.** *Given a joint density function $f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})$, which depends on the parameter vector $\boldsymbol{\theta}\in\Theta$, and given that the density functions comply with the boundary condition for θ_k (see Appendix A), where θ_k is the k-th component of $\boldsymbol{\theta}$, then:*

$$\begin{aligned} i_F\left(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} &= i_F\left(f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} + i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} & (23)\\ &= i_F\left(f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} + i_F\left(f_{\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} & (24)\end{aligned}$$

respectively.

Proof.

$$\begin{aligned} i_F\left(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} &\equiv \iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})\left(\frac{\partial \ln f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{\partial\theta_k}\right)^2 d\mathbf{x}d\mathbf{y} & (25)\\ &= \iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})\left(\frac{\partial \ln\left(f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x})f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})\right)}{\partial\theta_k}\right)^2 d\mathbf{x}d\mathbf{y} & (26)\\ &= \iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})\left(\frac{\partial \ln f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x})}{\partial\theta_k}+\frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k}\right)^2 d\mathbf{x}d\mathbf{y} & (27)\\ &= i_F\left(f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} + i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} + 2\iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})\frac{\partial \ln f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x})}{\partial\theta_k}\frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k}d\mathbf{x}d\mathbf{y} & (28)\end{aligned}$$

but,

$$\begin{aligned} \iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})\frac{\partial \ln f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x})}{\partial\theta_k}\frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k}d\mathbf{x}d\mathbf{y} &= \iint \frac{\partial f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x})}{\partial\theta_k}\frac{\partial f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k}d\mathbf{x}d\mathbf{y} & (29)\\ &= \int\left(\frac{\partial f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k}\int\frac{\partial f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x})}{\partial\theta_k}d\mathbf{y}\right)d\mathbf{x} & (30)\end{aligned}$$

If $f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x})$ complies with the boundary condition with respect to θ_k (see Appendix A), then:

$$\int\frac{\partial f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x})}{\partial\theta_k}d\mathbf{y} = \frac{\partial}{\partial\theta_k}\int d\mathbf{y}f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x}) = 0 \tag{31}$$

Therefore, the theorem is proven. The other result is proven analogously. □

When the chain rule is used to estimate the Fisher information associated with a parameter, it is important to take into account that all of the terms that come out after applying the chain rule contain derivatives with respect to the same parameter. Because some of these terms may be dependent on density functions that do not depend on the parameter, some of these terms may be equal to zero.

Example 3. *Given the random variable $Y = X + N$, where X is a Gaussian density function with mean μ and standard deviation η and N another Gaussian density function with mean zero and standard deviation ν, if the joint density function is available, and the parameter to be estimated is μ, then:*

$$\begin{aligned} i_F\left(f_{Y,X;\mu,\eta,\nu}\right)_\mu &= i_F\left(f_{Y|X;\mu,\eta,\nu}\right)_\mu + i_F\left(f_{X;\mu,\eta}\right)_\mu && (32)\\ &= i_F\left(f_{N;\nu}\right)_\mu + i_F\left(f_{X;\mu,\eta}\right)_\mu && (33)\\ &= i_F\left(f_{X;\mu,\eta}\right)_\mu && (34)\\ &= \frac{1}{\eta^2} && (35) \end{aligned}$$

The previous result implies that if the joint density function of the output Y and the input X is available, the noise does not affect the estimation process. This is not surprising, since Y is a corrupted version of X, and it cannot shed more information on μ than that contained in X. Because all of the information hidden in X is available through the joint density function, it makes sense to think that the Fisher information of the joint density function corresponds to that of the marginal distribution $f_{X;\mu,\eta}$.

Given the density functions mentioned above, it is possible to prove that:

$$f_{Y;\mu,\eta,\nu}(y) = \frac{1}{\sqrt{2\pi(\eta^2+\nu^2)}}\exp\left(-\frac{(y-\mu)^2}{2(\eta^2+\nu^2)}\right) \tag{36}$$

with Fisher information associated with μ equal to:

$$i_F\left(f_{Y;\mu,\eta,\nu}\right)_\mu = \frac{1}{\eta^2+\nu^2} \tag{37}$$

Using the other expression for the chain rule:

$$\begin{aligned} i_F\left(f_{Y,X;\mu,\eta,\nu}\right)_\mu &= i_F\left(f_{X|Y;\mu,\eta,\nu}\right)_\mu + i_F\left(f_{Y;\mu,\eta,\nu}\right)_\mu && (38)\\ &= i_F\left(f_{X|Y;\mu,\eta,\nu}\right)_\mu + \frac{1}{\eta^2+\nu^2} && (39) \end{aligned}$$

Using the previous results:

$$\frac{1}{\eta^2} = i_F\left(f_{X|Y;\mu,\eta,\nu}\right)_\mu + \frac{1}{\eta^2+\nu^2} \tag{40}$$

which implies:

$$i_F\left(f_{X|Y;\mu,\eta,\nu}\right)_\mu = \frac{\nu^2}{\eta^2(\eta^2+\nu^2)} \tag{41}$$

4.5. Chain Rule for Many Random Variables

In the case of more than two density functions:

Theorem 6 (Chain Rule for Many Random Variables)**.** *Given a set of n random variables $\mathbf{X}_1$, $\mathbf{X}_2$, ..., $\mathbf{X}_n$, all of them depending on θ_k, if the density functions comply with the boundary condition for θ_k (see Appendix A), then:*

$$i_F\left(f_{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n;\boldsymbol{\theta}}\right)_{\theta_k} = \sum_{k=1}^{n} i_F\left(f_{\mathbf{X}_k|\mathbf{X}_{k-1},\ldots,\mathbf{X}_1;\boldsymbol{\theta}}\right)_{\theta_k} \tag{42}$$

Proof.

$$\begin{aligned} i_F\left(f_{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n;\boldsymbol{\theta}}\right)_{\theta_k} &= i_F\left(f_{\mathbf{X}_n,\ldots,\mathbf{X}_2|\mathbf{X}_1;\boldsymbol{\theta}}\right)_{\theta_k} + i_F\left(f_{\mathbf{X}_1;\boldsymbol{\theta}}\right)_{\theta_k} & (43)\\ &= i_F\left(f_{\mathbf{X}_n,\ldots,\mathbf{X}_3|\mathbf{X}_2,\mathbf{X}_1;\boldsymbol{\theta}}\right)_{\theta_k} + i_F\left(f_{\mathbf{X}_2|\mathbf{X}_1;\boldsymbol{\theta}}\right)_{\theta_k} + i_F\left(f_{\mathbf{X}_1;\boldsymbol{\theta}}\right)_{\theta_k} & (44)\\ &\vdots \\ &= \sum_{k=1}^{n} i_F\left(f_{\mathbf{X}_k|\mathbf{X}_{k-1},\ldots,\mathbf{X}_1;\boldsymbol{\theta}}\right)_{\theta_k} & (45) \end{aligned}$$

□

If the n random variables in Theorem 6 are i.i.d., then $i_F\left(f_{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n;\boldsymbol{\theta}}\right)_{\theta_k} = n \cdot i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k}$.

5. Relative Fisher Information Type I

In the following, the relative Fisher information is defined. As far as it was possible to determine, the first definition of the relative Fisher information was given by Otto and Villani [36], who defined it for the translationally-invariant case. Furthermore, this expression has been rediscovered or simply used in many applications thereafter in different problems and fields [22,37–44]. Furthermore, it seems that the first general analysis of the relative Fisher information was presented by the author in [45]. The following sections focus on this latter general case, where there is no assumption of translational invariance.

Analogously to the Kullback–Leibler divergence [46], also known as as relative entropy, which was designed to established how much two density functions differed, the relative Fisher information of Type I is obtained when the ratio of two intervening density functions is replaced into Equation (1), as is shown in the following definition.

Definition 4. *The relative Fisher information Type I is defined by:*

$$d_{F(I)}(f_{\mathbf{X};\boldsymbol{\theta}}||f_{\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} \equiv \int f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) \left(\frac{\partial}{\partial \theta_k}\left(\ln\left(\frac{f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x})}\right)\right)\right)^2 d\mathbf{x} \tag{46}$$

The same mechanism can be used to generate a second definition for the relative Fisher information. The same ratio can be replaced into Equation (12), producing an alternative and equally valid expression, which is designated as relative Fisher information Type II. This second expression is studied in the following sections.

6. Information Correlation

Definition 5. *The information correlation with respect to θ_k is defined by:*

$$i_C(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} \equiv \iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}) \frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} \frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} d\mathbf{x} d\mathbf{y} \tag{47}$$

The name information correlation comes from the similarity between this definition and that of the classical correlation coefficient. It is important to keep in mind that it is different from the terms that fill the Fisher information matrix [23].

According to the definition $i_C(f_{\mathbf{X},\mathbf{X};\boldsymbol{\theta}})_{\theta_k} = i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k}$, and $i_C(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} = i_C(f_{\mathbf{Y},\mathbf{X};\boldsymbol{\theta}})_{\theta_k}$.

Example 4. *Continuing with the example where $Y = X + N$, the information correlation between Y and X is given by:*

$$i_C(f_{Y,X;\mu,\eta,\nu})_\mu = \iint f_{Y,X;\mu,\eta,\nu}(y,x) \frac{d \ln f_{Y;\mu,\eta,\nu}(y)}{d\mu} \frac{d \ln f_{X;\mu,\eta}(x)}{d\mu} dy dx \tag{48}$$

$$= \iint f_{Y|X;\mu,\eta,\nu}(y|x) f_{X;\mu,\eta}(x) \frac{d \ln f_{Y;\mu,\eta,\nu}(y)}{d\mu} \frac{d \ln f_{X;\mu,\eta}(x)}{d\mu} dy dx \tag{49}$$

$$= \iint f_{N;\nu}(y-x) f_{X;\mu,\eta}(x) \frac{d \ln f_{Y;\mu,\eta,\nu}(y)}{d\mu} \frac{d \ln f_{X;\mu,\eta}(x)}{d\mu} dy dx \tag{50}$$

$$= \iint \left(\frac{1}{\sqrt{2\pi}\nu} \exp\left(-\frac{(y-x)^2}{2\nu^2} \right) \right) \left(\frac{1}{\sqrt{2\pi}\eta} \exp\left(-\frac{(x-\mu)^2}{2\eta^2} \right) \right) \frac{d \ln f_{Y;\mu,\eta,\nu}(y)}{d\mu} \frac{d \ln f_{X;\mu,\eta}(x)}{d\mu} dy dx \tag{51}$$

$$= \frac{1}{2\pi\nu\eta} \iint \exp\left(-\frac{1}{2} \left(\frac{(y-x)^2}{\nu^2} + \frac{(x-\mu)^2}{\eta^2} \right) \right) \frac{d \ln f_{Y;\mu,\eta,\nu}(y)}{d\mu} \frac{d \ln f_{X;\mu,\eta}(x)}{d\mu} dy dx \tag{52}$$

where:

$$\frac{d \ln f_{Y;\mu,\eta,\nu}(y)}{d\mu} = \frac{d}{d\mu} \left(\ln \frac{1}{\sqrt{2\pi}\sqrt{\eta^2+\nu^2}} - \frac{(y-\mu)^2}{2(\eta^2+\nu^2)} \right) \tag{53}$$

$$= \frac{(y-\mu)}{\eta^2+\nu^2} \tag{54}$$

Analogously:

$$\frac{d \ln f_{X;\mu,\eta}(y)}{d\mu} = \frac{(x-\mu)}{\eta^2} \tag{55}$$

Replacing these derivatives into the information correlation expression:

$$i_C(f_{Y,X;\mu,\eta,\nu})_\mu = \frac{1}{2\pi\nu\eta} \iint \exp\left(-\frac{1}{2} \left(\frac{(y-x)^2}{\nu^2} + \frac{(x-\mu)^2}{\eta^2} \right) \right) \left(\frac{y-\mu}{\eta^2+\nu^2} \right) \left(\frac{x-\mu}{\eta^2} \right) dy dx \tag{56}$$

$$= \frac{1}{2\pi\nu\eta(\eta^2+\nu^2)\eta^2} \iint (y-\mu)(x-\mu) \exp\left(-\frac{1}{2} \left(\frac{(y-x)^2}{\nu^2} + \frac{(x-\mu)^2}{\eta^2} \right) \right) dy dx \tag{57}$$

$$= \frac{1}{2\pi\nu\eta(\eta^2+\nu^2)\eta^2} \int (x-\mu) \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\eta^2} \right) \left(\int (y-\mu) \exp\left(-\frac{1}{2} \frac{(y-x)^2}{\nu^2} \right) dy \right) dx \tag{58}$$

$$= \frac{1}{\eta^2+\nu^2} \tag{59}$$

Theorem 7. *The information correlation is bounded according to:*

$$\left(i_C(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k}\right)^2 \leq i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} i_F\left(f_{\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} \tag{60}$$

Proof.

$$0 \leq \iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}) \left(a\frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} + \frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} \right)^2 d\mathbf{x}d\mathbf{y} \tag{61}$$

which can be reexpressed as:

$$0 \leq a^2 i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} + 2a \cdot i_C(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} + i_F\left(f_{\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} \tag{62}$$

This is a second degree equation that is true for every possible a. Because this equation is always greater than zero, the discriminant of the equation has to comply with $4(i_C(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k})^2 - 4i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} i_F\left(f_{\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} \leq 0$, which proves the theorem. □

Definition 6. *The information correlation coefficient is defined by:*

$$\rho_F = \frac{i_C(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k}}{\sqrt{i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} i_F\left(f_{\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k}}} \tag{63}$$

Theorem 8. *The information correlation coefficient is limited by:*

$$-1 \leq \rho_F \leq 1 \tag{64}$$

Proof. This comes from the definition of the information correlation coefficient and Theorem 7. □

Theorem 9. *If at least one of the following conditions:*

(1) $f_{\mathbf{X};\boldsymbol{\theta}}$ and $f_{\mathbf{Y};\boldsymbol{\theta}}$ are independent.
(2) Either $f_{\mathbf{X};\boldsymbol{\theta}}$ or $f_{\mathbf{Y};\boldsymbol{\theta}}$ does not depend on θ_k.

is true, then:

$$i_C(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} = 0 \tag{65}$$

Proof. Examination of the information correlation definition clearly shows that compliance with the first and second cases directly implies that this quantity is zero. □

7. Mutual Fisher Information Type I

As happens in Shannon's differential entropy handling, in this work, mutual Fisher information is also defined as relative Fisher information Type I, where the argument is the ratio between a joint density function and the product of its marginals.

7.1. Definition

Definition 7. *The mutual Fisher information Type I is defined by:*

$$m_{F(I)}(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} \equiv \iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}) \left(\frac{\partial}{\partial \theta_k} \ln \left(\frac{f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})} \right) \right)^2 d\mathbf{x} d\mathbf{y} \tag{66}$$

From the definition, it is obvious that $m_{F(I)}(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} \geq 0$.

Theorem 10. *If the boundary condition (see Appendix A) with respect to θ_k holds for $f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})$, the mutual Fisher information Type I can be reformulated as a function of the Fisher information as follows:*

$$\begin{aligned} m_{F(I)}(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} &= i_F\left(f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} + 2 \cdot i_C(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} & (67) \\ &= i_F\left(f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} + 2 \cdot i_C(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} & (68) \end{aligned}$$

Proof.

$$\left(\frac{\partial}{\partial \theta_k} \ln \left(\frac{f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})} \right) \right)^2 = \left(\frac{\partial \ln f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{\partial \theta_k} - \frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} - \frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} \right)^2 \tag{69}$$

$$\begin{aligned} = \; & \left(\frac{\partial \ln f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{\partial \theta_k} \right)^2 + \left(\frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} \right)^2 + \left(\frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} \right)^2 - 2 \frac{\partial \ln f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{\partial \theta_k} \frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} \\ & -2 \frac{\partial \ln f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{\partial \theta_k} \frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} + 2 \frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} \frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} \end{aligned} \tag{70}$$

$$\begin{aligned} = \; & \left(\frac{\partial \ln f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})}{\partial \theta_k} \right)^2 + 2 \frac{\partial \ln f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})}{\partial \theta_k} \frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} + \left(\frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} \right)^2 \\ & + \left(\frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} \right)^2 + \left(\frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} \right)^2 \\ & -2 \frac{\partial \ln f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})}{\partial \theta_k} \frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} - 2 \frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} \frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} \\ & -2 \frac{\partial \ln f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})}{\partial \theta_k} \frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} - 2 \frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} \frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} \\ & +2 \frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} \frac{\partial \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial \theta_k} \end{aligned} \tag{71}$$

Simplifying:

$$\begin{aligned} \left(\frac{\partial}{\partial \theta_k} \ln \left(\frac{f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})} \right) \right)^2 &= \left(\frac{\partial \ln f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})}{\partial \theta_k} \right)^2 + \left(\frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} \right)^2 \\ & -2 \frac{\partial \ln f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})}{\partial \theta_k} \frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} \end{aligned} \tag{72}$$

Now,

$$\begin{aligned} & 2 \iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}) \frac{\partial \ln f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})}{\partial \theta_k} \frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} d\mathbf{x} d\mathbf{y} \\ = \; & 2 \iint \frac{f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})} \frac{\partial f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})}{\partial \theta_k} \frac{\partial f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} d\mathbf{x} d\mathbf{y} & (73) \\ = \; & 2 \int \left(\frac{1}{f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})} \frac{\partial f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \theta_k} \int f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y}) \frac{\partial f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})}{\partial \theta_k} d\mathbf{y} \right) d\mathbf{x} & (74) \end{aligned}$$

Assuming that $f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}$ complies with the boundary condition (see Appendix A) with respect to θ_k, then:

$$\begin{aligned}\frac{\partial f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k} &= \frac{\partial}{\partial\theta_k}\int f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})d\mathbf{y} &(75)\\ &= \int \frac{\partial f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial\theta_k}d\mathbf{y} &(76)\\ &= \int f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})\frac{\partial f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})}{\partial\theta_k}d\mathbf{y} + \int f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})\frac{\partial f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial\theta_k}d\mathbf{y} &(77)\end{aligned}$$

Hence,

$$\int f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})\frac{\partial f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})}{\partial\theta_k}d\mathbf{y} = \frac{\partial f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k} - \int f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})\frac{\partial f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial\theta_k}d\mathbf{y} \qquad (78)$$

Using the previous result, it is obtained:

$$\begin{aligned}&2\iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})\frac{\partial\ln f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})}{\partial\theta_k}\frac{\partial\ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k}d\mathbf{x}d\mathbf{y}\\ &= 2\int\frac{1}{f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}\frac{\partial f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k}\frac{\partial f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k}d\mathbf{x} - 2\int\frac{1}{f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}\frac{\partial f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k}\int d\mathbf{y} f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})\frac{\partial f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial\theta_k}d\mathbf{x} \qquad (79)\\ &= 2i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} - 2\iint\frac{f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}\frac{\partial f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k}\frac{\partial f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial\theta_k}d\mathbf{x}d\mathbf{y} \qquad (80)\end{aligned}$$

This implies:

$$m_{F(I)}(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} = i_F\left(f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} + 2\cdot i_C(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} \qquad (81)$$

The other result is obtained analogously. □

Example 5. *Continuing with the example where $Y = X + N$, the mutual Fisher information Type I is given by:*

$$\begin{aligned}m_{F(I)}(f_{Y,X;\mu,\eta,\nu})_\mu &= i_F\left(f_{Y|X}\right)_{\theta_k} - i_F\left(f_Y\right)_{\theta_k} + 2\cdot i_C(f_{Y,X;\mu,\eta,\nu})_\mu &(82)\\ &= i_F\left(f_N\right)_{\theta_k} - i_F\left(f_Y\right)_{\theta_k} + 2\cdot i_C(f_{Y,X;\mu,\eta,\nu})_\mu &(83)\\ &= \frac{1}{\nu^2} - \frac{1}{\eta^2+\nu^2} + 2\cdot i_C(f_{Y,X;\mu,\eta,\nu})_\mu &(84)\\ &= \frac{\eta^2}{\nu^2(\eta^2+\nu^2)} + 2\cdot i_C(f_{Y,X;\mu,\eta,\nu})_\mu &(85)\\ &= \frac{\eta^2}{\nu^2(\eta^2+\nu^2)} + \frac{2}{\eta^2+\nu^2} &(86)\\ &= \frac{1}{\eta^2+\nu^2} + \frac{1}{\nu^2} &(87)\end{aligned}$$

7.2. Conditional Mutual Fisher Information of Type I

Definition 8. *The conditional information correlation with respect to θ_k of random variables* $\mathbf{X}$ *and* $\mathbf{Y}$ *given random variable* $\mathbf{Z}$ *is defined by:*

$$i_C(f_{\mathbf{X},\mathbf{Y}|\mathbf{Z};\boldsymbol{\theta}})_{\theta_k} \equiv \iint f_{\mathbf{X},\mathbf{Y},\mathbf{Z};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y},\mathbf{z})\frac{\partial\ln f_{\mathbf{X}|\mathbf{Z};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})}{\partial\theta_k}\frac{\partial\ln f_{\mathbf{Y}|\mathbf{Z};\boldsymbol{\theta}}(\mathbf{y}|\mathbf{z})}{\partial\theta_k}d\mathbf{x}d\mathbf{y}d\mathbf{z} \qquad (88)$$

Definition 9. *The conditional mutual Fisher information of Type I of random variables* $\mathbf{X}$ *and* $\mathbf{Y}$ *given random variable* $\mathbf{Z}$ *is defined by:*

$$m_{F(I)}(f_{\mathbf{X},\mathbf{Y}|\mathbf{Z};\boldsymbol{\theta}})_{\theta_k} \equiv \iint f_{\mathbf{X},\mathbf{Y},\mathbf{Z};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y},\mathbf{z}) \left(\frac{\partial}{\partial \theta_k} \left(\ln \left(\frac{f_{\mathbf{X},\mathbf{Y}|\mathbf{Z};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}|\mathbf{z})}{f_{\mathbf{X}|\mathbf{Z};\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z}) f_{\mathbf{Y}|\mathbf{Z};\boldsymbol{\theta}}(\mathbf{y}|\mathbf{z})} \right) \right) \right)^2 d\mathbf{x} d\mathbf{y} d\mathbf{z} \tag{89}$$

Corollary 1. *If the boundary condition (see Appendix A) with respect to* θ_k *holds for* $f_{\mathbf{X},\mathbf{Y},\mathbf{Z};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y},\mathbf{z})$*, the conditional mutual Fisher information of Type I of random variables* $\mathbf{X}$ *and* $\mathbf{Y}$ *given random variable* $\mathbf{Z}$ *can be reformulated as a function of the Fisher information as follows:*

$$\begin{aligned} m_{F(I)}(f_{\mathbf{X},\mathbf{Y}|\mathbf{Z};\boldsymbol{\theta}})_{\theta_k} &= i_F\left(f_{\mathbf{X}|\mathbf{Y},\mathbf{Z};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{X}|\mathbf{Z};\boldsymbol{\theta}}\right)_{\theta_k} + 2 \cdot i_C(f_{\mathbf{X},\mathbf{Y}|\mathbf{Z};\boldsymbol{\theta}})_{\theta_k} & (90) \\ &= i_F\left(f_{\mathbf{Y}|\mathbf{X},\mathbf{Z};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{Y}|\mathbf{Z};\boldsymbol{\theta}}\right)_{\theta_k} + 2 \cdot i_C(f_{\mathbf{X},\mathbf{Y}|\mathbf{Z};\boldsymbol{\theta}})_{\theta_k} & (91) \end{aligned}$$

Proof. This follows analogously to that of the simpler case. □

8. Relative Fisher Information Type II

Given that there is an alternative expression for the Fisher information (check Equation (12)), there is another way of defining the relative Fisher information expression.

Definition 10. *The relative Fisher information Type II is defined by:*

$$d_{F(II)}(f_{\mathbf{X};\boldsymbol{\theta}}||f_{\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} \equiv -\int f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) \frac{\partial^2}{\partial \theta_k^2} \left(\ln \left(\frac{f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x})} \right) \right) d\mathbf{x} \tag{92}$$

Even though both definitions for the relative Fisher information are derived from equivalent expressions, they are not equivalent. Why is this so? This is because the argument of the Fisher information definition is a density function, whereas the argument of the relative Fisher information expression is a ratio of density functions, not a density function, thus their difference.

9. Mutual Fisher Information Type II

Analogously to the definition of the mutual Fisher information Type I, but in this case using the relative Fisher information of Type II, the following definition is obtained:

Definition 11. *The mutual Fisher information Type II is defined by:*

$$m_{F(II)}(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} \equiv -\iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}) \frac{\partial^2}{\partial \theta_k^2} \ln \left(\frac{f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})} \right) d\mathbf{x} d\mathbf{y} \tag{93}$$

Theorem 11. *The mutual Fisher information Type II can be reformulated as a function of the Fisher information as follows:*

$$m_{F(II)}(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} = i_F\left(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} \tag{94}$$

Proof.

$$\begin{aligned} m_{F(II)}(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} &= -\iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})\frac{\partial^2}{\partial\theta_k^2}\ln\left(\frac{f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}\right)d\mathbf{x}d\mathbf{y} \qquad (95)\\ &= -\iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})\frac{\partial^2 \ln f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})}{\partial\theta_k^2}d\mathbf{x}d\mathbf{y}\\ &\quad +\iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})\frac{\partial^2 \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial\theta_k^2}d\mathbf{x}d\mathbf{y}\\ &\quad +\iint f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}}(\mathbf{x},\mathbf{y})\frac{\partial^2 \ln f_{\mathbf{Y};\boldsymbol{\theta}}(\mathbf{y})}{\partial\theta_k^2}d\mathbf{x}d\mathbf{y} \qquad (96) \end{aligned}$$

from which the theorem follows. □

Corollary 2.

$$\begin{aligned} m_{F(II)}(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} &= i_F\left(f_{\mathbf{X}|\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} \qquad (97)\\ &= i_F\left(f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} \qquad (98) \end{aligned}$$

Proof. This comes from combining Theorem 11 and the chain rule for Fisher information. □

Example 6. *For the example where $Y = X + N$, the mutual Fisher information Type II is given by:*

$$\begin{aligned} m_{F(II)}(f_{Y,X;\mu,\eta,\nu})_{\theta_k} &= i_F\left(f_{Y|X}\right)_{\theta_k} - i_F\left(f_Y\right)_{\theta_k}\\ &= i_F\left(f_N\right)_{\theta_k} - i_F\left(f_Y\right)_{\theta_k} \qquad (99)\\ &= \frac{1}{\nu^2} - \frac{1}{\eta^2+\nu^2} \qquad (100)\\ &= \frac{\eta^2}{\nu^2(\eta^2+\nu^2)} \qquad (101) \end{aligned}$$

Corollary 3.

$$m_{F(I)}(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} = m_{F(II)}(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} + 2\cdot i_C(f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} \qquad (102)$$

Proof. This can be deduced from the mutual Fisher information theorems. □

Given that $m_{F(I)}$ is always greater than or equal to zero, the expression $m_{F(II)}$ can be positive or negative according to the value of the information correlation.

10. Other Properties

10.1. Lower Bound for Fisher Information

Stam's inequality [8,9,40,47–50] states a lower bound for Fisher information, which links Fisher information and Shannon's entropy power. However, this expression is limited to the special case where the parameters in the Fisher information expression correspond to a location parameter.

A more general result was recently proven by Stein *et al.* [51], which says that given a multidimensional random variable with density function $f_{\mathbf{X};\boldsymbol{\theta}}$ with:

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \int_{\mathcal{S}} \mathbf{x} f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) d\mathbf{x} \tag{103}$$

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \int_{\mathcal{S}} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^{\mathrm{T}} f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) d\mathbf{x} \tag{104}$$

If the Fisher information matrix is defined by:

$$\mathbf{F}(f_{\mathbf{X};\boldsymbol{\theta}}) = \int_{\mathcal{S}} f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) \left(\frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \left(\frac{\partial \ln f_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})}{\partial \boldsymbol{\theta}} \right) d\mathbf{x} \tag{105}$$

then:

$$\mathbf{F}(f_{\mathbf{X};\boldsymbol{\theta}}) \succeq \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \tag{106}$$

if $\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ exists. The authors of [51] explain that this is the same as saying that:

$$0 \leq \mathbf{x}^{\mathrm{T}} \left(\mathbf{F}(f_{\mathbf{X};\boldsymbol{\theta}}) - \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \left(\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \right) \mathbf{x} \tag{107}$$

The previous expression states that the difference of matrices between the large parenthesis is a positive semi-definite matrix. Thus, its diagonal elements are non-negative, and it can be stated:

Corollary 4. *The following lower bound for Fisher information holds:*

$$\sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial \mu_i}{\partial \theta_k} c_{ij}^{-1} \frac{\partial \mu_j}{\partial \theta_k} \leq i_F \left(f_{\mathbf{X};\boldsymbol{\theta}} \right)_{\theta_k} \tag{108}$$

and c_{ij}^{-1} stands for the ij-th element of $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})$.

10.2. In Some Cases, Conditioning Increases the Fisher Information

The following result states that in some cases, conditioning a random variable with another variable may increase the Fisher information. This result is a generalization of another published previously by Zamir [32].

Theorem 12 (Conditioning Increases Information)**.** *If $f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}$ depends on θ_k and $f_{\mathbf{X}}$ does not depend on it, then:*

$$i_F \left(f_{\mathbf{Y};\boldsymbol{\theta}} \right)_{\theta_k} \leq i_F \left(f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}} \right)_{\theta_k} \tag{109}$$

Proof. Thus, given that only $f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}$ depends on θ_k, Theorem 9 guarantees that:

$$i_C (f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} = 0 \tag{110}$$

Hence, from the previous mutual Fisher information expressions:

$$0 \leq m_{F(I)} (f_{\mathbf{X},\mathbf{Y};\boldsymbol{\theta}})_{\theta_k} = i_F \left(f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}} \right)_{\theta_k} - i_F \left(f_{\mathbf{Y};\boldsymbol{\theta}} \right)_{\theta_k} \tag{111}$$

Thus:

$$i_F\left(f_{\mathbf{Y};\boldsymbol{\theta}}\right)_{\theta_k} \leq i_F\left(f_{\mathbf{Y}|\mathbf{X};\boldsymbol{\theta}}\right)_{\theta_k} \tag{112}$$

□

10.3. Data Processing Inequality

Following the same analysis done by Cover and Thomas to present the data processing theorem for Shannon entropy [30] and continuing with the work done by Zamir [32], the case where the joint density function of the random variables $\mathbf{R}$, $\mathbf{S}$ and $\mathbf{T}$ can be expressed by $f_{\mathbf{R},\mathbf{S},\mathbf{T};\boldsymbol{\theta}} = f_{\mathbf{R};\boldsymbol{\theta}} \cdot f_{\mathbf{S}|\mathbf{R};\boldsymbol{\theta}} \cdot f_{\mathbf{T}|\mathbf{S};\boldsymbol{\theta}}$ is considered. In this case, they form a short Markov chain that is represented by $\mathbf{R} \rightarrow \mathbf{S} \rightarrow \mathbf{T}$. Because Markovicity implies conditional independence, then it is true that $f_{\mathbf{R},\mathbf{T}|\mathbf{S};\boldsymbol{\theta}} = f_{\mathbf{R}|\mathbf{S};\boldsymbol{\theta}} \cdot f_{\mathbf{T}|\mathbf{S};\boldsymbol{\theta}}$.

Theorem 13. *Given a Markov chain* $\mathbf{R} \rightarrow \mathbf{S} \rightarrow \mathbf{T}$*, where only* $f_{\mathbf{T}|\mathbf{S};\boldsymbol{\theta}}$ *depends on* θ_k*, then:*

$$m_{F(I)}(f_{\mathbf{R},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \leq m_{F(I)}(f_{\mathbf{S},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \tag{113}$$

Proof. From the previous results:

$$m_{F(I)}(f_{(\mathbf{R},\mathbf{S}),\mathbf{T};\boldsymbol{\theta}})_{\theta_k} = i_F\left(f_{\mathbf{R},\mathbf{S}|\mathbf{T};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{R},\mathbf{S};\boldsymbol{\theta}}\right)_{\theta_k} + 2 \cdot i_C(f_{(\mathbf{R},\mathbf{S}),\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \tag{114}$$

$$= i_F\left(f_{\mathbf{R}|\mathbf{S},\mathbf{T};\boldsymbol{\theta}}\right)_{\theta_k} + i_F\left(f_{\mathbf{S}|\mathbf{T};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{R}|\mathbf{S};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{S};\boldsymbol{\theta}}\right)_{\theta_k} + 2 \cdot i_C(f_{(\mathbf{R},\mathbf{S}),\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \tag{115}$$

$$\begin{aligned} = &\left(i_F\left(f_{\mathbf{R}|\mathbf{S},\mathbf{T};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{R}|\mathbf{S};\boldsymbol{\theta}}\right)_{\theta_k} + 2 \cdot i_C(f_{\mathbf{R},\mathbf{T}|\mathbf{S};\boldsymbol{\theta}})_{\theta_k}\right) - 2 \cdot i_C(f_{\mathbf{R},\mathbf{T}|\mathbf{S};\boldsymbol{\theta}})_{\theta_k} \\ &+ \left(i_F\left(f_{\mathbf{S}|\mathbf{T};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{S};\boldsymbol{\theta}}\right)_{\theta_k} + 2 \cdot i_C(f_{\mathbf{S},\mathbf{T};\boldsymbol{\theta}})_{\theta_k}\right) - 2 \cdot i_C(f_{\mathbf{S},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} + 2 \cdot i_C(f_{(\mathbf{R},\mathbf{S}),\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \end{aligned} \tag{116}$$

$$= m_{F(I)}(f_{\mathbf{R},\mathbf{T}|\mathbf{S};\boldsymbol{\theta}})_{\theta_k} + m_{F(I)}(f_{\mathbf{S},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} - 2 \cdot i_C(f_{\mathbf{R},\mathbf{T}|\mathbf{S};\boldsymbol{\theta}})_{\theta_k} - 2 \cdot i_C(f_{\mathbf{S},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} + 2 \cdot i_C(f_{(\mathbf{R},\mathbf{S}),\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \tag{117}$$

Analogously:

$$\begin{aligned} m_{F(I)}(f_{(\mathbf{R},\mathbf{S}),\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \;=\; & m_{F(I)}(f_{\mathbf{S},\mathbf{T}|\mathbf{R};\boldsymbol{\theta}})_{\theta_k} + m_{F(I)}(f_{\mathbf{R},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \\ & -2 \cdot i_C(f_{\mathbf{S},\mathbf{T}|\mathbf{R};\boldsymbol{\theta}})_{\theta_k} - 2 \cdot i_C(f_{\mathbf{R},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} + 2 \cdot i_C(f_{(\mathbf{R},\mathbf{S}),\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \end{aligned} \tag{118}$$

Because only $f_{\mathbf{T}|\mathbf{S};\boldsymbol{\theta}}$ depends on θ_k and all of the information correlation terms have derivatives of density functions that do not depend on this parameter, then all of the information correlation terms are zero. Hence:

$$m_{F(I)}(f_{\mathbf{R},\mathbf{T}|\mathbf{S};\boldsymbol{\theta}})_{\theta_k} + m_{F(I)}(f_{\mathbf{S},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} = m_{F(I)}(f_{\mathbf{S},\mathbf{T}|\mathbf{R};\boldsymbol{\theta}})_{\theta_k} + m_{F(I)}(f_{\mathbf{R},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \tag{119}$$

Given that $m_{F(I)}(f_{\mathbf{R},\mathbf{T}|\mathbf{S};\boldsymbol{\theta}})_{\theta_k} = 0$ because $\mathbf{R}$ and $\mathbf{T}$ are independent given $\mathbf{S}$, and $m_{F(I)}(f_{\mathbf{S},\mathbf{T}|\mathbf{R};\boldsymbol{\theta}})_{\theta_k} \geq 0$, then:

$$m_{F(I)}(f_{\mathbf{R},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \leq m_{F(I)}(f_{\mathbf{S},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \tag{120}$$

□

Given that in the previous proof, all of the information correlation terms are zero, then $m_{F(II)}(f_{\mathbf{R},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} = m_{F(I)}(f_{\mathbf{R},\mathbf{T};\boldsymbol{\theta}})_{\theta_k}$, and $m_{F(II)}(f_{\mathbf{S},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} = m_{F(I)}(f_{\mathbf{S},\mathbf{T};\boldsymbol{\theta}})_{\theta_k}$. Thus, the following corollary is obtained:

Corollary 5. *Given a Markov chain* $\mathbf{R} \to \mathbf{S} \to \mathbf{T}$*, where only* $f_{\mathbf{T}|\mathbf{S};\boldsymbol{\theta}}$ *depends on* θ_k*, then:*

$$m_{F(II)}(f_{\mathbf{R},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \leq m_{F(II)}(f_{\mathbf{S},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} \tag{121}$$

Proof. The conditional independence provided by the Markovicity of the random variables follows directly from the mutual Fisher information Type II definition, and in this case, the values of mutual Fisher information Type I and mutual Fisher information Type II are identical. □

Using the definition of mutual Fisher information Type II and the previous expression, it is readily obtained, in a simpler way, a result already proven by Plastino *et al.* [52]:

Corollary 6. *From the previous results, it is obvious that:*

$$i_F\left(f_{\mathbf{T}|\mathbf{R};\boldsymbol{\theta}}\right)_{\theta_k} \leq i_F\left(f_{\mathbf{T}|\mathbf{S};\boldsymbol{\theta}}\right)_{\theta_k} \tag{122}$$

Proof. From Equation (121):

$$\begin{aligned} m_{F(II)}(f_{\mathbf{R},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} &\leq m_{F(II)}(f_{\mathbf{S},\mathbf{T};\boldsymbol{\theta}})_{\theta_k} && (123)\\ i_F\left(f_{\mathbf{T}|\mathbf{R};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{T};\boldsymbol{\theta}}\right)_{\theta_k} &\leq i_F\left(f_{\mathbf{T}|\mathbf{S};\boldsymbol{\theta}}\right)_{\theta_k} - i_F\left(f_{\mathbf{T};\boldsymbol{\theta}}\right)_{\theta_k} && (124)\\ i_F\left(f_{\mathbf{T}|\mathbf{R};\boldsymbol{\theta}}\right)_{\theta_k} &\leq i_F\left(f_{\mathbf{T}|\mathbf{S};\boldsymbol{\theta}}\right)_{\theta_k} && (125) \end{aligned}$$

□

In other words, in any Markovian process, the further away that the random variables used by the estimator are, the larger is the variance of the estimated parameter.

10.4. Upper Bound on Estimation Error

A well-known result states that given a variance η, of all possible density functions, the one that maximizes the differential entropy is the Gaussian density function [30]. Hence, for an arbitrary density function f_X, some side information Y and an estimator $\hat{X}$, it is possible to obtain an estimation version of the Fano inequality [10] (p. 255):

$$\frac{1}{2\pi e} e^{2h_S(f_{X|Y})} \leq \mathbb{E}_X\left\{(X - \hat{X}(Y))^2\right\} \tag{126}$$

In the context of Fisher information, the same question arises: is it possible to bound the estimation error using this quantity as well? Surprisingly, the answer is yes, but in the form of an upper bound. Thus, Shannon entropy can be used to set error lower bounds and Fisher information upper ones. In order to establish this bound, the following setup is defined, where a random variable $\mathbf{R}$ is given, and a related random variable $\mathbf{Y}$ is observed, which, in turn, is used to calculate a function $\hat{\mathbf{R}} = \mathbf{g}(\mathbf{Y})$. It is desired to bound the probability that $(\mathbf{R} - \hat{\mathbf{R}})^2 > \epsilon$. It is important to note that $\mathbf{R} \to \mathbf{Y} \to \hat{\mathbf{R}}$ is a Markov chain and that $\hat{\mathbf{R}}$ depends on $\boldsymbol{\theta}$.

Theorem 14. *Given a random variable* $\mathbf{R}$ *and an estimator of it named* $\hat{\mathbf{R}}$*, the estimation error is defined by:*

$$E = (\mathbf{R} - \hat{\mathbf{R}})^2 \tag{127}$$

Then, the probability that the estimation error exceeds some ε *value:*

$$\mathcal{P}\{E > \varepsilon\} \leq \frac{i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}};\boldsymbol{\theta}}\right)_{\theta_k}}{i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}},E=\xi;\boldsymbol{\theta}}\right)_{\theta_k}} \tag{128}$$

for some $\xi \in [\varepsilon, \infty]$.

Proof. Using the chain rule for Fisher information:

$$i_F\left(f_{\mathbf{R},E|\hat{\mathbf{R}};\boldsymbol{\theta}}\right)_{\theta_k} = i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}},E;\boldsymbol{\theta}}\right)_{\theta_k} + i_F\left(f_{E|\hat{\mathbf{R}};\boldsymbol{\theta}}\right)_{\theta_k} = i_F\left(f_{E|\mathbf{R},\hat{\mathbf{R}};\boldsymbol{\theta}}\right)_{\theta_k} + i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}};\boldsymbol{\theta}}\right)_{\theta_k} \tag{129}$$

Using the fact that given $\mathbf{R}$ and $\hat{\mathbf{R}}$, then E is no longer a random variable, then:

$$i_F\left(f_{E|\mathbf{R},\hat{\mathbf{R}};\boldsymbol{\theta}}\right)_{\theta_k} = 0 \tag{130}$$

Hence,

$$i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}},E;\boldsymbol{\theta}}\right)_{\theta_k} + i_F\left(f_{E|\hat{\mathbf{R}};\boldsymbol{\theta}}\right)_{\theta_k} = i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}};\boldsymbol{\theta}}\right)_{\theta_k} \tag{131}$$

Neglecting $i_F\left(f_{E|\hat{\mathbf{R}};\boldsymbol{\theta}}\right)_{\theta_k}$, because it is always greater or equal to zero, it is obtained:

$$i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}},E;\boldsymbol{\theta}}\right)_{\theta_k} \leq i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}};\boldsymbol{\theta}}\right)_{\theta_k} \tag{132}$$

Moreover, the term:

$$i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}},E;\boldsymbol{\theta}}\right)_{\theta_k} = \int_{e\in E}^{\infty}\iint_{\mathbf{r}\in\mathbf{R},\hat{\mathbf{r}}\in\hat{\mathbf{R}}|(\mathbf{r}-\hat{\mathbf{r}})^2=e} f_{\mathbf{R},\hat{\mathbf{R}},E;\boldsymbol{\theta}}(\mathbf{r},\hat{\mathbf{r}},e)\left(\frac{\partial \ln f_{\mathbf{R}|\hat{\mathbf{R}},E;\boldsymbol{\theta}}(\mathbf{r}|\hat{\mathbf{r}},e)}{\partial\theta_k}\right)^2 d\mathbf{r}d\hat{\mathbf{r}}de \tag{133}$$

$$\begin{aligned} &= \int_{e\in E|e\leq\varepsilon}^{\varepsilon}\iint_{\mathbf{r}\in\mathbf{R},\hat{\mathbf{r}}\in\hat{\mathbf{R}}|(\mathbf{r}-\hat{\mathbf{r}})^2=e} f_{\mathbf{R},\hat{\mathbf{R}},E;\boldsymbol{\theta}}(\mathbf{r},\hat{\mathbf{r}},e)\left(\frac{\partial \ln f_{\mathbf{R}|\hat{\mathbf{R}},E;\boldsymbol{\theta}}(\mathbf{r}|\hat{\mathbf{r}},e)}{\partial\theta_k}\right)^2 d\mathbf{r}d\hat{\mathbf{r}}de \\ &\quad + \int_{e\in E|e>\varepsilon}^{\infty}\iint_{\mathbf{r}\in\mathbf{R},\hat{\mathbf{r}}\in\hat{\mathbf{R}}|(\mathbf{r}-\hat{\mathbf{r}})^2=e} f_{\mathbf{R},\hat{\mathbf{R}},E;\boldsymbol{\theta}}(\mathbf{r},\hat{\mathbf{r}},e)\left(\frac{\partial \ln f_{\mathbf{R}|\hat{\mathbf{R}},E;\boldsymbol{\theta}}(\mathbf{r}|\hat{\mathbf{r}},e)}{\partial\theta_k}\right)^2 d\mathbf{r}d\hat{\mathbf{r}}de \end{aligned} \tag{134}$$

$$\geq \int_{e\in E|e>\varepsilon}^{\infty}\iint_{\mathbf{r}\in\mathbf{R},\hat{\mathbf{r}}\in\hat{\mathbf{R}}|(\mathbf{r}-\hat{\mathbf{r}})^2=e} f_{\mathbf{R},\hat{\mathbf{R}},E;\boldsymbol{\theta}}(\mathbf{r},\hat{\mathbf{r}},e)\left(\frac{\partial \ln f_{\mathbf{R}|\hat{\mathbf{R}},E;\boldsymbol{\theta}}(\mathbf{r}|\hat{\mathbf{r}},e)}{\partial\theta_k}\right)^2 d\mathbf{r}d\hat{\mathbf{r}}de \tag{135}$$

$$= \int_{e\in E|e>\varepsilon}^{\infty}\iint_{\mathbf{r}\in\mathbf{R},\hat{\mathbf{r}}\in\hat{\mathbf{R}}|(\mathbf{r}-\hat{\mathbf{r}})^2=e} f_{\mathbf{R},\hat{\mathbf{R}}|E;\boldsymbol{\theta}}(\mathbf{r},\hat{\mathbf{r}}|e)f_{E;\boldsymbol{\theta}}(e)\left(\frac{\partial \ln f_{\mathbf{R}|\hat{\mathbf{R}},E;\boldsymbol{\theta}}(\mathbf{r}|\hat{\mathbf{r}},e)}{\partial\theta_k}\right)^2 d\mathbf{r}d\hat{\mathbf{r}}de \tag{136}$$

$$= \int_{e\in E|e>\varepsilon}^{\infty} f_{E;\boldsymbol{\theta}}(e)\iint_{\mathbf{r}\in\mathbf{R},\hat{\mathbf{r}}\in\hat{\mathbf{R}}|(\mathbf{r}-\hat{\mathbf{r}})^2=e} f_{\mathbf{R},\hat{\mathbf{R}}|E;\boldsymbol{\theta}}(\mathbf{r},\hat{\mathbf{r}}|e)\left(\frac{\partial \ln f_{\mathbf{R}|\hat{\mathbf{R}},E;\boldsymbol{\theta}}(\mathbf{r}|\hat{\mathbf{r}},e)}{\partial\theta_k}\right)^2 d\mathbf{r}d\hat{\mathbf{r}}de \tag{137}$$

$$= \int_{e\in E|e>\varepsilon}^{\infty} f_{E;\boldsymbol{\theta}}(e) i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}},E=e;\boldsymbol{\theta}}\right)_{\theta_k} de \tag{138}$$

Using the mean value theorem, for some $\xi \in [\varepsilon, \infty]$:

$$i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}},E=\xi;\boldsymbol{\theta}}\right)_{\theta_k}\int_{e\in E|e>\varepsilon}^{\infty} f_{E;\boldsymbol{\theta}}(e)de = i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}},E=\xi;\boldsymbol{\theta}}\right)_{\theta_k}\cdot\mathcal{P}\{E>\varepsilon\} \leq i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}},E;\boldsymbol{\theta}}\right)_{\theta_k} \leq i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}};\boldsymbol{\theta}}\right)_{\theta_k} \tag{139}$$

Hence:

$$\mathcal{P}\left\{E > \varepsilon\right\} \leq \frac{i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}};\boldsymbol{\theta}}\right)_{\theta_k}}{i_F\left(f_{\mathbf{R}|\hat{\mathbf{R}},E=\xi;\boldsymbol{\theta}}\right)_{\theta_k}} \tag{140}$$

□

11. Discussion

The Fisher information, which sets a bound on how precise the estimation of an unknown parameter of a density function can be, has an associated set of properties that are equivalent to those of Shannon's differential entropy. The properties presented in this work help to understand how to manipulate and use Fisher information in ways that so far have been exclusive to Shannon's differential entropy. These properties that are of special importance to the generalization of the mutual information concept for the Fisher information realm are a new version of the data processing theorem that shows that Fisher information decreases in a Markov chain and an upper bound of the estimation error of a random variable that is regulated by the Fisher information.

Acknowledgments

The author thanks Alexis Fuentes and Carlos Alarcón for reviewing this work, and helping to improve some expressions. The author also thanks CONICYT Chile for its grant FONDECYT 1120680.

Conflicts of Interest

The author declares no conflict of interest.

A. Boundary Condition

A general result from calculus establishes that for any function $g(x, \theta_k)$, the following is true:

$$\frac{\partial}{\partial\theta_k}\int_{l(\theta_k)}^{u(\theta_k)} g(x,\theta_k)dx = g(u(\theta_k),\theta_k)\frac{\partial u(\theta_k)}{\partial\theta_k} - g(l(\theta_k),\theta_k)\frac{\partial l(\theta_k)}{\partial\theta_k} + \int_{l(\theta_k)}^{u(\theta_k)}\frac{\partial g(x,\theta_k)}{\partial\theta_k}dx \tag{141}$$

In the case of a vector integral, the previous expression applies to all of the components without any loss of generality.

Some of the results in this work use the following condition:

Condition 1 (Boundary Condition)**.** *A function complies with the boundary condition if it is possible to neglect the boundary terms in Equation (141), such that:*

$$\frac{\partial}{\partial\theta_k}\int g(x,\theta_k)dx = \int\frac{\partial g(x,\theta_k)}{\partial\theta_k}dx \tag{142}$$

This condition corresponds to what sometimes are called regular cases [34] (p. 373).

It is important to keep in mind that not all density functions go along with this condition. As an example, in calculations that involve the uniform density function, where the parameters define the support, it is not possible to neglect the terms, and the boundary condition does not hold. Hence, it is always necessary to check whether the condition holds or not. If not, one may arrive at false results.

However, it is always possible to add a smooth function, one that does not change the original function too much, such that the new mathematical expression does comply with the boundary condition. In this way, functions, such as the uniform density function, as an example, can be adjusted to comply with this condition.

References

1. Shannon, C. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
2. Fisher, R. Theory of Statistical Estimation. *Proc. Camb. Philos. Soc.* **1925**, *22*, 700–725.
3. Rao, C.R. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81–89.
4. Cramer, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, USA, 1945.
5. Kullback, S. *Information Theory and Statistics*; Dover Publications Inc.: Mineola, NY, USA, 1968.
6. Blahut, R.E. *Principles and Practice of Information Theory*; Addison-Wesley Publishing Company: Boston, MA, USA, 1987.
7. Frieden, B.R. *Science from Fisher Information: A Unification*; Cambridge University Press: Cambridge, UK, 2004.
8. Stam, A.J. Some mathematical properties of quantities of information. Ph.D. Thesis, Technological University of Delft, Delft, The Netherlands, 1959.
9. Stam, A.J. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Control* **1959**, *2*, 101–112.
10. Cover, T.; Thomas, J. *Elements of Information Theory*; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 2006.
11. Narayanan, K.R.; Srinivasa, A.R. *On the Thermodynamic Temperature of a General Distribution*; Cornell University Library: Ithaca, NY, USA, 2007.
12. Guo, D. Relative Entropy and Score Function: New Information-Estimation Relationships through Arbitrary Additive Perturbation. In Proceedings of the IEEE International Symposium on Information Theory, Seoul, Korea, 28 June–3 July 2009; pp. 814–818.
13. Blachman, N.M. The Convolution Inequality for Entropy Powers. *IEEE Trans. Inf. Theory* **1965**, *11*, 267–271.
14. Costa, M.H.M.; Cover, T.M. *On the Similarity of the Entropy Power Inequality and the Brunn Minkowski Inequality*; Technical Report, Stanford University: Stanford, CA, USA, 1983.
15. Zamir, R.; Feder, M. A generalization of the entropy power inequality with applications. *IEEE Trans. Inf. Theory* **1993**, *39*, 1723–1728.
16. Lutwak, E.; Yang, D.; Zhang, G. CramerâĂŞRao and Moment-Entropy Inequalities for Renyi Entropy and Generalized Fisher Information. *IEEE Trans. Inf. Theory* **2005**, *51*, 473–478.
17. Frieden, B.R.; Plastino, A.; Plastino, A.R.; Soffer, B.H. Fisher-Based Thermodynamics: Its Legendre Transform and Concavity Properties. *Phys. Rev. E* **1999**, *60*, 48–53.
18. Frieden, B.R.; Plastino, A.; Plastino, A.R.; Soffer, B.H. Non-equilibrium thermodynamics and Fisher information: An illustrative example. *Phys. Lett. A* **2002**, *304*, 73–78.

19. Frieden, B.R.; Petri, M. Motion-dependent levels of order in a relativistic universe. *Phys. Rev. E* **2012**, *86*, 1–5.
20. Frieden, B.R.; Gatenby, R.A. Principle of maximum Fisher information from Hardy's axioms applied to statistical systems. *Phys. Rev. E* **2013**, *88*, 1–6.
21. Flego, S.; Olivares, F.; Plastino, A.; Casas, M. Extreme Fisher Information, Non-Equilibrium Thermodynamics and Reciprocity Relations. *Entropy* **2011**, *13*, 184–194.
22. Venkatesan, R.C.; Plastino, A. Legendre transform structure and extremal properties of the relative Fisher information. *Phys. Lett. A* **2014**, *378*, 1341–1345.
23. Van Trees, H.L. *Detection, Estimation, and Modulation Theory: Part 1*; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 2001.
24. Amari, S.I. Natural Gradient Works Efficiently in Learning. *Neural Comput.* **1998**, *10*, 251–276.
25. Pascanu, R.; Bengio, Y. *Revisiting Natural Gradient for Deep Networks*; Cornell University Library: Ithaca, NY, USA, 2014; pp. 1–18.
26. Luo, S. Maximum Shannon entropy, minimum Fisher information, and an elementary game. *Found. Phys.* **2002**, *32*, 1757–1772.
27. Langley, R.S. Probability Functionals for Self-Consistent and Invariant Inference: Entropy and Fisher Information. *IEEE Trans. Inf. Theory* **2013**, *59*, 4397–4407.
28. Zegers, P.; Fuentes, A.; Alarcon, C. Relative Entropy Derivative Bounds. *Entropy* **2013**, *15*, 2861–2873.
29. Cohen, M. The Fisher Information and Convexity. *IEEE Trans. Inf. Theory* **1968**, *14*, 591–592.
30. Cover, T.; Thomas, J. *Elements of Information Theory*; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 1991.
31. Frieden, B.R. *Physics from Fisher Information: A Unification*; Cambridge University Press: Cambridge, UK, 1998.
32. Zamir, R. A Proof of the Fisher Information Inequality Via a Data Processing Argument. *IEEE Trans. Inf. Theory* **1998**, *44*, 1246–1250.
33. Taubman, D.; Marcellin, M. *JPEG2000: Image Compression Fundamentals, Standards, and Practice*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2002.
34. Hogg, R.V.; Craig, A.T. *Introduction to Mathematical Statistics*; Prentice Hall: Upper Saddle River, NJ, USA, 1995.
35. Frieden, B.R. *Probability, Statistical Optics, and Data Testing*; Springer-Verlag: Berlin, Germany, 1991.
36. Otto, F.; Villani, C. Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality. *J. Funct. Anal.* **2000**, *173*, 361–400.
37. Yáñez, R.J.; Sánchez-Moreno, P.; Zarzo, A.; Dehesa, J.S. Fisher information of special functions and second-order differential equations. *J. Math. Phys.* **2008**, *49*, 082104.
38. Gianazza, U.; Savaré, G.; Toscani, G. The wasserstein gradient flow of the fisher information and the quantum drift-diffusion equation. *Arch. Ration. Mech. Anal.* **2009**, *194*, 133–220.
39. Verdú, S. Mismatched Estimation and Relative Entropy. *IEEE Trans. Inf. Theory* **2010**, *56*, 3712–3720.

40. Hirata, M.; Nemoto, A.; Yoshida, H. An integral representation of the relative entropy. *Entropy* **2012**, *14*, 1469–1477.
41. Sánchez-Moreno, P.; Zarzo, A.; Dehesa, J.S. Jensen divergence based on Fisher's information. *J. Phys. A: Math. Theor.* **2012**, *45*, 125305.
42. Yamano, T. Phase space gradient of dissipated work and information: A role of relative Fisher information. *J. Math. Phys.* **2013**, *54*, 1–9.
43. Yamano, T. De Bruijn-type identity for systems with flux. *Eur. Phys. J. B* **2013**, *86*, 363.
44. Bobkov, S.G.; Chistyakov, G.P.; Gotze, F. Fisher information and the central limit theorem. *Probab. Theory Relat. Fields* **2014**, *159*, 1–59.
45. Zegers, P. Some New Results on The Architecture, Training Process, and Estimation Error Bounds for Learning Machines. Ph.D. Thesis, The University of Arizona, Tucson, AZ, USA, 2002.
46. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.
47. Lutwak, E.; Yang, D.; Zhang, G. Renyi entropy and generalized Fisher information. *IEEE Trans. Inf. Theory* **2005**, *51*, 473–478 .
48. Kagan, A.; Yu, T. Some Inequalities Related to the Stam Inequality. *Appl. Math.* **2008**, *53*, 195–205.
49. Lutwak, E.; Lv, S.; Yang, D.; Zhang, G. Extensions of Fisher Information and Stam's Inequality. *IEEE Trans. Inf. Theory* **2012**, *58*, 1319–1327.
50. Bercher, J.F. *On Generalized Cramér-Rao Inequalities, and an Extension of the Shannon-Fisher-Gauss Setting*; Cornell University Library: Ithaca, NY, USA, 2014.
51. Stein, M.; Mezghani, A.; Nossek, J.A. A Lower Bound for the Fisher Information Measure. *IEEE Signal Process. Lett.* **2014**, *21*, 796–799.
52. Plastino, A.; Plastino, A. Symmetries of the Fokker-Planck equation and the Fisher-Frieden arrow of time. *Phys. Rev. E* **1996**, *54*, 4423–4426.