Probability in the Engineering and Informational Sciences (2024), **38**:2 403–427
doi:10.1017/S0269964823000189

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

On Jensen-χ_α^2 divergence measure

Omid Kharazmi[1] and Narayanaswamy Balakrishnan[2]

[1]Department of Statistics, Faculty of Mathematical Sciences, Vali-e-Asr University of Rafsanjan, Rafsanjan, Kerman, Iran
[2]Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada
Corresponding author: Narayanaswamy Balakrishnan; Email: bala@mcmaster.ca

Keywords: q-Fisher information; chi-square divergence; generalized chi-square divergence; Jensen-chi-square divergence; Jensen–Shannon divergence measure

Abstract
The purpose of this paper is twofold. The first part is to introduce relative-χ_α^2, Jensen-χ_α^2 and (p, w)-Jensen-χ_α^2 divergence measures and then examine their properties. In addition, we also explore possible connections between these divergence measures and Jensen–Shannon entropy measure. In the second part, we introduce (p, η)-mixture model and then show it to be an optimal solution to three different optimization problems based on χ_α^2 divergence measure. We further study the relative-χ_α^2 divergence measure for escort and arithmetic mixture densities. We also provide some results associated with relative-χ_α^2 divergence measure of mixed reliability systems. Finally, to demonstrate the usefulness of the Jensen-χ_α^2 divergence measure, we apply it to a real example in image processing and present some numerical results. Our findings in this regard show that the Jensen-χ_α^2 is an effective criteria for quantifying the similarity between two images.

1. Introduction

Information theory is one of the most important branches of science and engineering and has attracted significant attention of numerous researchers over the past seven decades. In information theory, several information-theoretic divergence measures between two probabilistic models have been introduced and then used in many fields including information theory, statistics, engineering and physics. Among the most important information, divergence measures are the Kullback–Leibler and chi-square divergence measures. These two information quantities have found many key applications in information theory, economics, statistics, physics and electrical engineering. In the literature, some extensions of Kullback–Leibler and chi-square divergence measures have appeared during the last three decades. For pertinent details, one may refer to [5, 7, 11, 18].

The chi-square divergence has several extensions, such as the symmetric chi-square, triangular divergence, generalized chi-square and Balakrishnan and Sanghvi divergence measures. Each of these measures has its own properties and applications in different fields.

In this work, we first consider chi-square (χ^2) and generalized chi-square (χ_α^2) divergence measures and then propose relative-χ_α^2 and two Jensen versions of χ_α^2 (Jensen-χ_α^2 and (p, w)-Jensen-χ_α^2) divergence measure. We further examine a possible connection between the proposed information measures and also discuss some potential applications of them.

The proposed relative-χ_α^2, $D_\alpha^\psi(f : g)$ divergence, provides a measure of the difference between two probability distributions, f and g, that is weighted by the density function $\psi(x)$. The weight density function $\psi(x)$ allows the divergence to be tailored to specific features and characteristics of the data, for the two models that are being compared.

The parameter α controls the sensitivity of the divergence to differences between f and g. For example, when $\alpha = 1$, the divergence reduces to the L_2 distance, which measures the difference between f and g in terms of their squared deviations. When $\alpha = 0$, the divergence measure reduces to half of the chi-square divergence measure. The weight density function $\psi(x)$ can be chosen to emphasize or de-emphasize certain regions of the data. For example, a weight function that down-weighs the tails of the distributions could be used to make the divergence more robust to outliers. Alternatively, a weight function that emphasizes a particular region of the data could be used to highlight differences in that region of the data.

Overall, the choice of α and the weight density function $\psi(x)$ can be tailored to suit the specific characteristics and features of the data for the two models that are being compared, allowing for greater sensitivity and flexibility in the comparison process, and the $D_\alpha^\psi(f : g)$ measure has potential uses in various fields, as listed below:

- Statistics: It can be used in goodness-of-fit tests and model selection criteria, for example, chi-square divergence ($\alpha = 0$) is commonly used in contingency table analysis.
- Machine learning: The proposed divergence measure can be used as a divergence measure in machine learning algorithms, such as clustering, classification and anomaly detection.
- Information theory: The proposed divergence can be used to measure the difference between probability distributions and to quantify the amount of information gained or lost in a data compression or transmission process.
- Signal processing: $D_\alpha^\psi(f : g)$ divergence measure can be used to compare the strength signals in signal processing applications.
- Image processing: The proposed $D_\alpha^\psi(f : g)$ divergence measure can be used to compare image histograms and textures in image processing applications.

One of the main motivations behind the development of $D_\alpha^\psi(f : g)$ divergence is that it encompasses several popular divergence measures as special cases, including the symmetric chi-square, triangular divergence, generalized chi-square, and Balakrishnan and Sanghvi divergence measures. This property makes the $D_\alpha^\psi(f : g)$ divergence measure a versatile tool for comparing probability distributions in a variety of fields and facilitates the integration of different divergence measures into a unified framework.

Furthermore, it should also be noted that the proposed Jensen-χ_α^2 and (p, w)-Jensen-χ_α^2 divergence measures are extensions of $D_\alpha^\psi(f : g)$ measure based on a convex combination. These extensions allow for the incorporation of additional divergence measures into the framework, further increasing the flexibility and applicability of the method. By combining different divergence measures in a convex form, these Jensen-type divergence measures can provide a more comprehensive and nuanced comparison of probability distributions.

In addition, in this paper, we also establish a new generalized mixture density and specifically show that the proposed model provides optimal information under three different optimization problems associated with χ_α^2 divergence measure. Moreover, some results on these information measures and their connections to other well-known information measures are also provided.

First, a diversity measure between two density functions f and g on common support $\mathcal{X}$, known as chi-square divergence, is defined as

$$\chi^2(f : g) = \int_{\mathcal{X}} \frac{(f(x) - g(x))^2}{f(x)} \, \mathrm{d}x. \tag{1.1}$$

Similarly, we can define $\chi^2(g : f)$.

A generalized version of χ^2 divergence measure, denoted by χ^2_α, between two densities f and g, for $\alpha \geq 0$, considered by [5], is defined as

$$\chi^2_\alpha(f : g) = \frac{\alpha+1}{2} \int_{\mathcal{X}} \frac{\left(f(x) - g(x)\right)^2}{f^{1-\alpha}(x)} \, \mathrm{d}x. \tag{1.2}$$

Balakrishnan and Sanghvi [4] introduced another version of the chi-square divergence in Eq. (1.1) as

$$\chi^2_{\mathrm{BS}}(f : g) = \int_{\mathcal{X}} \left(\frac{f(x) - g(x)}{f(x) + g(x)}\right)^2 f(x) \, \mathrm{d}x = E_f \left[\frac{f(X) - g(X)}{f(X) + g(X)}\right]^2, \tag{1.3}$$

where E denotes expectation taken with respect to density f on support $\mathcal{X}$, assuming it exists. This information measure is known as Balakrishnan–Sanghvi divergence measure.

Moreover, a symmetric version of chi-square divergence measure of the form

$$\chi^2_T(f : g) = \int_{\mathcal{X}} \frac{\left(f(x) - g(x)\right)^2}{f(x) + g(x)} \, \mathrm{d}x = 8E_h \left[\frac{f(X) - g(X)}{f(X) + g(X)}\right]^2 \tag{1.4}$$

has been introduced by [12]. Here, E denotes expectation under mixture density $h(x) = \frac{f(x)+g(x)}{2}$. The divergence measure in Eq. (1.4) is known as triangular divergence measure. Throughout this paper, we will suppress $\mathcal{X}$ in the integration with respect to X, unless a distinction becomes necessary.

The rest of this paper is organized as follows. In Section 2, we first examine the connection between χ^2_α divergence measure and q-Fisher information measure. Here, based on the χ^2_α divergence measure, we introduce a relative-χ^2_α divergence measure, which includes other well-known versions of chi-square divergence as special cases. We propose Jensen-χ^2_α divergence measure in Section 3. We then show that Jensen-χ^2_α divergence is a mixture of the proposed relative-χ^2_α divergence measures. Further, we show that a lower bound for Jensen-χ^2_α divergence can be given by Jensen–Shannon entropy measure. In Section 4, we first introduce (p, w)-Jensen-χ^2_α divergence measure and then discuss some of its properties. Next, the relative-χ^2_α divergence measure of escort and arithmetic densities are studied in Section 5. We then introduce (p, η)-mixture density in Section 6 and show that this mixture distribution involves optimal information under three different optimization problems associated with χ^2_α divergence measure. In Section 7, we study the relative-χ^2_α divergence measure of order statistics and mixed reliability systems. Next, in Section 8, we use a real example in image processing and present some numerical results in this regard in terms of Jensen-χ^2_α divergence measure. We specifically show that this divergence could serve as an useful measure of similarity between two images. Finally, we make some concluding remarks in Section 9.

2. Relative-χ^2_α divergence measure and connection between χ^2_α divergence measure and q-Fisher information

In this section, we first show that the χ^2_α divergence measure in Eq. (1.2) has a close connection to $q-$Fisher information of mixing parameter of a given arithmetic mixture distribution. Next, we introduce a relative-χ^2_α divergence measure and show that it includes some of the well-known chi-square-type divergence measures as special cases.

2.1. Connection between χ^2_α divergence measure and q-Fisher information

The q-Fisher information of a density function f_θ about parameter θ, defined by [14], is given by

$$\mathcal{I}_q(\theta) = \int \left(\frac{\partial \log_q f_\theta(x)}{\partial \theta}\right)^2 f_\theta(x)\,\mathrm{d}x, \tag{2.1}$$

where $\log_q(x)$ is the q-logarithmic function defined as

$$\log_q(x) = \frac{x^q - 1}{q} \quad \left(x \in \Re,\ q \neq 0\right) \tag{2.2}$$

for more details, see [9, 15, 23]. Then, we have the following result.

Theorem 2.1. *Let f_1 and f_2 be two density functions. Then, the q-information measure of mixing parameter p in the two-component mixture model*

$$f_p(x) = pf_1(x) + (1-p)f_2(x), \quad p \in (0,1), \tag{2.3}$$

is given by

$$\mathcal{I}_q(p) = \frac{8}{2q+1}\mathcal{M}_{\frac{1}{2}}\left(\chi^2_{2q}(f_p,f_1), \chi^2_{2q}(f_p,f_2)\right),$$

where $M_{\frac{1}{2}}(.,.)$ is the power mean with exponent $\frac{1}{2}$, defined as $\mathcal{M}_{\frac{1}{2}}(x,y) = \left(\frac{x^{\frac{1}{2}}}{2} + \frac{y^{\frac{1}{2}}}{2}\right)^2$ for positive x and y.

Proof. From the mixture model in Eq. (2.3), we readily see that

$$f_1(x) - f_2(x) = \frac{f_1(x) - f_p(x)}{1-p} = \frac{f_p(x) - f_2(x)}{p}. \tag{2.4}$$

Now, from the definition of q-Fisher information measure in Eq. (2.1), we find

$$\mathcal{I}_q(p) = \int \frac{\left(f_1(x) - f_2(x)\right)^2}{f_p^{1-2q}(x)}\,\mathrm{d}x = \begin{cases} \frac{2}{(1+2q)(1-p)^2}\chi^2_{2q}(f_p:f), & f = f_1, \\ \frac{2}{(1+2q)p^2}\chi^2_{2q}(f_p:f), & f = f_2, \end{cases} \tag{2.5}$$

which readily yields

$$\begin{aligned} \mathcal{I}_q(p) &= \frac{2}{1+2q}\left(\sqrt{\chi^2_{2q}(f_p:f_1)} + \sqrt{\chi^2_{2q}(f_p:f_2)}\right)^2 \\ &= \frac{8}{1+2q}\left(\frac{1}{2}\sqrt{\chi^2_{2q}(f_p:f_1)} + \frac{1}{2}\sqrt{\chi^2_{2q}(f_p:f_2)}\right)^2 \\ &= \frac{8}{1+2q}\mathcal{M}_{\frac{1}{2}}\left(\chi^2_{2q}(f_p:f_1), \chi^2_{2q}(f_p:f_2)\right), \end{aligned}$$

as required. □

2.2. Relative-χ^2_α divergence measure

In this subsection, we introduce a relative-χ^2_α divergence measure and show that it includes some of the well-known chi-square-type divergence measures as special cases. Further, we show that the special case of the proposed measure, when $\alpha = 0$, is connected to the variance of density ratios.

Definition 2.2. *Let f and g be two density functions on support $\mathcal{X}$. Then, a relative version of χ^2_α divergence measure between f and g with respect to density function ψ on support $\mathcal{X}$, denoted by R-χ^2_α, for $\alpha \geq 0$, is defined as*

$$D^{\psi}_{\alpha}(f:g) = \frac{1+\alpha}{2}\int_{\mathcal{X}} \frac{\left(f(x)-g(x)\right)^2}{\psi^{1-\alpha}(x)}\,\mathrm{d}x, \tag{2.6}$$

provided the involved integral exists. In addition, the special case of R-χ^2_α divergence measure, when $\alpha = 0$, is of the form

$$D^{\psi}_{\alpha=0}(f:g) = \frac{1}{2}\int_{\mathcal{X}} \frac{\left(f(x)-g(x)\right)^2}{\psi(x)}\,\mathrm{d}x. \tag{2.7}$$

Moreover, it is useful to note that $D^{\psi}_{\alpha}(f:g)$ reduces to $\chi^2_\alpha(f:g)$ when $\psi = f$. It is easily seen from Eq. (2.6) that R-χ^2_α divergence measure can be expressed based on two expectations under densities f and g to be

$$\begin{aligned} D^{\psi}_{\alpha}(f:g) &= \frac{1+\alpha}{2}\int_{\mathcal{X}} \frac{\left(f(x)-g(x)\right)^2}{\psi^{1-\alpha}(x)}\,\mathrm{d}x \\ &= \frac{1+\alpha}{2}E_f\left(\frac{f(X)-g(X)}{\psi^{1-\alpha}(X)}\right) + \frac{1+\alpha}{2}E_g\left(\frac{g(X)-f(X)}{\psi^{1-\alpha}(X)}\right). \end{aligned}$$

From the definition of $D^{\psi}_{\alpha}(f:g)$, the weight density function, $\psi(x)$, can be utilized to assign varying degrees of importance to different regions of the dataset. For instance, a weight function that places less emphasis on extreme values can be employed to make the divergence measure more robust to outliers. On the other hand, a weight function that highlights a specific region of the data can be used to detect dissimilarities within that region of the data.

In general, $D^{\psi}_{\alpha}(f:g)$ divergence provides a flexible and powerful framework for assessing the differences between probability distributions in a wide range of applications. The parameters α and $\psi(x)$ can be adjusted to suit the specific characteristics and features of the data for the two models that are being compared, offering greater sensitivity and flexibility in the comparison process.

Remark 2.3.

(i) If $\alpha = 1$, then $D^{\psi}_{\alpha=1}(f:g) = L_2(f:g) = \int \left(f(x)-g(x)\right)^2\mathrm{d}x$.

(ii) If $\psi(x) = f(x)$, then $D^{\psi}_{\alpha=0}(f:g) = \chi^2_0(f,g) = \frac{\chi^2(f,g)}{2}$.

(iii) If $\psi(x) = g(x)$, then $D^{\psi}_{\alpha=0}(f:g) = \chi^2_0(g,f) = \frac{\chi^2(g,f)}{2}$.

(iv) If $\psi(x) = pf(x) + (1-p)g(x)$, then $D^{\psi}_{\alpha=0}(f:g) = \frac{1}{2(1-p)^2}\chi^2(\psi:f) = \frac{1}{2p^2}\chi^2(\psi:g)$.

(v) If $\psi(x) = \frac{f(x)+g(x)}{2}$, then $D^{\psi}_{\alpha=0}(f:g) = \chi^2_T(f:g)$, where $\chi^2_T(f:g)$ is the triangular divergence defined in Eq. (1.4).

(vi) If $\psi(x) = \frac{f(x)+g(x)}{2}$, then $D^{\psi}_{\alpha=0}(f:g) = \chi^2_{\mathrm{BS}}(f:g) + \chi^2_{\mathrm{BS}}(g:f)$, where $D_{\mathrm{BS}}(f:g)$ is the Balakrishnan–Sanghvi divergence measure defined in Eq. (1.3).

Theorem 2.4. *Let ψ be a density function. Then, $D^{\psi}_{\alpha=0}(f:g)$ divergence measure in Eq. (2.7) can be expressed as*

$$D^{\psi}_{\alpha=0}(f:g) = \frac{Var_{\psi}\left(\frac{f(X)}{\psi(X)}\right) + Var_{\psi}\left(\frac{g(X)}{\psi(X)}\right)}{2} - E_{\psi}\left(\frac{f(X)g(X)}{\psi^2(X)}\right) + 1. \tag{2.8}$$

Proof. From the definition of $D^{\psi}_{\alpha=0}(f:g)$, we have

$$\begin{aligned} 2D^{\psi}_{\alpha=0}(f:g) &= \int \frac{f^2(x)}{\psi(x)}\,\mathrm{d}x - \left(\int f(x)\mathrm{d}x\right)^2 \\ &\quad + \int \frac{g^2(x)}{\psi(x)}\,\mathrm{d}x - \left(\int g(x)\mathrm{d}x\right)^2 - 2\int \frac{f(x)g(x)}{\psi(x)}\,\mathrm{d}x + 2 \\ &= Var_{\psi}\left(\frac{f(X)}{\psi(X)}\right) + Var_{\psi}\left(\frac{g(X)}{\psi(X)}\right) - 2E_{\psi}\left(\frac{f(X)g(X)}{\psi^2(X)}\right) + 2, \end{aligned}$$

as required. □

3. Jensen-χ^2_α divergence measure

In this section, we first introduce Jensen-χ^2_α divergence measure and then establish some of its properties.

In fact, the Jensen-χ^2_α divergence measure is an expansion of $D^{\psi}_{\alpha}(f:g)$ that is established based on a convex combination. This extension allows for the incorporation of additional divergence measures into the framework, further increasing the flexibility and applicability of the method.

Definition 3.1. *Let X_1, X_2 and Y be random variables with density functions f_1, f_2 and ψ, respectively, Then, the Jensen-χ^2_α (J-χ^2_α) divergence measure, for $p \in (0,1)$, is defined as*

$$\mathcal{J}^{\psi}_{\alpha}(f_1, f_2; \boldsymbol{P}) = p\chi^2_\alpha(\psi : f_1) + (1-p)\chi^2_\alpha(\psi : f_2) - \chi^2_\alpha(\psi : pf_1 + (1-p)f_2). \tag{3.1}$$

Lemma 3.2. *The J-χ^2_α divergence measure in Eq. (3.1) is non-negative.*

Proof. As $\phi(x) = x^2$ is a convex function, by using Jensen's inequality, we readily find

$$\begin{aligned} \frac{2}{1+\alpha}\mathcal{J}^{\psi}_{\alpha}(f_1, f_2; \mathbf{P}) &= p\int \frac{\left(f_1(x) - \psi(x)\right)^2}{\psi^{1-\alpha}(x)}\,\mathrm{d}x + (1-p)\int \frac{\left(f_2(x) - \psi(x)\right)^2}{\psi^{1-\alpha}(x)}\,\mathrm{d}x \\ &\quad - \int \frac{\left(pf_1(x) + (1-p)f_2(x) - \psi(x)\right)^2}{\psi^{1-\alpha}(x)}\,\mathrm{d}x \\ &= p\int \frac{f_1^2(x)}{\psi^{1-\alpha}(x)}\,\mathrm{d}x + (1-p)\int \frac{f_2^2(x)}{\psi^{1-\alpha}(x)}\,\mathrm{d}x - \int \frac{\left(pf_1(x) + (1-p)f_2(x)\right)^2}{\psi^{1-\alpha}(x)}\,\mathrm{d}x, \\ &\geq 0, \end{aligned}$$

where the last expression follows from the fact that

$$\left(pf_1(x) + (1-p)f_2(x)\right)^2 \leq pf_1^2(x) + (1-p)f_2^2(x).$$

□

 https://doi.org/10.1017/S0269964823000189

Theorem 3.3. *A representation for $\mathcal{J}^{\psi}_{\alpha=0}(f_1,f_2;\boldsymbol{P})$, based on variance of the ratio of densities, is given by*

$$\mathcal{J}^{\psi}_{\alpha=0}(f_1,f_2;\boldsymbol{P}) = \frac{1}{2}\left\{p\, Var_{\psi}\left(\frac{f_1(X)}{\psi(X)}\right) + (1-p)Var_{\psi}\left(\frac{f_2(X)}{\psi(X)}\right) - Var_{\psi}\left(\frac{pf_1(X)+(1-p)f_2(X)}{\psi(X)}\right)\right\}.$$

Proof. From the definition of $\mathcal{J}^{\psi}_{\alpha=0}(f_1,f_2;\mathbf{P})$, we have

$$\begin{aligned}
2\mathcal{J}^{\psi}_{\alpha=0}(f_1,f_2;\mathbf{P}) &= p\int \frac{\left(f_1(x)-\psi(x)\right)^2}{\psi(x)}\mathrm{d}x + (1-p)\int \frac{\left(f_2(x)-\psi(x)\right)^2}{\psi(x)}\mathrm{d}x \\
&\quad - \int \frac{\left(pf_1(x)+(1-p)f_2(x)-\psi(x)\right)^2}{\psi(x)}\mathrm{d}x \\
&= p\int \frac{f_1^2(x)}{\psi(x)}\mathrm{d}x + (1-p)\int \frac{f_2^2(x)}{\psi(x)}\mathrm{d}x - \int \frac{\left(pf_1(x)+(1-p)f_2(x)\right)^2}{\psi(x)}\mathrm{d}x \\
&= p\left\{\int \frac{f_1^2(x)}{\psi(x)}\mathrm{d}x - 1\right\} + (1-p)\left\{\int \frac{f_2^2(x)}{\psi(x)}\mathrm{d}x - 1\right\} \\
&\quad - \left\{\int \frac{\left(pf_1(x)+(1-p)f_2(x)\right)^2}{\psi(x)}\mathrm{d}x - 1\right\} \\
&= p\, Var_{\psi}\left(\frac{f_1(X)}{\psi(X)}\right) + (1-p)Var_{\psi}\left(\frac{f_2(X)}{\psi(X)}\right) - Var_{\psi}\left(\frac{pf_1(X)+(1-p)f_2(X)}{\psi(X)}\right),
\end{aligned}$$

as required. □

Theorem 3.4. *Let the random variables X_1 and X_2 have density functions f_1 and f_2, respectively. Then, $\mathcal{J}^{\psi}_{\alpha}(f_1,f_2;\boldsymbol{P})$ measure is a mixture of R-χ^2_{α} divergence measures of the form*

$$\mathcal{J}^{\psi}_{\alpha}(f_1,f_2;\boldsymbol{P}) = pD^{\psi}_{\alpha}(f_1:f_T) + (1-p)D^{\psi}_{\alpha}(f_2:f_T),$$

where $D^{\psi}_{\alpha}(f_i : f_T)$ is the divergence measure in Eq. (2.6), with $f_T = pf_1 + (1-p)f_2$ being the two-component mixture density.

Proof. With $f_T = pf_1 + (1-p)f_2$, we first find

$$\begin{aligned}
\frac{2}{1+\alpha}\mathcal{J}^{\psi}_{\alpha}(f_1,f_2;\mathbf{P}) &= p\int \frac{\left(f_1(x)-\psi(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x + (1-p)\int \frac{\left(f_2(x)-\psi(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x \\
&\quad - \int \frac{\left(pf_1(x)+(1-p)f_2(x)-\psi(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x \\
&= p\int \frac{f_1^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x + (1-p)\int \frac{f_2^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x - \int \frac{\left(pf_1(x)+(1-p)f_2(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x.
\end{aligned}$$

On the other hand, with $k = pD^{\psi}_{\alpha}(f_1:f_T) + (1-p)D^{\psi}_{\alpha}(f_2:f_T)$, we also have

$$\frac{2}{1+\alpha}k = p\int \frac{\left(f_1(x)-f_T(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x + (1-p)\int \frac{\left(f_2(x)-f_T(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x$$

$$
\begin{aligned}
&= p\int \frac{f_1^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x - 2p\int \frac{f_1(x)f_T(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x + p\int \frac{f_T^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x \\
&\quad +(1-p)\int \frac{f_2^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x - 2(1-p)\int \frac{f_2(x)f_T(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x + (1-p)\int \frac{f_T^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x \\
&= p\int \frac{f_1^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x + (1-p)\int \frac{f_2^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x - 2\int \frac{f_T^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x + \int \frac{f_T^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x \\
&= p\int \frac{f_1^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x + (1-p)\int \frac{f_2^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x - \int \frac{f_T^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x,
\end{aligned}
$$

which establishes the required result. □

Theorem 3.5. *A connection between $\mathcal{J}^{\psi}_{\alpha=0}(f_1,f_2;\boldsymbol{P})$ with $\psi = \frac{f_1+f_2}{2}$ and Balakrishnan–Sanghvi divergence measure is given by*

$$
\mathcal{J}^{\psi}_{\alpha=0}(f_1,f_2;\boldsymbol{P}) = p\left\{\chi^2_{\mathrm{BS}}(f_1:f_T) + \chi^2_{\mathrm{BS}}(f_T:f_1)\right\} + (1-p)\left\{\chi^2_{\mathrm{BS}}(f_2:f_T) + \chi^2_{\mathrm{BS}}(f_T:f_2)\right\},
$$

where $f_T = pf_1 + (1-p)f_2$ is the two-component mixture density.

Proof. With $f_T = pf_1 + (1-p)f_2$ and from Part (vi) of Remark 2.3 and Theorem 3.4, we have

$$
\begin{aligned}
\mathcal{J}^{\psi}_{\alpha=0}(f_1,f_2;\mathbf{P}) &= pD^{\psi}_{\alpha=0}(f_1:f_T) + (1-p)D^{\psi}_{\alpha=0}(f_2:f_T) \\
&= p\left\{\chi^2_{\mathrm{BS}}(f_1:f_T) + \chi^2_{\mathrm{BS}}(f_T:f_1)\right\} + (1-p)\left\{\chi^2_{\mathrm{BS}}(f_2:f_T) + \chi^2_{\mathrm{BS}}(f_T:f_2)\right\},
\end{aligned}
$$

as required. □

Theorem 3.6. *We have*

$$
\frac{-1}{2}\frac{\partial^2}{\partial p^2}\mathcal{J}^{\psi}_{\alpha}(f_1,f_2;\boldsymbol{P}) = D^{\psi}_{\alpha}(f_1:f_2). \tag{3.2}
$$

Proof. From the definition $\mathcal{J}^{\psi}_{\alpha}(f_1,f_2;\mathbf{P})$ in Eq. (3.1) and making use of the dominated convergence theorem, we have

$$
\begin{aligned}
\frac{-1}{2}\frac{\partial^2}{\partial p^2}\mathcal{J}^{\psi}_{\alpha}(f_1,f_2;\mathbf{P}) &= -\frac{1+\alpha}{4}\frac{\partial}{\partial p}\left(\int \frac{\left(f_1(x)-\psi(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x + \int \frac{\left(f_2(x)-\psi(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x\right) \\
&\quad + \frac{1+\alpha}{2}\frac{\partial}{\partial p}\left(\int \left(f_1(x)-f_2(x)\right)\frac{pf_1(x)+(1-p)f_2(x)-\psi(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x\right) \\
&= \frac{1+\alpha}{2}\int \frac{\left(f_1(x)-f_2(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x \\
&= D^{\psi}_{\alpha}(f_1:f_2),
\end{aligned}
$$

as required. □

We now extend the definition of Jensen-χ^2_α divergence measure in Eq. (3.1) to the case of $n+1$ random variables. Let $X_1,\ldots,X_n$ and Y be random variables with density functions $f_1,\ldots,f_n$ and ψ, respectively, and $p_1, \ldots, p_n$ be non-negative real numbers such that $\sum_{i=1}^{n} p_i = 1$. Then, the Jensen-χ_α^2 measure is defined as

$$\mathcal{J}_\alpha^\psi(f_1,\ldots,f_n;\mathbf{P}) = \sum_{i=1}^{n} p_i \chi_\alpha^2(\psi : f_i) - \chi_\alpha^2\left(\psi : \sum_{i=1}^{n} p_i f_i\right). \tag{3.3}$$

The special case of Jensen-χ_α^2 divergence measure, when $\alpha = 0$, has the representation

$$\begin{aligned}\mathcal{J}_{\alpha=0}^\psi(f_1,\ldots,f_n;\mathbf{P}) &= \frac{1}{2}\sum_{i=1}^{n} p_i \chi^2(\psi : f_i) - \frac{1}{2}\chi^2\left(\psi : \sum_{i=1}^{n} p_i f_i\right)\\ &= \frac{1}{2}\sum_{i=1}^{n} p_i Var_\psi\left(\frac{f_i(X)}{\psi(X)}\right) - \frac{1}{2}Var_\psi\left(\frac{\sum_{i=1}^{n} p_i f_i(X)}{\psi(X)}\right).\end{aligned}$$

Corollary 3.7. *The $\mathcal{J}_\alpha^\psi(f_1,\ldots,f_n;\boldsymbol{P})$ measure in Eq. (3.3) is a mixture of D_α^ψ measures in Eq. (2.6) of the form*

$$\mathcal{J}_\alpha^\psi(f_1,\ldots,f_n;\boldsymbol{P}) = \sum_{i=1}^{n} p_i D_\alpha^\psi(f_i : f_T).$$

Theorem 3.8. *The $\mathcal{J}_\alpha^\psi(f_1,\ldots,f_n;\boldsymbol{P})$ measure in Eq. (3.3) is a mixture of D_α^ψ measures in Eq. (2.6) of the form*

$$\mathcal{J}_\alpha^\psi(f_1,\ldots,f_n;\boldsymbol{P}) = 2\sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j D_\alpha^\psi(f_i : f_j).$$

Proof. From Corollary 3.7 and making use of the identity ([21], pp. 95–96)

$$\sum_{i=1}^{n} w_i (x_i - \bar{x}_w)^2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j (x_i - x_j)^2, \qquad \bar{x}_w = \sum_{i=1}^{n} w_i x_i, \ \sum_{i=1}^{n} w_i = 1,$$

we obtain

$$\begin{aligned}\mathcal{J}_\alpha^\psi(f_1,\ldots,f_n;\mathbf{P}) &= \sum_{i=1}^{n} p_i D_\alpha^\psi(f_i : f_T)\\ &= \frac{1+\alpha}{2}\sum_{i=1}^{n} p_i \int \frac{\left(f_i(x) - \sum_{j=1}^{n} p_j f_j(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x\\ &= \frac{(1+\alpha)}{4}\sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j \int \frac{\left(f_i(x) - f_j(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x\\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j D_\alpha^\psi(f_i : f_j),\end{aligned}$$

as required. □

Theorem 3.9. *Let* $f_i \geq \frac{\psi^{1-\alpha}}{2}, i = 1, \ldots, n$. *Then, a lower bound for* $\mathcal{J}_\alpha^\psi(f_1, \ldots, f_n; \boldsymbol{P})$ *is given by*

$$\mathcal{J}_\alpha^\psi(f_1, \ldots, f_n; \boldsymbol{P}) \geq \frac{1+\alpha}{4} JS_{\boldsymbol{P}}(f_1, \ldots, f_n),$$

where $JS_{\boldsymbol{P}}(f_1, \ldots, f_n)$ *is the Jensen–Shannon entropy; see [13].*

Proof. From the assumption, Theorem 3.8 and by making use of the identity

$$\sum_{i=1}^n w_i (x_i - \bar{x}_w)^2 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_i w_j (x_i - x_j)^2,$$

and then setting $w_i = p_i$, $w_j = p_j$, $x_i = f_i(x)$, $x_j = f_j(x)$ and $\bar{x}_w = \sum_{i=1}^n p_i f_i(x)$, we find

$$\begin{aligned}
\frac{2}{1+\alpha} \mathcal{J}_\alpha^\psi(f_1, \ldots, f_n; \mathbf{P}) &= \sum_{i=1}^n p_i \int \frac{\left(f_i(x) - \sum_{j=1}^n p_j f_j(x)\right)^2}{\psi^{1-\alpha}(x)} \mathrm{d}x \\
&= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n p_i p_j \int \frac{\left(f_i(x) - f_j(x)\right)^2}{\psi^{1-\alpha}(x)} \mathrm{d}x \\
&= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n p_i p_j \int \frac{\left(f_i(x) - f_j(x)\right)^2}{f_i(x)} \frac{f_i(x)}{\psi^{1-\alpha}(x)} \mathrm{d}x \\
&\geq \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n p_i p_j \int \frac{\left(f_i(x) - f_j(x)\right)^2}{f_i(x)} \mathrm{d}x \\
&\geq \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n p_i p_j \int f_i(x) \log\left(\frac{f_i(x)}{f_j(x)}\right) \mathrm{d}x \\
&= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n p_i p_j KL(f_i : f_j) \\
&\geq \frac{1}{2} JS_{\mathbf{P}}(f_1, \ldots, f_n),
\end{aligned}$$

where the second inequality follows from the fact that $\log(x) < x - 1, x > 0$, and the last inequality follows from [3]. □

4. (p, w)-Jensen-χ_α^2 divergence measure

In this section, we first review the definition of (p, w)-Jensen–Shannon divergence measure. Then, we introduce (p, w)-Jensen-χ_α^2 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. Then, the Kullback–Leibler divergence between f and g is defined as

$$KL(f, g) = \int f(x) \log\left(\frac{f(x)}{g(x)}\right) \mathrm{d}x,$$

where log denotes the natural logarithm. 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

$$JS_{(p,w)}(f_1, f_2) = H\big((1-\bar{s})f_1 + \bar{s}f_2\big) - wH\big((1-p)f_1 + pf_2\big) - (1-w)H\big(pf_1 + (1-p)f_2\big)$$

$$
\begin{aligned}
&= wKL\left((1-p)f_1+pf_2 : (1-\bar{s})f_1+\bar{s}f_2\right)\\
&\quad +(1-w)KL\left(pf_1+(1-p)f_2 : (1-\bar{s})f_1+\bar{s}f_2\right),
\end{aligned}
$$

where $\bar{s} = wp + (1-w)(1-p)$. For more details, one may refer to [16, 17].

Definition 4.1. *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-χ^2_α divergence measure, for w and $p \in (0,1)$, is defined as*

$$
\begin{aligned}
\mathcal{J}^{\psi}_{\alpha}(f_1,f_2;\boldsymbol{P},\boldsymbol{w}) &= w\chi^2_\alpha(\psi : (1-p)f_1+pf_2) + (1-w)\chi^2_\alpha(\psi : pf_1+(1-p)f_2)\\
&\quad -\chi^2_\alpha(\psi : (1-\bar{s})f_1+\bar{s}f_2),
\end{aligned}
$$

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

Theorem 4.2. *Let the random variables X_1 and X_2 have density functions f_1 and f_2, respectively. Then, $\mathcal{J}^{\psi}_{\alpha}(f_1,f_2;\boldsymbol{P},\boldsymbol{w})$ is a mixture of relative measures in Eq. (2.6) of the form*

$$
\mathcal{J}^{\psi}_{\alpha}(f_1,f_2;\boldsymbol{P},\boldsymbol{w}) = wD^{\psi}_{\alpha}\left((1-p)f_1+pf_2 : f^T_{\bar{s}}\right) + (1-w)D^{\psi}_{\alpha}\left(pf_1+(1-p)f_2 : f^T_{\bar{s}}\right), \tag{3.3}
$$

with $f^T_{\bar{s}} = (1-\bar{s})f_1+\bar{s}f_2$ is the two-component mixture density.

Proof. With $f^T_{\bar{s}} = (1-\bar{s})f_1+\bar{s}f_2$, we find

$$
\begin{aligned}
&\frac{2}{1+\alpha}\mathcal{J}^{\psi}_{\alpha}(f_1,f_2;\mathbf{P},\mathbf{w})\\
&= w\int \frac{\left((1-p)f_1(x)+pf_2(x)-\psi(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x + (1-w)\int \frac{\left(pf_1(x)+(1-p)f_2(x)-\psi(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x\\
&\quad -\int \frac{\left((1-\bar{s})f_1(x)+\bar{s}f_2(x)-\psi(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x\\
&= w\int \frac{\left((1-p)f_1(x)+pf_2(x)\right)^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x + (1-w)\int \frac{\left(pf_1(x)+(1-p)f_2(x)\right)^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x\\
&\quad -\int \frac{\left((1-\bar{s})f_1(x)+\bar{s}f_2(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x.
\end{aligned}
$$

On the other hand, letting

$$
\frac{1+\alpha}{2}k = wD^{\psi}_{\alpha}\left((1-p)f_1+pf_2 : f^T_{\bar{s}}\right) + (1-w)D^{\psi}_{\alpha}\left(pf_1+(1-p)f_2 : f^T_{\bar{s}}\right)
$$

and using the fact that

$$
\begin{aligned}
f^T_{\bar{s}}(x) &= (1-\bar{s})f_1(x)+\bar{s}f_2(x)\\
&= \left(1-(wp+(1-w)(1-p))\right)f_1(x) + (wp+(1-w)(1-p))f_2(x)\\
&= w\left((1-p)f_1(x)+pf_2(x)\right) + (1-w)\left(pf_1(x)+(1-p)f_2(x)\right),
\end{aligned}
$$

we find

$$
k = w\int \frac{\left((1-p)f_1(x)+pf_2(x)-f^T_{\bar{s}}(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x + (1-w)\int \frac{\left(pf_1(x)+(1-p)f_2(x)-f^T_{\bar{s}}(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x
$$

$$
\begin{aligned}
&= w\int \frac{\left((1-p)f_1(x)+pf_2(x)\right)^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x - 2w\int \frac{\left((1-p)f_1(x)+pf_2(x)\right)f_{\bar{s}}^T(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x + \int \frac{\left(f_{\bar{s}}^T(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x \\
&\quad +(1-w)\int \frac{\left((pf_1(x)+(1-p)f_2(x)\right)^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x - 2(1-w)\int \frac{\left(pf_1(x)+(1-p)f_2(x)\right)f_{\bar{s}}^T(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x \\
&= w\int \frac{\left((1-p)f_1(x)+pf_2(x)\right)^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x + (1-w)\int \frac{\left((pf_1(x)+(1-p)f_2(x)\right)^2(x)}{\psi^{1-\alpha}(x)}\mathrm{d}x \\
&\quad + \int \frac{\left(f_{\bar{s}}^T(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x.
\end{aligned}
$$

Now, from the above results, we have

$$
\mathcal{J}_\alpha^\psi(f_1,f_2;\mathbf{P},\mathbf{w}) = \frac{1+\alpha}{2}k = wD_\alpha^\psi\left((1-p)f_1+pf_2 : f_{\bar{s}}^T\right) + (1-w)D_\alpha^\psi\left(pf_1+(1-p)f_2 : f_{\bar{s}}^T\right),
$$

which establishes the required result. □

From Definitions 3.1 and 4.1, we readily have the following Corollary.

Corollary 4.3. *A connection between $\mathcal{J}_\alpha^\psi(f_1,f_2;\boldsymbol{P},\boldsymbol{w})$ and $\mathcal{J}_\alpha^\psi(f_1,f_2;\boldsymbol{w})$ measures is given by*

$$
\mathcal{J}_\alpha^\psi(f_1,f_2;\boldsymbol{P},\boldsymbol{w}) = \mathcal{J}_\alpha^\psi\Big((1-p)f_1+pf_2 : pf_1+(1-p)f_2;\boldsymbol{w}\Big).
$$

Theorem 4.4. *We have*

(i)

$$
\frac{-1}{2}\frac{\partial^2}{\partial w^2}\mathcal{J}_\alpha^\psi(f_1,f_2;\boldsymbol{P},\boldsymbol{w}) = D_\alpha^\psi\left((1-p)f_1+pf_2 : pf_1+(1-p)f_2\right);
$$

(ii)

$$
\frac{-1}{2}\frac{\partial^2}{\partial p^2}\mathcal{J}_\alpha^\psi(f_1,f_2;\boldsymbol{P},\boldsymbol{w}) = D_\alpha^\psi\left((1-w)f_1+wf_2 : wf_1+(1-w)f_2\right) - D_\alpha^\psi(f_1:f_2).
$$

Proof. From Theorem 3.6 and Corollary 4.3, we have

$$
\begin{aligned}
\frac{-1}{2}\frac{\partial^2}{\partial w^2}\mathcal{J}_\alpha^\psi(f_1,f_2;\mathbf{P},\mathbf{w}) &= \frac{-1}{2}\frac{\partial^2}{\partial w^2}\mathcal{J}_\alpha^\psi\left((1-p)f_1+pf_2, pf_2+(1-p)f_1,\mathbf{w}\right) \\
&= D_\alpha^\psi\left((1-p)f_1+pf_2 : pf_1+(1-p)f_2\right),
\end{aligned}
$$

which proves Part (i). From Corollary 4.3 and using the facts that

$$
f_{\bar{s}}^T(x) = w\big((1-p)f_1(x)+pf_2(x)\big) + (1-w)\big(pf_1(x)+(1-p)f_2(x)\big)
$$

and

$$
(1-w)f_1(x)+wf_2(x) - \big(wf_1(x)+(1-w)f_2(x)\big) = w(f_2(x)-f_1(x)) + (1-w)(f_1(x)-f_2(x)),
$$

we find

$$\begin{aligned}
\frac{-1}{2}\frac{\partial^2}{\partial p^2}\mathcal{J}_\alpha^\psi(f_1,f_2;\mathbf{P},\mathbf{w}) &= \frac{-1}{2}\frac{\partial^2}{\partial p^2}\mathcal{J}_\alpha^\psi\left((1-p)f_1+pf_2,pf_1+(1-p)f_2;\mathbf{w}\right)\\
&= -w\frac{\alpha+1}{4}\frac{\partial^2}{\partial p^2}\int\frac{\left((1-p)f_1(x)+pf_2(x)-\psi(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x\\
&\quad -(1-w)\frac{\alpha+1}{4}\frac{\partial^2}{\partial p^2}\int\frac{\left(pf_1(x)+(1-p)f_2(x)-\psi(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x\\
&\quad +\frac{\alpha+1}{4}\frac{\partial^2}{\partial p^2}\int\frac{\left(f_{\bar{s}}^T(x)-\psi(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x\\
&= \frac{\alpha+1}{2}\int\frac{\left(w(f_2(x)-f_1(x))+(1-w)(f_1(x)-f_2(x))\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x\\
&\quad -\frac{\alpha+1}{2}\int\frac{\left(f_2(x)-f_1(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x\\
&= \frac{\alpha+1}{2}\int\frac{\left((1-w)f_1(x)+wf_2(x)-\left(wf_1(x)+(1-w)f_2(x)\right)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x\\
&\quad -\frac{\alpha+1}{2}\int\frac{\left(f_2(x)-f_1(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x\\
&= D_\alpha^\psi\left((1-w)f_1+wf_2:wf_1+(1-w)f_2\right)-D_\alpha^\psi\left(f_1:f_2\right),
\end{aligned}$$

which proves Part (ii). Hence, the theorem. □

5. D_α^ψ divergence measure of escort and arithmetic mixture densities

In this section, we examine D_α^ψ divergence measure of escort and arithmetic mixture densities.

5.1. D_α^ψ divergence measure of escort and generalized escort densities

The escort distribution is a key concept in nonextensive statistical mechanics and coding theory and is closely associated with Tsallis and Rényi entropy measures. Bercher [6] studied some connections between coding theory and the measure of complexity in nonextensive statistical mechanics in terms of escort distributions.

Let f be a density function. Then, the escort density with order $\eta > 0$, associated with f, is defined as

$$f_\eta(x) = \frac{f^\eta(x)}{\int f^\eta(x)\mathrm{d}x}. \tag{5.1}$$

Theorem 5.1. *Let f and g be two density functions and f_α be the escort density corresponding to f. Then, for $0 \le \eta \le 1$ and $\psi(x) = f_\eta(x)$, we have*

$$D_\alpha^\psi(f:g) = \frac{1+\alpha}{1+\beta}G_\eta^{1-\alpha}(f)\chi_\beta^2(f:g), \tag{5.2}$$

where $\beta = 1 - \eta(1-\alpha)$ and $G_\eta(f)$ is the information generating function of density f with order η defined as

$$G_\eta(f) = \int f^\eta(x)\mathrm{d}x. \tag{5.3}$$

Proof. From the definition of $D^\psi_\alpha(f:g)$ and the assumption that $\psi(x) = f_\eta(x)$, we have

$$\begin{aligned}
D^\psi_\alpha(f:g) &= \frac{1+\alpha}{2}\int \frac{\big(f(x)-g(x)\big)^2}{f_\eta^{1-\alpha}(x)}\mathrm{d}x \\
&= \frac{1+\alpha}{2}\left(\int f^\eta(x)\mathrm{d}x\right)^{1-\alpha}\int \frac{\big(f(x)-g(x)\big)^2}{f^{\eta(1-\alpha)}(x)}\mathrm{d}x \\
&= \frac{1+\alpha}{2}G_\eta^{1-\alpha}(f)\int \frac{\big(f(x)-g(x)\big)^2}{f^{\eta(1-\alpha)}(x)}\mathrm{d}x \\
&= \frac{1+\alpha}{1+\beta}G_\eta^{1-\alpha}(f)\chi^2_\beta(f:g),
\end{aligned}$$

where $\beta = 1 - \eta(1-\alpha)$, as desired.

Next, let f and g be two density functions. Then, the generalized escort density, for $1 > \eta > 0$, is defined as

$$h_\eta(x) = \frac{f^\eta(x)g^{1-\eta}(x)}{\int f^\eta(x)g^{1-\eta}(x)\mathrm{d}x}. \tag{5.4}$$

Let $\psi(x) = h_\eta(x)$. We then have

$$\begin{aligned}
2D^\psi_{\alpha=0}(f:g) &= \int \frac{\big(f(x)-g(x)\big)^2}{h_\eta(x)}\mathrm{d}x = \int \frac{\big(f(x)-g(x)\big)^2}{\frac{f^\eta(x)g^{1-\eta}(x)}{\int f^\eta(x)g^{1-\eta}(x)\mathrm{d}x}}\mathrm{d}x \\
&= \left(\int f^\eta(x)g^{1-\eta}(x)\mathrm{d}x\right)\int \frac{\big(f(x)-g(x)\big)^2}{f^\eta(x)g^{1-\eta}(x)}\mathrm{d}x \\
&= R_\eta(f:g)\int \frac{\big(f(x)-g(x)\big)^2}{f^\eta(x)g^{1-\eta}(x)}\mathrm{d}x,
\end{aligned} \tag{5.5}$$

where $R_\eta(f:g)$ is the relative information-generating function between density functions f and g defined as

$$R_\eta(f:g) = \int f^\eta(x)g^{1-\eta}(x)\mathrm{d}x. \tag{5.6}$$

□

Theorem 5.2. *A lower bound for $D^\psi_{\alpha=0}(f:g)$ in Eq. (5.5) is given by*

$$D^\psi_{\alpha=0}(f:g) \geq \frac{R_\eta(f,g)}{2(1-\eta)^2}\chi^2(f_\eta:f),$$

where $f_\eta = \eta f + (1-\eta)g$ is the two-component mixture density, $\chi^2(.:.)$ is the chi-square divergence, and $R_\eta(f:g)$ is as defined in Eq. (5.6).

Proof. From the definition of $D^{\psi}_{\alpha}(f:g)$ and the assumption that

$$\psi(x) = h_\eta(x) = \frac{f^\eta(x)g^{1-\eta}(x)}{\int f^\eta(x)g^{1-\eta}(x)\mathrm{d}x},$$

for $0 \le \eta \le 1$, and using the geometric mean-arithmetic mean inequality between densities f and g given by

$$f^\eta(x)g^{1-\eta}(x) \le \eta f(x) + (1-\eta)g(x),$$

and the fact that $g(x) - f(x) = \frac{1}{1-\eta}(f_\eta(x) - f(x))$, we obtain

$$\begin{aligned}
2D^{\psi}_{\alpha=0}(f:g) &= \int \frac{\left(f(x)-g(x)\right)^2}{h_\eta(x)}\mathrm{d}x \\
&= \left(\int f^\eta(x)g^{1-\eta}(x)\mathrm{d}x\right)\int \frac{\left(f(x)-g(x)\right)^2}{f^\eta(x)g^{1-\eta}(x)}\mathrm{d}x \\
&= R_\eta(f:g)\int \frac{\left(f(x)-g(x)\right)^2}{f^\eta(x)g^{1-\eta}(x)}\mathrm{d}x \\
&\ge R_\eta(f:g)\int \frac{\left(f(x)-g(x)\right)^2}{\eta f(x)+(1-\eta)g(x)}\mathrm{d}x \\
&= \frac{R_\eta(f:g)}{(1-\eta)^2}\int \frac{\left(f_\eta(x)-f(x)\right)^2}{f_\eta(x)}\mathrm{d}x \\
&= \frac{R_\eta(f:g)}{(1-\eta)^2}\chi^2(f_\eta:f),
\end{aligned}$$

as required. □

5.2. D^{ψ}_{α} divergence measure between two arithmetic mixture densities

In this subsection, we study D^{ψ}_{α} divergence measure between two arithmetic mixture densities. Consider two mixture density functions $f_m(x) = \sum_{i=1}^n p_i f_i(x)$ and $g_m(x) = \sum_{i=1}^n p_i g_i(x)$. Then, we have

$$\begin{aligned}
D^{\psi}_{\alpha}(f_m:g_m) &= \frac{1+\alpha}{2}\int \frac{\left(f_m(x)-g_m(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x = \frac{1+\alpha}{2}\int \frac{\left(\sum_{i=1}^n p_i f_i(x) - \sum_{i=1}^n p_i g_i(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x \\
&= \frac{1+\alpha}{2}\int \frac{\left(\sum_{i=1}^n p_i\left(f_i(x)-g_i(x)\right)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x \\
&= \frac{1+\alpha}{2}\int \left\{\sum_{i=1}^n p_i^2 \frac{\left(f_i(x)-g_i(x)\right)^2}{\psi^{1-\alpha}(x)} + 2\sum_{\substack{i=1 \\ i<j}}^n \sum_{j=1}^n p_i p_j \frac{\left(f_i(x)-g_i(x)\right)\left(f_j(x)-g_j(x)\right)}{\psi^{1-\alpha}(x)}\right\}\mathrm{d}x \\
&= \frac{1+\alpha}{2}\sum_{i=1}^n p_i^2 \int \frac{\left(f_i(x)-g_i(x)\right)^2}{\psi^{1-\alpha}(x)}\mathrm{d}x \\
&\quad +(1+\alpha)\sum_{\substack{i=1 \\ i<j}}^n \sum_{j=1}^n p_i p_j \int \frac{\left(f_i(x)-g_i(x)\right)\left(f_j(x)-g_j(x)\right)}{\psi^{1-\alpha}(x)}\mathrm{d}x
\end{aligned}$$

$$= \frac{1+\alpha}{2}\sum_{i=1}^{n} p_i^2 D_\alpha^\psi(f_i : g_i) + (1+\alpha)\sum_{\substack{i=1\\ i<j}}^{n}\sum_{j=1}^{n} p_i p_j \int \frac{\left(f_i(x)-g_i(x)\right)\left(f_j(x)-g_j(x)\right)}{\psi^{1-\alpha}(x)}\mathrm{d}x.$$

Theorem 5.3. *Let $f_1, \ldots, f_n$ be n density functions. Now, consider the probability mixing vector $\boldsymbol{P} = (p_1, \ldots, p_n)$ and its corresponding negation probability vector*

$$\bar{\boldsymbol{P}} = (\bar{p}_1, \ldots, \bar{p}_n) = \left(\frac{1-p_1}{n-1}, \ldots, \frac{1-p_n}{n-1}\right).$$

Then, we have the lower bound for $D_{\alpha=0}^\psi$ as

$$D_{\alpha=0}^\psi\left(\sum_{i=1}^n p_i f_i : \sum_{i=1}^n \bar{p}_i f_i\right) \geq \frac{1}{2}\sum_{i=1}^n \left(\frac{np_i - 1}{n-1}\right)^2 \left(KL(f_i : \psi) + 1\right) + L,$$

where $L = \sum_{\substack{i=1\\ i<j}}^n \sum_{j=1}^n \frac{(np_i-1)(np_j-1)}{(n-1)^2}\int \frac{f_i(x)f_j(x)}{\psi^2(x)}\mathrm{d}x$. For more details about negation probability, see [22].

Proof. From the definition of $D_{\alpha=0}^\psi$ divergence measure between mixture densities $\sum_{i=1}^n p_i f_i$ and $\sum_{i=1}^n \bar{p}_i f_i$ and upon setting

$$L = \sum_{\substack{i=1\\ i<j}}^{n}\sum_{j=1}^{n} \frac{(np_i - 1)(np_j - 1)}{(n-1)^2}\int \frac{f_i(x)f_j(x)}{\psi^2(x)}\mathrm{d}x,$$

we find

$$\begin{aligned}
D_{\alpha=0}^\psi\left(\sum_{i=1}^n p_i f_i : \sum_{i=1}^n \bar{p}_i f_i\right) &= \frac{1}{2}\int \frac{\left(\sum_{i=1}^n p_i f_i(x) - \sum_{i=1}^n \bar{p}_i f_i(x)\right)^2}{\psi(x)}\mathrm{d}x \\
&= \frac{1}{2}\sum_{i=1}^n \left(\frac{np_i - 1}{n-1}\right)^2 \int \frac{f_i^2(x)\mathrm{d}x}{\psi(x)}\mathrm{d}x + L \\
&= \frac{1}{2}\sum_{i=1}^n \left(\frac{np_i - 1}{n-1}\right)^2 \left(\chi^2(f_i : \psi) + 1\right) + L \\
&\geq \frac{1}{2}\sum_{i=1}^n \left(\frac{np_i - 1}{n-1}\right)^2 \left(KL(f_i : \psi) + 1\right) + L,
\end{aligned}$$

where the last inequality follows from the inequality between Kullback–Leibler and chi-square divergence measures. This proves the required result. □

6. Optimal information under χ_α^2 divergence measure

In this section, we first introduce (p, η)-mixture density as a generalization of arithmetic and harmonic mixture densities. Then, we examine optimal information property of (p, η)-mixture density. To follow this, we consider optimization problem for χ_α^2 divergence under three types of constraints. For more details about optimal information properties of some mixture distributions (arithmetic, geometric and $\alpha-$mixture distributions), one may refer to [2] and the references therein.

6.1. (p, η)*-mixture density*

Definition 6.1. *Let f_0 and f_1 be two density functions. Then, a generalized mixture density, called the (p, η)-mixture density, is defined as*

$$f_m(x) = \frac{pf_0^{\eta}(x) + (1-p)f_1^{\eta}(x)}{pf_0^{\eta-1}(x) + (1-p)f_1^{\eta-1}(x)} \left(\int \frac{pf_0^{\eta}(x) + (1-p)f_1^{\eta}(x)}{pf_0^{\eta-1}(x) + (1-p)f_1^{\eta-1}(x)} \mathrm{d}x \right)^{-1}.$$

The (p, η)-mixture density provides arithmetic and harmonic mixture densities as special cases:

(i) If $p = 0$, then $f_m(x) = f_1(x)$.
(ii) If $p = 1$, then $f_m(x) = f_0(x)$.
(iii) If $\eta = 1$, then $f_m(x) = pf_0(x) + (1-p)f_1(x)$ is the arithmetic mixture density.
(iv) If $\eta = 0$, then $f_m(x) = \frac{\left(\frac{p}{f_0(x)} + \frac{1-p}{f_1(x)}\right)^{-1}}{\int \left(\frac{p}{f_0(x)} + \frac{1-p}{f_1(x)}\right)^{-1} \mathrm{d}x}$ is the harmonic mixture density.

6.2. Optimal information property of (p, η)*-mixture density*

Theorem 6.2. *Let f, f_0 and f_1 be three density functions. Then, the solution to the optimization problem*

$$\min_f \chi^2_\alpha(f_0 : f) \ \textit{ subject to } \chi^2_\alpha(f_1 : f) = \eta, \int f(x)\mathrm{d}x = 1 \tag{6.1}$$

is the (p, η)-mixture density with $\eta = \alpha$ and mixing parameter $p = \frac{1}{1+\lambda_0}$, and $\lambda_0 > 0$ is the Lagrangian multiplier.

Proof. We use the Lagrangian multiplier technique for finding the solution of the optimization problem in Eq. (6.1). Thus, we have

$$L(f, \lambda_0, \lambda_1) = \frac{1+\alpha}{2} \int \frac{(f(x) - f_0(x))^2}{f_0^{1-\alpha}(x)} \mathrm{d}x + \frac{1+\alpha}{2} \lambda_0 \int \frac{(f(x) - f_1(x))^2}{f_1^{1-\alpha}(x)} \mathrm{d}x + \lambda_1 \int f(x)\mathrm{d}x.$$

Now, differentiating with respect to f, we obtain

$$\frac{\partial}{\partial f} L(f, \lambda_0, \lambda_1) = (1+\alpha) \frac{f(x) - f_0(x)}{f_0^{1-\alpha}(x)} + (1+\alpha)\lambda_0 \frac{f(x) - f_1(x)}{f_1^{1-\alpha}(x)} + \lambda_1. \tag{6.2}$$

Setting Eq. (6.2) to zero, we get the optimal density function to be

$$f(x) = \frac{pf_0^{\alpha}(x) + (1-p)f_1^{\alpha}(x)}{pf_0^{\alpha-1}(x) + (1-p)f_1^{\alpha-1}(x)} \left(\int \frac{pf_0^{\alpha}(x) + (1-p)f_1^{\alpha}(x)}{pf_0^{\alpha-1}(x) + (1-p)f_1^{\alpha-1}(x)} \mathrm{d}x \right)^{-1},$$

where $p = \frac{1}{1+\lambda_0}$, as required. □

Theorem 6.3. *Let f, f_0 and f_1 be three density functions. Then, the solution to the optimization problem,*

$$\min_f \{w\chi^2_\alpha(f_0 : f) + (1-w)\chi^2_\alpha(f_1 : f)\} \quad \textit{subject to} \int f(x)\mathrm{d}x = 1, \quad 0 \le w \le 1, \tag{6.3}$$

is the (p, η)*-mixture density with mixing parameter* $p = w$.

Proof. Making use of the Lagrangian multiplier technique in the same way as in Theorem 6.1, the required result is obtained. □

Theorem 6.4. *Let f, f_0 and f_1 be three density functions and* $T_\alpha(X) = \frac{f(X)}{f_1^{1-\alpha}(X)}$*. Then, the solution to the optimization problem,*

$$\min_f \chi^2_\alpha(f_0 : f) \quad \textit{subject to } E_f(T_\alpha(X)) = \eta, \quad \int f(x)\mathrm{d}x = 1, \tag{6.4}$$

is the (p, η)*-mixture density with mixing parameter* $p = \frac{1}{1+\lambda_0}$ *and* $\lambda_0 > 0$ *is the Lagrangian multiplier.*

Proof. Making use of the Lagrangian multiplier technique in the same way as in Theorem 6.1, the required result is obtained.

Now, we extend Theorem 6.2 to the case of $n+2$ density functions. □

Theorem 6.5. *Let f, $f_0, \ldots, f_n$ be $n+2$ density functions. Then, the solution to the optimization problem,*

$$\min_f \chi^2_\alpha(f_0 : f) \quad \textit{subject to } \chi^2_\alpha(f_i : f) = \eta_i, i = 1, \ldots, n, \quad \int f(x)\mathrm{d}x = 1, \tag{6.5}$$

is the extended (p, η)*-mixture density with* $\eta = \alpha$ *and mixing parameters* $p_i = \frac{\lambda_i}{1+\sum_{i=0}^{n-1}\lambda_i}$ *and* $\lambda_i > 0$, $i = 0, \ldots, n$*, is the Lagrangian multiplier.*

Proof. We use the Lagrangian multiplier technique for finding the solution to the optimization problem in Eq. (6.5). Thus, we have

$$\begin{aligned} L(f, \lambda_0, \ldots, \lambda_n) &= \frac{1+\alpha}{2}\int \frac{(f(x) - f_0(x))^2}{f_0^{1-\alpha}(x)}\mathrm{d}x + \sum_{i=0}^{n-1}\frac{\lambda_i(1+\alpha)}{2}\int \frac{(f(x) - f_{i+1}(x))^2}{f_{i+1}^{1-\alpha}(x)}\mathrm{d}x \\ &\quad + \lambda_n \int f(x)\mathrm{d}x. \end{aligned}$$

Now, differentiating with respect to f, we obtain

$$\frac{\partial}{\partial f}L(f, \lambda_0, \ldots, \lambda_n) = (1+\alpha)\frac{f(x) - f_0}{f_0^{1-\alpha}} + (1+\alpha)\sum_{i=0}^{n-1}\lambda_i\frac{f(x) - f_{i+1}}{f_{i+1}^{1-\alpha}} + \lambda_n. \tag{6.6}$$

Setting Eq. (6.6) to zero, we get the optimal density function to be

$$f(x) = k\frac{(1 - \sum_{i=1}^n p_i)f_0^\alpha(x) + \sum_{i=1}^n p_i f_i^\alpha(x)}{(1 - \sum_{i=1}^n p_i)f_0^{\alpha-1}(x) + \sum_{i=1}^n p_i f_i^{\alpha-1}(x)},$$

where

$$k = \int \left(\frac{(1 - \sum_{i=1}^{n} p_i) f_0^{\alpha}(x) + \sum_{i=1}^{n} p_i f_i^{\alpha}(x)}{(1 - \sum_{i=1}^{n} p_i) f_0^{\alpha-1}(x) + \sum_{i=1}^{n} p_i f_i^{\alpha-1}(x)} \right)^{-1} \mathrm{d}x$$

and $p_i = \frac{\lambda_i}{1+\sum_{i=0}^{n-1} \lambda_i}$, as required. □

7. Relative-χ^2_α divergence measure of mixed reliability systems

Consider a system with component lifetimes $X_1, \ldots, X_n$, which are independent and identically distributed (i.i.d.) with a common lifetime cumulative distribution function (c.d.f.) F and a probability density function (p.d.f.) f. Then, the system lifetime $T = \phi(X_1, \ldots, X_n)$, where ϕ is referred to as the system's structure function, is connected to signature vector $\mathbf{s} = (s_1, \ldots, s_n)$ through

$$s_i = P(T = X_{i:n}) = \frac{n_i}{n!}, \ i = 1, \ldots, n,$$

where $X_{1:n}, \ldots, X_{n:n}$ are the order statistics of component lifetimes and n_i is the number of ways that component lifetimes can be arranged such that $T = \phi(X_1, \ldots, X_n) = X_{i:n}$; for more details, see [20]. Then, the reliability function of T can be expressed as a mixture of reliability functions of $X_{i:n}, i = 1, \ldots, n$, as

$$\bar{F}_T(t) = \sum_{i=1}^{n} s_i \bar{F}_{i:n}(t).$$

Consequently, the corresponding p.d.f. of T is

$$f_T(t) = \sum_{i=1}^{n} s_i f_{i:n}(t), \tag{7.1}$$

where $f_{i:n}$ is the p.d.f. of $X_{i:n}$, given by

$$f_{i:n}(x) = \frac{n!}{(i-1)!(n-i)!} f(x) F^{i-1}(x)(1 - F(x))^{n-i};$$

see [1].

7.1. D_α^ψ measure for order statistics

Suppose $X_1, \ldots, X_n$ are i.i.d. variables from an absolutely continuous c.d.f. F and p.d.f. f, and $X_{1:n}, \ldots, X_{n:n}$ are the corresponding order statistics.

Theorem 7.1. *The D_α^ψ divergence measure between densities $f_{i:n}$ and f is given by*

$$D_\alpha^\psi(f_{i:n} : f) = \frac{1+\alpha}{2} \int_0^1 \frac{f(F^{-1}(u))}{\psi^{1-\alpha}\big(F^{-1}(u)\big)} \big(f_U(u) - f_{U_{i:n}(u)}\big)^2 \mathrm{d}u, \tag{7.2}$$

where the random variables U and $U_{i:n}$ are uniform and Beta$(i, n-i+1)$ random variables on $(0, 1)$ with density functions f_U and $f_{U_{i:n}}$, respectively.

Proof. By using the definition of D_α^ψ divergence measure and the transformation $u = F(x)$, we obtain

$$\begin{aligned} D_\alpha^\psi(f_{i:n} : f) &= \frac{1+\alpha}{2}\int_0^\infty \frac{(f_{i:n}(x) - f(x))^2}{\psi^{1-\alpha}(x)}\mathrm{d}x \\ &= \frac{1+\alpha}{2}\int_0^1 \frac{f(F^{-1}(u))}{\psi^{1-\alpha}(F^{-1}(u))}\left(1 - \frac{n!}{(i-1)!(n-i)!}u^{i-1}(1-u)^{n-i}\right)^2 \mathrm{d}u \\ &= \frac{1+\alpha}{2}\int_0^1 \frac{f(F^{-1}(u))}{\psi^{1-\alpha}(F^{-1}(u))}(f_U(u) - f_{U_{i:n}}(u))^2\mathrm{d}u, \end{aligned}$$

as required. □

Corollary 7.2. *From Theorem 7.1, we readily deduce the following:*

(i) If $\psi(x) = f(x)$, then

$$D_\alpha^\psi(f_{i:n} : f) = \chi_\alpha^2(f_U : f_{U_{i:n}}).$$

(ii) If $\psi(x) = f_{i:n}(x)$, then

$$D_\alpha^\psi(f_{i:n} : f) = \chi_\alpha^2(f_{U_{i:n}} : f_U).$$

From Corollary 7.2, it is immediately seen that under the imposed assumptions, $D_\alpha^\psi(f_{i:n} : f)$ divergence is free of the baseline distribution.

Theorem 7.3. *The D_α^ψ divergence measure between two density functions $f_{i:n}$ and $f_{j:n}$ is given by*

$$D_\alpha^\psi(f_{i:n} : f_{j:n}) = \frac{1+\alpha}{2}\int_0^1 \frac{f(f^{-1}(u))}{\psi^{1-\alpha}(f^{-1}(u))}(f_{U_{i:n}}(u) - f_{U_{j:n}}(u))^2\mathrm{d}u. \tag{7.3}$$

Proof. By using the definition of D_α^ψ divergence measure and the transformation $u = F(x)$ in the same way as in the proof of Theorem 7.1, the required result is obtained.

In the special case when $\psi(x) = f_{i:n}(x)$, we find that

$$D_\alpha^\psi(f_{i:n} : f_{j:n}) = \chi_\alpha^2(f_{U_{i:n}} : f_{U_{j:n}}). \tag{7.4}$$

□

7.2. D_α^ψ measure for mixed systems

In this section, we examine the D_α^ψ divergence measure associated with mixed reliability systems.

Theorem 7.4. *If $\psi(x) = f_{i:n}(x)$, then the $D_\alpha^\psi(f_T : f_{i:n})$ divergence measure is given by*

$$D_\alpha^\psi(f_T : f_{i:n}) = \chi_\alpha^2\left(f_{U_{i:n}} : \sum_{i=1}^n s_i f_{U_{i:n}}\right). \tag{7.5}$$

 https://doi.org/10.1017/S0269964823000189

Proof. From the assumption that $\psi(x) = f_{i:n}(x)$ and the definition of $D^{\psi}_{\alpha}(f_T : f_{i:n})$ measure, and making use of the transformation $u = F(x)$, we have

$$\begin{aligned} D^{\psi}_{\alpha}(f_T : f_{i:n}) &= \frac{1+\alpha}{2}\int_0^{\infty} \frac{\left(\sum_{j=1}^n \alpha_j f_{j:n}(x) - f_{i:n}(x)\right)^2}{f_{i:n}^{1-\alpha}(x)}\mathrm{d}x \\ &= \frac{1+\alpha}{2}\int_0^{1} \frac{\left(\sum_{j=1}^n s_j f_{U_{j:n}}(u) - f_{U_{i:n}}(u)\right)^2}{f_{U_{i:n}}^{1-\alpha}(u)}\mathrm{d}u \\ &= \chi^2_{\alpha}\left(f_{U_{i:n}} : \sum_{i=1}^n s_j f_{U_{j:n}}\right), \end{aligned}$$

as required. □

Theorem 7.5. *Let T_1 and T_2 be the lifetimes of two mixed systems with signatures $\mathbf{s}$ and $\mathbf{s}'$ consisting of n i.i.d. components having common c.d.f. F and p.d.f. f. Then, if $\psi(x) = f(x)$, we have*

$$D^{\psi}_{\alpha}(f_{T_1} : f_{T_2}) = \frac{1+\alpha}{2}\int_0^1 \left(\sum_{i=1}^n f_{U_{i:n}}(u)\left(s_i - s_i'\right)\right)^2 f^{\alpha}(F^{-1}(u))\mathrm{d}u,$$

where $f_{U_{i:n}}(u)$ is the p.d.f. of a beta distribution with parameters i and $n-i+1$.

Proof. From the assumption made and use of the transformation $u = F(x)$, we have

$$\begin{aligned} D^{\psi}_{\alpha}(f_{T_1} : f_{T_2}) &= \frac{1+\alpha}{2}\int_0^{\infty} \frac{\left(\sum_{i=1}^n s_i f_{i:n}(x) - \sum_{i=1}^n s_i' f_{i:n}(x)\right)^2}{f^{1-\alpha}(x)}\mathrm{d}x \\ &= \frac{1+\alpha}{2}\int_0^1 \left(\sum_{i=1}^n s_i f_{U_{i:n}}(u) - \sum_{i=1}^n s_i' f_{U_{i:n}}(u)\right)^2 f^{\alpha}(F^{-1}(u))\mathrm{d}u \\ &= \frac{1+\alpha}{2}\int_0^1 \left(\sum_{i=1}^n f_{U_{i:n}}(u)\left(s_i - s_i'\right)\right)^2 f^{\alpha}(F^{-1}(u))\mathrm{d}u, \end{aligned}$$

as required. □

8. Application to image processing

In this section, we present an application of Jensen-χ^2_{α} measure in the framework of image quality assessment. For pertinent details about image quality assessment, see [10].

Figure 1 shows the original lake image that includes 512×512 cells, and the level of the color gray of each cell assumes a value in the interval $[0, 1]$ (0 for black and 1 for white). It depicts the image labeled as X and three adjusted versions of it labeled as $Y(= X+0.3)$ (increasing brightness), $Z(= \sqrt{2\times X})$ (with increased contrast and gamma correction) and $W(= \sqrt{X})$ (gamma corrected). For pertinent details, see **EBImage** package in **R** software [19].

The extracted histograms with the corresponding empirical densities for images X, Y, Z and W are plotted in Figure 2.

We can see from Figures 1 and 2 that the highest degree of similarity is first related to W and then to Y, whereas Z has the highest degree of divergence from the original image X.

Figure 1. *The original lake image and its three adjusted versions. Image* X *(top-left corner), Image* Y *(top-right corner), Image* Z *(bottom-left corner) and Image* W *(bottom-left corner).*

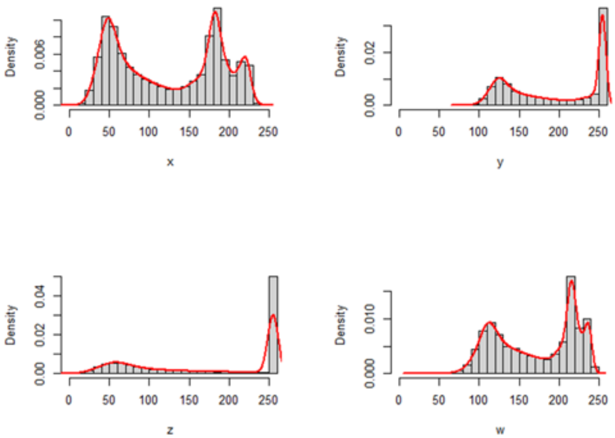

Figure 2. *The histograms and the corresponding empirical densities for lake image (*X*) and its three adjusted versions (*Y, Z *and* W*).*

8.1. Nonparametric estimation of the Jensen-χ^2_α divergence measure

Let f_1, f_2 and ψ be probability density functions. Suppose we draw independent and identically distributed random samples from each of these distributions, obtaining samples of sizes n_1, n_2 and n_ψ, respectively. Denote the resulting samples by $X_1^{(1)}, \ldots, X_{n_1}^{(1)}$ for f_1, $X_1^{(2)}, \ldots, X_{n_2}^{(2)}$ for f_2 and $X_1^{\psi}, \ldots, X_{n_\psi}^{\psi}$ for ψ.

To estimate the underlying probability density functions f_1, f_2 and ψ using kernel density estimation, we can use the following functions:

Let $\hat{f}_1(x)$ be the kernel density estimate of f_1, based on the sample $X_1^{(1)}, \ldots, X_{n_1}^{(1)}$. Then, we have

Table 1. *The Jensen-χ^2_α divergence measure between each pair of adjusted images with respected to the original image for the choices $\alpha = 0.5, 1.5$ and* p = *0.5*

Measure	Jensen-$\chi^2_{0.5}$	Jensen-$\chi^2_{1.5}$
$(Y, Z \parallel X)$	0.016992	0.000058
$(Y, W \parallel X)$	0.0337651	0.000071
$(Z, W \parallel X)$	0.0562018	0.000117

$$\hat{f}_1(x) = \frac{1}{n_1 h_1} \sum_{i=1}^{n} K\left(\frac{x - X_i^{(1)}}{h_1}\right).$$

where $K(\cdot)$ is a kernel function, typically chosen to be a symmetric probability density function, and h_1 is a bandwidth parameter that controls the smoothness of the estimate.

Similarly, let $\hat{f}_2(x)$ be the kernel density estimate for f_2, based on the sample $X_1^{(2)}, \ldots, X_{n_2}^{(2)}$. Then, we can write

$$\hat{f}_2(x) = \frac{1}{n_2 h_2} \sum_{i=1}^{n_2} K\left(\frac{x - X_i^{(2)}}{h_2}\right),$$

where h_2 is a bandwidth parameter for the kernel density estimate of f_2. Finally, let $\hat{\psi}(x)$ be the kernel density estimate for ψ, based on the sample $x_1^{\psi}, \ldots, x_{n_\psi}^{\psi}$. Then, we have

$$\hat{\psi}(x) = \frac{1}{n_\psi h_\psi} \sum_{i=1}^{n_\psi} K\left(\frac{x - X_i^{\psi}}{h_\psi}\right),$$

where h_ψ is a bandwidth parameter for the kernel density estimate of ψ.

For more details, see [8].

Using these estimates based on Gaussian kernel, $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$, we can compute the integrated nonparametric estimate of the Jensen-χ^2_α measure, for $0 < p < 1$, as

$$\widehat{\mathcal{J}^{\psi}_{\alpha}}(f_1, f_2; \mathbf{P}) \quad = \quad p\chi^2_\alpha(\widehat{\psi} : \widehat{f_1}) + (1-p)\chi^2_\alpha(\widehat{\psi} : \widehat{f_2}) - \chi^2_\alpha\left(\widehat{\psi} : p\widehat{f_1} + (1-p)\widehat{f_2}\right).$$

We have computed the Jensen-χ^2_α information measure for each pair of adjusted images with respect to the original lake image, and these are presented in Table 1. The results demonstrate that the Jensen-χ^2_α divergence is an effective measure of similarity between each pair of adjusted images and the reference original image. Specifically, the Jensen-χ^2_α divergence highlights the high degree of similarity between images Y and Z with respect to the original image (X). Furthermore, the results in Table 1 indicate that the comparison of images Z and W with respect to the reference image X results in low similarity. Therefore, the Jensen-χ^2_α information measure can be considered as an efficient criteria for comparing the similarity between each pair of adjusted images with respect to the reference image.

9. Concluding remarks

In this paper, by considering the χ^2_α divergence measure, we have proposed relative-χ^2_α, Jensen-χ^2_α and (p, w)-Jensen-χ^2_α divergence measures. We have first shown that the χ^2_α divergence measure has a close relationship with q-Fisher information of mixing parameter of an arithmetic mixture distribution. We have then shown that the proposed relative-χ^2_α divergence measure includes some other well-known versions of chi-square divergence such as the usual chi-square (χ^2), generalized-χ^2 (χ^2_α), triangular and Balakrishnan–Sanghavi divergence measures all as special cases. We have shown that the Jensen-χ^2_α divergence is a mixture of relative-χ^2_α divergence measures. A lower bound for Jensen-χ^2_α divergence has been obtained in terms of Jensen–Shannon entropy measure. We have also introduced (p, w)-Jensen-χ^2_α divergence measure and have then established some of its properties. Further, we have studied the relative-χ^2_α divergence measure of escort and arithmetic mixture densities. Next, we have introduced (p, η)-mixture density, which includes arithmetic-mixture and harmonic-mixture densities as special cases. Interestingly, we have shown that the proposed mixture density possesses optimal information under three different optimization problems associated with the χ^2_α divergence measure. We have also provided a discussion about the relative-χ^2_α divergence measure of order statistics and mixed reliability systems. Finally, we have described an application of the Jensen-χ^2_α measure in image processing.

In summary, in this paper, some extensions of the chi-square divergence measure such as the relative-χ^2_α, D^ψ_α, Jensen-χ^2_α and (p, w)-Jensen-χ^2_α divergence measures have been proposed. Particularly, it has been shown that the relative-χ^2_α divergence measure includes the well-known divergence measures, such as L_2, χ^2, triangular, symmetric χ^2, χ^2_α and Balakrishnan–Sanghvi divergence measures, all as special cases, and provides a flexible and powerful divergence measure for comparing probability distributions in a wide rage of problems. The choice of α and the weight function $\psi(x)$ can be tailored to suit the specific characteristics and the features of the data for the models that are being compared, allowing for greater sensitivity and flexibility in the comparison process.

Furthermore, the proposed Jensen-χ^2_α and (p, w)-Jensen-χ^2_α divergence measures are extensions of $D^\psi_\alpha(f : g)$ that are based on a convex combination. These extensions allow for the incorporation of additional divergence measures into the framework, further increasing the flexibility and applicability of the method.

There are, of course, several areas of the proposed information measures that require more study with regard to its theoretical as well as experimental analysis. Additionally, with the incorporation of the idea of relative-χ^2_α divergence and Jensen-χ^2_α divergence measures, there is an opportunity to broaden and explore the discrete and cumulative versions of the established divergence measures, utilizing the properties of convexity or concavity. It will also be of great interest to study cumulative versions of these measures, and we plan to do this in our future work. Finally, there is also a potential to extend the idea to relative Fisher information measure. We are currently working on these problems and hope to report the findings in a future paper.

Acknowledgements. The authors express their sincere thanks to the Editor and the anonymous reviewers for their useful comments and suggestion on the earlier version of this manuscript, which resulted in this much improved version.

Competing interests. On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] Arnold, B.C., Balakrishnan, N., & Nagaraja, H.N. (1992). *A first course in order statistics*. New York: Wiley.
[2] Asadi, M., Ebrahimi, N., Kharazmi, O., & Soofi, E.S. (2018). Mixture models, Bayes Fisher information, and divergence measures. *IEEE Transactions on Information Theory* 65(4): 2316–2321.
[3] Asadi, M., Ebrahimi, N., Soofi, E.S., & Zohrevand, Y. (2016). Jensen–Shannon information of the coherent system lifetime. *Reliability Engineering & System Safety* 156: 244–255.
[4] Balakrishnan, V. & Sanghvi, L.D. (1968). Distance between populations on the basis of attribute data. *Biometrics* 859–865.
[5] Basu, A., Harris, I.R., Hjort, N.L., & Jones, M.C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika* 85: 549–559.
[6] Bercher, J.F. (2009). Source coding with escort distributions and Rényi entropy bounds. *Physics Letters A* 373: 3235–3238.
[7] Cover, T.M. & Thomas, J.A. (2006). *Elements of information theory*. Hoboken, New Jersey: Wiley.
[8] Duong, T., Duong, M.T., & Suggests, M.A.S.S. (2022). Package 'ks'. *R Package Version* 1(5).

[9] Furuichi, S. (2009). On the maximum entropy principle and the minimization of the Fisher information in Tsallis statistics. *Journal of Mathematical Physics* 50: 013303.
[10] Gonzalez, R.C. (2009). *Digital image processing*. New York: Prentice-Hall.
[11] Kharazmi, O. & Balakrishnan, N. (2021). Cumulative residual and relative cumulative residual Fisher information and their properties. *IEEE Transactions on Information Theory* 67: 6306–6312.
[12] Le Cam, L. (1986). *Asymptotic methods in statistical decision theory*. New York: Springer.
[13] Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory* 37: 145–151.
[14] Lutwak, E., Lv, S., Yang, D., & Zhang, G. (2012). Extensions of Fisher information and Stam's inequality. *IEEE Transactions on Information Theory* 58: 1319–1327.
[15] Masi, M. (2005). A step beyond Tsallis and Rényi entropies. *Physics Letters A* 338: 217–224.
[16] 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.
[17] Nielsen, F. (2020). On a generalization of the Jensen–Shannon divergence and the Jensen-Shannon centroid. *Entropy* 22: 221.
[18] Nielsen, F. & Nock, R. (2013). On the chi square and higher-order chi distances for approximating f-divergences. *IEEE Signal Processing Letters* 21: 10–13.
[19] Oles, A., Pau, G., Smith, M., Sklyar, O., & Hube, W. (2020). Image processing and analysis toolbox for R. *R Package Version* 4: 1–52.
[20] Samaniego, F.J. (2007). *System signatures and their applications in engineering reliability*. New York: Springer.
[21] Steele, J.M. (2004). *The Cauchy-Schwarz master class: An introduction to the art of mathematical inequalities*. Cambridge, England: Cambridge University Press.
[22] Wu, Q., Deng, Y. & Xiong, N. (2022). Exponential negation of a probability distribution. *Soft Computing* 26: 2147–2156.
[23] Yamano, T. (2002). Some properties of q-logarithm and q-exponential functions in Tsallis statistics. *Physica A: Statistical Mechanics and its Applications* 305: 486–496.

Cite this article: Kharazmi O. and Balakrishnan N. (2024). On Jensen-χ^2_α divergence measure. *Probability in the Engineering and Informational Sciences* 38(2): 403–427. https://doi.org/10.1017/S0269964823000189