Filomat 35:13 (2021), 4287–4303
https://doi.org/10.2298/FIL2113287K

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

Informational Properties of Transmuted Distributions

Omid Kharazmi[a], Narayanaswamy Balakrishnan[b]

[a]*Department of Statistics, Faculty of Mathematical Sciences, Vali-e-Asr University of Rafsanjan, Iran*
[b]*Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada*

Abstract. In this paper, we establish some informational properties of transmuted distributions. Specifically, we derive Shannon entropy, Gini's mean difference, and Fisher and Bayes Fisher information measures of a transmuted distribution. Two extensions of Shannon entropy and Gini's mean difference information measures are also provided. Finally, the distances between transmuted distribution and its components based on the Kullback-Leibler, chi-square and energy distance divergences are all derived.

1. The first section

In information theory, there are different criteria to measure the uncertainty of a probabilistic model. Moreover, various divergences measures have been developed in the literature for measuring similarity (closeness) between two probability distributions. Shannon entropy, Fisher information and Gini's mean difference are three most important information measures that have been used in many different fields. For more details, see Shannon (1948), Fisher (1929) and Gini (1912). More recently, these information measures have been generalized based on Jensen inequality, which have come to be knownas Jensen-Shannon, Jensen-Fisher and Jensen-Gini information measures, respectively. For pertinent details, one may see Lin (1991), Sánchez-Moreno et al. (2016) and Mehrali et al. (2018).

Recently, considerable work in distribution theory has focused on the family of transmuted distributions derived through a quadratic rank transmutation map. Although the transmuted distributions are of particular importance in modeling various data sources, these models have not been studied from information theory viewpoint so far.

With this in mind, our main interest here is to establish some informational properties, including Shannon entropy, Fisher information, Bayes Fisher information and Gini's mean difference for transmuted distributions. These models get formulated by transforming a parent distribution into its generalized counterpart. Transmuted distribution was introduced by Shaw and Buckley (2009) in an unpublished report and applied to uniform, exponential and normal distributions. During the past decade, considerable efforts have been directed at developing more flexible distributions with the use of this construct. Many standard probability distributions have been generalized and developed into some flexible models in this manner; see, for example, Sarabia et al. (2020) and the references therein. Kozubowski and Podgórski

2020 *Mathematics Subject Classification*. Primary 62B10, 81P45
Keywords. Shannon entropy, Fisher information, Bayes Fisher information, Transmuted distribution, Kullback-Leibler divergence, Chi-square divergence, Energy distance divergence.
Received: 27 September 2020; Revised: 11 January 2021; Accepted: 22 April 2021
Communicated by Biljana Popović
Corresponding author: Omid Kharazmi
Email addresses: omid.kharazmi@vru.ac.ir (Omid Kharazmi), bala@mcmaster.ca (Narayanaswamy Balakrishnan)

(2016) have provided an interesting link between transmutation distribution and random extrema concept (maxima (or minima)) of a random number N of independent and identically distributed variables with the base distribution, where N has a Bernoulli distribution shifted by one. In fact, this approach provides a theoretical interpretation of the construction through transmuted mapping. Recently, Balakrishnan and He (2021) have proposed analogous transmuted distributions based on the theory of records and examined their hazard properties.

In the framework of transmutation map, from a given base cumulative distribution function (*CDF*) F, a new probability *CDF* F_{X_T} is defined as

$$F_{X_T}(x) = (1+\lambda)F(x) - \lambda F^2(x),\ x \in \mathcal{R},\ |\lambda| \leq 1. \tag{1}$$

Model (1) is referred to as a transmuted distribution. For a stochastic representation of (1), let X_1 and X_2 be a random sample of size two from the absolutely continuous *CDF* F with probability density function (*PDF*) f. The density functions associated with extreme random variables $min(X_1, X_2)$ and $max(X_1, X_2)$ are known to be $f_{min}(x) = 2f(x)(1 - F(x))$ and $f_{max}(x) = 2f(x)F(x)$, respectively. Moreover, the corresponding *CDFs* are given by $F_{min}(x) = 1 - (1 - F(x))^2$ and $F_{max}(x) = F^2(x)$, respectively; see, for example, Arnold et al. (1992). Hence, we can rewrite the *CDF* in (1) based on the components F and F_{max}, as follows:

$$\begin{aligned} F_{X_T}(x) &= (1+\lambda)F(x) - \lambda F^2(x) \\ &= (1+\lambda)F(x) - \lambda F_{max}(x). \end{aligned} \tag{2}$$

Further, the corresponding *PDF* is given by

$$\begin{aligned} f_{X_T}(x) &= (1+\lambda)f(x) - 2\lambda f(x)F(x) \\ &= (1+\lambda)f(x) - \lambda f_{max}(x). \end{aligned} \tag{3}$$

It is worthwhile to note that (2) and (3) can also be stated based on the variable $min(X_1, X_2)$ as

$$F_{X_T}(x) = (1-\lambda)F(x) + \lambda F_{min}(x)$$

and

$$f_{X_T}(x) = (1-\lambda)f(x) + \lambda f_{min}(x),$$

respectively. In order to avoid repetitive results within the present work, we only consider the transmuted distribution in (2).

If the baseline distribution in (3) is considered as Uniform on $(0, 1)$, then the density function of transmuted uniform variable U_T is given by

$$f_{U_T}(u) = (1+\lambda) - 2\lambda u, \quad u \in (0, 1). \tag{4}$$

We now briefly introduce some informational measures that will be used in the sequel. Let X be an absolutely continuous random variable with *CDF* F and density function f. Then, the Shannon entropy of X (or density f) is defined as

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

where "log" stands for the natural logarithm. For more details, see the pioneering paper of Shannon (1948). To simplify notation, we suppress $\mathcal{X}$ for integration with respect to dx throughout the paper, unless a distinction becomes necessary.

Kullback-Leibler (*KL*) discrimination information is introduced next. Let X and Y be two continuous random variables with absolutely continuous density functions f and g, respectively. Then, the Kullback-Leibler distance between X and Y (or f and g) is defined as

$$KL(X||Y) = KL(f, g) = \int f(x) \log \frac{f(x)}{g(x)} dx.$$

The Kullback-Leibler discrimination between Y and X can be defined analogously. For more details, see Kullback and Leibler (1951).

Another important diversity measure between two density functions f and g is the chi-square divergence, defined as

$$\chi^2(f, g) = \int \frac{(f(x) - g(x))^2}{g(x)} dx.$$

In the same way, we can define $\chi^2(g, f)$. For more details, see Nielsen and Nock (2013).

The Fisher information of a random variable X, or its PDF $f(x;\theta)$, about the parameter θ is defined as

$$\mathcal{I}(\theta) = \int \left[\frac{\partial \log f(x;\theta)}{\partial \theta}\right]^2 f(x;\theta)dx.$$

It is assumed that θ lies in an open interval in the real line and that $f(x;\theta) > 0$ for all values of θ in the parameter space and is differentiable with respect to θ. In Bayesian statistics, it is assumed that the parameter θ is endowed with a prior $\pi(\theta)$. Then, the expected Fisher information $\tilde{\mathcal{I}}(\theta) = E_\pi[\mathcal{I}(\Theta)]$ is called *Bayes Fisher information*. For more details, see Asadi et al. (2018) and the references therein.

Gini's mean difference (*GMD*) and energy distance are two other prominent information measures that will be considered in this work. Let X be a continuous random variable with distribution function F. Then, the *GMD* associated with X is defined as

$$GMD(F) = 2\int F(x)\bar{F}(x)dx,$$

where $\bar{F}(x) = 1 - F(x)$ denotes the survival function of X. Yitzhaki (2003) stated that the *GMD*, as a measure of variability, shares many properties of the variance of X, and is more suitable for distributions that are far from normality. Moreover, the energy distance (L^2) between two CDFs F_1 and F_2 is given by

$$D(F_1, F_2) = \int (F_1(x) - F_2(x))^2 dx.$$

For more details, see Mehrali et al. (2018).

In this paper, we establish some important informational properties of transmuted distributions. In Section 2, we study the Shannon entropy of transmuted density functions and also consider the Kullback-Libeler divergence between transmuted distribution and the densities of each of its components. We then propose transmuted Shannon entropy, which is an extension of Shannon entropy of transmuted distribution, as well as Jensen-Shannon entropy. It is shown that the *KL* divergence between a general transmuted model and its components is free of the parent distribution. The proposed transmuted Shannon entropy is further expressed based on the transmuted structure of two *KL* divergences of the general model and its components. In Section 3, by considering the transmuted distribution, we obtain Gini's mean difference and also discuss the energy divergence between transmuted distribution and its components. The transmuted Gini's mean difference is also proposed and it is shown that this information measure can be expressed in terms of a quadratic rank structure of two energy distances of the transmuted distribution and its components. Section 4 discusses the chi-square divergence between the general transmuted density function and the density functions of each of its components. In Section 5, the Fisher information for parameter λ is derived. It is shown that the Fisher information of the transmuted model about parameter λ is connected to chi-square divergence. A detailed discussion about Bayes Fisher information for the parameter λ of transmuted distribution under different prior distributions is also discussed in this section. It is specifically shown that the Bayes Fisher information measures for transmuted model under uniform prior are the same as Jeffreys' divergence. Another result shows that, under triangular and Beta priors, the measures for these two models are different Kullback-Leibler based measures. Finally, Section 6 presents some concluding remarks.

2. Entropic measures

In this section, we first derive the Shannon entropy of the transmuted random variable X_T with *PDF* as in (3), and then examine the Kullback-Leibler divergence between the transmuted distribution and each of its components. We further propose an extension of the Shannon entropy that is suitable for transmuted distributions.

2.1. Shannon entropy

Suppose random variables X_1 and X_2 are independent and identically distributed as X with common *PDF* f and *CDF* F. We now derive the Shannon entropy of the transmuted random variable X_T.

Theorem 2.1. *For the transmuted random variable X_T with PDF as in (3), the Shannon entropy is given by*

$$H(X_T) = (1+\lambda)H(X) - \lambda H(max(X_1, X_2)) + \lambda H(V) + H(U_T), \tag{5}$$

where the random variable V has $Beta(2,1)$ distribution with PDF

$$f_V(v) = 2v, \qquad v \in (0,1),$$

and the random variable U_T is the transmuted uniform random variable on $(0,1)$ with its PDF as in (4).

Proof: By the definition of Shannon entropy, we have

$$\begin{aligned}
H(X_T) &= -\int_{-\infty}^{\infty} f_{X_T}(x)\log f_{X_T}(x)dx \\
&= -\int_{-\infty}^{\infty} \Big\{(1+\lambda)f(x) - 2\lambda f(x)F(x)\Big\}\log f(x)dx \\
&\quad -\int_{-\infty}^{\infty} \Big\{(1+\lambda)f(x) - 2\lambda f(x)F(x)\Big\}\log\Big\{(1+\lambda) - 2\lambda F(x)\Big\}dx \\
&= (1+\lambda)H(X) - \lambda H(max(X_1, X_2)) - \lambda\int_0^1 2v\log(2v)dv \\
&\quad -\int_0^1 \Big\{(1+\lambda) - 2\lambda u\Big\}\log\Big\{(1+\lambda) - 2\lambda u\Big\}du \\
&= (1+\lambda)H(X) - \lambda H(max(X_1, X_2)) + \lambda H(V) + H(U_T),
\end{aligned}$$

as required.

If we now define $\phi(\lambda) = \lambda H(V) + H(U_T)$, then by performing some algebraic manipulations, we can obtain

$$\phi(\lambda) = -\lambda\Big(\log 2 - \frac{1}{2}\Big) + \frac{2\lambda + (\lambda-1)^2\log(1-\lambda) - (\lambda+1)^2\log(\lambda+1)}{4\lambda}. \tag{6}$$

In order to examine the shape of the function $\phi(\lambda)$, we get its first derivative to be

$$\begin{aligned}
\frac{\partial\phi(\lambda)}{\partial\lambda} &= \frac{1-\lambda^2}{4\lambda}\log\Big(\frac{1+\lambda}{1-\lambda}\Big) - \frac{0.5+0.1931\lambda}{\lambda} \\
&= \frac{1-\lambda^2}{2\lambda}\tanh^{-1}(\lambda) - \frac{0.5+0.1931\lambda}{\lambda}.
\end{aligned} \tag{7}$$

The root of (7) is found to be $\lambda = -0.543$. Further, the second derivative of $\phi(\lambda)$ is also obtained to be

$$\frac{\partial^2\phi(\lambda)}{\partial\lambda^2} = \frac{1}{\lambda^2}\Big(1 - \frac{\tanh^{-1}(\lambda)}{\lambda}\Big).$$

Because $\tanh^{-1}(\lambda) \ge \lambda$ for λ in $(0,1)$ and $\tanh^{-1}(\lambda) \le \lambda$ for λ in $(-1,0)$, we always have $\frac{\partial^2\phi(\lambda)}{\partial\lambda^2} \le 0$. A plot of the function $\phi(\lambda)$, for λ in the interval $[-1,1]$, is presented in Figure 1. From Figure 1, it easy to see that for $\lambda \in [-1,0]$, the function $\phi(\lambda)$ is positive, while for $\lambda \in (0,1]$, $\phi(\lambda)$ is negative.

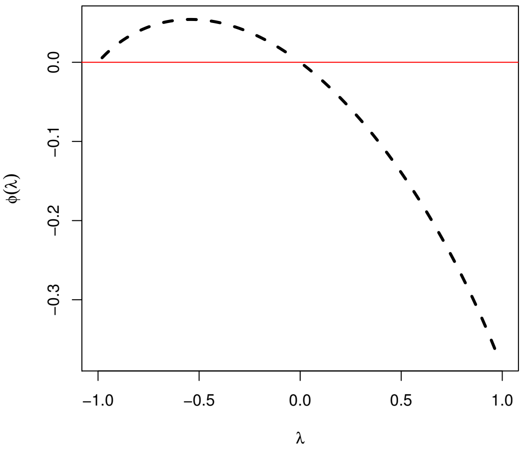

Fig 1. Plot of the function $\phi(\lambda)$ for values of λ in $[-1, 1]$.

Example 2.2. *Suppose X_1 and X_2 are independent and identically distributed as an exponential variable X, with PDF $f(x) = \alpha e^{-\alpha x}$, $x > 0$. Then, the Shannon entropy of the corresponding transmuted random variable X_T is given by*

$$\begin{aligned} H(X_T) &= (1+\lambda)H(X) - \lambda H(max(X_1, X_2)) + \lambda H(V) + H(U_T) \\ &= (1+\lambda)(1-\log\alpha) + \lambda\Big\{\log(2\alpha) - 2\Big\} + \phi(\lambda), \end{aligned}$$

where $\phi(\lambda)$ is as given in (6).

2.2. Kullback-Leibler divergence

In this subsection, we present some results concerning Kullbak-Leibler divergence between the transmuted density f_{X_T} and the density functions of each of its components, f and f_{max}. For this purpose, we first derive in the following theorem the Kullback-Leibler divergence between the transmuted uniform random variable and the uniform random variable in the interval $(0, 1)$.

Theorem 2.3. *Let variable X have its PDF as $f(x)$ and the corresponding transmuted random variable X_T have its PDF as in (3). Then, the Kullbak-Leibler divergences between f and f_T are given, respectively, by*

$$(i)\ \ KL(f_{X_T}, f) = KL(f_{U_T}, f_U),$$

$$(ii)\ \ KL(f, f_{X_T}) = KL(f_U, f_{U_T}),$$

where the random variables U and U_T are uniform and transmuted uniform random variables on (0,1) with density functions f_U and f_{U_T}, respectively.

Proof: By using the definition of Kullback-Leibler divergence and setting $u = F(x)$, we have

$$\begin{aligned} KL(f_{X_T}, f) &= \int_{-\infty}^{\infty} f_{X_T}(x)\log\Big\{\frac{f_{X_T}(x)}{f(x)}\Big\}dx \\ &= \int_{-\infty}^{\infty} f_{X_T}(x)\log\Big\{\frac{f(x)(1+\lambda-2\lambda F(x))}{f(x)}\Big\}dx \\ &= \int_0^1 (1+\lambda-2\lambda u)\log(1+\lambda-2\lambda u)du \\ &= KL(f_{U_T}, f_U), \end{aligned}$$

as required for Part (i). Part (ii) can be proved similarly.

Theorem 2.4. *Let X_1 and X_2 be independent and identically distributed as X with PDF f. Then, the Kullbak-Leibler divergences between f_{max} and f_{X_T} are given by*

(i) $KL(f_{X_T}, f_{max(X_1,X_2)}) = KL(f_{U_T}, f_V)$,

(ii) $KL(f_{max(X_1,X_2)}, f_{X_T}) = KL(f_V, f_{U_T})$,

where the random variables V and U_T are as defined earlier in Theorem 2.1.

Proof: By using the definition of Kullback-Leibler divergence and using the transformation $u = F(x)$, we get

$$\begin{aligned} KL(f_{X_T}, f_{max}) &= \int_{-\infty}^{\infty} f_{X_T}(x) \log\left\{\frac{f_{X_T}(x)}{f_{max}(x)}\right\} dx \\ &= \int_{-\infty}^{\infty} f_{X_T}(x) \log\left\{\frac{f(x)(1+\lambda-2\lambda F(x))}{2f(x)F(x)}\right\} dx \\ &= \int_0^1 (1+\lambda-2\lambda u) \log\frac{1+\lambda-2\lambda u}{2u} du \\ &= KL(f_{U_T}, f_V), \end{aligned}$$

as in Part (i). Part (ii) follows similarly.

Remark 2.5. *From Theorems 2.3 and 2.4, it is clear that the KL distances between f_{X_T} and each of its components f and f_{max} are free of the underlying distribution F.*

Remark 2.6. *From Theorems 2.3 and 2.4, Jeffreys' divergences $J(f, f_{X_T})$ and $J(f_{max}, f_{X_T})$ can be readily obtained as*

$$J(f, f_{X_T}) = KL(f_U, f_{U_T}) + KL(f_{U_T}, f_U) = J(f_U, f_{U_T})$$

and

$$J(f_{max}, f_{X_T}) = KL(f_V, f_{U_T}) + KL(f_{U_T}, f_V) = J(f_V, f_{U_T}),$$

respectively. Thus, Jeffreys' divergences are also free of distribution F.

With regard to the results in Theorems 2 and 3, it should be noted that these results can also be presented in terms of the invariant properties of Kullback-Leibler divergence under invertible transformations. For more details, see Qiao and Minematsu (2010). Making use of this property of KL divergence and considering the transformations $X \stackrel{d}{=} F^{-1}(U)$, $X_T \stackrel{d}{=} F^{-1}(U_T)$ and $max(X_1, X_2) \stackrel{d}{=} F^{-1}(V)$, the results of Theorems 2 and 3 can be obtained.

Remark 2.7. *Let the variables U, V and U_T be distributed as uniform, $Beta(2, 1)$ and transmuted uniform on the interval $(0, 1)$, respectively. Then:*

(i) $KL(f_U, f_{U_T}) = \frac{2\lambda+(1-\lambda)\log(1-\lambda)-(\lambda+1)\log(\lambda+1)}{2\lambda}$,

(ii) $KL(f_{U_T}, f_U) = \frac{-2\lambda-(\lambda-1)^2\log(1-\lambda)+(\lambda+1)^2\log(\lambda+1)}{4\lambda}$,

(iii) $KL(f_V, f_{U_T}) = \log 2 - \frac{1}{2} - \frac{(3\lambda^2-2\lambda-1)\log(1-\lambda)-2\lambda(2\lambda+1)+(1+\lambda)^2\log(1+\lambda)}{4\lambda^2}$,

(iv) $KL(f_{U_T}, f_V) = \frac{-2\lambda-(\lambda-1)^2\log(1-\lambda)+(\lambda+1)^2\log(\lambda+1)}{4\lambda} + \frac{\lambda}{2} + 1 - \log 2$.

Proof: By the definition of Kullback-Leibler divergence, we have

$$\begin{aligned} KL(f_U, f_{U_T}) &= -\int_0^1 \log(1+\lambda-2\lambda u) du \\ &= \frac{2\lambda+(1-\lambda)\log(1-\lambda)-(\lambda+1)\log(\lambda+1)}{2\lambda}, \end{aligned}$$

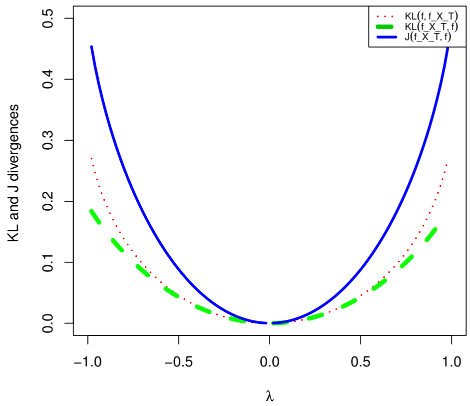

Fig 2. Plot of the divergence measures $KL(f, f_{X_T})$, $KL(f_{X_T}, f)$ and $J(f_{X_T}, f)$ for λ in $[-1, 1]$.

which yields Part (i). Parts (ii), (iii) and (iv) can be derived in a similar manner.

The divergence measures $KL(f, f_{X_T})$, $KL(f_{X_T}, f)$ and $J(f_{X_T}, f)$ are all plotted in Figure 2. From Figure 2, we observe that, for $\lambda \in [-1, 1]$,

$$KL(f_{X_T}, f) \leq KL(f, f_{X_T}) \leq J(f_{X_T}, f).$$

It is possible to prove formally the above inequalities. As the arguments are quite similar for all cases, we shall demonstrate it just for $KL(f_{X_T}, f) \leq KL(f, f_{X_T})$ in the case when $\lambda \in [0, 1]$. Then, from the corresponding expressions from Lemma 2.7, it is equivalent to showing that

$$6\lambda + (\lambda^2 - 4\lambda + 3)\log(1-\lambda) - (\lambda^2 + 4\lambda + 3)\log(1+\lambda) \geq 0$$

for $\lambda \in [0, 1]$. Clearly, the left-hand side is 0 when $\lambda = 0$. Moreover, its derivative with respect to λ, namely, $(2\lambda - 4)\log(1-\lambda) - (2\lambda + 4)\log(1+\lambda)$, is clearly positive, meaning it is an increasing function of λ, thus proving its non-negativity.

Next, let us consider a general mixture distribution with *PDF* $f_m(x) = pf(x) + 2(1-p)f(x)F(x)$, $p \in [0, 1]$, and the transmuted density f_{X_T} in (3). We then derive the *KL* divergence between these two densities, generalizing Theorems 2.3 and 2.4.

Example 2.8. *The KL divergence between the mixture PDF f_m and the transmuted PDF f_{X_T} in (3) is*

$$\begin{aligned}
KL(f_m, f_{X_T}) &= \int (pf(x) + 2(1-p)f(x)F(x)) \log\left\{\frac{pf(x) + 2(1-p)f(x)F(x)}{(1+\lambda)f(x) - 2\lambda f(x)F(x)}\right\} dx \\
&= \int_0^1 (p + 2(1-p)u) \log\left\{\frac{p + 2(1-p)u}{(1+\lambda) - 2\lambda u}\right\} du \\
&= p\frac{(2 - 2p + (p-2)\log(2-p) + p\log p)}{p-1} \\
&\quad + 2(1-p)\frac{-4p^2 + p^2\log p + 6p + (p-2)(3p-2)\log(2-p) - 2}{8(p-1)^2} \\
&\quad - p\frac{-2\lambda + (\lambda-1)\log(1-\lambda) + (\lambda+1)\log(1+\lambda)}{2\lambda} \\
&\quad - 2(1-p)\frac{(3\lambda^2 - 2\lambda - 1)\log(1-\lambda) - 2\lambda(2\lambda+1) + (\lambda+1)^2\log(\lambda+1)}{8\lambda^2}.
\end{aligned}$$

Remark 2.9. *The above result shows that the KL distance between f_m and f_{X_T} is also free of the underlying distribution F.*

A 3D-plot of $KL(f_m, f_{X_T})$ is presented in Figure 3, from which we observe that the KL divergence gets minimized when $p = 1$ (i.e., for the parent density f) and gets maximized when $p = 0$ (i.e., for the density f_{max}).

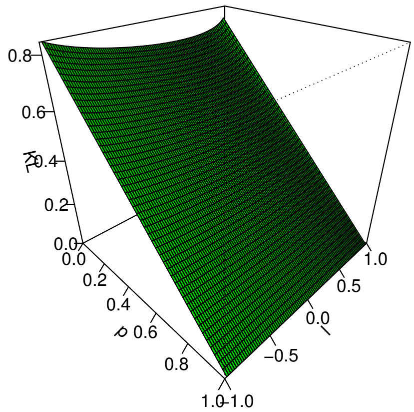

Fig 3. 3D-plot of the KL divergence between the mixture PDF f_m and the transmuted PDF f_{X_T}.

2.3. *Transmuted Shannon entropy*

We now present an extension of Shannon entropy of transmuted distribution as well as Jensen-Shannon divergence, and we refer to it as transmuted Shannon (TS) entropy.

Definition 2.10. *Let X be a continuous random variable with CDF F, and the corresponding transmuted random variable X_T have its PDF as in (3). Then, the transmuted Shannon entropy between components f and f_{max} is defined as*

$$TS(f, f_{max}; \lambda) \quad = \quad H(f_{X_T}) - \left[(1+\lambda)H(f) - \lambda H(f_{max})\right], \ \lambda \in [-1, 1].$$

Remark 2.11. *The TS entropy is given by the formula*

$$TS(f, f_{max}; \lambda) \quad = \quad (1+\lambda)KL(f, f_{X_T}) - \lambda KL(f_{max}, f_{X_T}).$$

Proof: Upon using Theorem 2.1, it is sufficient to show that

$$(1+\lambda)KL(f, f_{X_T}) - \lambda KL(f_{max}, f_{X_T}) = \lambda H(V) + H(U_T).$$

From Theorems 2.3 and 2.4, we have

$$KL(f, f_{X_T}) = KL(f_U, f_{U_T})$$

and

$$KL(f_{max}, f_{X_T}) = KL(f_V, f_{U_T}),$$

where U, V and U_T are as defined in Lemma 2.7. Thus, we have

$$(1+\lambda)KL(f, f_{X_T}) - \lambda KL(f_{max}, f_{X_T}) = (1+\lambda)KL(f_U, f_{U_T}) - \lambda KL(f_V, f_{U_T}),$$

from which we obtain

$$\begin{aligned}(1+\lambda)KL(f_U, f_{U_T}) - \lambda KL(f_V, f_{U_T}) &= -(1+\lambda)\int_0^1 \log\Big\{(1+\lambda) - 2\lambda u\Big\}du \\ &\quad -\lambda\int_0^1 2u\log\Big\{\frac{2u}{(1+\lambda)-2\lambda u}\Big\}du \\ &= -\lambda\int_0^1 2v\log 2v dv \\ &\quad -\int_0^1 \Big\{(1+\lambda) - 2\lambda u\Big\}\log\Big\{(1+\lambda) - 2\lambda u\Big\}du \\ &= \lambda H(V) + H(U_T),\end{aligned}$$

as required.

Remark 2.12. *It is easily seen that Jensen-Shannon entropy between f and f_{max} is a special case of TS entropy when $\lambda \in [-1, 0]$ and using the re-parametrization $p = 1 + \lambda$, that is,*

$$TS(f, f_{max}; \lambda) = JS(f, f_{max}; p),$$

where $JS(f, f_{max}; p) = H\Big(pf + (1-p)f_{max}\Big) - \Big\{pH(f) + (1-p)H(f_{max})\Big\}$.

Remark 2.13. *From Theorems 2.3 and 2.4 and Lemma 2.7, we see that the TS entropy is free of the underlying distribution F as well, and the exact value of TS is given by*

$$\begin{aligned}TS(f, f_{max}; \lambda) &= (1+\lambda)KL(f_U, f_{U_T}) - \lambda KL(f_V, f_{U_T}) \\ &= (1+\lambda)\frac{2\lambda + (1-\lambda)\log(1-\lambda) - (\lambda+1)\log(\lambda+1)}{2\lambda} \\ &\quad + \frac{(3\lambda^2 - 2\lambda - 1)\log(1-\lambda) - 2\lambda(2\lambda+1) + (1+\lambda)^2\log(1+\lambda)}{4\lambda} \\ &\quad -\lambda\log 2 + \frac{\lambda}{2}. \qquad (8)\end{aligned}$$

Remark 2.14. *The TS entropy is concave with respect to parameter λ.*

Proof: We need to show that $\frac{\partial^2 TS(f, f_{max}; \lambda)}{\partial\lambda^2} \leq 0$. The second derivative of $TS(f, f_{max}; \lambda)$ is obtained from (8) to be

$$\frac{\partial^2 TS(f, f_{max}; \lambda)}{\partial\lambda^2} = \frac{2\lambda + \log(1-\lambda) - \log(1+\lambda)}{2\lambda^3}.$$

Now, upon using the fact that $\tanh^{-1}(\lambda) = \frac{1}{2}\log\left\{\frac{1+\lambda}{1-\lambda}\right\}$ in the above equation, we find

$$\begin{aligned}\frac{\partial^2 TS(f, f_{max}; \lambda)}{\partial \lambda^2} &= \frac{1}{\lambda^2} - \frac{\tanh^{-1}(\lambda)}{\lambda^3}\\ &= \frac{1}{\lambda^2}\Big(1 - \frac{\tanh^{-1}(\lambda)}{\lambda}\Big).\end{aligned}$$

Because $\tanh^{-1}(\lambda) \geq \lambda$ for λ in $(0, 1)$ and $\tanh^{-1}(\lambda) \leq \lambda$ for λ in $(-1, 0)$, we obtain

$$\frac{\partial^2 TS(f, f_{max}; \lambda)}{\partial \lambda^2} = \frac{1}{\lambda^2}\Big(1 - \frac{\tanh^{-1}(\lambda)}{\lambda}\Big) \leq 0,$$

as required.

Figure 4 plots the TS entropy as a function of the mixture parameter λ, and from it we observe that the TS entropy is indeed a concave function, supporting Lemma 2.14.

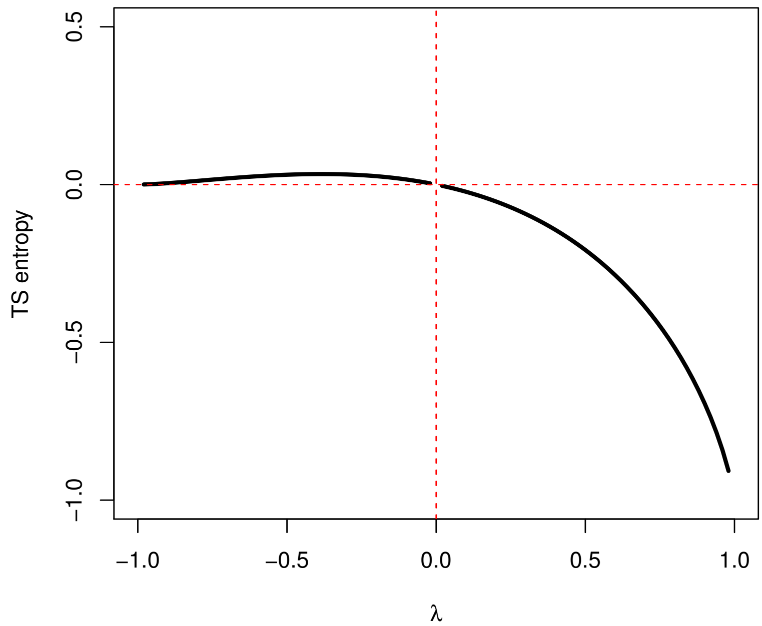

Fig 4. Plot of the TS entropy as a function of λ.

3. Connection with Gini information

Let X be a continuous random variable with CDF F. Then, Gini's mean difference (GMD) associated with X is defined as

$$GMD(F) = 2\int F(x)\bar{F}(x)dx, \tag{9}$$

where $\bar{F}(x) = 1 - F(x)$ denotes the survival function of X. It has been argued by Yitzhaki (2003) that *GMD*, as a measure of variability, shares many properties of the variance of X, and is suitable for distributions that are non-normal.

If X and Y are independent random variables on $\mathbb{R}$ with *CDFs* F and G, respectively, then the energy distance between them is defined as

$$D(F,G) = \int (F(x) - G(x))^2\, dx. \tag{10}$$

3.1. Gini's mean difference of transmuted distribution

In this subsection, we derive an expression for *GMD* of transmuted distribution function F_{X_T}.

Theorem 3.1. *Suppose F_{X_T} is the transmuted distribution given in (2) based on the components F and F_{max}. Then, the GMD associated with F_{X_T} is given by*

$$GMD(F_{X_T}) = (1+\lambda)^2 GMD(F) + \lambda^2 GMD(F_{max}) - 2\lambda(1+\lambda)\int_0^1 \frac{u(1-u)(1+2u)}{f(F^{-1}(u))du}. \tag{11}$$

Proof: From the definition of *GMD* in (9), upon setting $u = F(x)$, we get

$$\begin{aligned}
\frac{1}{2}GMD(F_{X_T}) &= \int \bar{F}_{X_T}(x)F_{X_T}(x)dx \\
&= \int \Big\{(1+\lambda)\bar{F}(x) - \lambda(1-F^2(x))\Big\}\Big\{1 - \Big((1+\lambda)\bar{F}(x) - \lambda(1-F^2(x))\Big)\Big\}dx \\
&= (1+\lambda)\int \bar{F}(x)dx - \lambda\int (1-F^2(x))dx - (1+\lambda)^2\int \bar{F}^2(x)dx \\
&- 2\lambda(1+\lambda)\int \bar{F}(x)(1-F^2(x))dx - \lambda^2\int (1-F^2(x))^2dx \\
&= (1+\lambda)^2\int \bar{F}(x)F(x)dx + \lambda^2\int F^2(x)(1-F^2(x))dx \\
&\quad -\lambda(1+\lambda)\int_0^1 \frac{u(1-u)(1+2u)}{f(F^{-1}(u))du}.
\end{aligned}$$

A rearrangement of the last equation yields the expression in (11).

Example 3.2. *Let X have a Beta$(\alpha, 1)$ distribution with CDF $F(x) = x^\alpha$, $0 < x < 1$, $\alpha > 0$. Then, the GMD of the transmuted distribution based on X is*

$$\begin{aligned}
GMD(F_{X_T}) &= (1+\lambda)^2\frac{\alpha}{2\alpha^2+3\alpha+1} + \lambda^2\frac{2\alpha}{8\alpha^2+6\alpha+1} \\
&\quad -2\lambda(1+\lambda)\Big\{\frac{1}{2\alpha+1} - \frac{2}{3\alpha+1} + \frac{1}{\alpha+1}\Big\}.
\end{aligned}$$

The following theorem gives lower and upper bounds for $GMD(F_{X_T})$ based on the *GMD* of the component distributions.

Theorem 3.3. *Suppose F, F_{max} and F_{X_T} have their GMD measures as $GMD(F)$, $GMD(F_{max})$ and $GMD(F_{X_T})$, respectively. Then:*

(i) If $\lambda \in [-1, 0]$, then

$$GMD(F_{X_T}) \geq (1+\lambda)^2 GMD(F) + \lambda^2 GMD(F_{max});$$

(i) If $\lambda \in [0,1]$, *then*

$$GMD(F_{X_T}) \leq (1+\lambda)^2 GMD(F) + \lambda^2 GMD(F_{max}).$$

Proof: From Theorem 3.1, the above inequalities are obtained readily.

Example 3.4. *For the distribution considered in Example 3.2, we find from Theorem 3.3 the following:*

(i) If $\lambda \in [-1,0]$,

$$GMD(F_{X_T}) \geq (1+\lambda)^2 \frac{\alpha}{2\alpha^2+3\alpha+1} + \lambda^2 \frac{2\alpha}{8\alpha^2+6\alpha+1};$$

(ii) If $\lambda \in [0,1]$,

$$GMD(F_{X_T}) \leq (1+\lambda)^2 \frac{\alpha}{2\alpha^2+3\alpha+1} + \lambda^2 \frac{2\alpha}{8\alpha^2+6\alpha+1}.$$

Figure 5 plots the *GMD* measure and its corresponding bound based on Examples 3.2 and 3.4, for $\alpha = 2$ and λ in $[-1,1]$.

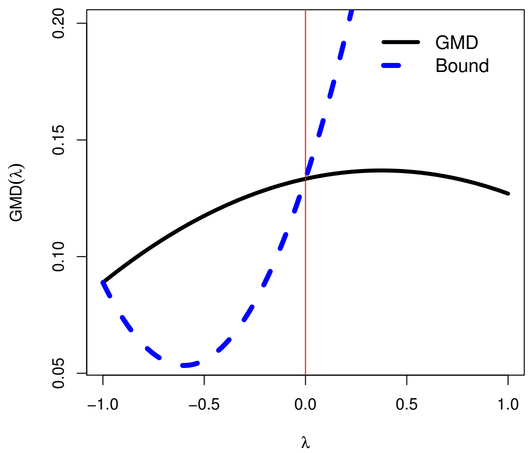

Fig 5. Plot of the *GMD* measure and its corresponding bound for values of λ in $[-1,1]$ and $\alpha = 2$.

3.2. *Transmuted Gini's mean difference and its connection with energy distance*

In this subsection, we introduce an extension of Gini's mean difference of transmuted distribution as well as Jensen-Gini (*JG*) measure of divergence, and we refer to it as transmuted Gini (*TG*) measure of divergence.

Definition 3.5. *Let X be a continuous random variable with CDF F, and the corresponding transmuted random variable* X_T *have its PDF as in (3). Then, the transmuted Gini divergence between components F and* F_{max} *is defined as*

$$TG(F, F_{max}; \lambda) = GMD(F_{X_T}) - [(1+\lambda)GMD(F) - \lambda GMD(F_{max})], \ \lambda \in [-1,1]. \tag{12}$$

The following theorem provides a connection between *TG* divergence and energy distance defined in (10).

Theorem 3.6. *The TG divergence can be expressed in terms of energy distance as*

$$TG(F, F_{max}; \lambda) = (1 + \lambda)D(F, F_{X_T}) - \lambda D(F_{max}, F_{X_T}),$$

where $D(H, F_{X_T})$ is the energy distance between distribution functions H and F_{X_T}, with H being either F or F_{max}.

Proof: From the definition of TG divergence in (12), we have

$$\begin{aligned} TG(F, F_{max}; \lambda) &= GMD(F_{X_T}) - [(1 + \lambda)GMD(F) - \lambda GMD(F_{max})] \\ &= \int \bar{F}_{X_T}(x)F_{X_T}(x)dx - \left[(1 + \lambda)\int \bar{F}(x)F(x)dx - \lambda \int \bar{F}_{max}(x)F_{max}(x)dx\right] \\ &= (1 + \lambda)\int \bar{F}^2(x)dx - \lambda \int \left\{1 - F^2(x)\right\}^2 dx \\ &\quad - \int \left\{(1 + \lambda)\bar{F}(x) - \lambda(1 - F^2(x))\right\}^2 dx. \end{aligned}$$

By denoting $K = (1 + \lambda)D(F, F_{X_T}) - \lambda D(F_{max}, F_{X_T})$, we have

$$\begin{aligned} K &= (1 + \lambda)\int \left[\bar{F}(x) - \bar{F}_{X_T}(x)\right]^2 dx - \lambda \int \left[\bar{F}_{max}(x) - \bar{F}_{X_T}(x)\right]^2 dx \\ &= (1 + \lambda)\int \left[\bar{F}(x) - \left[(1 + \lambda)\bar{F}(x) - \lambda(1 - F^2(x))\right]\right]^2 dx \\ &\quad -\lambda \int \left\{(1 - F^2(x)) - \left[(1 + \lambda)\bar{F}(x) - \lambda(1 - F^2(x))\right]\right\} dx \\ &= (1 + \lambda)\int \bar{F}^2(x)dx - \lambda \int \left\{1 - F^2(x)\right\}^2 dx \\ &\quad - \int \left\{(1 + \lambda)\bar{F}(x) - \lambda(1 - F^2(x))\right\}^2 dx, \end{aligned}$$

which is exactly the same expression derived above for $TG(F, F_{max}; \lambda)$. Thus,

$$\begin{aligned} TG(F, F_{max}; \lambda) &= GMD(F_{X_T}) - [(1 + \lambda)GMD(F) - \lambda GMD(F_{max})] \\ &= (1 + \lambda)D(F, F_{X_T}) - \lambda D(F_{max}, F_{X_T}) \\ &= (1 + \lambda)\int \bar{F}^2(x)dx - \lambda \int (1 - F^2(x))^2 dx \\ &\quad - \int \left\{(1 + \lambda)\bar{F}(x) - \lambda(1 - F^2(x))\right\}^2 dx, \end{aligned}$$

proving the theorem.

4. Chi-square divergence between *PDF* f_{X_T} and its components f and f_{max}

Let f_1 and f_2 be *PDFs* of random variables X_1 and X_2, respectively. Then, the chi-square divergence between f_1 and f_2 is defined as

$$\chi^2(f_1, f_2) = \int \frac{[f_1(x) - f_2(x)]^2}{f_2(x)} dx. \tag{13}$$

Theorem 4.1. *Let X_1 and X_2 be two independent random variables identically distributed as X with PDF $f(x)$. Then, the chi-square divergence between f_{max} and f_{X_T} (and f and f_{X_T}) are given by*

(i) $\chi^2\Big(f_{max}, f_{X_T}\Big) = \chi^2(f_V, f_{U_T}) = (1+\lambda)^2\, \frac{\tanh^{-1}(\lambda)-\lambda}{\lambda^3}$;

(ii) $\chi^2(f_{X_T}, f) = \chi^2(f_{U_T}, f_U) = \frac{\lambda^3}{3}$;

(iii) $\chi^2(f, f_{X_T}) = \chi^2(f_U, f_{U_T}) = \frac{\tanh^{-1}(\lambda)}{\lambda} - 1$,

where the random variables U, V and U_T are as defined in Lemma 2.7.

Proof: From the definition of χ^2 divergence in (13), upon setting $u = F(x)$, we get

$$\begin{aligned}\chi^2(f_{max}, f_{X_T}) &= \int_{-\infty}^{\infty} \frac{[f_{max}(x) - f_{X_T}(x)]^2}{f_{X_T}(x)} dx \\ &= \int_{-\infty}^{\infty} \frac{\Big\{2f(x)F(x) - (1+\lambda)f(x) + 2\lambda f(x)F(x)\Big\}^2}{(1+\lambda)f(x) - 2\lambda f(x)F(x)} dx \\ &= \chi^2(f_V, f_{U_T}) \\ &= (1+\lambda)^2\, \frac{\tanh^{-1}(\lambda) - \lambda}{\lambda^3},\end{aligned}$$

which proves Part (i). Parts (ii) and (iii) can be proved in a similar manner.

Remark 4.2. *From Theorem 4.1, we see easily that the chi-square divergence between transmuted PDF f_{X_T} and its component f_{max} is free of the underlying distribution F.*

5. Fisher information and Bayes Fisher information for λ

In this section, we derive some results about Fisher information and Bayes Fisher information for the parameter λ of the transmuted distribution.

5.1. Fisher information for parameter λ of transmuted distribution

Theorem 5.1. *The Fisher information for λ of the transmuted PDF in (3) is given by*

$$\mathcal{I}(\lambda) = \int \frac{[f(x) - f_{max}(x)]^2}{f_{X_T}(x)} dx = \frac{1}{(1+\lambda)^2} \chi^2(f_{max}, f_{X_T}).$$

Proof: From the definition of Fisher information, we have

$$\begin{aligned}\mathcal{I}(\lambda) &= E\left[\frac{\partial}{\partial\lambda} \log f_{X_T}(X)\right]^2 = E\left[\frac{f(X) - 2f(X)F(X)}{f_{X_T}(X)}\right]^2 \\ &= \int \left(\frac{f(x) - 2f(x)F(x)}{f_{X_T}(x)}\right)^2 f_{X_T}(x) dx \\ &= \frac{1}{(1+\lambda)^2} \chi^2(f_{max}, f_{X_T}),\end{aligned}$$

as required.

Remark 5.2. *From Theorem 5.1, we readily obtain*

$$\mathcal{I}(\lambda) = \frac{1}{(1+\lambda)^2} \chi^2(f_V, f_{U_T}) = \frac{tanh^{-1}(\lambda) - \lambda}{\lambda^3}. \tag{14}$$

From this expression, it is clear that $\mathcal{I}(\lambda)$ is symmetric and is free of the underlying distribution F.

Remark 5.3. *The Fisher Information $\mathcal{I}(\lambda)$ in (14) can be expressed based on the TS entropy as follows:*

$$\mathcal{I}(\lambda) = -\frac{\partial^2 TS(f, f_{max}; \lambda)}{\partial \lambda^2}.$$

Proof: From Lemma 2.14, we have

$$\mathcal{I}(\lambda) = -\frac{\partial^2 TS(f, f_{max}; \lambda)}{\partial \lambda^2} = -\frac{1}{\lambda^2}\Big(1 - \frac{\tanh^{-1}(\lambda)}{\lambda}\Big),$$

as required.

Figure 6 plots the Fisher information for parameter λ, from which we observe that $\mathcal{I}(\lambda)$ is minimized when $\lambda = 0$ with $\mathcal{I}(0) = 0.5$ and gets maximized when λ tends to ± 1.

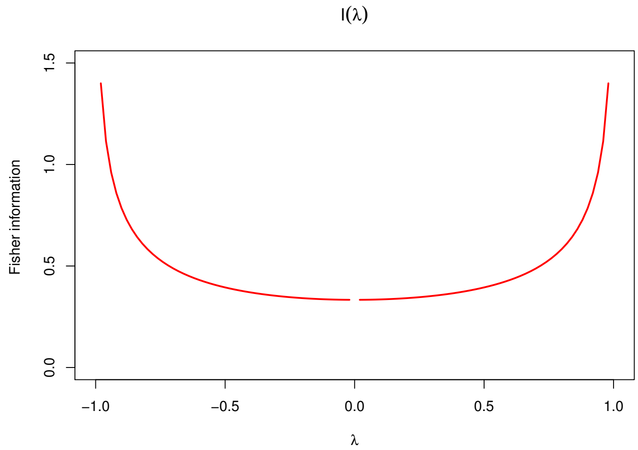

Fig 6. Fisher information for λ of the transmuted f_{X_T} with underlying density f.

5.2. Bayes Fisher information for λ of transmuted distribution

Theorem 5.4. *The Bayes Fisher information for parameter λ of the transmuted PDF in (3), under the uniform prior on $[-1, 1]$, is given by*

$$\tilde{\mathcal{I}}(\lambda) = \frac{1}{2} KL(f_{max}, f_{min}) = \frac{1}{2}. \tag{15}$$

Proof: Upon setting $u = F(x)$, we get

$$\begin{aligned}
\tilde{\mathcal{I}}(\lambda) &= E\left[\mathcal{I}(\Lambda)\right] = \frac{1}{2}\int_{-1}^{1} \mathcal{I}(\lambda) d\lambda = \frac{1}{2}\int_{-1}^{1}\int_{-\infty}^{\infty} \frac{[f(x) - f_{max}(x)]^2}{f_{X_T}(x)} dx d\lambda \\
&= \frac{1}{2}\int_{0}^{1} \log\Big(\frac{1-u}{u}\Big) du + \int_{-\infty}^{\infty} f(x)F(x) \log\Big\{\frac{2f(x)F(x)}{2f(x)\bar{F}(x)}\Big\} dx \\
&= \frac{1}{2}\int_{-\infty}^{\infty} f_{max}(x) \log\Big\{\frac{f_{max}(x)}{f_{min}(x)}\Big\} dx \\
&= \frac{1}{2} KL(f_{max}, f_{min}) \\
&= \frac{1}{2},
\end{aligned}$$

where the fourth equality holds due to the fact that $\int_0^1 \log\left(\frac{1-u}{u}\right)du = 0$.

Suppose we consider the following general triangular prior for the parameter λ:

$$\pi_\alpha(\lambda) = \begin{cases} \frac{2\lambda}{\alpha}, & 0 < \lambda \leq \alpha, \\ \frac{2(1-\lambda)}{1-\alpha}, & \alpha \leq \lambda < 1. \end{cases} \tag{16}$$

Theorem 5.5. *The Bayes Fisher information for parameter λ in the transmuted distribution with PDF in (3), under the general triangular prior with density $\pi_\alpha(\lambda)$ as in (16), we have*

$$\tilde{\mathcal{I}}(\lambda) = \frac{2}{\alpha(1-\alpha)}\left[\alpha KL(f_{min}, f_\alpha) + (1-\alpha)KL(f, f_\alpha)\right],$$

where f_α is a transmuted density with PDF

$$f_\alpha(x) = (1+\alpha)f(x) - \alpha f_{max}(x).$$

Proof: From (14) and (16), we have

$$\begin{aligned}
\tilde{\mathcal{I}}(\lambda) &= E\left[\mathcal{I}(\Lambda)\right] = \int_0^\alpha \mathcal{I}(\lambda)\pi_\alpha d\lambda + \int_\alpha^1 \mathcal{I}(\lambda)\pi_\alpha d\lambda \\
&= \frac{2}{\alpha}\int_{-\infty}^{\infty}(f(x) - f_{max}(x))\left[\int_0^\alpha \frac{\lambda(f(x) - f_{max}(x))}{f_{X_T}(x)}d\lambda\right]dx \\
&\quad + \frac{2}{1-\alpha}\int_{-\infty}^{\infty}(f(x) - f_{max}(x))\left[\int_\alpha^1 \frac{(1-\lambda)(f(x) - f_{max}(x))}{f_{X_T}(x)}d\lambda\right]dx \\
&= \frac{2}{\alpha}\int_{-\infty}^{\infty}(f(x) - f_{max}(x))\left[\alpha - \frac{f(x)}{f(x) - f_{max}(x)}\log\left\{\frac{f_\alpha(x)}{f(x)}\right\}\right]dx \\
&= \frac{2}{1-\alpha}\int_{-\infty}^{\infty}(f(x) - f_{max}(x))\left[\alpha - 1 - \frac{2f(x)\bar{F}(x)}{f(x) - f_{max}(x)}\log\left\{\frac{f_\alpha(x)}{2f(x)\bar{F}(x)}\right\}\right]dx \\
&= \frac{2}{\alpha}\int_{-\infty}^{\infty} f(x)\log\left\{\frac{f(x)}{f_\alpha(x)}\right\}dx + \frac{2}{1-\alpha}\int_{-\infty}^{\infty} f_{min}(x)\log\left\{\frac{f_{min}(x)}{f_\alpha(x)}\right\}dx \\
&= \frac{2}{\alpha(1-\alpha)}\left[\alpha KL(f_{min}, f_\alpha) + (1-\alpha)KL(f, f_\alpha)\right],
\end{aligned}$$

as required.

Corollary 5.6. *The Bayes Fisher information for λ under the triangular prior distribution with PDF $\pi(\lambda) = 1 - |\lambda|, |\lambda| \in [0,1]$, is*

$$\tilde{\mathcal{I}}(\lambda) = KL(f_{max}, f_{min}) - \left[KL(f, f_{max}) - KL(f, f_{min})\right].$$

Proof: From Theorem 5.5, this is readily obtained.

6. Conclusions

In this paper, some informational properties of the transmuted distributions, such as Shannon entropy, Fisher information, Bayes Fisher information and Gini's mean difference, have been derived. Two new information measures: transmuted Shannon entropy and transmuted Gini's mean difference, have also been proposed. It has been shown that the transmuted Shannon entropy is based on the transmuted structure of two *KL* divergences between the general model and its components. A similar result has been provided for transmuted Gini's mean difference and it has been shown that it can be given in terms of a quadratic rank structure of two energy distances between transmuted distribution and its components. Some interesting results about Fisher information and Bayes Fisher information for parameter λ have been presented under different prior distributions including uniform, beta and triangular. Finally, the divergence between the general transmuted distribution and the densities of its components have been studied with the use of Kullback-Leibler, chi-square and energy distances. It is of interest to mention that Granzotto et al. (2017) recently introduced a more general family of transmuted distributions, called cubic rank transmuted distributions, and we plan to extend the results here to this family. We hope to report these findings in a future paper.

Acknowledgments
The authors thank the Editor and the anonymous reviewers for their useful comments and suggestions on an earlier version of this manuscript which led to this improved version.

References

[1] Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1992). A First Course in Order Statistics. John Wiley & Sons, New York.
[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, 2316-2321.
[3] Balakrishnan, N., He, M. (2021). A record-based transmuted family of distributions. In: I. Ghosh et al. (eds.), Advances in Statistics - Theory and Applications, Springer, New York, $https://doi.org/10.1007/978-3-030-62900-7_1$.
[4] Fisher, R. A. (1929). Tests of significance in harmonic analysis. Proceedings of the Royal Society of London, Series A, 125, 54-59.
[5] Gini, C. (1912). Variabilità e mutabilità. Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi.
[6] Granzotto, D. C. T., Louzada, F., & Balakrishnan, N. (2017). Cubic rank transmuted distributions: inferential issues and applications. Journal of Statistical Computation and Simulation, 87, 2760-2778.
[7] Kozubowski, T. J., & Podgórski, K. (2016). Transmuted distributions and random extrema. Statistics & Probability Letters, 116, 6-8.
[8] Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. Annals of Mathematical Statistics, 22, 79-86.
[9] Lin, J. (1991). Divergence measures based on the Shannon entropy. IEEE Transactions on Information theory, 37, 145-151.
[10] Mehrali, Y., Asadi, M., & Kharazmi, O. (2018). A Jensen-Gini measure of divergence with application in parameter estimation. Metron, 76, 115-131.
[11] 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.
[12] Qiao, Y., & Minematsu, N. (2010). A study on invariance of f-divergence and its application to speech recognition. IEEE Transactions on Signal Processing, 58, 3884-3890.
[13] Sánchez-Moreno, P., Zarzo, A., & Dehesa, J. S. (2012). Jensen divergence based on Fisher's information. Journal of Physics A: Mathematical and Theoretical, 45, 125305.
[14] Sarabia, J. M., Raja, A. V., & Asha, G. (2020). Bivariate distributions with transmuted conditionals: Models and applications. Communications in Statistics-Theory and Methods, 49, 221-242.
[15] Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379-423.
[16] Shaw, W. T., & Buckley, I. R. (2009). The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. arXiv preprint arXiv: 0901.0434.
[17] Yitzhaki, S. (1982). Stochastic dominance, mean variance, and Gini's mean difference. The American Economic Review, 72, 178-185.
[18] Yitzhaki S, Wodon QT (2003) Mobility, inequality, and horizontal equity. Available at SSRN 419322.