Filomat 39:9 (2025), 3127–3144
https://doi.org/10.2298/FIL2509127K

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Mixture Fisher-Shannon information measure: Extensions and application

Omid Kharazmi[a,*], Javier E. Contreras-Reyes[b]

[a]*Department of Statistics, Faculty of Mathematical Sciences, Vali-e-Asr University of Rafsanjan, Iran*
[b]*Instituto de Matemática, Física y Estadística, Facultad de Ingeniería y Negocios, Universidad de Las Américas, Sede Viña del Mar, 7 Norte 1348, Viña del Mar, Chile*

Abstract. The purpose of this paper is two-fold. In the first part, we introduce a novel information measure known as the mixture Fisher–Shannon information measure, motivated by de Bruijn's identity. We also propose and study a specific case of this measure called the difference information measure along with its Jensen version. Subsequently, the paper delves into an examination of their properties. In the second part, we introduce (p, η)-Jensen difference Fisher–Shannon information measure. Additionally, we explore possible connections between this divergence measure and Jensen–Shannon entropy and Jensen–Fisher information measures. Our analysis not only examines theoretical foundations but also extends to practical applications. Specifically, we apply these measures to analyze time series data concerning the fish condition factor index, providing valuable insights into data interpretation.

1. Introduction

Information theory stands as one of the paramount branches of science, finding applications across diverse fields such as statistics, physics, economics, and engineering. Over the past seven decades, this theory has captivated the attention of numerous researchers. In particular, Shannon entropy [28] and Fisher information [14] measures are two fundamental information-theoretic concepts that play crucial roles in various scientific and engineering disciplines. Shannon entropy characterizes the uncertainty associated with a continuous random variable X possessing a probability density function (PDF) f defined on the support $\mathcal{X}$. Shannon entropy is defined by

$$H(X) \equiv \mathcal{H}(f) = -\int_{\mathcal{X}} f(x) \log f(x) dx, \tag{1}$$

where log denotes the natural logarithm.

2020 *Mathematics Subject Classification*. Primary 94A15; Secondary 62B10.
Keywords. Fisher information; Shannon entropy; Extropy measure; de Bruijn's identity; Jensen–Fisher information; Jensen–Shannon entropy.
Received: 10 April 2024; Accepted: 29 October 2024
Communicated by Biljana Č. Popović
* Correspondng author: Omid Kharazmi
Email address: omid.kharazmi@vru.ac.ir (Omid Kharazmi)
ORCID iD: https://orcid.org/0000-0003-4176-9708 (Omid Kharazmi), https://orcid.org/0000-0003-1172-5456 (Javier E. Contreras-Reyes)

[14] proposed a quantity of information, describing the interior properties of a probabilistic model, that has become vital to likelihood-based inferential methods. Fisher information as well as Shannon entropy are very important and fundamental criteria in statistical inference, physics, thermodynamics, information theory, and some other disciplines. For more details, see [3], [31], [15], [16, 17], [9, 10] and [20, 21]. The Fisher information of a random variable X, about parameter θ, is defined as

$$\mathcal{I}(\theta) \quad = \quad \int_{\mathcal{X}} \left(\frac{\partial \log f_\theta(x)}{\partial \theta} \right)^2 f_\theta(x) dx. \tag{2}$$

For some recent studies on its properties, one may refer to [15] and [1]. In order to simplify the notation, we suppress $\mathcal{X}$ for integration with respect to dx throughout the paper, unless a distinction becomes necessary. There is another kind of Fisher information, known as Fisher information of the density itself. Let X be a continuous random variable with density function f and $\rho(x) = \frac{\partial \log f(x)}{\partial x}$. Then, the Fisher information of density f itself is defined as

$$\mathcal{I}(X) = \mathcal{I}(f) = \int \left\{ \rho(x) \right\}^2 f(x) dx. \tag{3}$$

Incidentally, the Fisher information measures in (2) and (3) are identical when θ is a location parameter, or equivalently, when the density f belongs to the location family of distributions [13].

[5] introduced a k-generalized Fisher information, extending (2) for $k > 0$, in the form

$$\mathcal{I}_k(\theta) = \int \left| \frac{\partial \log f_\theta(x)}{\partial \theta} \right|^k f_\theta(x) dx. \tag{4}$$

Subsequently, [5] demonstrated the applicability of this measure within the context of non-extensive thermostatistics and provided insightful results in this domain. More recently, [6] explored the density version of $I_k(f_\theta)$ in (3), expressed as

$$\mathcal{I}_k(f) = \int \left| \rho(x) \right|^k f(x) dx, \tag{5}$$

where $\rho(x) = \frac{f'(x)}{f(x)}$ represents the score function corresponding to the density f. To simplify notation, we suppress $\mathcal{X}$ for integration with respect to dx throughout the paper, unless a distinction is necessary.

de Bruijn's identity links two fundamental concepts in information theory: entropy and Fisher information. Let X be a random variable with finite variance and density function $f(x)$ and Z be an independent normal variable with zero mean and unit variance. By denoting $X_t = X + \sqrt{t}Z$, de Bruijn's identity states

$$\frac{\partial}{\partial t} H(X_t) = \frac{1}{2} \mathcal{I}(X_t), \tag{6}$$

where H is the differential entropy and $\mathcal{I}$ is the Fisher information. Further, if the limit exists as $t \to 0$, then

$$\frac{\partial}{\partial t} H(X_t) \Big|_{t=0} = \frac{1}{2} \mathcal{I}(X). \tag{7}$$

The identity in (6) plays a key role in signal processing; see, for example, [13]. Moreover, an integral representation of de Bruijn's identity can be obtained between entropy of X and Fisher information of X_t [24] as

$$H(X) = \frac{1}{2} \log(2\pi e) - \frac{1}{2} \int_0^\infty \left[\mathcal{I}(X_t) - \frac{1}{1+t} \right] dt. \tag{8}$$

This paper is motivated by a quest to explore information measures, focusing on Fisher and Shannon information. Initially, the study delves into comprehending these fundamental concepts within information theory. However, the ambition transcends mere understanding, aiming to innovate by proposing a new information measure called the mixture Fisher–Shannon information measure. This measure is designed to amalgamate insights from both Fisher and Shannon information measures. The motivation for this innovation arises from de Bruijn's identity [24], which highlights a connection between Fisher information and Shannon entropy. This identity serves as a guiding principle for the proposed measure. Building upon this motivation, the paper defines the mixture information measure, $D_F^S(f;\alpha)$, incorporating Fisher information and Shannon entropy. The formulation includes a weight parameter α, enabling adaptable adjustments reflecting the relative importance attributed to each component. Through this exploration, the study seeks to unravel the properties and implications of this novel information measure, thus enriching the broader understanding of information theory.

The rest of this paper is organized as follows. In Section **??**, we propose a mixture Fisher-Shannon information measure by considering de Bruijn's identity and investigate its properties. This section also examines a special case of the mixture Fisher-Shannon information measure, known as the difference information measure, which is defined based on the difference between Fisher and Shannon entropy measures. Section 3 is devoted to the Jensen-difference Fisher-Shannon information measure. We begin by establishing and exploring its fundamental properties. Subsequently, we introduce two extensions of this measure and present key results related to these extensions. The practical implications of our work are examined in Section 4, where we apply the developed concepts to analyze time series data related to the fatness condition factor (CF) index of anchovies. By employing mixture information measures and associated divergences, we revisit previous analyses conducted in references [8, 9, 11]. Finally, we present some concluding remarks in Section 5.

2. Mixture Fisher–Shannon information measure

In this section, we first consider Fisher and Shannon information measures, and then define a novel information measure, subsequently examining its properties. From de Bruijn's identity in (7), we have

$$\lim_{t\longrightarrow 0}\left(\frac{1}{2}\mathcal{I}(X)-\frac{\partial}{\partial t}H(X_t)\right)=0. \tag{9}$$

This relationship serves as motivation to introduce the following definition.

Definition 2.1. *Let X be a continuous random variable with density f defined on support $\mathcal{X}$, having Fisher information $\mathcal{I}(f)$ and Shannon entropy $\mathcal{H}(f)$. Then, a mixture information measure in terms of $\mathcal{I}(f)$ and $\mathcal{H}(f)$ for $0 \le \alpha \le 2$, denoted by $D_F^S(f;\alpha)$ (or $D_F^S(X;\alpha)$), is defined as*

$$\mathcal{D}_F^S(f;\alpha)=\alpha\frac{\mathcal{I}(f)}{2}+(1-\alpha)\mathcal{H}(f). \tag{10}$$

For a specific PDF f, the measure (10) can be interpreted as a linear equation, i.e., it could have a positive trend when α increases if the Fisher information is larger than Shannon entropy, and have a negative trend when α increases if the Shannon entropy is larger than Fisher information. It seen that the proposed measure (10) can be derived based on the expectation of the self-information measure, as follows.

$$\begin{aligned}\mathcal{D}_F^S(f;\alpha) &= \int\frac{\alpha}{2}\Big\{\rho(x)\Big\}^2 f(x)\,dx-(1-\alpha)\int f(x)\log f(x)\,dx\\ &= \int\left[\frac{\alpha}{2}\left\{\frac{\partial\log f(x)}{\partial x}\right\}^2-(1-\alpha)\log f(x)\right]f(x)\,dx\\ &= E\Big[\frac{\alpha}{2}\Big(S'(X)\Big)^2+(1-\alpha)S(X)\Big],\end{aligned}$$

where $S(x)=-\log f(x)$ represents the self-information measure of the density f at point x.

The positivity or negativity of $\mathcal{D}_F^S(f;\alpha)$ depends on the relative magnitudes of Fisher information and Shannon entropy, as well as the chosen weight α. A higher weight given to Fisher information tends to make the measure positive, while a lower weight may result in a negative measure due to the dominance of Shannon entropy

If $0 \leq \alpha \leq 1$, from the relation (10), it is evident that the mixture information measure $D_F^S(f;\alpha)$ is a weighted combination of the Fisher information $\mathcal{I}(f)$ and the Shannon entropy $\mathcal{H}(f)$, where the weight parameter α controls the contribution of each component.

- When $\alpha = 1$, the measure is solely based on the Fisher information, indicating a focus on the amount of information about the density itself.
- When $\alpha = 0$, the measure relies only on the Shannon entropy, implying a focus on the uncertainty or randomness in the distribution.
- For values of α between 0 and 1, the measure balances between the two components, reflecting a trade-off between information content and uncertainty.

This weighted approach offers a smooth transition between the two types of information, enabling us to explore the behavior of this information measure in different scenarios (global vs. local). Additionally, our formulation is consistent with other well-known combined measures, such as the Fisher-Shannon complexity, which also blends these measures to provide insight into both local and global characteristics. In practical terms, the mixture information measure can be interpreted as a generalization that captures a spectrum of information, from purely global (entropy-based) to purely local (Fisher-based), allowing for a richer analysis than either measure would provide on its own.

If $1 \leq \alpha \leq 2$, the mixture information measure can be considered as a negative weighted mixture information measure, and the special case $\alpha = 2$ possesses some interesting results that will be studied in the next sections particularly.

Theorem 2.2. *Let $\mathcal{H}(f) > 0$, then a lower bound for the $\mathcal{D}_F^S(f;\alpha)$ for $0 < \alpha < 1$ is given by*

$$\mathcal{D}_F^S(f;\alpha) \geq max\Big\{0^+, \Big(\frac{\sqrt{\mathcal{I}(f)}}{\mathcal{H}(f)}\Big)^{\alpha}\Big(1 + 2J(f)\Big)\Big\}, \tag{11}$$

where $J(f)$ is extropy measure and defined as

$$J(f) = -\frac{1}{2}\int f^2(x)dx, \tag{12}$$

and the notation 0^+ is commonly used to denote a value that approaches zero from the positive side, indicating a limit as a variable approaches zero without actually reaching zero itself

Proof: From the definition of $\mathcal{D}_F^S(f;\alpha)$ and by employing the inequality known as the arithmetic mean-geometric mean inequality, we have

$$\begin{aligned}
\mathcal{D}_F^S(f;\alpha) &= \alpha\frac{\mathcal{I}(f)}{2} + (1-\alpha)\mathcal{H}(f) \\
&\geq \Big(\sqrt{\mathcal{I}(f)}\Big)^{\alpha}\mathcal{H}(f)^{1-\alpha} \\
&= \Big(\frac{\sqrt{\mathcal{I}(f)}}{\mathcal{H}(f)}\Big)^{\alpha}\mathcal{H}(f) \\
&= \Big(\frac{\sqrt{\mathcal{I}(f)}}{\mathcal{H}(f)}\Big)^{\alpha}\Big(-\int_{\mathcal{X}} f(x)\log f(x)dx\Big) \\
&\geq \Big(\frac{\sqrt{\mathcal{I}(f)}}{\mathcal{H}(f)}\Big)^{\alpha}\Big(1-\int_{\mathcal{X}} f^2(x)dx\Big) \\
&= \Big(\frac{\sqrt{\mathcal{I}(f)}}{\mathcal{H}(f)}\Big)^{\alpha}\Big(1 + 2J(f)\Big),
\end{aligned}$$

where the last inequality follows from the inequality $-\log(x) \geq 1 - x$, valid for $x > 0$.

On the other hand, from the assumption $\mathcal{H}(f) > 0$ and $0 < \alpha < 1$, it is seen that

$$\mathcal{D}_F^S(f;\alpha) = \alpha\frac{\mathcal{I}(f)}{2} + (1-\alpha)\mathcal{H}(f) > 0.$$

Now from the above results, we have

$$\mathcal{D}_F^S(f;\alpha) \geq max\Big\{0^+, \Big(\frac{\sqrt{\mathcal{I}(f)}}{\mathcal{H}(f)}\Big)^{\alpha}\Big(1 + 2J(f)\Big)\Big\},$$

as required.

Corollary 2.3. *From Theorem 2.2 and under the assumption $\mathcal{H}(f) > 0$, for $0 < \alpha < 1$, we have*

$$\mathcal{D}_F^S(f;\alpha) \geq \Big(\frac{\sqrt{\mathcal{I}(f)}}{\mathcal{H}(f)}\Big)^{\alpha}. \tag{13}$$

To further explore the concept of mixture information measure in (10), we turn our attention to a specific case defined by equation (10), where the parameter α takes the value of 2. This equation represents the mixture information measure $\mathcal{D}_F^S(f;\alpha)$, which combines the Fisher information $\mathcal{I}(f)$ and the Shannon entropy $\mathcal{H}(f)$ of a density function f. In the specific case where $\alpha = 2$, the emphasis is solely on the difference between the Fisher information and the Shannon entropy, providing a clear measure of the net information content in the density function.

Definition 2.4. *Let X be a continuous random variable with density f defined on support $\mathcal{X}$, having Fisher information $\mathcal{I}(f)$ and Shannon entropy $\mathcal{H}(f)$. Then, a difference information measure between $\mathcal{I}(f)$ and $\mathcal{H}(f)$, denoted by $D_F^S(f)$ (or $D_F^S(X)$), is defined as*

$$\mathcal{D}_F^S(f) \equiv \mathcal{D}_F^S(f;\alpha)\Big|_{\alpha=2} = \mathcal{I}(f) - \mathcal{H}(f). \tag{14}$$

In this definition, we introduce a concept called the "difference information measure" denoted by $D_F^S(f)$, which compares the Fisher information $\mathcal{I}(f)$ and the Shannon entropy $\mathcal{H}(f)$ of a density function f on the support $\mathcal{X}$.

When $\mathcal{I}(f)$ is greater than $\mathcal{H}(f)$, $D_F^S(f)$ will be positive, indicating that the density function carries more information about the parameters of interest than the randomness in its distribution. Conversely, when $\mathcal{I}(f)$ is less than $\mathcal{H}(f)$, $D_F^S(f)$ will be negative, suggesting that the randomness in the distribution dominates the information content.

Therefore, $D_F^S(f)$ serves as a useful metric for quantifying the relative importance of information content and uncertainty in a probability distribution described by the density function f.

Lemma 2.5. *Let X be a continuous random variable with density function f. Then, for a positive real value a, we have*

$$\mathcal{D}_F^S(aX) = \mathcal{D}_F^S(X) + \frac{1-a^2}{a^2}\mathcal{I}(X) - \log(a). \tag{15}$$

Proof: From the definition of $\mathcal{D}_F^S(X)$, for $a > 0$, we have

$$\begin{aligned}\mathcal{D}_F^S(aX) &= \mathcal{I}(aX) - H(aX)\\ &= \frac{1}{a^2}\mathcal{I}(f) - \mathcal{H}(f) - \log(a)\\ &= \mathcal{D}_F^S(X) + \frac{1-a^2}{a^2}\mathcal{I}(X) - \log(a),\end{aligned}$$

as required.

2.1. Upper bounds for DFS information measure

In this subsection, we first provide some bounds for the convolution of two continuous independent random variables. Then, we examine an upper bound for the absolute value of the difference information measure with order $p > 0$.

Let x and y be two real values. The arithmetic mean and harmonic mean are defined as $m_A(x, y) = \frac{x+y}{2}$ and $m_H(x, y) = \frac{2}{\frac{1}{x}+\frac{1}{y}}$, respectively.

Theorem 2.6. *Let X and Y be two independent random variables. Then, an upper bound for the difference information measure of the convolution random variable X + Y is given by*

$$\mathcal{D}_F^S(X+Y) \leq m_A\left(m_H\Big(\mathcal{I}(X), \mathcal{I}(Y)\Big), m_H\Big(\frac{1}{\mathcal{N}(X)}, \frac{1}{\mathcal{N}(Y)}\Big)\right), \tag{16}$$

where $\mathcal{N}(X)$ is the power entropy associated with random variable X and is defined as $\mathcal{N}(X) = e^{2\mathcal{H}(X)}$. For more details, refer to [7].

Proof: Proof: From the definition of $\mathcal{D}_F^S(X+Y)$ and making use of the following inequalities ([13] and [7]):

$$\frac{1}{\mathcal{I}(X+Y)} \geq \frac{1}{\mathcal{I}(X)} + \frac{1}{\mathcal{I}(Y)}$$

and

$$\mathcal{N}(X+Y) \geq \mathcal{N}(X) + \mathcal{N}(Y),$$

we have

$$\begin{aligned}
\mathcal{D}_F^S(X+Y) &= \mathcal{I}\Big(X+Y\Big) - \mathcal{H}\Big(X+Y\Big) \\
&\leq \frac{1}{\frac{1}{\mathcal{I}(X)} + \frac{1}{\mathcal{I}(Y)}} - \frac{1}{2}\log\left(e^{2\mathcal{H}(X)} + e^{2\mathcal{H}(Y)}\right) \\
&\leq \frac{1}{2} m_H\Big(\mathcal{I}(X), \mathcal{I}(Y)\Big) + \frac{1}{2}\log\left(\frac{1}{\mathcal{N}(X) + \mathcal{N}(Y)}\right) \\
&= \frac{1}{2} m_H\Big(\mathcal{I}(X), \mathcal{I}(Y)\Big) + \frac{1}{2}\log\left(\frac{1}{\mathcal{N}(X) + \mathcal{N}(Y)}\right) \\
&\leq \frac{1}{2} m_H\Big(\mathcal{I}(X), \mathcal{I}(Y)\Big) + \frac{1}{2} m_H\Big(\frac{1}{\mathcal{N}(X)}, \frac{1}{\mathcal{N}(Y)}\Big),
\end{aligned}$$

where the last inequality follows from the inequality $\log(x) \leq x - 1$, valid for $x > 0$.

Let X and Y be two independent random variables. In information theory, it is widely recognized that the following inequalities hold:

$$\mathcal{I}(X+Y) \leq min\Big(\mathcal{I}(X), \mathcal{I}(Y)\Big) \leq \mathcal{I}(X) + \mathcal{I}(Y), \tag{17}$$

$$max\Big(\mathcal{H}(X), \mathcal{H}(Y)\Big) \leq \mathcal{H}(X+Y) \leq \mathcal{H}(X) + \mathcal{H}(Y), \tag{18}$$

where $I(\cdot)$ denotes the Fisher information and $H(\cdot)$ represents the entropy.

These inequalities reveal the predictable impact on information and uncertainty when combining independent random variables through addition. Specifically, the Fisher information tends to decrease, while the entropy tends to increase. Based on these observations, we can obtain an upper bound for $\mathcal{D}_F^S(X+Y)$ as

$$\mathcal{D}_F^S(X+Y) \leq \min\Big(\mathcal{I}(X), \mathcal{I}(Y)\Big) - \max\Big(\mathcal{H}(X), \mathcal{H}(Y)\Big). \tag{19}$$

The inequality indicates that the difference information measure is no greater than the difference between the minimum Fisher information and the maximum entropy of the individual variables.

Theorem 2.7. *Let X and Y be two independent random variables. Then, an upper bound for the difference information measure of the convolution random variable X + Y is given by*

$$\mathcal{D}_F^S(X+Y) \leq \frac{1}{2}\mathcal{D}_F^S(X) + \frac{1}{2}\mathcal{D}_F^S(Y).$$

Proof: From the definition of difference information measure and upon making use relation (19) and the facts that

$$min\Big(\mathcal{I}(X), \mathcal{I}(Y)\Big) = \frac{\mathcal{I}(X)+\mathcal{I}(Y)}{2} - \frac{|\mathcal{I}(X)-\mathcal{I}(Y)|}{2},$$

$$max\Big(\mathcal{H}(X), \mathcal{H}(Y)\Big) = \frac{\mathcal{H}(X)+\mathcal{H}(Y)}{2} + \frac{|\mathcal{H}(X)-\mathcal{H}(Y)|}{2},$$

we get

$$\begin{aligned}
\mathcal{D}_F^S(X+Y) &= \mathcal{I}\Big(X+Y\Big) - \mathcal{H}\Big(X+Y\Big) \\
&= \min\Big(\mathcal{I}(X), \mathcal{I}(Y)\Big) - \max\Big(\mathcal{H}(X), \mathcal{H}(Y)\Big) \\
&\leq \frac{\mathcal{I}(X)+\mathcal{I}(Y)}{2} - \frac{\mathcal{H}(X)+\mathcal{H}(Y)}{2} \\
&= \frac{\mathcal{I}(X)-\mathcal{H}(X)}{2} + \frac{\mathcal{I}(Y)-\mathcal{H}(Y)}{2} \\
&= \frac{1}{2}\mathcal{D}_F^S(X) + \frac{1}{2}\mathcal{D}_F^S(Y),
\end{aligned}$$

as required.

Theorem 2.8. *Let X and Y be two independent random variables. Then, an upper bound for the difference information measure of the mixture random variable $\alpha X + (1-\alpha)Y$, $\alpha \in [0,1]$, we have*

$$\mathcal{D}_F^S(\alpha X + (1-\alpha)Y) \leq \frac{1}{2}\mathcal{D}_F^S(X) + \frac{1}{2}\mathcal{D}_F^S(Y) + \mathcal{K}_\alpha(X,Y),$$

where $\mathcal{K}_\alpha(X,Y) = \frac{1-\alpha^2}{\alpha^2}\mathcal{I}(X) + \frac{1-(1-\alpha)^2}{(1-\alpha)^2}\mathcal{I}(Y) - \log(\alpha(1-\alpha))$.

Proof: From the definition of difference information measure of of the mixture random variable $\alpha X+(1-\alpha)Y$, $\alpha \in [0,1]$ and upon making use Theorem 2.7, and Lemma 2.5, we have

$$\begin{aligned}
\mathcal{D}_F^S(\alpha X + (1-\alpha)Y) &\leq \frac{1}{2}\mathcal{D}_F^S(\alpha X) + \frac{1}{2}\mathcal{D}_F^S((1-\alpha)Y) \\
&= \mathcal{D}_F^S(X) + \frac{1-\alpha^2}{\alpha^2}\mathcal{I}(X) - \log(\alpha) \\
&\quad + \mathcal{D}_F^S(Y) + \frac{1-(1-\alpha)^2}{(1-\alpha)^2}\mathcal{I}(Y) - \log(1-\alpha)
\end{aligned}$$

as required.

Theorem 2.9. *An upper bound for the $|\mathcal{D}_F^S(X)|^p$, for $p > 0$ is given by*

$$|\mathcal{D}_F^S(X)|^p \leq \gamma_p\Big\{\mathcal{I}^p(X) + \frac{\gamma_p}{2^p}\Big(\Big(2\pi e\Big)^p V^p(X) + 1\Big)\Big\}, \tag{20}$$

where $\gamma_p = max(1, 2^{p-1})$ and $\mathcal{I}(X)$ and $V(X)$ are Fisher information and variance of an arbitrary random variable on support $\mathcal{X} = R$.

Proof: From the inequality

$$|a-b|^p \leq \gamma_p\Big(|a|^p+|b|^b\Big), \text{with } \gamma_p = max(1, 2^{p-1}), \quad p>0,$$

we readily have

$$|\mathcal{D}_F^S(X)|^p \;=\; |\mathcal{I}(X)-H(X)|^p \leq \gamma_p\Big\{\mathcal{I}^p(X)+|H(X)|^p\Big\}. \tag{21}$$

From the Lemma 1 of [27], we have an upper bound for any arbitrary random variable X on support R as

$$H(X) \leq \frac{1}{2}\log(2\pi e V(X)). \tag{22}$$

Now, from (21) and (22) and also making use the inequality $\log(x) \leq x-1, x>0$, we find that

$$\begin{aligned}
|\mathcal{D}_F^S(X)|^p &= |\mathcal{I}(X)-H(X)|^p \leq \gamma_p\Big\{\mathcal{I}^p(X)+|H(X)|^p\Big\} \\
&\leq \gamma_p\Big\{\mathcal{I}^p(X)+\Big|\frac{1}{2}\log(2\pi e V(X))\Big|^p\Big\} \\
&\leq \gamma_p\Big\{\mathcal{I}^p(X)+\frac{1}{2^p}\Big|2\pi e V(X)-1\Big|^p\Big\} \\
&\leq \gamma_p\Big\{\mathcal{I}^p(X)+\frac{\gamma_p}{2^p}\Big(\Big(2\pi e\Big)^p V^p(X)+1\Big)\Big\},
\end{aligned}$$

as required.

The special case of Theorem 2.9 when $\alpha = 0$, has the representation

$$|\mathcal{D}_F^S(X)| \leq \mathcal{I}(X)+(\pi e)V(X)+\frac{1}{2}. \tag{23}$$

3. Jensen-difference Fisher–Shannon information measure

In this section, we initially establish and explore several properties of a Jensen-difference Fisher-Shannon information measure. Subsequently, we introduce two extensions of this informational measure and provide some results in this regard.

Definition 3.1. *Let f_0 and f_1 be two density functions on support $\mathcal{X}$. Then, the Jensen-difference Fisher–Shannon information measure, $\mathcal{JD}(f_0, f_1)$, is defined as*

$$\mathcal{JD}(f_0, f_1) = \frac{1}{2}\mathcal{D}_F^S(f_0)+\frac{1}{2}\mathcal{D}_F^S(f_1)-\mathcal{D}_F^S\Big(\frac{f_0+f_1}{2}\Big). \tag{24}$$

Lemma 3.2. *The $\mathcal{JD}$ divergence measure in (24) is non-negative.*

Proof: From the definition of difference Fisher–Shannon information measure in (10) and making use the convexity properties of Fisher information with respect to density f, we have that

$$\mathcal{I}\Big(\frac{f_0+f_1}{2}\Big) \leq \frac{1}{2}\mathcal{I}(f_0)+\frac{1}{2}\mathcal{I}(f_1)$$

and concavity property of Shannon entropy with respect to density f

$$\frac{1}{2}\mathcal{H}(f_0)+\frac{1}{2}\mathcal{H}(f_1) \leq \mathcal{H}\Big(\frac{f_0+f_1}{2}\Big),$$

we have

$$\mathcal{I}\Big(\frac{f_0+f_1}{2}\Big)-\mathcal{H}\Big(\frac{f_0+f_1}{2}\Big)\leq\frac{1}{2}\mathcal{I}(f_0)+\frac{1}{2}\mathcal{I}(f_1)-\frac{1}{2}\mathcal{H}(f_0)-\frac{1}{2}\mathcal{H}(f_1).$$

From this we find that

$$\begin{aligned}\mathcal{JD}(f_0,f_1) &= \frac{1}{2}\mathcal{I}(f_0)+\frac{1}{2}\mathcal{I}(f_1)-\mathcal{I}\Big(\frac{f_0+f_1}{2}\Big)\\ &\quad+\mathcal{H}\Big(\frac{f_0+f_1}{2}\Big)-\frac{1}{2}\mathcal{H}(f_0)-\frac{1}{2}\mathcal{H}(f_1)\\ &= \mathcal{JF}(f_0,f_1)+\mathcal{JS}(f_0,f_1) \qquad (25)\\ &\geq 0,\end{aligned}$$

as required, where $\mathcal{JF}(f,g)$ is the Jensen–Fisher information divergence [21] between two density functions f and g defined as

$$\mathcal{JF}(f,g)=\frac{1}{2}D\Big(f,\frac{f+g}{2}\Big)+D\Big(g,\frac{f+g}{2}\Big), \tag{26}$$

$D(f,g)$ is relative Fisher information divergence [9, 30] between two density functions f and g defined as

$$D(f,g)=\frac{1}{2}\int_{\mathcal{X}}\Big(\frac{\partial}{\partial x}\log f(x)-\frac{\partial}{\partial x}\log g(x)\Big)^2 f(x)dx,$$

$\mathcal{JS}(f,g)$ is the Jensen–Shannon divergence [8, 22] between two density functions f and g which can be defined in terms of Shannon entropies or Kullback–Leibler divergences [13] as

$$\begin{aligned}\mathcal{JS}(f,g) &= H\Big(\frac{f+g}{2}\Big)-\frac{1}{2}[H(f)+H(g)]\\ &= KL\Big(f,\frac{f+g}{2}\Big)+KL\Big(g,\frac{f+g}{2}\Big),\end{aligned}$$

and $KL(f,g)$ is the Kullback–Leibler divergence between two density functions f and g defined as

$$KL(f,g)=\int_{\mathcal{X}}f(x)\log\Big(\frac{f(x)}{g(x)}\Big)dx.$$

Corollary 3.3. *Let f_0 and f_1 be two density functions on support $\mathcal{X}$. Then, from the proof of Lemma 3.2, we have*

$$\mathcal{JD}(f_0,f_1)=\frac{\mathcal{SD}(f_0,f_1)+\mathcal{SKL}(f_0,f_1)}{2},$$

where

$$\mathcal{SD}(f_0,f_1)=D\Big(f_0,\frac{f_0+f_1}{2}\Big)+D\Big(f_1,\frac{f_0+f_1}{2}\Big),$$

and

$$\mathcal{SKL}(f_0,f_1)=KL\Big(f_0,\frac{f_0+f_1}{2}\Big)+KL\Big(f_1,\frac{f_0+f_1}{2}\Big).$$

3.1. Extension to k random variables

We now extend the definition of Jensen-difference divergence measure in (24) to the case of k random variables. Let $X_1,\ldots,X_k$ be random variables with density functions $f_1,\ldots,f_k$, respectively, and $\boldsymbol{p}=(p_1,\ldots,p_k)$ be a k-dimensional vector with non-negative real numbers such that $\sum_{i=1}^k p_i=1$. Then, the Jensen-difference measure is defined as

$$\mathcal{JD}\big(f_1,\ldots,f_k;\boldsymbol{p}\big)=\sum_{i=1}^k p_i\mathcal{D}_F^S(f_i)-\mathcal{D}_F^S\Big(\sum_{i=1}^k p_if_i\Big). \tag{27}$$

Lemma 3.4. *The $\mathcal{JD}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big)$ information measure can be represented based on Fisher information distance and Kullback–Leibler divergence measure as*

$$\mathcal{JD}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big) = \sum_{i=1}^{k} p_i\Big(D(f_i,f_T)+KL(f_i,f_T)\Big),$$

where $f_T=\sum_{i=1}^{n} p_i f_i$ is mixture density.

Proof: From the definition of the Jensen-difference measure in (27) and upon making use the proof of Lemma 3.2, we have

$$\begin{aligned}\mathcal{JD}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big) &= \sum_{i=1}^{k} p_i\mathcal{D}_F^S(f_i)-\mathcal{D}_F^S\Big(\sum_{i=1}^{k} p_i f_i\Big)\\ &= \mathcal{JF}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big)+\mathcal{JS}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big)\\ &= \sum_{i=1}^{k} p_i\Big(D(f_i,f_T)+KL(f_i,f_T)\Big)\end{aligned}$$

where $\mathcal{JF}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big)$ and $\mathcal{JS}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big)$ are Jensen-Fisher and Jensen-Shannon information measures based on k density functions, defined as

$$\mathcal{JF}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big)=\sum_{i=1}^{k} p_i\mathcal{I}(f_i)-\mathcal{I}\Big(\sum_{i=1}^{k} p_i f_i\Big)$$

and

$$\mathcal{JS}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big)=\mathcal{H}\Big(\sum_{i=1}^{k} p_i f_i\Big)-\sum_{i=1}^{k} p_i\mathcal{H}_i(f_i). \tag{28}$$

Theorem 3.5. *An upper bound for the $\mathcal{JD}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big)$ in (27) is given by*

$$\mathcal{JD}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big)\le\sum_{i=1}^{k}\sum_{j=1}^{k} p_ip_j\Big\{\mathcal{I}(f_i)+\chi^2(f_i,f_j)\Big\}, \tag{29}$$

where $\chi^2(f_i,f_j)=\int\frac{(f_i(x)-f_j(x))^2}{f_i(x)}dx$ is known as chi-square between two density functions f_i and f_j; for more details, one may refer to [18].

Proof: From the definitions of $\mathcal{JD}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big)$ (27) and $\mathcal{JS}\Big(f_1,\dots,f_k;\boldsymbol{p}\Big)$ in (28), and using Lemma 3.4, we have

$$\begin{aligned}\mathcal{JD}\Big(f_1,\ldots,f_k;\boldsymbol{p}\Big) &= \sum_{i=1}^{k} p_i\mathcal{D}_F^S(f_i)-\mathcal{D}_F^S\Big(\sum_{i=1}^{k} p_i f_i\Big)\\ &= \sum_{i=1}^{k} p_i\mathcal{I}_i(f_i)-\mathcal{I}\Big(\sum_{i=1}^{k} p_i f_i\Big)+\sum_{i=1}^{k} p_i KL(f_i,f_T)\\ &\le \sum_{i=1}^{k} p_i\mathcal{I}_i(f_i)+\sum_{i=1}^{k}\sum_{j=1}^{k} p_i p_j KL(f_i,f_j)\\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} p_i p_j\Big\{\mathcal{I}(f_i)+KL(f_i,f_j)\Big\}\\ &\le \sum_{i=1}^{k}\sum_{j=1}^{k} p_i p_j\Big\{\mathcal{I}(f_i)+\chi^2(f_i,f_j)\Big\},\end{aligned}$$

where the first inequality follows from the fact that $KL(f_i, f_T) \le \sum_{j=1}^{k} p_j KL(f_i, f_j)$, and the last inequality follows from the fact that $KL(f_i, f_j) \le \chi^2(f_i, f_j)$.

3.2. (p, w)-Jensen difference information measure

In this subsection, we first review the definition of (p, w)-Jensen–Shannon divergence measure. Then, we introduce (p, w)-Jensen–Fisher divergence measure and (p, w)-Jensen–difference divergence measure in a way similar to (p, w)-Jensen–Shannon divergence. Furthermore, we establish some results for this extended divergence measure.

Let f and g be two density functions. The (p, w)-Jensen–Shannon divergence between two density functions f_1 and f_2, for α and $p \in (0, 1)$, is defined as

$$\begin{aligned}JS\Big(f_1,f_2;\boldsymbol{p},\boldsymbol{w}\Big) &= H\Big((1-\bar{s})f_1+\bar{s}f_2\Big)-wH\Big((1-p)f_1+pf_2\Big)\\ &\quad -(1-w)H\Big(pf_1+(1-p)f_2\Big)\\ &= wKL\Big((1-p)f_1+pf_2:(1-\bar{s})f_1+\bar{s}f_2\Big)\\ &\quad +(1-w)KL\Big(pf_1+(1-p)f_2:(1-\bar{s})f_1+\bar{s}f_2\Big),\end{aligned} \tag{30}$$

where $\bar{s} = wp + (1 - w)(1 - p)$. For more details, one may refer to [26] and [25].

3.3. (p, w)-Jensen–Fisher information

Definition 3.6. *Let X_1 and X_2 be two random variables with density functions f_1 and f_2, respectively, Then, the (p, w)-Jensen–Fisher information measure, for $w, p \in (0, 1)$, is defined as*

$$\begin{aligned}\mathcal{JF}\Big(f_1,f_2;\boldsymbol{p},\boldsymbol{w}\Big) &= w\mathcal{I}\Big((1-p)f_1+pf_2\Big)-(1-w)\mathcal{I}\Big(pf_1+(1-p)f_2\Big)\\ &\quad -\mathcal{I}\Big((1-\bar{s})f_1+\bar{s}f_2\Big),\end{aligned} \tag{31}$$

where $\bar{s} = wp + (1 - w)(1 - p)$.

Lemma 3.7. *The (p, w)-Jensen–Fisher information measure in (31) can be represented as*

$$\begin{aligned}\mathcal{JF}\Big(f_1,f_2;\boldsymbol{p},\boldsymbol{w}\Big) &= wD\Big((1-p)f_1+pf_2:(1-\bar{s})f_1+\bar{s}f_2\Big)\\ &\quad +(1-w)D\Big(pf_1+(1-p)f_2:(1-\bar{s})f_1+\bar{s}f_2\Big),\end{aligned}$$

where $\bar{s} = wp + (1-w)(1-p)$.

Proof: Let us consider $h_1(x) = (1-p)f_1(x) + pf_2(x)$, $h_2(x) = pf_1(x) + (1-p)f_2(x)$, and

$$h_T(x) = wh_1(x) + (1-w)h_2(x) = (1-\bar{s})f_1(x) + \bar{s}f_2(x),$$

then by letting $C = wD\Big(h_1 : h_T(x)\Big) + (1-w)D\Big(h_2 : h_T\Big)$, we have

$$\begin{aligned}
C &= w\int\left(\frac{h_1'(x)}{h_1(x)} - \frac{h_T'(x)}{h_T(x)}\right)^2 h_1(x)dx + (1-w)\int\left(\frac{h_2'(x)}{h_2(x)} - \frac{h_T'(x)}{h_T(x)}\right)^2 h_2(x)dx \\
&= w\int\left(\frac{h_1'(x)}{h_1(x)}\right)^2 h_1(x)dx + (1-w)\int\left(\frac{h_2'(x)}{h_2(x)}\right)^2 h_2(x)dx + \int\left(\frac{h_T'(x)}{h_T(x)}\right)^2 h_T(x)dx \\
&\quad -2\int\frac{wh_1'(x)h_T'(x)}{h_T(x)}dx - 2\int\frac{(1-w)h_2'(x)h_T'(x)}{h_T(x)}dx \\
&= w\int\left(\frac{h_1'(x)}{h_1(x)}\right)^2 h_1(x)dx + (1-w)\int\left(\frac{h_2'(x)}{h_2(x)}\right)^2 h_2(x)dx + \int\left(\frac{h_T'(x)}{h_T(x)}\right)^2 h_T(x)dx \\
&\quad -2\int\left(\frac{h_T'(x)}{h_T(x)}\right)^2 h_T(x)dx \\
&= w\int\left(\frac{h_1'(x)}{h_1(x)}\right)^2 h_1(x)dx + (1-w)\int\left(\frac{h_2'(x)}{h_2(x)}\right)^2 h_2(x)dx - \int\left(\frac{h_T'(x)}{h_T(x)}\right)^2 h_T(x)dx \\
&= w\mathcal{I}\Big(h_1\Big) + (1-w)\mathcal{I}\Big(h_2\Big) - \mathcal{I}\Big(h_T\Big) \\
&= \mathcal{JF}\Big(f_1, f_2; \boldsymbol{p}, \boldsymbol{w}\Big),
\end{aligned}$$

as required.

3.4. (p, w)*–Jensen difference Fisher–Shannon information measure*

Definition 3.8. *Let* X_1, X_2 *and* Y *be random variables with density functions* f_1, f_2 *and* ψ*, respectively. Then, the* (p, w)*-Jensen difference information measure, for* $w, p \in (0, 1)$*, is defined as*

$$\begin{aligned}
\mathcal{JD}\Big(f_1, f_2; \boldsymbol{p}, \boldsymbol{w}\Big) &= w\mathcal{D}_F^S\Big((1-p)f_1 + pf_2\Big) + (1-w)\mathcal{D}_F^S\Big(pf_1 + (1-p)f_2\Big) \\
&\quad -\mathcal{D}_F^S\Big((1-\bar{s})f_1 + \bar{s}f_2\Big),
\end{aligned} \tag{32}$$

where $\bar{s} = wp + (1-w)(1-p)$.

Theorem 3.9. *The* (p, w)*-Jensen difference information measure can be expressed as*

$$\mathcal{JD}\Big(f_1, f_2; \boldsymbol{p}, \boldsymbol{w}\Big) = JS\Big(f_1, f_2; \boldsymbol{p}, \boldsymbol{w}\Big) + \mathcal{JF}\Big(f_1, f_2; \boldsymbol{p}, \boldsymbol{w}\Big). \tag{33}$$

Proof: From the definitions of the (p, w)-Jensen difference information measure in (32) and (p, w)-Jensen-Shannon information measure in (30), and (p, w)-Jensen-Fisher information measure in (31) we have

$$\begin{aligned}
\mathcal{JD}\Big(f_1, f_2; \boldsymbol{p}, \boldsymbol{w}\Big) &= w\mathcal{D}_F^S\Big((1-p)f_1 + pf_2\Big) + (1-w)\mathcal{D}_F^S\Big(pf_1 + (1-p)f_2\Big) \\
&\quad -\mathcal{D}_F^S\Big((1-\bar{s})f_1 + \bar{s}f_2\Big) \\
&= w\mathcal{I}\Big((1-p)f_1 + pf_2\Big) - (1-w)\mathcal{I}\Big(pf_1 + (1-p)f_2\Big) \\
&\quad -\mathcal{I}\Big((1-\bar{s})f_1 + \bar{s}f_2\Big) - H\Big((1-\bar{s})f_1 + \bar{s}f_2\Big) \\
&\quad +wH\Big((1-p)f_1 + pf_2\Big) + (1-w)H\Big(pf_1 + (1-p)f_2\Big) \\
&= JS\Big(f_1, f_2; \boldsymbol{p}, \boldsymbol{w}\Big) + \mathcal{JF}\Big(f_1, f_2; \boldsymbol{p}, \boldsymbol{w}\Big),
\end{aligned}$$

as required.

4. Application

We say that Z is standardized skew-normal distributed [2, 8–10] with shape parameter η and PDF given by:

$$f(z) = 2\phi(z)\Phi(\eta z), \quad z \in \mathbb{R}. \tag{34}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$, are respectively the PDF and cumulative distribution function of the standardized gaussian distribution. If Z is standardized skew-normal distributed, then is denoted as $Z \sim SN(\eta)$.

Proposition 4.1. *Let $Z \sim SN(\theta)$ with PDF defined in (34), then*

i) [9] The Shannon entropy of Z is

$$\mathcal{H}(f) = \mathcal{H}(f_0) - E[\log\{2\Phi(\eta Z)\}], \tag{35}$$

where f_0 is the standardized gaussian density function with Shannon entropy $\mathcal{H}(f_0) = (1/2)\log(2\pi e)$.

ii) [10] A suitable approximation of Fisher information of Z is

$$\mathcal{I}(f) \approx 1 + \frac{(b\eta)^2}{\sqrt{1 + 2b^4\eta^2}}, \tag{36}$$

with $b = \sqrt{2/\pi}$.

In the part (i) of Proposition 4.1, the expected value of (35) is an integral that must to be evaluated numerically using, for example, the `QUADPACK` routine. The Shannon entropy and approximated Fisher information given in Proposition 4.1 could be replaced directly in (10) to obtain the mixture information measure. In the left panel of Figure 1 is plotted this measure for several values of α and η. Note that the information increases when $|\eta| \gg 0$ and $\alpha \to 2$, and tends to 1 when $|\eta| \to 0$ and $\alpha \to 0$. Minimum information is obtained at $|\eta| = 0$ and $\alpha = 2$. The right panel of Figure 1 illustrates the particular case of $\alpha = 2$ for several values of η, to clarify the latter observations.

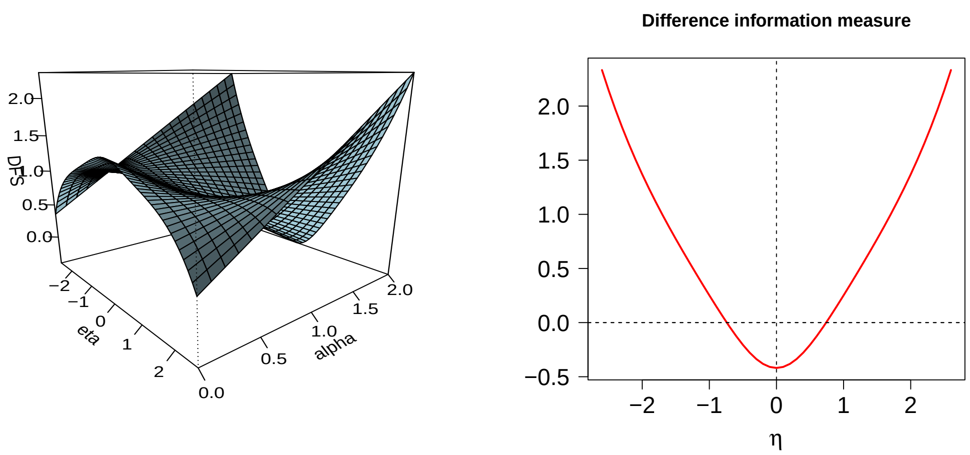

Figure 1: Left: Mixture information measure of a standarized skew-normal random variable with skewness parameter η and $0 \le \alpha \le 2$. Right: Case $\alpha = 2$ (difference information measure) for the same random variable.

Proposition 4.2. *Let $X_1 \sim SN(\eta_1)$ and $X_2 \sim SN(\eta_2)$, be two skew-gaussian random variables with respective PDFs f_1 and f_2 defined as in (34). Then, the Jensen-difference Fisher–Shannon information measure between X_1 and X_2 given in (25) is the sum of the following two divergences:*

i) [8] The Jensen–Shannon divergence between X and Y is

$$\mathcal{JS}(f_1,f_2) = \frac{1}{2}E\left[\log\left\{\frac{2\Phi(\eta_1 X_1)}{\Phi(\eta_1 X_1)+\Phi(\eta_2 X_1)}\right\}\right] + \frac{1}{2}E\left[\log\left\{\frac{2\Phi(\eta_1 X_2)}{\Phi(\eta_1 X_2)+\Phi(\eta_2 X_2)}\right\}\right]. \quad (37)$$

ii) [9, 10] A suitable approximation of Jensen–Fisher divergence between X_1 and X_2 is

$$\mathcal{JF}(f_1,f_2) \approx \frac{(b\eta_1)^2}{\sqrt{1+2b^4\eta_1^2}} + \frac{(b\eta_2)^2}{\sqrt{1+2b^4\eta_2^2}} - \frac{2\eta_1\eta_2 b^2}{\sqrt{1+\eta_1^2+2b^4\eta_2^2-\eta_2^2}}, \quad (38)$$

under condition $1+\eta_1^2 > \eta_2^2(1-2b^4)$.

As in the Shannon entropy case, the expected values of (37) must to be evaluated numerically using the `QUADPACK` routine. In Figure 2 is illustrated the Jensen-difference Fisher–Shannon information measure for several values of η_i, $i = 1,2$. It is observed in Figure 2 that $\mathcal{JD}(f_1, f_2; p_1, p_2) = 0$ when $\eta_1 = \eta_2$. In addition, the distance increases when skewness parameters increase such as in Jensen–Shannon divergence [8] and Jensen–Fisher divergence [9] cases. However, it can be observed that information is mainly affected by the Jensen–Fisher divergence where, for $\eta_1 \neq \eta_2$, we observe that measure has the highest values.

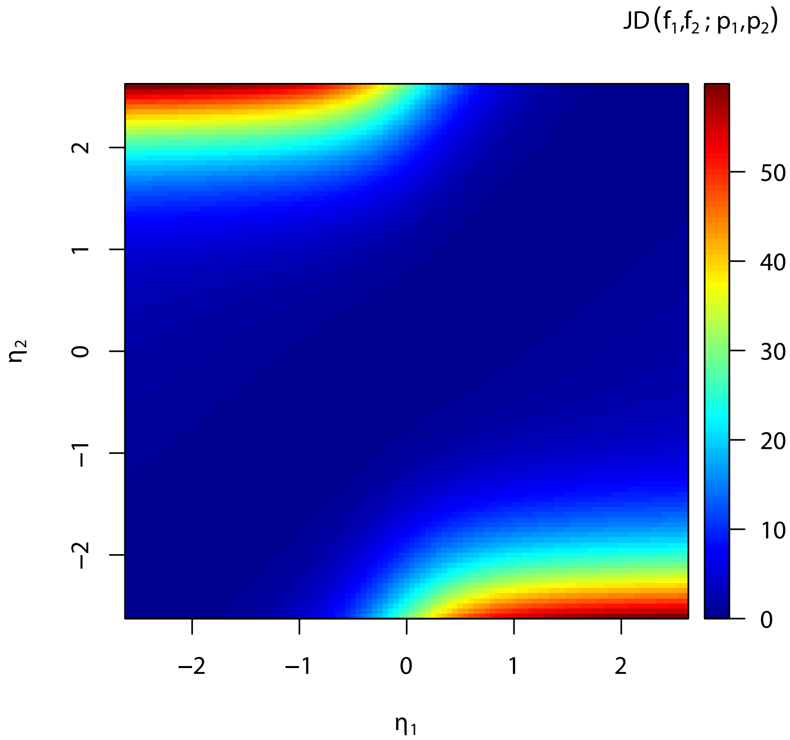

Figure 2: Jensen-difference Fisher–Shannon information measure for two independent skew-normal random variables with skewness parameters η_1 and η_2.

4.1. *Maximum likelihood estimator*

For this application, is considered the maximum likelihood estimator (MLE) of shape parameter η_i computed in [8] with the sn library of R software's. Let $z_1, \ldots, z_n$ be a sample of size n from a skew-normal random variable, the MLE $\widehat{\eta_i}$ is obtained numerically by maximizing the log-likelihood function:

$$\ell(\eta_i) = -\frac{1}{2}\sum_{t=1}^{n} z_t^2 + \sum_{t=1}^{n} \log\, \Phi\{\eta_i z_t\},$$

i.e., $\widehat{\eta_i} = \arg\max_{\eta_i}\{\ell(\eta_i)\}$. Moreover, MLE $\widehat{\eta_i}$ could be replaced in (35)–(38) as a plug-in estimator to obtain the required measures.

4.2. *Fish condition factor time series*

In this section, we revisited the time series analyzed in [8, 9, 11] about fatness condition factor (CF) index of anchovies from the perspective of mixture information measure and related divergence. The CF index is a biological indicator for the morphometric relationships of the weight and length. Therefore we consider skew-normal-distributed CF monthly time series (from 01/1990 to 12/2010), denoted as $Z \sim SN(\eta_i)$, for lengths $12, \ldots, 18$ (cm) and genre groups (males and females) [see Figure 5 of 8]. Total number of lengths are 7 and the sample size of each time series is $n = 264$. Table 1 shows the maximum likelihood estimations of skewness parameters based on the skew-normal density fits and related to each time series and group.

Table 1: Shape parameter estimates of standardized skew-normal densities with its respective standard errors in parenthesis, for each length class and group (males or females). Source: [8].

Length	Males	Females
12	-1.065 (0.147)	0.552 (0.196)
13	1.778 (0.134)	2.267 (0.128)
14	-1.086 (0.166)	1.242 (0.157)
15	-0.442 (0.150)	0.728 (0.174)
16	1.616 (0.149)	1.702 (0.155)
17	-3.092 (0.080)	-0.689 (0.149)
18	1.368 (0.201)	-1.581 (0.136)

Mixture information measure is considered first to compares between length classes of CF time series for several orders. Figure 3 shows the mixture information measure values for $\alpha \in [0, 2]$, where information for males are quite similar to than one of females. However, some interesting difference between groups can be highlighted depending on length class. For example, for males, only length classes 13 and 17 have positive trends, while for females, only length class 13 have positive trend.

Given that mixture information measure has the advantage of detect predominance of Fisher information or Shannon entropy for CF time series, results obtained from this measures allow new light on the interpretation of the data. In particular, for latter length classes, Fisher information is always predominant over Shannon entropy. This means that increments the whole information in CF data are produced by the presence of asymmetry, which is modelled by skewness parameter of SN model. In addition, uncertainty considered by Shannon entropy not affect enough the whole information. This predominance of Fisher information over Shannon entropy was not detected by traditional tools, such as Fisher–Shannon information plane [9, 12], which consider the product of both measures.

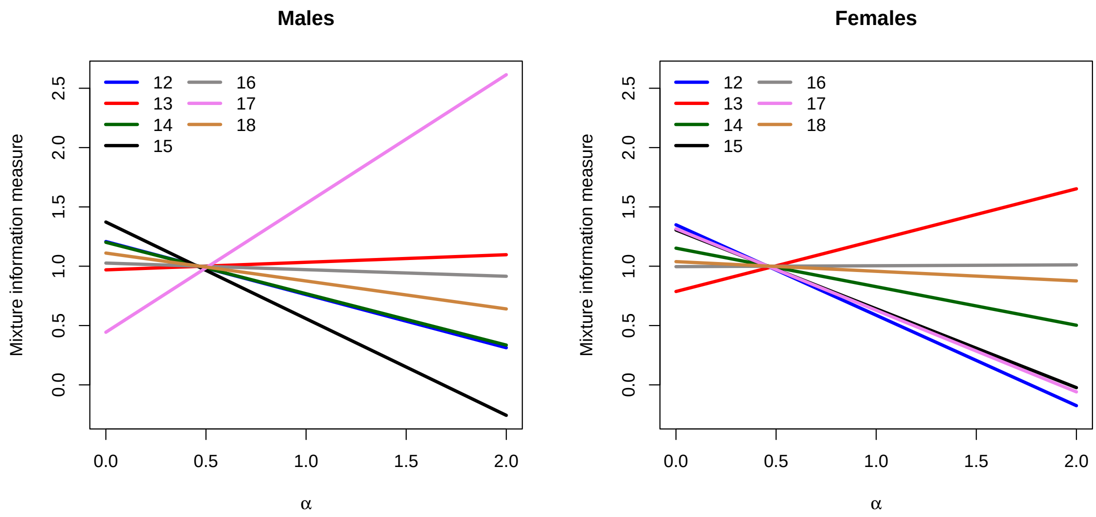

Figure 3: Mixture information measure for both genres and all lengths $(12, 13, \dots, 18)$ based on standardized skew-normal random variables with estimated skewness parameters of Table 1 and order $\alpha \in [0, 2]$.

Jensen-difference Fisher–Shannon information measure between length classes of males are given by

$$\begin{bmatrix} 0.000 & 15.271 & 0.000 & 0.214 & 11.446 & 7.144 & 7.178 \\ 4.808 & 0.000 & 4.991 & 1.806 & 0.015 & 107.575 & 0.086 \\ 0.000 & 15.341 & 0.000 & 0.226 & 11.503 & 6.835 & 7.221 \\ 0.424 & 11.843 & 0.459 & 0.000 & 8.744 & 25.676 & 5.297 \\ 4.634 & 0.017 & 4.815 & 1.664 & 0.000 & 106.648 & 0.035 \\ 1.086 & 18.391 & 1.068 & 1.663 & 14.226 & 0.000 & 9.566 \\ 4.342 & 0.131 & 4.520 & 1.439 & 0.045 & 104.692 & 0.000 \end{bmatrix}$$

and for females are given by

$$\begin{bmatrix} 0.000 & 6.211 & 0.544 & 0.026 & 2.014 & 1.195 & 8.789 \\ 0.949 & 0.000 & 0.401 & 0.796 & 0.137 & 2.987 & 12.658 \\ 0.248 & 1.102 & 0.000 & 0.147 & 0.176 & 2.030 & 11.159 \\ 0.021 & 4.154 & 0.265 & 0.000 & 1.234 & 1.429 & 9.614 \\ 0.546 & 0.229 & 0.108 & 0.413 & 0.000 & 2.486 & 11.962 \\ 1.001 & 30.364 & 4.779 & 1.510 & 11.841 & 0.000 & 1.001 \\ 1.913 & 35.869 & 6.371 & 2.529 & 14.544 & 0.364 & 0.000 \end{bmatrix}$$

From the latter measures, specifically for males, the length class $L = 13$ and 17 cm produces the higher values of Jensen-difference Fisher–Shannon information measure with the pairs of length classes $(13, 14)$, $(13, 17)$, $(16, 17)$ and $(17, 18)$; since the high value of estimated skewness parameter for $L = 13$ and 17 cm (Table 1). For females, Jensen-difference Fisher–Shannon information measure produces clear discrepancies for the length classes $L = 13$ and 18 cm with respect to other classes, such as those detected by Jensen–Shannon divergence [8]. Specifically, the Jensen-difference Fisher–Shannon information measure highlights the discrepancy of the pairs of length classes $(13, 17)$, $(16, 18)$ and $(13, 18)$; since the high value of estimated skewness parameter for $L = 13$ and 18 cm (Table 1). Similar results for both groups were detected by Fisher–Shannon information plane based on skew-normal densities [9] and belief Fisher–Shannon information plane [12]. These discrepancies were also detected for these length classes with respect to other ones by the Jensen–Shannon divergence, the Kullback–Leibler divergence-based hypothesis tests [8, 11], and the Jensen-variance distance [19].

5. Conclusions

In this paper, we have introduced and investigated various information measures, focusing primarily on the Fisher-Shannon framework. Beginning with the exploration of foundational concepts such as Fisher and Shannon information measures, we extended our inquiry to propose a novel mixture Fisher-Shannon information measure. Motivated by de Bruijn's identity, we established its properties and examined a specific case known as the difference information measure. Moving forward, we delved into the Jensen-difference Fisher-Shannon information measure, elucidating its fundamental properties and introducing two extensions to broaden its applicability. By rigorously exploring these measures, we provided valuable insights into their theoretical foundations and practical implications. Furthermore, our analysis extended beyond theoretical considerations to practical applications, as demonstrated in the examination of time series data related to the fatness condition factor (CF) index of anchovies. Through the application of mixture information measures and associated divergences, we revisited previous analyses, shedding new light on the interpretation of the data.

It will also be of great interest to study cumulative versions of these measures, and we plan to do this in our future work. Additionally, there is potential to extend the idea based on generalized Fisher information and the generalized extensions of Shannon entropy (Rényi and Tsallis) measures [23]. We are currently working on these problems and hope to report our findings in a future paper.

Acknowledgements

The authors thank the editor and an anonymous referee for their helpful comments and suggestions.

References

[1] Asadi, M., Ebrahimi, N., Kharazmi, O., Soofi, E.S. (2018). Mixture models, Bayes Fisher information, and divergence measures. IEEE Transactions on Information Theory 65, 2316–2321.
[2] Azzalini, A. (2013). The Skew-normal and Related Families. Vol. 3, Cambridge University Press, Cambridge, UK.
[3] Balakrishnan, N., Stepanov, A. (2006). On the Fisher information in record data. Statistics & Probability Letters 76, 537–545.
[4] Bercher, J. F. (2009). Source coding with escort distributions and Rényi entropy bounds. Physics Letters A 373, 3235–3238.
[5] Bercher, J.F. (2013). Some properties of generalized Fisher information in the context of nonextensive thermostatistics. Physica A 392, 3140–3154.
[6] Bobkov, S.G. (2019). Moments of the scores. IEEE Transactions on Information Theory 65, 5294–5301.
[7] Bukal, M. (2022). The concavity of generalized entropy powers. IEEE Transactions on Information Theory, 68(11), 7054-7059.
[8] Contreras-Reyes, J.E. (2016). Analyzing fish condition factor index through skew-gaussian information theory quantifiers. Fluctuation and Noise Letters 15, 1650013.
[9] Contreras-Reyes, J.E. (2021). Fisher information and uncertainty principle for skew-gaussian random variables. Fluctuation and Noise Letters 20, 2150039.
[10] Contreras-Reyes, J.E. (2022). Information-Theoretic Aspects of Location Parameter Estimation under Skew-Normal Settings. Entropy 24, 399.
[11] Contreras-Reyes, J. E. (2023). Information quantity evaluation of nonlinear time series processes and applications. Physica D 445, 133620.
[12] Contreras-Reyes, J.E., Kharazmi, O. (2023). Belief Fisher–Shannon information plane: Properties, extensions, and applications to time series analysis. Chaos, Solitons & Fractals 177, 114271.
[13] Cover, T.M., Thomas, J.A. (2006). Elements of Information Theory. Wiley & Son, Inc., New York, USA.
[14] Fisher, R.A. (1929). Tests of significance in harmonic analysis. Proceedings of the Royal Society of London: Series A 125, 54–59.
[15] Kharazmi, O., Asadi, M. (2018). On the time-dependent Fisher information of a density function. Brazilian Journal of Probability and Statistics 32, 795–814.
[16] Kharazmi, O., Balakrishnan, N. (2021). Cumulative residual and relative cumulative residual Fisher information and their properties. IEEE Transactions on Information Theory 67, 6306–6312.
[17] Kharazmi, O., Balakrishnan, N. (2022). Generating function for generalized Fisher information measure and its application to finite mixture models. Hacettepe Journal of Mathematics and Statistics 51, 1472–1483.
[18] Kharazmi, O., & Balakrishnan, N. (2023). On Jensen-χ^2 divergence measure. Probability in the Engineering and Informational Sciences, 1-25.

[19] Kharazmi, O., Contreras-Reyes, J.E., Basirpour, M.B. (2024). Jensen-variance distance measure: A unified framework for statistical and information measures. Computational and Applied Mathematics 43, 144.
[20] Kharazmi, O., Contreras-Reyes, J.E., Balakrishnan, N. (2023a). Jensen–Fisher information and Jensen–Shannon entropy measures based on complementary discrete distributions with an application to Conway's game of life. Physica D 453, 133822.
[21] Kharazmi, O., Jamali, H., Contreras-Reyes, J.E. (2023b). Fisher information and its extensions based on infinite mixture density functions. Physica A 624, 128959.
[22] Lin, J. (1991). Divergence measures based on the Shannon entropy. IEEE Transactions on Information Theory 37, 145–151.
[23] Masi, M. (2005). A step beyond Tsallis and Rényi entropies. Physics Letters A 338, 217–224.
[24] Madiman, M., Barron, A. (2007). Generalized entropy power inequalities and monotonicity properties of information. IEEE Transactions on Information Theory 53, 2317-2329.
[25] Melbourne, J., Talukdar, S., Bhaban, S., Madiman, M., Salapaka, M.V. (2022). The differential entropy of mixtures: New bounds and applications. IEEE Transactions on Information Theory 68, 2123–2146.
[26] Nielsen, F. (2020). On a generalization of the Jensen–Shannon divergence and the Jensen–Shannon centroid. Entropy 22, 221.
[27] Nielsen, F., Nock, R. (2017). MaxEnt upper bounds for the differential entropy of univariate continuous distributions. IEEE Signal Processing Letters 24, 402–406.
[28] Shannon, C.E. (1948). A mathematical theory of communication. The Bell System Technical Journal 27, 379–423.
[29] Van Erven, T., Harremos, P. (2014). Rényi divergence and Kullback-Leibler divergence. IEEE Transactions on Information Theory 60, 3797–3820.
[30] Yáñez, R.J., Sánchez-Moreno, P., Zarzo, A., Dehesa, J.S. (2008). Fisher information of special functions and second-order differential equations. Journal of Mathematical Physics 49, 082104.
[31] Zegers, P. (2015). Fisher information properties. Entropy 17, 4918–4939.