Jensen-information generating function and its connections to some well-known information measures

ARTICLE INFO

Keywords:
Information generating function
Shannon entropy
Jensen-Shannon entropy
Jensen-extropy
Kullback-Leibler divergence

ABSTRACT

In this work, we consider the information generating function measure and develop some new results associated with it. We specifically propose two new divergence measures and show that some of the well-known information divergences such as Jensen-Shannon, Jensen-extropy and Jensen-Taneja divergence measures are all special cases of it. Finally, we also discuss the information generating function for residual lifetime variables.

1. Introduction

In information theory, several criteria have been proposed for measuring the uncertainty of a probabilistic model. Shannon entropy is the most important information measure that has been used in different branches of science and engineering. It originated from the pioneering work of Shannon (1948), based on a study of the behavior of systems described by probability density (or mass) functions.

In distribution theory, successive derivatives of the moment generating function at point 0 give successive moments of a probability distribution, which are specifically used for extracting characteristics such as mean, variance, skewness and kurtosis. In information theory, generating functions have also been defined for probability densities to determine information quantities such as Shannon information and Kullback-Leibler divergence.

Golomb (1966) proposed information generating function of a probability density function $f(x)$, whose derivatives, evaluated at 1, yield some statistical information measures for the probability distribution. Suppose the variable X has a density function $f(x)$. Then, the information generating (IG) function of density $f(x)$, for any $\alpha > 0$, is defined as

$$G_\alpha(f) = \int_{\mathcal{X}} f^\alpha(x)dx, \tag{1}$$

provided the integral exists. To simplify notation, we suppress $\mathcal{X}$ for integration with respect to dx throughout the paper, unless a distinction is needed. Golomb (1966) then showed the following properties of $G_\alpha(f)$ in (1):

$$(i) \quad G_1(f) = 1; \qquad (ii) \quad \frac{\partial}{\partial\alpha}G_\alpha(f)|_{\alpha=1} = -H(f), \tag{2}$$

where $H(f)$ is the Shannon differential entropy defined as $H(f) = -\int f(x) \log f(x) dx$. In particular, when $\alpha = 2$, the IG measure is reduced to $\int_{\mathcal{X}} f^2(x)dx$, which is known as informational energy (IE) function. Onicescu (1966) introduced a discrete version of informational energy measure into information theory by analogy to kinetic energy in mechanics. Let $\boldsymbol{p} = (p_1, ..., p_n)$ be a probability mass function (PMF). Then, Onicescu information energy is defined by $IE(\boldsymbol{p}) = \sum_{i=1}^{n} p_i^2$. Moreover, Theodorescu (1977) has defined a generalization of the IE measure for a PMF $\boldsymbol{p} = (p_1, ..., p_n)$ of order α as $IE_\alpha(\boldsymbol{p}) = \sum_{i=1}^{n} p_i^\alpha$. For more details, see for example Bhatia (1997). Oh and Ho (2019) have shown some application of Onicescu information energy to analyze the ground state helium by using highly correlated Hylleraas wave functions. Both discrete and continuous versions of informational energy measures have been used extensively in physics and chemistry to investigate the complex structure of a physical or chemical system that can be described with a probabilistic model for both continuous and discrete atomic properties; see López-Ruiz et al. (1995). Oh and Ho (2019) and Flores-Gallegos (2016). In fact, the Shannon entropy and IE measure are the complementary and uncertainty quantities to study the complex structure of a physical or chemical system in statistical thermodynamics and quantum theory. Moreover, with the recent fusion between chemistry and information theory, many concepts of chemistry can be explained using concepts of Shannon entropy and related informational measures such as IE measure and their extensions. Flores-Gallegos (2016) has introduced a definition of the IE discrete informational energy, and showed that the proposed measure can be used as a correlation measure in atomic and molecular systems. This author has also provided some practical analysis of correlation effects atoms of a set of 1864 molecules. As mentioned above, these information measures are used in physics, chemistry and materials science to investigate the atomic structure of a given system. Moreover, sometimes IG measure, known as entropic moment in physics and chemistry, plays an important role in non-extensive thermodynamics and chaos theory. This is closely connected to the so-called Rényi and Tsallis entropies. The entropic moment measure of uncertainty, similar to information entropy, also gives the degree of the spread of probability distribution of observables; see Bercher (2013).

Recently, research on information generating function has been carried out by Clark (2019), who provided an analogous information generating function for stochastic processes to facilitate the derivation of information statistics for point processes. Relative entropy (based on Golomb's information function) has been defined by Guiasu and Reischer (1985); its first derivative at 1 yields Kullback-Leibler divergence (Kullback and Leibler (1951)). Let X and Y be two random variables with density functions f and g, respectively. Then, the relative information generating function, for any $\alpha > 0$, is defined as

$$R_\alpha(f, g) = \int f^\alpha(x) g^{1-\alpha}(x) dx \tag{3}$$

provided the integral exists. Evidently, $R_1(f, g) = 1$, and $\frac{\partial^n}{\partial \alpha^n} R_\alpha(f, g) = \int \left(\log \frac{f(x)}{g(x)}\right)^n f^\alpha(x) g^{1-\alpha}(x) dx$. In particular, the Kullback-Leibler divergence is obtained from the first derivative as $\frac{\partial}{\partial\alpha}R_\alpha(f,g)|_{\alpha=1} = \int f(x)\left(\log\frac{f(x)}{g(x)}\right)dx$.

The purpose of this work is two-fold. The first part is to establish some new interesting properties of the IG measure. In the second part, we propose Jensen-information generating (JIG) function, whose derivatives generate the well-known Jensen-Shannon, Jensen-Taneja and Jensen-Extropy information measures. We further propose a new divergence measure between two density functions f_1 and f_2, so that the Jensen-information generating function can be expressed as a mixture of two versions of this new divergence measure. Finally, we develop some results for $G_\alpha(f)$ function for the residual lifetime distribution.

The rest of this paper is organized as follows. In Section 2, we consider the information generating function and establish some new properties that reveal its connections to some other well-known information measures. It is also shown that the IG measure can be expressed based on different orders of fractional Shannon entropy. Moreover, we obtain the information generating function for escort density functions and some results associated with stochastic ordering are also provided. The new Jensen-information generating (JIG) function is proposed in Section 3. We show that the Jensen-information generating function yields some well-known information measures such as Jensen-Shannon and Jensen-Tanja entropies and Jensen-extropy information measure as special cases. Also, a new divergence measure between two densities f_1 and f_2 is proposed. In particular, we show that the JIG function can be expressed as a mixture of two versions of the new divergence measure. Section 4 is devoted to the study of information generating function for residual lifetime distributions, wherein we provide some new results for residual information generating function connected with well-known reliability measures such as hazard function and mean residual lifetime. Finally, Section 5 presents some concluding remarks.

2. IG measure

We first present the definition of information generating function originally given by Colomb (1966), which facilitates representing different orders of fractional Shannon entropy, as shown by Ubriaco (2009). It enables us to derive some informational measures by using repeated derivatives of the IG function. Inspired by the definition of fractional Shannon entropy, we give the following lemma.

Lemma 1. *Suppose the random variable X has density function f. Then, a new representation for IG measure is given by $G_\alpha(f) = \sum_{n=0}^{\infty}\frac{(1-\alpha)^n}{n!}H_n(f)$, where $H_n(f)$ is the fractional Shannon entropy of order n defined as $H_n(f) = \int\left\{-\log f(x)\right\}^n f(x)dx$. For more details about fractional Shannon entropy, see Xiong et al. (2019).*

Proof: From the definition of IG measure in (1) and by using Maclaurin expansion and Fubini's theorem, we have

$$G_\alpha(f) = E\left[e^{(\alpha-1)\log f(X)}\right] = \sum_{n=0}^{\infty}\frac{(1-\alpha)^n}{n!}\int\left\{-\log f(x)\right\}^n f(x)dx = \sum_{n=0}^{\infty}\frac{(1-\alpha)^n}{n!}H_n(f).$$

From the properties of $G_\alpha(f)$ in (2), we obtain Shannon entropy as $\frac{\partial}{\partial\alpha}G_\alpha(f)|_{\alpha=1} = \int f(x)\log f(x)dx = -H(f)$.

Moreover, when $\alpha = 2$, we get $-\frac{1}{2}G_2(f) = -\frac{1}{2}\int f^2(x)dx = J(f)$, where $J(f)$ is known as extropy or informational energy function.

2.1. Some results on IG measure

We now provide some new results for the IG measure.

Theorem 2. *Let X be a nonnegative random variable with an absolutely continuous density function f. Then, for $\alpha \geq \frac{1}{2}$, we have $\left\{G_{\frac{\alpha+1}{2}}(f)\right\}^2 \leq G_\alpha(f) \leq \left\{G_{2\alpha-1}(f)\right\}^{\frac{1}{2}}$.*

Proof: To prove the first inequality $\left\{G_{\frac{\alpha+1}{2}}(f)\right\}^2 \leq G_\alpha(f)$, we use Cauchy-Schwartz inequality. If $g(x)$ and $h(x)$ are two real integrable functions, then the Cauchy-Schwartz inequality between $g(x)$ and $h(x)$, is given by

$$\left\{\int g(x)h(x)dx\right\}^2 \leq \int g^2(x)dx \int h^2(x)dx$$

provided all three integrals exist. Now, for density function $f(x)$, setting $g(x) = f^{\frac{\alpha}{2}}(x)$ and $h(x) = f^{\frac{1}{2}}(x)$, we obtain

$$\left\{G_{\frac{\alpha+1}{2}}(f)\right\}^2 = \left\{\int f^{\frac{\alpha}{2}}(x)f^{\frac{1}{2}}(x)dx\right\}^2 \leq \int f^\alpha(x)dx \int f(x)dx = \int f^2(x)dx = G_\alpha(f),$$

which proves the first inequality. If $h(x)$ is a non-negative function such that $\int h(x)dx = 1$, and $g(x)$ is a real function and $\phi(x)$ is a convex function, then the Jensen inequality is given by $\phi\left(\int g(x)h(x)dx\right) \leq \int \phi\big(g(x)\big)h(x)dx$. Now, upon setting $h(x) = f(x)$, $g(x) = f^{\alpha-1}(x)$ and $\phi(x) = x^2$, then the second inequality, for $\alpha \geq \frac{1}{2}$, follows since

$$G_\alpha^2(f) = \left\{\int f^{\alpha-1}(x)f(x)dx\right\}^2 \leq \int f^{2(\alpha-1)}(x)f(x)dx = \int f^{2\alpha-1}(x)dx = G_{2\alpha-1}(f).$$

Theorem 3. *Let X be a variable with an absolutely continuous density function f, and ϕ be an increasing, differentiable and invertible function. Then, we have*

$$G_\alpha(f_\phi) = \int_{-\infty}^{\infty}\left\{\frac{f(\phi^{-1}(x))}{\phi'(\phi^{-1}(x))}\right\}^\alpha dx = \int_L^U \frac{f^\alpha(x)}{(\phi'(x))^{\alpha+1}}dx,$$

where $L = \phi^{-1}(-\infty)$ and $U = \phi^{-1}(\infty)$.

Proof: Let X have its density function f with support $(-\infty, +\infty)$. For the function ϕ under conditions $\phi(-\infty) = -\infty$ and $\phi(+\infty) = +\infty$, the density of the transformed variable $\phi(X)$ is readily obtained as

$$f_{\phi(X)}(x) = \left|\frac{\partial\phi^{-1}(x)}{\partial x}\right| f\left(\phi^{-1}(x)\right) = \frac{f(\phi^{-1}(x))}{\frac{\partial\phi(\phi^{-1}(x))}{\partial x}} = \frac{f(\phi^{-1}(x))}{\phi'(\phi^{-1}(x))}, \quad -\infty < x < +\infty.$$

Hence, by using the definition of IG measure for density f_ϕ, we have $G_\alpha(f_\phi) = \int_{-\infty}^{\infty} \left\{ \frac{f(\phi^{-1}(x))}{\phi'(\phi^{-1}(x))} \right\}^\alpha dx = \int_L^U \frac{f^\alpha(x)}{(\phi'(x))^{\alpha+1}} dx$, where $L = \phi^{-1}(-\infty)$ and $U = \phi^{-1}(\infty)$. The second equality holds due to integral transformation $\phi^{-1}(x)$.

The following example investigates the amount of IG measure for the most popular Weibull distribution.

Example 1. Let X have a two-parameter Weibull distribution with density $f(x) = \frac{\beta}{\lambda^\beta} x^{\beta-1} \exp\left\{-(\frac{x}{\lambda})^\beta\right\}$, $x > 0, \beta, \lambda > 0$. Then, for $\beta > 1 - \frac{1}{\alpha}$, we have $G_\alpha(f) = \frac{\beta^{\alpha-1}\lambda^{1-\alpha}}{\alpha^{\frac{\alpha(\beta-1)+1}{\beta}}} \Gamma\left(\frac{\alpha(\beta-1)+1}{\beta}\right)$, *where* $\Gamma(.)$ *denotes the complete gamma function.*

From Example 1, we can obtain the Shannon entropy and informational energy measures as $H(f) = \left.\frac{\partial G_\alpha(f)}{\partial \alpha}\right|_{\alpha=0} = \gamma(1-\frac{1}{\beta}) + \log\left(\frac{\lambda^{-\frac{1}{\beta}}}{\beta}\right) + 1$, and $IE(f) = G_2(f) = \frac{\beta\lambda^{-1}}{2^{\frac{2(\beta-1)+1}{\beta}}} \Gamma\left(\frac{2(\beta-1)+1}{\beta}\right)$, respectively, where γ is the Euler-Mascheroni constant. The following theorem provides lower and upper bounds for IG measure based on the Shannon entropy and hazard rate function.

Theorem 4. *Let X be a non-negative variable with Shannon entropy, information generating measure and hazard rate function as $H(f)$, $G_\alpha(f)$ and r_F, respectively. Then, for any $\alpha > 0$, we have*

$$max(0, 1 - \alpha H(f)) \leq G_{\alpha+1}(f) \leq E\left[r_F^\alpha(X)\right], \tag{4}$$

where the hazard rate function is given by $r_F = \frac{f(x)}{\bar{F}(x)}$ and $\bar{F}(x)$ is the survival function of X.

Proof: By using the inequality $\alpha \log(x) \leq x^\alpha - 1$, the lower bound in (4) is readily obtained. The upper bound is obtained as follows: $G_{\alpha+1}(f) = \int_0^\infty f^{\alpha+1}(x)dx = \int_0^\infty \left(r_F(x)\bar{F}(x)\right)^\alpha f(x)dx \leq \int_0^\infty r_F(x)^\alpha f(x)dx = E\left[r_F^\alpha(X)\right]$.

It is worthwhile to mention that in the study of informational measures in reliability theory, an interesting issue is to find relations between these measures and fundamental quantities of reliability such as hazard rate and mean residual lifetime. For more details, see, for example, Kharazmi and Asadi (2018) and Asadi and Zohrevand (2007) and the references therein.

Corollary 1. *When $\alpha = 1$, we obtain from Theorem 4 the corresponding bounds for extropy measure $J(f) = -\frac{1}{2}\int f^2(x)dx$ as $-\frac{1}{2}E[r_F(X)] \leq J(X) \leq \frac{1}{2}(H(f) - 1)$.*

Example 2. Let X have a scaled-exponential distribution with density $f(x) = \lambda e^{-\lambda x}$, $x > 0, \lambda > 0$. Then, $G_{\alpha+1}(f) = \frac{\lambda^\alpha}{(\alpha+1)}$, $E[r^\alpha(X)] = \lambda^\alpha$, and the Shannon entropy is $H(f) = 1 - \log\lambda$.

2.2. IG measure for escort and generalized escort distributions

Let X be a variable with density function f. Then, the variable X_e is said to be the escort random variable corresponding to X if, for any positive real number η, the density function of X_e is given by $f_\eta(x) = \frac{f^\eta(x)}{\int f^\eta(x)dx}$, provided $\int f^\eta(x)dx < \infty$. It should be noted that this is a special case of weighted distributions in which the weight function has been chosen to be $w(x) = f^{\eta-1}(x)$. An extension of escort distribution is known as generalized escort distribution. Assume that variables X and Y have density functions f and g, respectively. Then, the corresponding generalized escort density is given by $h_\eta(x) = \frac{f^\eta(x)g^{1-\eta}(x)}{\int f^\eta(x)g^{1-\eta}(x)dx}$.

Note that this again is a special case of weighted distributions in which the weight function has been chosen to be $w(x) = \left\{\frac{f(x)}{g(x)}\right\}^{\eta-1}$. Escort distributions have found applications in different areas. Beck and Schögl (1993) have shown that they play a key role in various analogies between chaos theory and thermodynamics. Bercher (2012) has shown their use in coding theory, who further proved that escort distributions appear as solution of a minimum Kullback-Leibler discrimination associated with state transition model. For more details, see Asadi et al. (2018). On the other hand, both discrete and continuous versions of informational energy measures have been used extensively in physics and chemistry to investigate the complex structure of a physical or chemical system that can be described with a probabilistic model; see López-Ruiz et al. (1995), Oh and Ho (2019) and Flores-Gallegos (2016). Moreover, with the recent fusion between chemistry and information theory, many concepts in chemistry can be explained using concepts of Shannon entropy and related informational measures such as IG measure and their extensions. Because of importance of these concepts in physics and chemistry, we are interested in to obtain IG measure of escort distributions. The following theorem presents IG measures for escort and generalized escort distributions.

Theorem 5. *Suppose variables X and Y have density functions f and g, respectively. Then:*

(i) For any $\alpha,\ \eta > 0$, $G_{\alpha\eta}(f) = G_\alpha(f_\eta)G^\alpha_\eta(f)$;

(ii) For any $\eta > 0$ and $\alpha \geq 1$, $G_\alpha(h_\eta) = \frac{R_\eta(f_\alpha,g_\alpha)}{R^\alpha_\eta(f,g)}G^\eta_\alpha(f)G^{1-\eta}_\alpha(g)$,

where $R_\alpha(f,g)$ is as in (3), and f_α and g_α are escort densities of order α to corresponding f and g, respectively.

Proof: By the definition of information generating function in (1), we define it for the escort distribution of order η to be $G_\alpha(f_\eta) = \int f^\alpha_\eta(x)dx = \frac{\int f^{\alpha\eta}(x)dx}{\left\{\int f^\eta(x)dx\right\}^\alpha} = \frac{G_{\alpha\eta}(f)}{G^\alpha_\eta(f)}$, which proves Part (i). Next, we have

$$\begin{aligned} G_\alpha(h_\eta) &= \int h^\alpha_\eta(x)dx = \frac{G^\eta_\alpha(f)G^{1-\eta}_\alpha(g)}{R^\alpha_\eta(f,g)}\int\left\{\frac{f^\alpha(x)}{\int f^\alpha(x)dx}\right\}^\eta\left\{\frac{g^\alpha(x)}{\int g^\alpha(x)dx}\right\}^{1-\eta}dx \\ &= \frac{G^\eta_\alpha(f)G^{1-\eta}_\alpha(g)}{R^\alpha_\eta(f,g)}\int\left\{f_\alpha(x)\right\}^\eta\left\{g_\alpha(x)\right\}^{1-\eta}dx \\ &= \frac{R_\eta(f_\alpha,g_\alpha)}{R^\alpha_\eta(f,g)}G^\eta_\alpha(f)G^{1-\eta}_\alpha(g), \end{aligned}$$

which proves Part (ii).

2.3. IG stochastic ordering

One important criteria for comparing diversity (or variation) in probability distributions is dispersive ordering. Suppose X and Y are two continuous random variables with distributions F and G and density functions f and g, respectively. Then, we say that X is less dispersed than Y (denoted by $X \leq_{disp} Y$) if $g(G^{-1}(x)) \leq f(F^{-1}(x))$ for all $x \in (0,1)$; see, for example, Shaked and Shanthikumar (2007) for pertinent details.

Definition 1. Let X and Y be two variables with IG measures $G_\alpha(f)$ and $G_\alpha(g)$, respectively. Then, X is said to be less than Y in information generating function, denoted by $X \leq_{IG} Y$, if $G_\alpha(f) \leq G_\alpha(g)$.

Theorem 6. *Suppose $X \leq_{disp} Y$. Then:*

(i) If $\alpha \leq 1$, $X \leq_{IG} Y$;

(ii) If $\alpha \geq 1$, $Y \leq_{IG} X$.

Proof: By the definition in (1), we have $G_\alpha(f) = \int_{-\infty}^{\infty} f^\alpha(x)dx = \int_0^1 f^{\alpha-1}(F^{-1}(v))dv$. Because $X \leq_{disp} Y$, we have $f(F^{-1}(v)) \geq g(G^{-1}(v))$ for all $v \in (0,1)$. So, for $\alpha \leq 1$, we get $f^{\alpha-1}(F^{-1}(v)) \leq g^{\alpha-1}(G^{-1}(v))$, and upon integrating both sides of this inequality, we obtain $G_\alpha(f) = \int_0^1 f^{\alpha-1}(F^{-1}(v))dv \leq \int_0^1 g^{\alpha-1}(G^{-1}(v))dv = G_\alpha(g)$, which proves Part (i). Part (ii) can be proved in an analogous manner.

Theorem 7. *Suppose $X \leq_{disp} Y$. Then:*

(i) If ϕ is convex and $\alpha \leq 1$, $\phi(X) \leq_{IG} \phi(Y)$;

(ii) If ϕ is concave and $\alpha \geq 1$, $\phi(Y) \leq_{IG} \phi(X)$.

Proof: By the definition in (1) and under the conditions $\phi^{-1}(-\infty) = -\infty$ and $\phi^{-1}(\infty) = \infty$, we have

$$G_\alpha(f_\phi) = \int_{-\infty}^{\infty} \left(\frac{f(\phi^{-1}(x))}{\phi'(\phi^{-1}(x))} \right)^\alpha dx = \int_{-\infty}^{\infty} \frac{f^\alpha(x)}{(\phi'(x))^{\alpha+1}} dx = \int_0^1 \frac{f^{\alpha-1}(F^{-1}(v))}{\left(\phi'(F^{-1}(v))\right)^{\alpha+1}} dv.$$

Because $X \leq_{disp} Y$, we have $f(F^{-1}(v)) \geq g(G^{-1}(v))$ for all $v \in (0,1)$. So, for $\alpha \leq 1$, we get $f^{\alpha-1}(F^{-1}(v)) \leq g^{\alpha-1}(G^{-1}(v))$. As $\phi(x)$ is convex, $\phi'(x)$ is increasing and so $\phi'(f(F^{-1}(v))) \geq \phi'(g(G^{-1}(v)))$. Hence,

$$G_\alpha(f_\phi) = \int_0^1 \frac{f^{\alpha-1}(F^{-1}(v))}{\left(\phi'(F^{-1}(v))\right)^{\alpha+1}} dv \leq \int_0^1 \frac{g^{\alpha-1}(G^{-1}(v))}{\left(\phi'(G^{-1}(v))\right)^{\alpha+1}} dv = G_\alpha(g_\phi),$$

which proves Part (i). Part (ii) can be proved in an analogous manner.

2.4. IG measure for convolution of independent random variables

The convolution operation occurs frequently in practice. Although the operation of convolution is elementary in its concept, it is rather onerous to implement on information theory, physics and engineering for several reasons that have been stated in the paper. For example, the de Bruijn identity, which relates entropy with Fisher information, can be obtained as a particular case of an immediate generalization of Price's theorem, which is a tool used in the analysis of nonlinear memoryless systems with Gaussian inputs; see Park et al. (2012).

Here, we provide some results associated with convolution of some random variables and examine the IG measure of convoluted variable with its components. We now establish an upper bound for IG measure for convolution of two independent random variables based on IG measures of each of its components.

Theorem 8. *Let X and Y be two non-negative independent random variables with density functions f and g, respectively. Then, $G_\alpha(h) \leq min(G_\alpha(f), G_\alpha(g))$, where h is the density function of the convoluted variable $X + Y$.*

Proof: The density of function of random variable $X + Y$ is given by $h(x) = \int_0^x f(x-t)g(t)dt$. By using the Jensen inequality for $\alpha \geq 1$, we have $h^\alpha(x) = \left[\int_0^\infty f(x-t)g(t)I(x \geq t)dt\right]^\alpha \leq \int_0^\infty f^\alpha(x-t)g(t)I(x \geq t)dt$, where $I(x \geq t)$ is the indicator function. Upon integrating both sides of the above inequality with respect to x from 0 to ∞, we obtain

$$\begin{aligned} G_\alpha(h) &= \int_0^\infty h^\alpha(x)dx \leq \int_0^\infty \left(\int_0^\infty f^\alpha(x-t)g(t)I(x \geq t)dt\right) dx = \int_0^\infty \left(\int_0^\infty f^\alpha(x-t)I(x \geq t)dx\right) g(t)dt \\ &= \int_0^\infty \left(\int_t^\infty f^\alpha(x-t)dx\right) g(t)dt = \int_0^\infty f^\alpha(u)du \int_0^\infty g(t)dt = \int_0^\infty f^\alpha(u)du = G_\alpha(f). \end{aligned}$$

In a similar way, we can show that $G_\alpha(h) \leq G_\alpha(f)$, thus establishing that $G_\alpha(h) \leq min(G_\alpha(f), G_\alpha(g))$.

From Theorem 8, we simply observe that convolution decreases the IG information. Theorem 8 can be extended readily to the convolution of n independent random variables as follows.

Corollary 2. *Let $X_1, ..., X_n$ be non-negative independent random variables with density functions $f_1, ..., f_n$, respectively. Then,*

$$G_\alpha(h) \leq min(G_\alpha(f_1), ..., G_\alpha(f_n)), \tag{5}$$

where h is the density function of $\sum_{i=1}^n X_i$.

Proof: By using Theorem 8 with $X = \sum_{i=1}^{n-1} X_i$ and $Y = X_n$, we immediately have $G_\alpha(h) \leq min(G_\alpha(h_{n-1}), G_\alpha(f_n))$, where $h_{n-1}(x)$ is the density function of X. By using this inequality recursively, we obtain (5).

3. Jensen-information generating function

The important and interesting properties of IG measure is that, it works as generator function to produce several well-known informational measures such as Shannon entropy, informational energy function and fractional Shannon entropy of different order. The class of Jensen divergences have been discussed greatly in recent years. For example, measures such as Jensen-Shannon (JS) entropy, Jensen-Fisher (JF) information and Jensen-Gini (JG) divergence have been suggested in different applications; see, for example, Mehrali et al. (2018) and the references therein. Asadi et al. (2016) have applied Jensen-Shannon (JS) entropy to examine the complexity of coherent mixed systems. Specifically, they showed that JS measure can be used to explain the remaining uncertainty about the system lifetime. We examine this idea for the IG function as a main source of content information and so introduced Jensen IG measure.

In this section, we propose a measure associated with IG based on Jensen inequality. For this purpose, we first define a new divergence measure between two density functions as follows.

Definition 2. *Let f_1 and f_2 be two density functions with IG measures $G_\alpha(f_1)$ and $G_\alpha(f_2)$, respectively. Then, a new divergence measure between f_1 and f_2, $D_\alpha(f_1 : f_2)$, is defined as*

$$D_\alpha(f_1 : f_2) = (\alpha - 1)\left\{ \int f_1^\alpha(x)\bar{L}_{\frac{1}{\alpha}}\left(\frac{f_1^\alpha(x)}{f_2^\alpha(x)}\right)dx - \left(G_\alpha(f_1) - G_\alpha(f_2)\right) \right\}, \tag{6}$$

where $\bar{L}_q(z)$ is generalized logarithm function defined as

$$\bar{L}_q(z) = \begin{cases} \frac{z^{q-1}-1}{q-1}, & q \in [0,1) \cup (1,\infty), \\ \log z, & q = 1. \end{cases} \tag{7}$$

In the same way, we can define $D_\alpha(f_2 : f_1)$.

Theorem 9. *Let f_1 and f_2 be two density functions with information generating functions $G_\alpha(f_1)$ and $G_\alpha(f_2)$, respectively. Then, $D_\alpha(f_1 : f_2)$ is non-negative.*

Proof: Using Lemma 1 of Asadi et al. (2017), the result follows readily.

Example 3. *Let X_1 and X_2 be exponential random variables with density functions $f_1(x) = \lambda_1 e^{-\lambda_1 x}$ and $f_2(x) = \lambda_2 e^{-\lambda_2 x}$, respectively. Then, the D_α divergence between f_1 and f_2 is given by*

$$D_\alpha(f_1 : f_2) = \frac{-\alpha\lambda_1\lambda_2^{\alpha-1}}{\lambda_1 + (\alpha-1)\lambda_2} + (\alpha-1)\left(\frac{\lambda_2^{\alpha-1}}{\alpha} - \frac{\lambda_1^{\alpha-1}}{\alpha}\right) + \lambda_1^{\alpha-1}. \tag{8}$$

Definition 3. *Let X_1 and X_2 be variables with density functions f_1 and f_2, respectively. Then, the Jensen-information generating function (JIG), for any $\alpha \geq 1$ and $0 < p < 1$, is defined as*

$$JG_\alpha(f_1, f_2; p) \;=\; pG_\alpha(f_1) + (1-p)G_\alpha(f_2) - G_\alpha\big(pf_1 + (1-p)f_2\big). \tag{9}$$

Theorem 10. *Let variables X_1 and X_2 have density functions f_1 and f_2, respectively. Then, the Jensen-information generating function, $JG_\alpha(f_1, f_2, p)$, is a mixture of the measure in (6), given by*

$$JG_\alpha(f_1, f_2; p) = pD_\alpha(f_1 : f_T) + (1-p)D_\alpha(f_2 : f_T), \tag{10}$$

where $D_\alpha(f_i : f_T)$, $i = 1, 2$, is the divergence measure in (6), and $f_T = pf_1 + (1-p)f_2$ is the mixture density based on components f_1 and f_2.

Proof: With $f_T = pf_1 + (1-p)f_2$, being the mixture density, for $\alpha \neq 1$, we have

$$\begin{aligned} JG_\alpha(f_1, f_2, p) &= p\int f_1^\alpha(x)dx + (1-p)\int f_2^\alpha(x)dx - \int \Big\{pf_1(x) + (1-p)f_2(x)\Big\}^\alpha dx \\ &= \int \Big\{pf_1^\alpha(x) + (1-p)f_2^\alpha(x) - f_T^\alpha(x)\Big\}dx \\ &= \frac{(\alpha-1)}{\alpha}p\left[\int f_1^\alpha(x)\bar{L}_{\frac{1}{\alpha}}\left(\frac{f_1^\alpha(x)}{f_T^\alpha(x)}\right)dx - (G_\alpha(f_1) - G_\alpha(f_T))\right] \\ &\quad + \frac{(\alpha-1)}{\alpha}(1-p)\left[\int f_2^\alpha(x)\bar{L}_{\frac{1}{\alpha}}\left(\frac{f_2^\alpha(x)}{f_T^\alpha(x)}\right)dx - (G_\alpha(f_2) - G_\alpha(f_T))\right] \\ &\quad + \frac{\alpha-1}{\alpha}JG_\alpha(f_1, f_2, p) \\ &= \frac{1}{\alpha}\big(pD_\alpha(f_1 : f_T) + (1-p)D_\alpha(f_2 : f_T)\big) + \frac{\alpha-1}{\alpha}JG_\alpha(f_1, f_2, p). \end{aligned}$$

By rearranging the above equation, we obtain (10), as required.

Theorem 11. *Let the variables $X_1, X_2, .., X_n$ have density functions $f_1, f_2, ..., f_n$, respectively. Then, the Jensen-information generating function $JG_\alpha(f_1, f_2, ..., f_n; \mathbf{p})$ is a mixture of $D_\alpha(f_i : f_T)$ measures in (6), given by*

$$JG_\alpha(f_1, f_2, ..., f_n; \mathbf{p}) \;=\; \sum_{i=1}^n p_i G_\alpha(f_i) - G_\alpha\Big(\sum_{i=1}^n p_i f_i\Big) = \sum_{i=1}^n p_i D_\alpha(f_i : f_T),$$

where D_α is the divergence measure defined in (6), and $f_T = \sum_{i=1}^{n} p_i f_i$ is the mixture density based on the components $f_1, f_2, ..., f_n$ and $\mathbf{p} = (p_1, ..., p_n)$, $p_i \geq 0$, $\sum p_i = 1$.

Proof: By definition, we have

$$\begin{aligned} JG_\alpha(f_1, f_2, ..., f_n; \mathbf{p}) &= \sum_{i=1}^{n} p_i G_\alpha(f_i) - G_\alpha\left(\sum_{i=1}^{n} p_i f_i\right) = \sum_{i=1}^{n} p_i \int f_i^\alpha(x)dx - \int \left(\sum_{i=1}^{n} p_i f_i(x)\right)^\alpha dx \\ &= \sum_{i=1}^{n} p_i \int f_i^\alpha(x)dx - \int f_T^\alpha(x)dx. \end{aligned}$$

But, on the other hand, we have

$$\begin{aligned} \sum_{i=1}^{n} p_i D_\alpha(f_i : f_T) &= \sum_{i=1}^{n} p_i \int f_i^\alpha(x)\left(1 - \left(\frac{f_T(x)}{f_i(x)}\right)^{\alpha-1}\right) dx = \sum_{i=1}^{n} p_i \int f_i^\alpha(x) - \int f_T^{\alpha-1}(x) \sum_{i=1}^{n} p_i f_i(x)dx \\ &= \sum_{i=1}^{n} p_i \int f_i^\alpha(x) - \int f_T^\alpha(x)dx, \end{aligned}$$

as required.

The following theorem shows the relations between Jensen-Taneja, Jensen-Shannon and Jensen-extropy measures with $JG_\alpha(f_1, f_2, ..., f_n; \mathbf{p})$. For more details, see Sharma and Taneja (1975) and Lin (1991).

Theorem 12. *Let the variables $X_1, X_2, .., X_n$ have density functions $f_1, f_2, ..., f_n$, respectively, and $JG_\alpha(f_1, f_2, ..., f_n; \mathbf{p})$ be the corresponding Jensen-information generating function. Then:*

(i) $\frac{\partial}{\partial\alpha} JG_\alpha(f_1, f_2, ..., f_n; \mathbf{p}) = \frac{1}{2^{\alpha-1}} JT(f_1, f_2, ..., f_n, \mathbf{p})$;

(ii) $\frac{\partial}{\partial\alpha} JG_\alpha(f_1, f_2, ..., f_n; \mathbf{p})|_{\alpha=1} = JS(f_1, f_2, ..., f_n; \mathbf{p})$;

(iii) $\frac{1}{2} JG_2(f_1, f_2, ..., f_n; \mathbf{p}) = JE(f_1, f_2, ..., f_n; \mathbf{p})$,

where $JT(f_1, f_2, ..., f_n; \mathbf{p})$, $JS(f_1, f_2, ..., f_n; \mathbf{p})$ and $JE(f_1, f_2, ..., f_n; \mathbf{p})$ are Jensen-Taneja, Jensen-Shannon and Jensen-extropy measures, respectively.

Proof: From the definition of Taneja entropy $H_r^T(f) = -2^{r-1} \int \left(\log f(x)\right)^r f(x)dx$, $r > 1$, and by using the definition of $JG_\alpha(f_1, f_2, ..., f_n; \mathbf{p})$ and differentiating with respect to α, we obtain

$$\begin{aligned} \frac{\partial}{\partial\alpha} JG_\alpha(f_1, f_2, ..., f_n; \mathbf{p}) &= \frac{\partial}{\partial\alpha}\left\{\sum_{i=1}^{n} p_i G_\alpha(f_i) - G_\alpha\left(\sum_{i=1}^{n} p_i f_i\right)\right\} \\ &= \sum_{i=1}^{n} p_i \int f^\alpha(x) \log f(x)dx - \int \left(\sum_{i=1}^{n} p_i f_i(x)\right)^\alpha \log\left(\sum_{i=1}^{n} p_i f_i(x)\right) dx \\ &= \frac{1}{2^{\alpha-1}}\left\{H_\alpha^T\left(\sum_{i=1}^{n} p_i f_i\right) - \sum_{i=1}^{n} p_i H_\alpha^T(f_i)\right\}, \end{aligned}$$

which proves Part (i). Setting $\alpha = 1$ in Part (i), we get

$$\frac{\partial}{\partial\alpha} JG_\alpha(f_1, f_2, ..., f_n; \mathbf{p})|_{\alpha=1} = \sum_{i=1}^{n} p_i \int f(x) \log f(x)dx - \int \left(\sum_{i=1}^{n} p_i f_i(x)\right) \log\left(\sum_{i=1}^{n} p_i f_i(x)\right) dx$$

$$= H\left(\sum_{i=1}^{n} p_i f_i\right) - \sum_{i=1}^{n} p_i H(f_i) = JS(f_1, f_2, ..., f_n; \mathbf{p}),$$

which proves Part (ii). Next, we have

$$\begin{aligned} JG_2(f_1, ..., f_n; \mathbf{p}) &= \left\{ \sum_{i=1}^{n} p_i G_2(f_i) - G_2(\sum_{i=1}^{n} p_i f_i) \right\} = \sum_{i=1}^{n} p_i \int f_i^2(x)dx - \int \left(\sum_{i=1}^{n} p_i f_i(x) \right)^2 dx \\ &= 2\left\{ J\left(\sum_{i=1}^{n} p_i f_i(x) \right) - \sum_{i=1}^{n} p_i J(f_i) \right\} = 2JE(f_1, f_2, ..., f_n; \mathbf{p}), \end{aligned}$$

which proves Part (iii).

4. Information generating function for residual lifetime

Duration study is a subject of interest in many branches of science such as reliability, survival analysis, actuary, economics and business. Let X be a nonnegative random variable denoting a duration such as a lifetime having distribution function F and the probability density function f. Capturing effects of the age t of an individual or a device under study on the information about the remaining lifetime is important for different reasons. For example, when a component or a system of components is working at time t, one will be interested in the study of information of the density of the lifetime of component or system beyond t, see Kharazmi and Asadi (2018). For this reason, we are interested in investigating the IG measure and its properties for residual lifetime distribution of a given system. Let X be the lifetime of a system and $X_t = (X - t|X > t)$ be the residual lifetime of the system at time t. Then, the density and survival functions of X_t are given by $f_t(x) = \frac{f(x)}{\bar{F}(t)}$, $x > t$, and $\bar{F}_t(x) = \frac{\bar{F}(x)}{\bar{F}(t)}$, respectively. For more details, see Barlow and Proschan (1975).

Definition 4. Let X_t be a residual life variable with an absolutely continuous density function $f_t(x)$. Then, the residual information generating (RIG) function of $f_t(x)$ is defined as

$$G_\alpha(f;t) = \frac{\int_t^\infty f^\alpha(x)dx}{\bar{F}^\alpha(t)} = \frac{E\left(f^{\alpha-1}(X)|X > t\right)}{\bar{F}^{\alpha-1}(t)} = \frac{E[r_F^{\alpha-1}(X_\alpha)|X_\alpha > t]}{\alpha}, \tag{11}$$

where X_α is proportional hazards random variable corresponding to X with survival function $\bar{F}^\alpha(x)$, and $r_F(x) = \frac{f(x)}{\bar{F}(x)}$ is the hazard function of X.

Definition 5. Let X and Y be two variables with residual information generating functions $G_\alpha(f;t)$ and $G_\alpha(g;t)$, respectively. Then, X is said to be less than Y in residual information generating function, denoted by $X \leq_{RIG} Y$, if $G_\alpha(f;t) \leq G_\alpha(g;t)$, for all $t \geq 0$.

Let X and Y be two non-negative continuous random variables having distribution functions F and G, densities functions f and g, and hazard functions r_F and r_G, respectively. We say that X has larger hazard rate than Y (denoted by $Y \leq_{hr} X$) if $r_G(t) \leq r_F(t)$ for all $t \geq 0$; see, for example Shaked and Shanthikumar (2007) for pertinent details.

Theorem 13. *If $Y \leq_{hr} X$ and either F or G has decreasing failure rate (DFR), then for $\alpha \geq 1$, $Y \leq_{RIG} X$.*

Proof: Because $Y \leq_{hr} X$, we have $r_F(t) \leq r_G(t)$ for all $t \geq 0$. In addition, $Y \leq_{hr} X$ implies that $Y_t \leq_{st} X_t$, meaning that for all $x \geq t \geq 0$, $\frac{\bar{F}(x)}{\bar{F}(t)} \geq \frac{\bar{G}(x)}{\bar{G}(t)}$. Now for $\alpha \geq 1$, it can be seen that $X_{\alpha,t} = (X_\alpha - t|X_\alpha > t) \geq_{st} (Y_\alpha - t|Y_\alpha > t) = Y_{\alpha,t}$, where X_α and Y_α are proportional hazards variables corresponding to X and Y, respectively. For $\alpha \geq 1$, by the assumption that F is DFR, we obtain

$$G_\alpha(f;t) = \frac{E\left[r_F^{\alpha-1}(X_\alpha)|X_\alpha > t\right]}{\alpha} \leq \frac{E\left[r_F^{\alpha-1}(Y_\alpha)|Y_\alpha > t\right]}{\alpha} \leq \frac{E\left[r_G^{\alpha-1}(Y_\alpha)|Y_\alpha > t\right]}{\alpha} = G_\alpha(g;t).$$

Let $\bar{F}(x)$ be a survival function of a nonnegative continuous random variable X with finite mean μ. Then, the random variable X_e is said to be the equilibrium random variable corresponding to X if the density function of X_e is given by

$$f_e(x) = \frac{\bar{F}(x)}{\mu}, \quad x > 0; \tag{12}$$

see Unnikrishnan Nair et al. (2013) for further details. The following theorem shows a representation for the residual information generating function of equilibrium distribution.

Theorem 14. *Let X_e be the equlibrium variable corresponding to X with density function f_e as given in (12). Then,*

$$G_\alpha(f_e,t) = \frac{E\left[X_\alpha - t|X_\alpha > t\right]}{\left\{E\left[X - t|X > t\right]\right\}^\alpha} = \frac{m_{X_\alpha}(t)}{m_X^\alpha(t)}, \tag{13}$$

where X_α is proportional hazards random variable corresponding to X with survival function $\bar{F}^\alpha(x)$, and $m_X(t)$ and $m_{X_\alpha}(t)$ are mean residual lifetimes corresponding to variables X and X_α, respectively.

Proof: By the definition of IG function in (1), we have $G_\alpha(f_e,t) = \int_t^\infty \left\{\frac{f_e(x)}{\bar{F}_e(t)}\right\}^\alpha dx = \frac{\int_t^\infty \frac{\bar{F}^\alpha(x)}{\bar{F}^\alpha(t)}dx}{\left\{\int_t^\infty \frac{\bar{F}(u)}{\bar{F}(t)}du\right\}^\alpha} = \frac{m_{X_\alpha}(t)}{m_X^\alpha(t)}$, as required.

The quantity $R(x) = \int_0^x r_F(t)dt = -\log \bar{F}(x)$ is known in reliability engineering as cumulative hazard rate, and plays an important role in the study of ageing properties of lifetime variables.

Theorem 15. *We have*

$$E[f^{\alpha-1}(X)R(X)] = \frac{E_{X_\alpha}[G_\alpha(f,X_\alpha)]}{\alpha}, \tag{14}$$

where X_α is proportional hazards random variable corresponding to X.

Proof: From (11), we have

$$E[f^{\alpha-1}(X)R(X)] = \int_0^\infty f(u)E[f^{\alpha-1}(X)|X > u]du = \int_0^\infty G_\alpha(f,u)f(u)\bar{F}^{\alpha-1}(u)du = \frac{E_{X_\alpha}[G_\alpha(f,X_\alpha)]}{\alpha},$$

as required.

From (14), we find, for example, $Cov(f^{\alpha-1}(X), R(X)) = \frac{E_{X_\alpha}[G_\alpha(f,X_\alpha)]}{\alpha} - G_\alpha(f)$. We also have the correlation coefficient between $f^{\alpha-1}(X)$ and $R(X)$ as $Corr(f^{\alpha-1}(X), R(X)) = \frac{Cov(f^{\alpha-1}(X),R(X))}{Var(f^{\alpha-1}(X))} = \frac{\frac{E_{X_\alpha}[G_\alpha(f,X_\alpha)]}{\alpha} - G_\alpha(f)}{\sqrt{G_{2\alpha-1}(f) - G_\alpha^2(f)}}$.

5. Concluding remarks

In this paper, we have considered information generating function and established some new properties of it. A new divergence measure based on the IG function has been proposed to measure the closeness between two density functions as well as Kullback-Leibler divergence. We have also proposed Jensen-information generating function, whose derivatives generate the well-known Jensen-Shannon, Jensen-Taneja and Jensen-Extropy information measures. We have shown that the Jensen-information generating function can be expressed as a mixture of two versions of the proposed new divergence measure. Finally, we have provided a discussion on the IG measure of residual lifetime variable.