REVSTAT – Statistical Journal
Volume 21, Number 3, July 2023, 303–319
https://doi.org/10.57805/revstat.v21i3.360

On an Induced Distribution and its Statistical Properties

Authors: Brijesh P. Singh
– Department of Statistics, Institute of Science, Banaras Hindu University, India
brijesh@bhu.ac.in

Utpal Dhar Das ✉
– Department of Mathematics, School of Basic Sciences and Research, Sharda University, India
utpal.statmath@gmail.com

Received: December 2020 Revised: November 2021 Accepted: November 2021

Abstract:

- In this study an attempt has been made to propose a way to develop new distribution. For this purpose, we need only idea about distribution function. Some important statistical properties of the new distribution like moments, cumulants, hazard and survival function has been derived. The Rényi entropy, Shannon entropy has been obtained. Also ML estimate of parameter of the distribution is obtained, that is not closed form. Therefore, numerical technique is used to estimate the parameter. Some real data sets are used to check the suitability of the proposed distribution over some other existing one parameter lifetime distributions. The various diagnostic tools such as −2LL, AIC, BIC and K-S test shows that the proposed distribution provides better fit than other distributions for the considered data sets.

Keywords:

- *induced distribution; Bonferroni and Gini index; entropy; generating function; hazard function; MLE; MRLF; order statistics.*

AMS Subject Classification:

- 49A05, 78B26.

✉ Corresponding author.

1. INTRODUCTION

Almost all applied sciences including, biomedical science, engineering, finance, demography, environmental and agricultural sciences, there is a need of statistical analysis and modeling of the data. A number of continuous distributions for modeling lifetime data have been introduced in statistical literature such as Exponential, Lindley, Gamma, Lognormal and Weibull. Among these Gamma and Lognormal distributions are less popular because their survival functions cannot be expressed in closed forms and both require numerical integration. Researchers in probability distribution theory often use a probability distributions based on either their mathematical simplicity or because of their flexibility. Several parametric models are used in the analysis of lifetime data and in the problems associated with the modeling of the failure process. The Exponential distribution is often used to model the time interval between successive random events but Gamma and Weibull distribution is the most widely used model for lifetime distribution due to its flexibility. The exponential distribution is a particular case of the Gamma and Weibull distribution. In order to increase the suitability of the well-known distributions, many authors have proposed different transformations to generate new distributions, it has been an increased interest in defining new generators for univariate continuous distributions by introducing one or more additional shape parameter(s) to the baseline model. This improves the goodness-of-fit of the proposed generated distribution.

In the context of increasing flexibility in distribution, many generalization or transformation methods are available in the literature based on baseline distribution. Ghitany *et al.* [6] developed a two-parameter weighted Lindley distribution and discussed its applications to survival data. Zakerzadeh and Dolati [26] obtained a generalized Lindley distribution and discussed its various properties and applications. Shaw and Buckley [23] proposed a new transformation method by adding one extra parameter and Kumaraswamy [9] gives another method of proposing new distribution by taking baseline distribution. A families of distributions for the median of a random sample drawn from an arbitrary lifetime distribution is introduced by Abd-Elrahman [1]. Since its failure rate function is monotonically increasing with finite limit for this they generalize distribution by making transformation $X = \left(\frac{Y-\delta}{\theta}\right)^{\lambda}$, the parameter δ is a threshold parameter, θ and λ are the scale and the shape parameters, respectively. Gupta *et al.* [7] proposed an exponentiated type distribution by adding one more shape parameter. A new generalization of Lindley distribution, i.e. SSD distribution, appear in Singh *et al.* [25]. In very recent compounded exponential lindley distribution (CEL) has been studied by Singh *et al.* [24]. A new class of distribution by adding two additional shape parameters is found (see Cordeiro *et al.* [4]). Also some well-known generators are the beta-G by Eugene *et al.* [5], gamma-G by Zografos and Balakrishnan [27], the Zografos-Balakrishnan-G family by Nadarajah *et al.* [13].

2. GENESIS OF THE DISTRIBUTION

In this study, an attempt has been made to develop a new continuous distribution using concept discussed by Gupta and Kirmani [8]. Let X be a continuous random variable with the cumulative distribution function (cdf) $F(x)$ and expectation $E(X)$. It is worthwhile to mention here that the $E(X)$ can be defined in terms of cdf of any distribution as follows:

$$E(X) = \int_0^\infty [1 - F(x)]dx.$$

Let us have, for positive x,

$$\int_0^\infty [1 - F(x)]dx = \lim_{k\to\infty} \int_0^k 1.[1 - F(x)]dx.$$

Now, integrating by parts, we have

$$\lim_{k\to\infty} [\{1 - F(k)\}k] + \lim_{k\to\infty} \int_0^k xf(x)dx, \quad \text{where} \quad \frac{d}{dx}[F(x)] = f(x). \tag{2.1}$$

Since $F(\infty) = 1$, $\lim_{k\to\infty} [\{1 - F(k)\}k] = 0$, then

$$\int_0^\infty [1 - F(x)]dx = \lim_{k\to\infty} \int_0^k xf(x)dx = \int_0^\infty xf(x)dx = E(X). \tag{2.2}$$

Keeping the above concept into mind we define a pdf $g^*(x)$ as

$$g^*(x) = \frac{1 - F(x)}{E(X)}, \qquad x > 0. \tag{2.3}$$

If $g^*(x)$ is a pdf then its integration over the range should be equal to 1. Now we have

$$\int_0^\infty g^*(x)dx = \int_0^\infty \frac{[1 - F(x)]}{E(X)}dx = \frac{1}{E(X)} \int_0^\infty [1 - F(x)]dx = \frac{E(X)}{E(X)} = 1.$$

Therefore the generated pdf using the above transformation technique will be a valid pdf. This $g^*(x)$ may be called an induced or equilibrium distribution. Actually this distribution is a particular case of weighted distribution defined by Patil and Rao [14]. According to the Patil and Rao [14], if $f(x;\theta)$ be the probability distribution function of random variable X and the unknown parameter θ the weighted distribution is defined as:

$$f^*(x;\theta) = \frac{w(x)f(x;\theta)}{E[w(x)]}, \qquad x \in \mathbb{R}, \quad \theta > 0,$$

where $w(x)$ is the weight function, and $f(x;\theta)$ is the base line distribution. We know that $1 - F(x) = S(x) = \frac{f(x)}{h(x)}$, i.e. if we take $w(x) = h(x)^{-1}$, we can get the induced distribution defined above in equation number (2.3). This distribution is well connected to its parent distribution and many of the statistical properties can be easily studied.

2.1. Proposed distribution

We consider cdf of one parameter Lindley distribution and using the idea of induced distribution given in the equation (2.3), the pdf and cdf of transformed distribution is given in equations (2.4) and (2.5) respectively:

$$f(x;\theta) = \frac{\theta}{\theta+2}(1+\theta+\theta x)e^{-\theta x}, \tag{2.4}$$

$$F(x;\theta) = 1 - \left[1 + \frac{\theta x}{\theta+2}\right]e^{-\theta x}, \qquad x > 0, \quad \theta > 0. \tag{2.5}$$

In fact this distribution is Garima distribution and already discussed by Shanker [20], which is a mixture of Exponential (θ) and Gamma (2, θ) distribution with mixing proportion $\frac{\theta+1}{\theta+2}$. Also he discussed its various statistical properties.

Therefore in this paper, we consider cdf $F(x)$ of Garima distribution as a base line distribution and try to develop a new distribution. The pdf and cdf of the new distribution is as follows:

$$g(x;\theta) = \frac{\theta}{\theta+3}(2+\theta+\theta x)e^{-\theta x}, \qquad x > 0, \quad \theta > 0, \tag{2.6}$$

and the corresponding cdf is

$$G(x;\theta) = 1 - \left[1 + \frac{\theta x}{\theta+3}\right]e^{-\theta x}, \qquad x > 0, \quad \theta > 0. \tag{2.7}$$

The above distribution is similar to the base line distribution and develop using concept of induced distribution thus named as induced Garima (i-Garima) distribution. This distribution can also be consider as second order induced Lindley distribution. The cdf of i-Garima distribution is displayed in Figure (1).

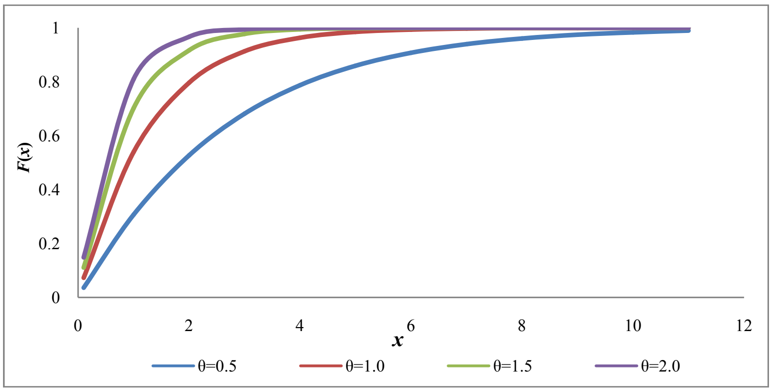

Figure 1: Cumulative distribution function of i-Garima distribution.

The proposed distribution, i.e. i-Garima distribution, can be easily expressed as a mixture of Exponential (θ) and Gamma $(2,\theta)$ as

$$f(x;\theta) = pg_1(x) + (1-p)g_2(x), \tag{2.8}$$

where $p = \frac{\theta+2}{\theta+3}$, $g_1(x) = \theta e^{-\theta x}$, and $g_2(x) = \theta^2 x e^{-\theta x}$.

3. PROPERTIES

The r-th order moments about origin is given by

$$E(X^r) = \int_0^\infty x^r g(x)dx = \frac{\theta}{\theta+3}\int_0^\infty x^r e^{-\theta x}(2+\theta+\theta x)dx.$$

Hence,

$$\mu_r' = \frac{r!}{\theta^r}\frac{(\theta+r+3)}{(\theta+3)}, \qquad r = 1, 2, 3, ... \tag{3.1}$$

First four moments about origin are obtained as:

$$\mu_1' = \frac{1}{\theta}\frac{(\theta+4)}{(\theta+3)}, \qquad \mu_2' = \frac{2}{\theta^2}\frac{(\theta+5)}{(\theta+3)}, \qquad \mu_3' = \frac{6}{\theta^3}\frac{(\theta+6)}{(\theta+3)}, \qquad \mu_4' = \frac{24}{\theta^4}\frac{(\theta+7)}{(\theta+3)}.$$

Using the above expression we get the four moments about mean, i.e. central moments of the proposed distribution are given by

$$\mu_1 = \frac{\theta+4}{\theta(\theta+3)}, \qquad \mu_2 = \frac{\theta^2+8\theta+14}{\theta^2(\theta+3)^2},$$
$$\mu_3 = \frac{2\left(\theta^3+12\theta^2+42\theta+46\right)}{\theta^3(\theta+3)^3}, \qquad \mu_4 = \frac{3\left(3\theta^4+48\theta^3+260\theta^2+592\theta+488\right)}{\theta^4(\theta+3)^4}.$$

The coefficient of variation (CV), coefficient of skewness $\sqrt{\beta_1}$, coefficient of kurtosis β_2 and index of dispersion γ of proposed distribution are obtained as:

$$CV = \frac{\sigma}{\mu_1} = \frac{\sqrt{\theta^2+8\theta+14}}{\theta+4}, \qquad \sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{2\left(\theta^3+12\theta^2+42\theta+46\right)}{\left(\theta^2+8\theta+14\right)^{\frac{3}{2}}},$$
$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\left(3\theta^4+48\theta^3+260\theta^2+592\theta+488\right)}{\left(\theta^2+8\theta+14\right)^2}, \qquad \gamma = \frac{\mu_2}{\mu_1} = \frac{\left(\theta^2+8\theta+14\right)}{\theta(\theta+3)(\theta+4)}.$$

The coefficient of variation (CV), index of dispersion (γ), coefficient of skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) are calculated for different values of θ. Coefficient of variation (CV) is observed less than 1 for all values of θ. Coefficient of skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) are found more than 1 and 3 respectively for different values of θ, therefore the proposed distribution is positively skewed and leptokurtic. The index of dispersion (γ) shows that the proposed distribution is under-dispersed as well as over-dispersed. It is observed that for θ= 1.1474, the value of γ is 1. For $\theta > 1.1474$, the distribution is under-dispersed and for $\theta < 1.1474$, it is over-dispersed. The graph for CV, γ, $\sqrt{\beta_1}$ and β_2 for different values of θ are shown in Figure 2.

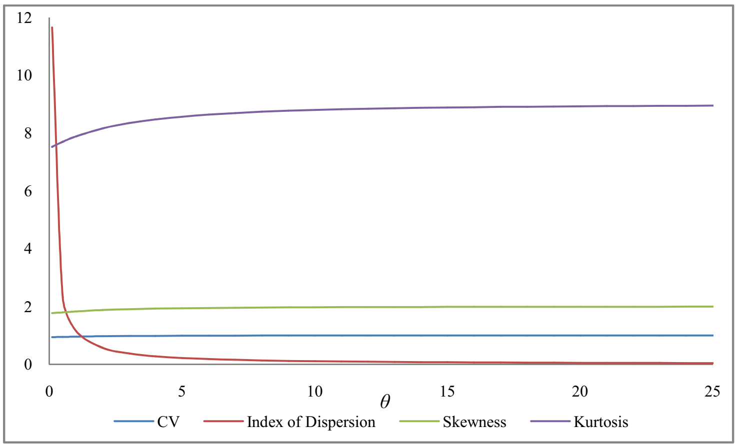

Figure 2: Graph of the CV, γ, β_1 and β_2 for different values of θ.

3.1. Generating functions

The moment generating function $M_x(t)$, characteristic function $\Phi_x(t)$ and cumulant generating function $\kappa_x(t)$ of proposed distribution are given by:

$$M_x(t) = \left[1 - \frac{(2+\theta)t}{(3+\theta)\theta}\right]\left(1 - \frac{t}{\theta}\right)^{-2}, \qquad \left|\frac{t}{\theta}\right| < 1, \tag{3.2}$$

$$\Phi_x(t) = \left[1 - \frac{(2+\theta)it}{(3+\theta)\theta}\right]\left(1 - \frac{it}{\theta}\right)^{-2}, \qquad i = \sqrt{-1}, \tag{3.3}$$

$$\kappa_x(t) = \log\left(1 - \frac{(2+\theta)it}{(3+\theta)\theta}\right) - 2\log\left(1 - \frac{it}{\theta}\right). \tag{3.4}$$

By series expansion of $\log(1-x) = -\sum_{r=0}^{\infty} \frac{x^r}{r}$, we get

$$\begin{aligned}\kappa_x(t) &= -\sum_{r=0}^{\infty}\left(\frac{(2+\theta)}{(3+\theta)\theta}\right)^r \frac{(it)^r}{r} + 2\sum_{r=0}^{\infty}\frac{\left(\frac{it}{\theta}\right)^r}{r} \\ &= 2\sum_{r=0}^{\infty}\frac{(r-1)!}{\theta^r}\frac{(it)^r}{r!} - \sum_{r=0}^{\infty}(r-1)!\left[\frac{\theta+2}{\theta(\theta+3)}\right]^r\frac{(it)^r}{r!}.\end{aligned}$$

Hence r-th cumulant of i-Garima distribution is given by

$$\begin{aligned}\kappa_r &= \text{coefficient of } \frac{(it)^r}{r!} \text{ in } \kappa_x(t) \\ &= 2\frac{(r-1)!}{\theta^r} - \frac{(r-1)!(\theta+2)^r}{[\theta(\theta+3)]^r}, \qquad r = 1, 2, 3, \ldots\end{aligned}$$

From the above equation we have four moments, that are the same as obtained earlier by equation (3.1):

$$\mu_1 = \kappa_1 = \frac{\theta+4}{\theta(\theta+3)}, \qquad \mu_2 = \kappa_2 = \frac{\theta^2+8\theta+14}{\theta^2(\theta+3)^2},$$

$$\mu_3 = \kappa_3 = \frac{2\left(\theta^3+12\theta^2+42\theta+46\right)}{\theta^3(\theta+3)^3}, \qquad \mu_4 = \kappa_4 + 3\kappa_2^2 = \frac{3\left(3\theta^4+48\theta^3+260\theta^2+592\theta+488\right)}{\theta^4(\theta+3)^4}.$$

3.2. Hazard rate and mean residual life function

Let X be a random variable with pdf $g(x)$ and cdf $G(x)$. The hazard function is given as

$$h(x) = \lim_{\Delta x \to \infty} \frac{P(X < x + \Delta x | X > x)}{\Delta x} = \frac{g(x;\theta)}{1 - G(x;\theta)}. \tag{3.5}$$

After using pdf and cdf of i-Garima distribution in above expression we get the hazard rate function $h(x)$ of i-Garima distribution as

$$h(x) = \frac{\theta(2 + \theta + \theta x)}{(3 + \theta + \theta x)}, \tag{3.6}$$

taking limit as $x \to 0$ in (3.6), we get

$$\lim_{x \to 0} h(x) = \lim_{x \to 0} \theta \left[1 - \frac{1}{(3 + \theta + \theta x)}\right] = \theta \left[1 - \frac{1}{(3 + \theta)}\right] > 0, \qquad \theta \in \mathbb{R}^+,$$

and for $x \to \infty$ we get

$$\lim_{x \to \infty} h(x) = \lim_{x \to \infty} \theta \left[1 - \frac{1}{(3 + \theta + \theta x)}\right] = \theta > 0, \qquad \theta \in \mathbb{R}^+.$$

Hence, $h(x) > 0$ for $x > 0$, $\theta > 0$. Therefore, $h(x)$ is an increasing function. The figure of hazard function is displayed in the Figure (3).

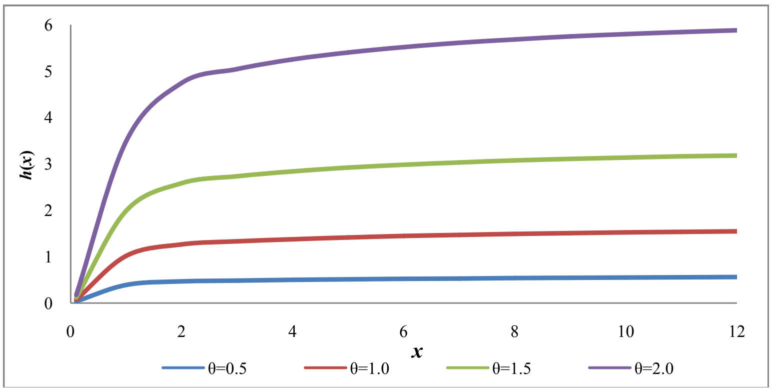

Figure 3: Hazard function of i-Garima distribution.

Now the mean residual life function (MRLF) is given as (3.7). We know that if a component of age t, the remaining lifetime after age t will be random. The expected value of the random life time is called the mean residual life and the mathematical form is known as MRLF. This may be more relevant than the hazard rate function in the study of repairable or replacement time. The MRLF provide idea about the entire residual life distribution or life expectancy, whereas the hazard rate is related only to the risk of immediate failure.

We have

$$m(x) = E[X - x|X > x] = \frac{1}{1 - G(x;\theta)} \int_x^\infty [1 - G(t;\theta)]dt,$$

$$m(x) = \frac{(4 + \theta + \theta x)}{\theta(3 + \theta + \theta x)}. \tag{3.7}$$

If $x = 0$, we get, $m(0) = \frac{\theta+4}{\theta(\theta+3)}$ which is $E(X)$ of the proposed distribution and also $m(x)$ is decreasing function for all $x > 0$ and $\theta > 0$. The graph of MRLF of i-Garima distribution is given in the Figure (4), which is decreasing type.

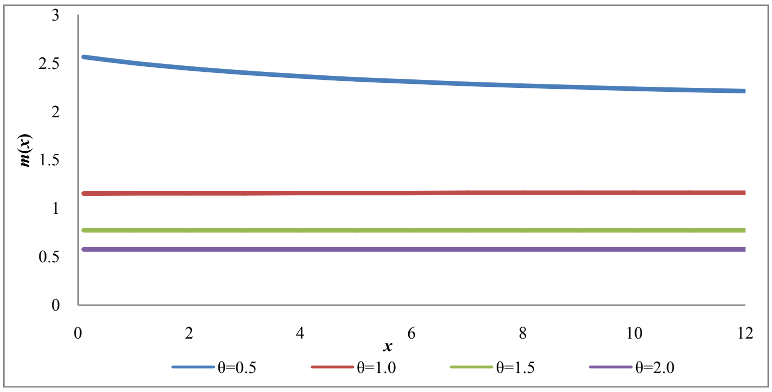

Figure 4: Mean residual life function (MRLF) of i-Garima distribution.

3.3. Quantile function

Theorem 3.1. *If* $X \sim i$-Garima(θ), *then Quantile function of* X *is defined as*

$$Q(p) = -1 - \frac{3}{\theta} - \frac{1}{\theta} W_{-1}\Big(-(1-p)(\theta+3)e^{-(-\theta+3)}\Big),$$

where $p \in (0,1)$ *and* W_{-1} *is the negative branch of the Lambert* W *function.*

Proof: Let

$$Q(p) = F^{-1}(p), \qquad p \in (0,1).$$

The quantile function, say $q(p)$, defined by $G(Q(p)) = p$ is the root of the equation

$$1 - \left(1 + \frac{\theta Q(p)}{\theta + 3}\right)e^{-\theta Q(p)} = p,$$

$$[3 + \theta + \theta Q(p)]e^{-\theta Q(p)} = (1-p)(\theta + 3).$$

Multiplying both sides by $-e^{-(\theta+3)}$ we get

$$-[3+\theta+\theta Q(p)]e^{-(3+\theta+\theta Q(p))} = -(1-p)(\theta+3)e^{-(3+\theta)}.$$

Now $(3+\theta+\theta Q(p)) > 1$, $\forall\theta > 0$, $Q(p) > 0$. By applying W-function defined as the solution of the equation $w(z)e^{W(z)} = z$, the above equation can be written as

$$W_{-1}\Big(-(1-p)(\theta+3)e^{-(-\theta+3)}\Big) = -(3+\theta+\theta Q(p)),$$

where and $W_{-1}(\cdot)$ is the negative branch of the Lambert W function and we get the required result:

$$Q(p) = -1 - \frac{3}{\theta} - \frac{1}{\theta}W_{-1}\Big(-(1-p)(\theta+3)e^{-(-\theta+3)}\Big). \tag{3.8}$$ □

3.4. Stochastic ordering

Stochastic ordering of a continuous random variable is an important tool to judging their comparative behaviour. A random variable X is said to be smaller than a random variable Y, when:

- **(i)** Stochastic order $X \leq_{st} Y$ if $F_X(x) \geq F_Y(x)$ for all x;
- **(ii)** Hazard rate order $X \leq_{hr} Y$ if $h_X(x) \geq h_Y(x)$ for all x;
- **(iii)** Mean residual life order $X \leq_{mrl} Y$ if $m_X(x) \geq m_Y(x)$ for all x;
- **(iv)** Likelihood ratio order $X \leq_{lr} Y$ if $\frac{f_X(x)}{f_Y(x)}$ decreases in x.

The following results by Shaked and Shanthikumar [16] are well known for introducing stochastic ordering of distributions:

$$\begin{array}{ccccc} X \leq_{lr} Y & \Longrightarrow & X \leq_{hr} Y & \Longrightarrow & X \leq_{mrl} Y \\ & & \Downarrow & & \\ & & X \leq_{st} Y. & & \end{array}$$

With the help of the following theorem we claim that i-Garima distribution is ordered with respect to strongest likelihood ratio ordering.

Theorem 3.2. *Let $X \sim i$-Garima(θ_1) distribution and $Y \sim i$-Garima(θ_2) distribution. If $\theta_1 > \theta_2$ then $X \leq_{lr} Y$ and therefore $X \leq_{hr} Y$, $X \leq_{mrl} Y$ and $X \leq_{st} Y$.*

Proof: We have

$$\frac{f_X(x)}{f_Y(x)} = \frac{\theta_1(\theta_2+3)}{\theta_2(\theta_1+3)}\left(\frac{2+\theta_1+\theta_1 x}{2+\theta_2+\theta_2 x}\right)e^{-(\theta_1-\theta_2)x}, \qquad x > 0.$$

Now taking log both sides we get

$$\log\left[\frac{f_X(x)}{f_Y(x)}\right] = \log\left[\frac{\theta_1(\theta_2+3)}{\theta_2(\theta_1+3)}\right] \log\left[\frac{2+\theta_1+\theta_1 x}{2+\theta_2+\theta_2 x}\right] - (\theta_1-\theta_2)x.$$

By differentiating both sides we get

$$\frac{d}{dx}\log\left[\frac{f_X(x)}{f_Y(x)}\right] = \frac{2(\theta_1-\theta_2)}{(2+\theta_1+\theta_1 x)(2+\theta_2+\theta_2 x)} - (\theta_1-\theta_2).$$

Thus, for $\theta_1 > \theta_2$, $\frac{d}{dx}\log\left[\frac{f_X(x)}{f_Y(x)}\right] < 0$. This means that $X \leq_{lr} Y$ and hence $X \leq_{hr} Y$, $X \leq_{mrl} Y$ and $X \leq_{st} Y$. □

3.5. Order statistics

Let $X_1, X_2, ..., X_m$ be a random sample of size m from i-Garima distribution and also let $X_{(1)}, X_{(2)}, ..., X_{(m)}$ be the corresponding order statistics. The pdf and cdf of r-th order statistics, say $Y = X_{(r)}$, are given by

$$\begin{aligned} f_{(r:m)}(y) &= \frac{m!}{(r-1)!(m-r)!}F^{r-1}(y)[1-F(y)]^{m-r}f(y) \\ &= \frac{m!}{(r-1)!(m-r)!}\sum_{l=0}^{m-r}\binom{m-r}{l}(-1)^l F^{r+l-1}(y)f(y) \end{aligned} \tag{3.9}$$

and

$$\begin{aligned} F_{(r:m)}(y) &= \sum_{j=r}^{m}\binom{m}{j}F^j(y)[1-F(y)]^{m-j} \\ &= \sum_{j=r}^{m}\sum_{l=0}^{m-j}\binom{m}{j}\binom{m-j}{l}(-1)^l F^{j+l}(y) \end{aligned} \tag{3.10}$$

respectively, for $r = 1(1)\,m$.

Based on equations (3.9) and (3.10) the pdf and cdf of r-th order statistics of i-Garima distribution is given in equations (3.11) and (3.12):

$$f_{(r:m)}(y) = \frac{m!\theta(3+\theta+\theta x)e^{-\theta x}}{(\theta+3)(r-1)!(m-r)!}\sum_{l=0}^{m-r}\binom{m-r}{l}\left[1-\frac{\theta x+(\theta+3)}{(\theta+3)}e^{-\theta x}\right]^{r+l-1} \tag{3.11}$$

and

$$F_{(r:m)}(y) = \sum_{j=r}^{m}\sum_{l=0}^{m-j}\binom{m}{j}\binom{m-j}{l}\left[1-\frac{\theta x+(\theta+3)}{(\theta+3)}e^{-\theta x}\right]^{j+l}. \tag{3.12}$$

3.6. Bonferroni and Lorenz curves

Let the random variable X is non-negative with a continuous and twice differentiable cumulative function. The Bonferroni [3] curve of the random variable X is defined as

$$B(p) = \frac{1}{p\mu}\int_0^q xg(x)dx = \frac{1}{p\mu}\left[\int_0^\infty xg(x)dx - \int_q^\infty xg(x)dx\right] = \frac{1}{p\mu}\left[\mu - \int_q^\infty xg(x)dx\right] \tag{3.13}$$

and the Lorenz curve (see Lorenz [12]) is defined by

$$L(p) = \frac{1}{\mu}\int_0^q xg(x)dx = \frac{1}{\mu}\left[\int_0^\infty xg(x)dx - \int_q^\infty xg(x)dx\right] = \frac{1}{\mu}\left[\mu - \int_q^\infty xg(x)dx\right] \tag{3.14}$$

where $q = G^{-1}(p)$ and $\mu = E(X)$, $p \in (0, 1]$.

The Gini index is given by

$$G = 1 - \frac{1}{\mu}\int_0^\infty (1 - G(x))^2 dx = \frac{1}{\mu}\int_0^\infty G(x)(1 - G(x))dx. \tag{3.15}$$

The Bonferroni, Lorenz curve and Gini index have application not only in economics to study income and poverty, but also in other fields like reliability, population studies, insurance, and medicine. Using the equations (3.13), (3.14) and (3.15) we get the Bonferroni curve, Lorenz curve and the Gini index as:

$$B(p) = \frac{1}{p}\left[1 - \frac{\{\theta^2 q^2 + (\theta^2 + 4\theta)q + (\theta + 4)\}e^{-\theta q}}{\theta + 4}\right], \tag{3.16}$$

$$L(p) = 1 - \frac{\{\theta^2 q^2 + (\theta^2 + 4\theta)q + (\theta + 4)\}e^{-\theta q}}{\theta + 4}, \tag{3.17}$$

$$G = \frac{2\theta^2 + 16\theta + 29}{4(\theta + 3)(\theta + 4)}. \tag{3.18}$$

4. ENTROPIES

Entropy, measures the variation in uncertainties associated with a random variable of a probability distributions. Rényi's and Shannon entropy are widely used to understand the uncertainty involved in random variables.

4.1. Rényi entropy

If X is a continuous random variable having probability density function $g(\cdot)$, then the Rényi Entropy (see Rényi [15]) is defined as

$$e(\eta) = \frac{1}{1 - \eta}\log\left[\int_0^\infty g^\eta(x)dx\right], \tag{4.1}$$

where $\eta > 0$ and $\eta = 0$.

The Rényi entropy for the i-Garima distribution is defined as

$$e(\eta) = \frac{1}{1-\eta}\log\left[\int_0^\infty \left(\frac{\theta}{\theta+3}\right)^\eta (2+\theta+\theta x)^\eta e^{-\eta\theta x} dx\right]$$

$$(4.2) \qquad = \frac{1}{1-\eta}\log\left[\int_0^\infty \frac{\theta^\eta(\theta+2)^\eta}{(\theta+3)^\eta}\left(1+\frac{\theta x}{\theta+2}\right)^\eta e^{-\eta\theta x} dx\right].$$

Now from the above equation (4.2), applying binomial expansion $(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$, we get

$$\frac{1}{1-\eta}\log\left[\int_0^\infty \frac{\theta^\eta(\theta+2)^\eta}{(\theta+3)^\eta}\sum_{j=0}^{\eta}\binom{\eta}{j}\left(\frac{\theta x}{\theta+2}\right)^j e^{-\eta\theta x} dx\right],$$

$$(4.3) \qquad \text{i.e.} \quad \frac{1}{1-\eta}\log\left[\sum_{j=0}^{\eta}\binom{\eta}{j}\frac{\theta^{\eta+j}(\theta+2)^{\eta-j}}{(\theta+3)^\eta}\int_0^\infty x^j e^{-\eta\theta x} dx\right].$$

After solving equation (4.3), we get the required results in equation (4.4):

$$(4.4) \qquad = \frac{1}{1-\eta}\log\left[\sum_{j=0}^{\eta}\binom{\eta}{j}\frac{\theta^{\eta-1}(\theta+2)^{\eta-j}}{(\theta+3)^\eta}\frac{\Gamma(j+1)}{(\eta)^{j+1}}\right],$$

since $\int_0^\infty x^{n-1}e^{-\theta x}dx = \frac{\Gamma(n)}{\theta^n}$.

4.2. Shannon entropy

The Shannon entropy (see Shannon [22]) of i-Garima distribution is given as

$$\Omega = E(-\log x) = -\int_0^\infty \log(f(x))f(x)dx$$

$$= -\log\left(\frac{\theta}{\theta+3}\right)\int_0^\infty f(x)dx - \int_0^\infty \log(2+\theta+\theta x)f(x)dx + \int_0^\infty \theta x f(x)dx$$

$$(4.5) \qquad = -\log\left(\frac{\theta}{\theta+3}\right) - \log(\theta+2) - \int_0^\infty \log\left(1+\frac{\theta x}{\theta+2}\right)f(x)dx + \theta E(x).$$

Here, $E(X) = \frac{\theta+4}{\theta(\theta+3)}$, mean of the distribution. Applying $\log(1+x) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{x^n}{n}$, $|x|<1$, in equation (4.5), we get

$$= -\log\left(\frac{\theta(\theta+2)}{\theta+3}\right) + \left(\frac{\theta+4}{\theta+3}\right) - \frac{\theta}{\theta+3}\int_0^\infty\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\left(\frac{\theta x}{\theta+2}\right)^k(2+\theta+\theta x)e^{-\theta x}dx$$

$$= -\log\left(\frac{\theta(\theta+2)}{\theta+3}\right) + \left(\frac{\theta+4}{\theta+3}\right) - \frac{\theta}{\theta+3}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\left(\frac{\theta}{\theta+2}\right)^k\int_0^\infty x^k(2+\theta+\theta x)e^{-\theta x}dx.$$

After the simplification above, we obtained Shannon entropy as

$$\Omega = \left(\frac{\theta+4}{\theta+3}\right) - \log\left(\frac{\theta(\theta+2)}{\theta+3}\right) - \frac{1}{\theta+3}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\frac{k!(\theta+k+3)}{(\theta+2)^k}. \tag{4.6}$$

5. STRESS-STRENGTH RELIABILITY

Stress-strength model describes the life of a system of component having a random strength X and random stress Y. If stress is more than strength, the system of component fails immediately. The measure of system reliability $R = P(Y < X)$ is also known as stress-strength parameter. It is used in engineering science such as deterioration of any structures, motors, static fatigue of ceramic components and aging of concrete pressure vessels.

Let X and Y be independently distributed, with $X \sim i$-Garima(θ_1) and $Y \sim i$-Garima(θ_2). The CDF F_1 of X and pdf f_2 of Y are obtained from equations (2.7) and (2.6), respectively. Then stress-strength reliability R is obtained as

$$\begin{aligned} R &= P(Y < X) = \int_0^\infty P(Y < X|X = x) f_x(X)dx = \int_0^\infty f(x;\theta_1)F(x;\theta_2)dx \\ &= 1 - \frac{\theta_1\left[(\theta_1\theta_2 + 3\theta_1 + 2\theta_2 + 6)(\theta_1+\theta_2)^2 + (2\theta_1\theta_2 + 3\theta_1 + 2\theta_2)(\theta_1+\theta_2) + 2\theta_1\theta_2\right]}{(\theta_1+3)(\theta_2+3)(\theta_1+\theta_2)^3}. \end{aligned} \tag{5.1}$$

6. MAXIMUM LIKELIHOOD ESTIMATION

Let $(x_1, x_2, ..., x_n)$ be a random sample from $X \sim i$-Garima(θ). The likelihood function, L, is obtained as

$$L = \left(\frac{\theta}{\theta+3}\right)^n \prod_{i=1}^{n}(2+\theta+\theta x_i)e^{-\theta\sum_{i=1}^{n}x_i}. \tag{6.1}$$

Taking log both sides of equation (6.1) we get

$$\log L = n\log\left(\frac{\theta}{\theta+3}\right) + \sum_{i=1}^{n}\log(2+\theta+\theta x_i) - \theta\sum_{i=1}^{n}x_i. \tag{6.2}$$

Now differentiating both sides of (6.2) by θ we get

$$\frac{d(\log L)}{d\theta} = \frac{3n}{\theta^2+3\theta} + \sum_{i=1}^{n}\frac{1+x_i}{2+\theta+\theta x_i} - n\bar{x} = 0, \tag{6.3}$$

where $\bar{x}$ is the sample mean. The maximum likelihood estimate $(\hat{\theta})$ of θ is the solution of the equation (6.3). Since this is a non-linear equation, thus we solve this by numerical method.

7. EMPIRICAL ILLUSTRATIONS AND GOODNESS OF FIT

In this section, we present applications of the proposed distribution and their competent models for two real data sets to illustrate their potentiality. We estimate the unknown parameters of the model by the maximum likelihood estimation (MLE) using Newton–Raphson method. First data is about vinyl chloride obtained from clean up gradient monitoring wells in mg/l, provided by Bhaumik *et al.* [2], and second data set represents completed remission times (in months) of a random sample of 128 bladder cancer patients reported in Lee and Wang [10]. The summary measures of the two data sets are given below in Table 1.

Table 1: Summary measures of two data sets.

Datasets	n	mean	sd	median	skewness	kurtosis	min	max
1st data set	34	1.953	1.879	1.150	1.604	5.005	0.10	8.000
2nd data set	128	9.209	10.40	6.280	3.399	19.39	0.08	79.05

Table 1 reveals that both data sets are positively skewed and leptokurtic. First data set is under-dispersed however second data set is over-dispersed. We applied the i-Garima distributions for the above data sets and compared the results with some other competent distributions (see Lindley [11], Shanker [17, 18, 19, 20, 21]).

The goodness of fit of the i-Garima distribution has been explained for two real data sets using $-2LL$ (-2log likelihood), AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and K-S Statistic (Kolmogorov-Smirnov Statistic). The estimate of these have been computed and shown in Tables 2 and 3, respectively. Smaller values of the AIC and BIC indicates better model fittings. The formulae for computing AIC, BIC, and K-S Statistics are as follows:

$$AIC = -2LL + 2k, \qquad BIC = -2LL + k\log n, \qquad D = \sup_x |F_n(x) - F_0(x)|,$$

where k = the number of parameters, n = the sample size, and $F_n(x)$ = empirical distribution function.

Table 2 and 3 reveals that i-Garima distribution provides closer fit for both data sets as it has lower −2LL, AIC, BIC, K-S values and higher p-values corresponding to K-S statistics than the other competitor models. Therefor, the proposed distribution i-Garima will consider as a potential alternative in modeling life time data and can be recommended for modelling data from engineering, medical, biological science and other applied sciences.

Table 2: MLE's, −2LL, AIC, BIC, K-S and p-values of the fitted distributions for the vinyl chloride dataset given by Bhaumik *et al.* [2].

Distribution	Estimate	−2LL	AIC	BIC	K-S	p-value
i-Garima	0.674	111.18	113.18	114.71	0.1039	0.8567
Garima	0.723	111.50	113.50	115.03	0.1135	0.7731
Aradhana	1.133	116.06	118.06	119.59	0.1695	0.2826
Sujatha	1.146	115.54	117.54	119.07	0.1640	0.3196
Akash	1.166	115.15	117.15	118.68	0.1564	0.3762
Shanker	0.853	112.91	114.91	116.44	0.1308	0.6062
Lindley	0.199	112.61	114.61	116.13	0.1326	0.5881

Table 3: MLE's, −2LL, AIC, BIC, K-S and p-values of the fitted distributions for the bladder cancer patients data given by Lee and Wang [10].

Distribution	Estimate	−2LL	AIC	BIC	K-S	p-value
i-Garima	0.143	825.57	827.57	830.42	0.0768	0.4374
Garima	0.158	826.49	828.49	831.34	0.0873	0.2835
Aradhana	0.295	868.28	870.28	873.13	0.1713	0.0011
Sujatha	0.303	873.22	875.22	878.08	0.1792	0.0005
Akash	0.315	881.04	883.04	885.89	0.1904	0.0002
Shanker	0.214	841.68	843.68	846.53	0.1243	0.0382
Lindley	0.199	833.79	835.79	838.64	0.1114	0.0832

8. CONCLUSION

Better modeling of the survival data is a major concern for statisticians and applied researchers. As a consequence, a significant progress has been made towards the extension of lifetime models and their application to various data sets. The present study suggests a technique for developing new probability distribution. A Single parameter distribution named i-Garima, is suggested and investigated in this study. Different statistical properties have been derived and studied for the proposed model. Moments about origin and mean have been obtained. The nature of pdf, cdf, hazard rate function and mean residual life function have been measured. The expression of stress-strength reliability is obtained, we can calculate system reliability when stress and strength parameter is known. Bonferroni, Lorenz curves and Gini index of the i-Garima are also measured. Maximum likelihood estimator of the model parameter is derived and obtained through Newton-Raphson method. The Rényi and Shannon entropies, order statistics and stochastic ordering are derived. An application of i-Garima distribution is given using two real lifetime data sets to show the suitability and the goodness of fit. Although the second data set have some censored cases but here we use only completed cases for the analysis. i-Garima provides a better fit over Garima, Aradhana, Sujatha, Akash, Shanker and Lindley distributions. It is realized that the proposed distribution in this study will consider some data sets in view of different censored mechanisms when specific interest comes into survival or reliability aspects. The article also opens a scope for studying under Bayesian paradigm of the parameters under different loss functions. The work in this direction will perform in near future.

ACKNOWLEDGMENTS

Authors extend their appreciation to the associated editor and anonymous referees for their productive suggestions to improve the quality of manuscript.

REFERENCES

[1] Abd-Elrahman, A.M. (2013). Utilizing ordered statistics in lifetime distributions production: a new lifetime distribution and applications, *Journal of Probability and Statistical Science*, **11**(2), 153–164.

[2] Bhaumik, D.K.; Kapur, K. and Gibbons, R.D. (2009). Testing parameters of a gamma distribution for small samples, *Technometrics*, **51**(3), 326–334.

[3] Bonferroni, C.E. (1936). Statistical theory of classification and calculation of the probability, *Publication of the R Institute Superiore of the Science of Economy and Commerce of Florence*, **8**, 3–62.

[4] Cordeiro, G.M.; Castellares, F.; Montenegro, L.C. and de Castro, M. (2013). The beta generalized gamma distribution, *Statistics*, **47**(4), 888–900.

[5] Eugene, N.; Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications, *Communications in Statistics – Theory and Methods*, **31**(4), 497–512.

[6] Ghitany, M.E.; Alqallaf, F.; Al-Mutairi, D.K. and Husain, H.A. (2011). A two-parameter weighted Lindley distribution and its applications to survival data, *Mathematics and Computers in Simulation*, **81**(6), 1190–1201.

[7] Gupta, R.C.; Gupta, P.L. and Gupta, R.D. (1998). Modeling failure time data by Lehman alternatives, *Communications in Statistics – Theory and Methods*, **27**(4), 887–904.

[8] Gupta, R.C. and Kirmani, S.N.U.A. (1990). The role of weighted distributions in stochastic modeling, *Communications in Statistics – Theory and Methods*, **19**(9), 3147–3162.

[9] Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes, *Journal of Hydrology*, **46**(1-2), 79–88.

[10] Lee, E.T. and Wang, J.W. (2003). *Statistical Methods for Survival Data Analysis*, 3rd ed., Hoboken, NJ, Wiley.

[11] Lindley, D.V. (1958). Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society. Series B (Methodological)*, 102–107.

[12] Lorenz, M.O. (1905). Methods of measuring the concentration of wealth, *Publications of the American Statistical Association*, **9**(70), 209–219.

[13] Nadarajah, S.; Cordeiro, G.M. and Ortega, E.M. (2015). The Zografos–Balakrishnan-G family of distributions: mathematical properties and applications, *Communications in Statistics – Theory and Methods*, **44**(1), 186–215.

[14] PATIL, G.P. and RAO, C.R. (1977). The weighted distributions: a survey of their applications, *Applications of Statistics*, 383–405.

[15] RÉNYI, A. (1961). On measures of entropy and information, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1: "Contributions to the Theory of Statistics", 547–561, University of California Press.

[16] SHAKED, M. and SHANTHIKUMAR, J.G. (2007). *Stochastic Orders*, Springer Science & Business Media.

[17] SHANKER, R. (2015). Akash distribution and its applications, *International Journal of Probability and Statistics*, **4**(3), 65–75.

[18] SHANKER, R. (2015). Shanker distribution and its applications, *International Journal of Statistics and Applications*, **5**(6), 338–348.

[19] SHANKER, R. (2015). Aradhana distribution and its applications, *International Journal of Sstatistics and Applications*, **6**(1), 23–28.

[20] SHANKER, R. (2016). Garima distribution and its application to model behavioral science data, *Biometrics & Biostatistics International Journal*, **4**(7), 275–281.

[21] SHANKER, R. (2016). Sujatha distribution and its applications, *Statistics in Transition. New Series*, **17**(3), 391–410.

[22] SHANNON, C.E. (1951). Prediction and entropy of printed English, *Bell System Technical Journal*, **30**(1), 50–64.

[23] SHAW, W.T. and BUCKLEY, I.R. (2007). The alchemy of probability distributions: beyond Gram-Charlier & Cornish-Fisher expansions, and skew-normal or kurtotic-normal distributions, *Submitted, Feb 7*, **64**.

[24] SINGH, B.P.; DHAR DAS, U. and SINGH, S. (2021). A compounded probability model for decreasing hazard and its inferential properties, *Reliability: Theory & Applications*, **16**(2(62)), 230–246.

[25] SINGH, B.P.; SINGH, S. and DHAR DAS, U. (2020). A general class of new continuous mixture distribution and application, *Journal of Mathematical and Computational Science*, **11**(1), 585–602.

[26] ZAKERZADEH, H. and DOLATI, A. (2009). Generalized Lindley distribution, *Journal of Mathematical Extension*, **3**(2), 13–25.

[27] ZOGRAFOS, K. and BALAKRISHNAN, N. (2009). On families of beta-and generalized gamma-generated distributions and associated inference, *Statistical Methodology*, **6**(4), 344–362.