Applied Mathematics E-Notes, 22(2022), 529-542 © ISSN 1607-2510
Available free at mirror sites of http://www.math.nthu.edu.tw/∼amen/

On A Proposal For A New One-Parameter Survival Distribution*

Farrukh Jamal†, Christophe Chesneau‡, Salma Abbas§

Received 15 January 2021

Abstract

This article is devoted to a new one-parameter survival distribution which presents interesting features for statistical modeling. First, we show that it possesses very flexible probability density and hazard rate functions. Specifically, the probability density function can decrease with a heavy right tail, or be unimodal with an upside-down shape along with a possible light left tail. For its part, the hazard rate function has all monotonic forms; it can be increasing, constant or decreasing. The proposed distribution is also connected, in some senses, to the exponential, Weibull and linear failure rate distributions. This connection is highlighted by proving several first-order stochastic ordering results. Then, the moments are discussed both theoretically and numerically. A part is devoted to the related order statistics, with a focus on the two extreme statistics. For the statistical study, an estimate of the parameter is proposed using the maximum likelihood method. Then, the new distribution is fitted to two data sets to check its goodness of fit. Comparisons are made with other one-parameter distributions, namely the exponential, Lindley, one-parameter Weibull and one-parameter linear failure rate distributions, with favorable indicators for the proposed distribution.

1 Introduction

The development of one-parameter survival distributions is still an active topic in distribution theory. Such distributions allow the construction of simple statistical models which can help to understand complex survival phenomena in reliability, biology, medicine, finance, and engineering. Also, they can serve as generators to construct more general families of distributions through diverse mathematical schemes, as developed in [3] and [4]. The most useful examples include the exponential, Maxwell-Boltzmann, chi-squared, inverse-chi-squared, half-normal, Lindley and Rayleigh distributions. Their good modeling behavior motivates the creation of new one-parameter survival distributions, achieving different statistical objectives. In addition, the favorable goodness of fit of some new distributions over conventional distributions with the available datasets gives more credit for conducting research in this direction. In this spirit, the years after 2010 welcome the length-biased exponential distribution in [9], Akash distribution in [19], Ishita distribution in [21], Sujatha distribution in [20], Pranav distribution in [23], modified Lindley distribution in [7], inverse modified Lindley distribution in [8], Prakaamy distribution in [22], Odoma distribution in [15], 2S and 2D Lindley distributions in [6], new generalized Lindley distribution in [2], generalized Weibull-Lindley distribution in [14], and power Ailamujia distribution in [12]. Their theory and applicability are available in the respective papers.

In this article, we take a new line of research on this topic and study a new one-parameter survival distribution. It is defined by the following cumulative distribution function (cdf):

$$F(x;k) = 1 - e^{-x(1+x)^{-k}}, \quad x \geq 0, \tag{1}$$

and $F(x;k) = 0$ for $x < 0$, with $k < 1$ which allows the negative values. For the purposes of this study, it is denoted as distribution JCA or JCA(k) with reference to the initials of the names of the authors.

*Mathematics Subject Classifications: 60E05, 62E15, 62F10.

†Department of Statistics, The Islamia University of Bahawalpur, Punjab, Pakistan

‡Université de Caen, LMNO, Campus II, Science 3, 14032, Caen, France

§Department of Statistics, The Islamia University of Bahawalpur, Punjab, Pakistan

The motivations for the JCA distribution are discussed below. First, in relation to the existing literature, one can notice that $F(x;k) = F_*\left(x(1+x)^{-k}\right)$, where $F_*(x)$ refers to the cdf of the standard exponential distribution with a parameter of 1 denoted by $\mathcal{E}(1)$. Hence, by taking $k=0$, the JCA distribution is reduced to the exponential distribution with parameter 1. Also, by taking $k=-1$, the JCA distribution becomes the linear failure rate distribution with polynomial coefficients in the exponential term equal to 1 (see [17]). In this sense, the parameter k in the JCA distribution allows us to unify these two conventional distributions. Moreover, the following features and results are developed in this study. The probability density function (pdf) of the JCA distribution can decrease with a heavy right tail or be unimodal with an upside down shape. As a good surprise, in this last case, a light left tail may be observed for large values of $-k$. Thus, the pdf of the JCA distribution has an excellent level in terms of modeling flexibility. In addition, the corresponding hazard rate function (hrf) can be increasing, constant or decreasing. On these functional aspects, the JCA distribution is more flexible than a large number of one-parameter survival distributions, including the famous exponential and Lindley distributions. For the theory, we establish a clear hierarchy on the first-order stochastic dominance of the JCA distribution with the exponential, one-parameter Weibull and one-parameter linear failure rate distributions. In addition, the moments of the JCA distribution are derived from theoretical and computational points of view. The related skewness and kurtosis are also discussed. The theory ends with the treatment of the order statistics. In a subsequent part, we will show how the JCA distribution can be used for data analysis. This aspect is initiated by performing a parametric estimation using the maximum likelihood method. Then, the JCA distribution is fitted to two data sets of interest. Its goodness of fit is measured by standard statistical tools: Akaike information criterion (AIC), corrected AIC (CAIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion (HQIC), Anderson-Darling (AD) statistic, Cramér-von Mises (CVM) statistic, Kolmogorov-Smirnov (KS) statistic with its related p-value (see [13]). The results obtained confirm the good modeling quality of the JCA distribution; for the considered data set, its goodness of fit is superior to those of the exponential, Lindley, one-parameter Weibull and one-parameter linear failure rate distributions.

We organize the rest of the paper as follows: Section 2 gives the details of the functions of the JCA distribution. The main theoretical results are contained in Section 3. Section 4 is devoted to the statistical applicability of the JCA distribution, with data fitting. Conclusion is provided in Section 5.

2 JCA Distribution

This section completes the presentation of the JCA distribution and discusses its modeling behavior for data fitting purposes.

2.1 Functions of Interest

Several functions of the JCA distribution play potentially important roles in various branches of probability and statistics. These functions are described in detail in Table 1.

Note that the qf of the JCA distribution is determined by a nonlinear equation. However, it can have an analytical expression for some values of k, such as the relative integer values.

All of the functions in Table 1 are interesting enough to receive special treatment, but here we focus on the pdf and hrf, whose behaviors are of great interest for data fitting.

2.2 Behavior of the PDF

As indicated in Table 1, the pdf of the JCA distribution is

$$f(x;k) = (1+x)^{-k-1}(1+(1-k)x)e^{-x(1+x)^{-k}}, \quad x \geq 0,$$

and $f(x;k)=0$ for $x<0$. Then, for any $k<1$, we have $f(0;k)=1$ and $\lim_{x\to+\infty} f(x;k)=0$. In addition, the following result holds.

Table 1: Important functions of the JCA distribution

Function	Abbreviation	Formula	Expression ($x \geq 0$ or $y \in (0,1)$)
Survival function	sf	$1 - F(x;k)$	$S(x;k) = e^{-x(1+x)^{-k}}$
Odds function	of	$\frac{S(x;k)}{F(x;k)}$	$O(x;k) = \left(e^{x(1+x)^{-k}} - 1\right)^{-1}$
Cumulative hazard rate function	chrf	$-\log[S(x;k)]$	$H(x;k) = x(1+x)^{-k}$
Probability density function	pdf	$F(x;k)'$	$f(x;k) = (1+x)^{-k-1}(1+(1-k)x)e^{-x(1+x)^{-k}}$
Hazard rate function	hrf	$\frac{f(x;k)}{S(x;k)}$	$h(x;k) = (1+x)^{-k-1}(1+(1-k)x)$
Reverse hazard rate function	rhrf	$\frac{f(x;k)}{F(x;k)}$	$h(x;k) = (1+x)^{-k-1}(1+(1-k)x)O(x;k)$
Quantile function	qf	$F^{-1}(x;k)$	$Q(y;k) \Rightarrow Q(y;k)\,(1+Q(y;k))^{-k} = -\log(1-y)$

Proposition 1

- *For $k \in [0,1)$, $f(x;k)$ is decreasing and the JCA distribution is heavy right-tailed.*
- *For $k < 0$, $f(x;k)$ has a maximum point and is upside-down shaped; in this case, the JCA distribution is unimodal.*

Proof. For $x \geq 0$, the derivative of $f(x;k)$ can be expressed as

$$f(x;k)' = -e^{-x(1+x)^{-k}}(1+x)^{-2(k+1)}\left[((1-k)x+2)(k(1+x)^k + (1-k)x) + 1\right]. \quad (2)$$

Therefore, for $k \in [0,1)$, the term in brackets is positive, implying that $f(x;k)' \leq 0$ and this entails that $f(x;k)$ is decreasing. Moreover, for any $t > 0$, we have $\lim_{x\to+\infty} e^{tx} f(x;k) = +\infty$ because the term e^{tx} is dominant, implying that $\int_0^{+\infty} e^{tx} f(x;k)dx = +\infty$; the JCA distribution is heavy right-tailed.

For $k < 0$, the term in brackets in (2) can vanish at a point x_m, that is, $((1-k)x_m + 2)(k(1+x_m)^k + (1-k)x_m) + 1 = 0$. Moreover, for $x < x_m$, we have $f(x;k)' > 0$ and for $x > x_m$, we get $f(x;k)' < 0$. Thus, $f(x;k)$ is maximal at x_m; it corresponds to the mode of the JCA distribution, and $f(x;k)$ is upside-down shaped. ■

The possible mode of the JCA distribution does not have an analytical expression, but can be determined numerically for a given value for k. The curve behavior of $f(x;k)$ is represented in Figure 1 with diverse plots for selected values of k such that Figure 1(i) shows some decreasing or unimodal curves with a dominant right-tail and Figure 1(ii) focuses on unimodal curves with a "non-negligible left tail". From this figure, various degrees of asymmetry, peakedness, and tailedness for the pdf are observed. In this aspect, the JCA distribution is clearly more versatile than the exponential and Lindley distributions. Thus, with a single parameter, we see that the pdf of the JCA distribution has high modeling power, allowing the fit of data with decreasing or unimodal with various skewed histogram shapes.

2.3 Behavior of the HRF

As developed in [1], the flexibility of the hrf of a survival distribution is of importance for statistical modeling. This feature is now being examined for the JCA distribution. From Table 1, the hrf of the JCA distribution is

$$h(x;k) = (1+x)^{-k-1}(1+(1-k)x), \quad x \geq 0,$$

and $h(x;k) = 0$ for $x < 0$. Therefore, we have $h(0;k) = 1$ and, when $x \to +\infty$, the following function equivalence holds: $h(x;k) \sim (1-k)x^{-k}$. Hence, for $k \in [0,1)$, we have $h(x;k) = 0$, and for $k < 0$, we have $h(x;k) = +\infty$. Furthermore, note that $h(x;k)$ is constant for $k = 0$ with $h(x;0) = 1$. In addition, the following result holds.

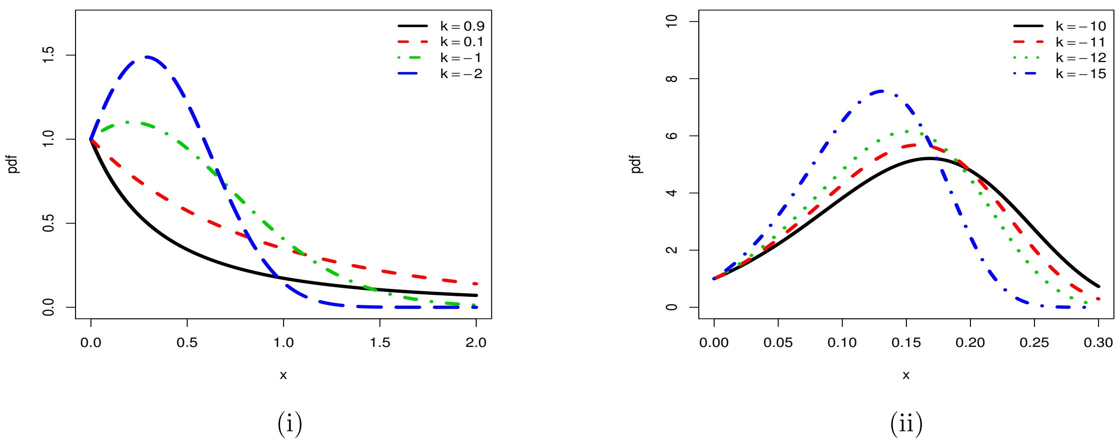


Figure 1: Plots of the pdf of the JCA distribution with various values of k by distinguishing those which are (i) mainly right-tailed or (ii) lightly left-tailed.

Proposition 2 *For $k \in (0,1)$, the hrf of the JCA distribution is decreasing and, for $k < 0$, it is increasing.*

Proof. For $x \geq 0$, the derivative of $h(x;k)$ can be simply expressed as

$$h(x;k)' = -k(1+x)^{-k-2}((1-k)x+2).$$

Then, it is clear that, for $k \in (0,1)$, $(1-k)x+2 \geq 0$, implying that $h(x;k)' \leq 0$ and that $h(x;k)$ is decreasing. For $k < 0$, we have $h(x;k)' \geq 0$ implying that $h(x;k)$ is increasing. This ends the proof of Proposition 2. ■

Through the selection of specific values of k, curves of $h(x;k)$ with different characteristics are shown in Figure 2. Figure 2 illustrates Proposition 2; the hrf can increase, decrease, and be constant; it benefits from flexible monotonic properties. For comparison, the shapes of this hrf are more diverse than those of the exponential distribution, which is only constant, and the Lindley distribution, which is only increasing.

3 Properties of Interest

In addition to the analytical functionalities of its pdf and hrf, the JCA distribution satisfies certain properties of interest, which are the subjects of this section.

3.1 Stochastic Ordering

Here, we discuss some ordering connections that exist between the JCA, exponential, Weibull and linear failure rate distributions, under some particular configurations of the parameters. The concept of first-order stochastic dominance is employed. In this regard, we may refer to [18].

We recall that the $\mathcal{E}(1)$ distribution corresponds to the JCA(0) distribution. Moreover, the following first-order stochastic dominance involving the JA and $\mathcal{E}(1)$ distributions holds.

Proposition 3 *For $k \leq 0$ and $x \in \mathbb{R}$, we have*

$$S(x;k) \leq S(x;0);$$

the $\mathcal{E}(1)$ distribution first-order stochastically dominates the JCA(k) distribution. For $k \in (0,1)$, the contrary holds.

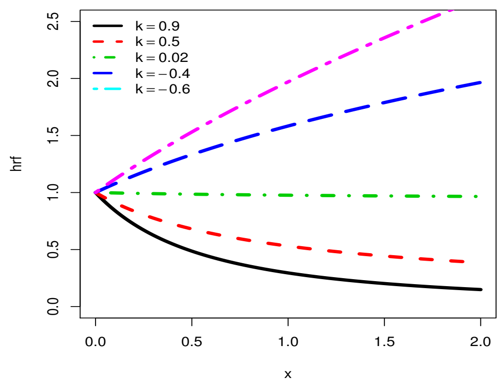


Figure 2: Plots of the hrf of the JCA distribution with various values of k

Proof. In the case $k \leq 0$, for $x \geq 0$, we have $1 + x \geq 1$, implying that $(1+x)^{-k} \geq 1$, so $-x(1+x)^{-k} \leq -x$, hence $e^{-x(1+x)^{-k}} \leq e^{-x}$ which is equivalent to $S(x;k) \leq S(x;0)$. The equality obviously holds for $x < 0$; we have $S(x;k) = S(x;0) = 1$. In the case $k \in (0,1)$, for $x \geq 0$, we have $1 + x \geq 1$, implying that $(1+x)^{-k} \leq 1$, so $-x(1+x)^{-k} \geq -x$, hence $e^{-x(1+x)^{-k}} \geq e^{-x}$ which corresponds to $S(x;k) \geq S(x;0)$. For $x < 0$, the equality is reached. This ends the proof of Proposition 3. ■

The following result exhibits a stochastic ordering result between the JCA(k) distribution and the Weibull distribution with scale parameter of 1 and shape parameter of $1-k$, denoted by W$(1, 1-k)$.

Proposition 4 *First, we specify that the* $W(1, 1-k)$ *distribution has the following sf:* $S_*(x;k) = e^{-x^{1-k}}$ *for* $x \geq 0$ *and* $S_*(x;k) = 1$ *for* $x < 0$. *Then, for* $k \leq 0$ *and* $x \in \mathbb{R}$, *we have*

$$S(x;k) \leq S_*(x;k);$$

the $W(1, 1-k)$ *distribution first-order stochastically dominates the* $JCA(k)$ *distribution. For* $k \in (0,1)$, *the contrary holds.*

Proof. In the case $k \leq 0$, for $x \geq 0$, we have $1 + x \geq x$, implying that $(1+x)^{-k} \geq x^{-k}$, so $-x(1+x)^{-k} \leq -x^{1-k}$, hence $e^{-x(1+x)^{-k}} \leq e^{-x^{1-k}}$ which is equivalent to $S(x;k) \leq S_*(x;k)$. The equality obviously holds for $x < 0$. In the case $k \in (0,1)$, for $x \geq 0$, we have $1 + x \geq x$, implying that $(1+x)^{-k} \leq x^{-k}$, so $-x(1+x)^{-k} \geq -x^{1-k}$, hence $e^{-x(1+x)^{-k}} \geq e^{-x^{1-k}}$ which corresponds to $S(x;k) \geq S_*(x;k)$. For $x < 0$, we have an immediate equality. This ends the proof of Proposition 4. ■

Now, we highlight a first-order stochastic result involving the JCA(k) distribution and the linear failure rate distribution with parameters of 1 and $1-k$, denoted by LFR$(1, -k)$.

Proposition 5 *First, we specify that the* $LFR(1, -k)$ *distribution has the following sf:* $S_{**}(x;k) = e^{-x+kx^2}$ *for* $x \geq 0$ *and* $S_{**}(x;k) = 1$ *for* $x < 0$. *Then, for* $k \in (-\infty, -1] \cup [0, 1)$ *and* $x \in \mathbb{R}$, *we have*

$$S(x;k) \leq S_{**}(x;k);$$

the $LFR(1, -k)$ *distribution first-order stochastically dominates the* $JCA(k)$ *distribution. For* $k \in (-1, 0)$, *the contrary holds.*

Proof. In the case $k \in (-\infty, -1] \cup [0, 1)$, for $x \geq 0$, the Bernoulli inequality ensures that $(1+x)^{-k} \geq 1 - kx$, so $-x(1+x)^{-k} \leq -x + kx^2$, hence $e^{-x(1+x)^{-k}} \leq e^{-x+kx^2}$ which corresponds to $S(x;k) \leq S_{**}(x;k)$. The equality obviously holds for $x < 0$. In the case $k \in (-1, 0)$, for $x \geq 0$, the (reverse) Bernoulli inequality gives $(1+x)^{-k} \leq 1 - kx$, so $-x(1+x)^{-k} \geq -x + kx^2$, hence $e^{-x(1+x)^{-k}} \geq e^{-x+kx^2}$ which is equivalent to $S(x; k) \geq S_{**}(x; k)$. The equality of these sfs is also satisfied for $x < 0$. The proof of Proposition 5 ends. ■

From Propositions 3, 4 and 5, the JCA(k) distribution can be viewed as an alternative to the mentioned one-parameter Weibull and linear failure rate distributions, in the first-order stochastic ordering sense. For this reason, these distributions will be considered as competing statistical models in the application section of this study.

3.2 Moments

The moments related to the JCA distribution are now examined. For this, we consider a random variable X following the JCA distribution. Then, for $r \geq 1$, the rth moment of X is basically obtained as

$$\mu'_r = E(X^r) = \int_0^{+\infty} x^r f(x; k) dx = \int_0^{+\infty} x^r (1+x)^{-k-1}(1 + (1-k)x) e^{-x(1+x)^{-k}} dx.$$

However, since X is continuous non-negative, it can be expressed in a more simple way as

$$\mu'_r = r \int_0^{+\infty} x^{r-1} S(x; k) dx = r \int_0^{+\infty} x^{r-1} e^{-x(1+x)^{-k}} dx. \quad (3)$$

The use of digital software makes it possible to calculate it numerically for given r and k. As a mathematical approach, a series expansion involving the upper incomplete gamma function is provided below.

Proposition 6 *The rth moment of X can be expanded as*

$$\mu'_r = \frac{r}{1-k} \sum_{m=0}^{+\infty} \sum_{\ell=0}^{r-1} \binom{r-1}{\ell} (-1)^{r-1-\ell} \frac{1}{m!} \Gamma\left(\frac{1}{1-k}(\ell - (m-1)k) + 1, 1\right),$$

where $\Gamma(s, x) = \int_x^{+\infty} t^{s-1} e^{-t} dt$ denotes the upper incomplete gamma function at s and x.

Proof. Based on (3), by setting $y = 1 + x$ with a change of the lower bound of the integration, by applying the binomial formula to $(y-1)^{r-1}$, by using the exponential series expansion for $e^{y^{-k}}$ and by taking $z = y^{1-k}$, the following development is obtained:

$$\begin{aligned}
\mu'_r &= r \int_1^{+\infty} (y-1)^{r-1} e^{-(y-1)y^{-k}} dy = r \sum_{\ell=0}^{r-1} \binom{r-1}{\ell} (-1)^{r-1-\ell} \int_1^{+\infty} y^{\ell} e^{-y^{1-k} + y^{-k}} dy \\
&= r \sum_{\ell=0}^{r-1} \binom{r-1}{\ell} (-1)^{r-1-\ell} \int_1^{+\infty} \sum_{m=0}^{+\infty} \frac{1}{m!} y^{\ell - mk} e^{-y^{1-k}} dy \\
&= \frac{r}{1-k} \sum_{\ell=0}^{r-1} \binom{r-1}{\ell} (-1)^{r-1-\ell} \sum_{m=0}^{+\infty} \frac{1}{m!} \int_1^{+\infty} y^{\ell-(m-1)k} (1-k) y^{-k} e^{-y^{1-k}} dy \\
&= \frac{r}{1-k} \sum_{\ell=0}^{r-1} \binom{r-1}{\ell} (-1)^{r-1-\ell} \sum_{m=0}^{+\infty} \frac{1}{m!} \int_1^{+\infty} z^{[\ell-(m-1)k]/(1-k)+1-1} e^{-z} dz \\
&= \frac{r}{1-k} \sum_{m=0}^{+\infty} \sum_{\ell=0}^{r-1} \binom{r-1}{\ell} (-1)^{r-1-\ell} \frac{1}{m!} \Gamma\left(\frac{1}{1-k}(\ell - (m-1)k) + 1, 1\right).
\end{aligned}$$

This ends the proof. ■

Table 2: Moments and related measures of the JCA distribution for several values of k.

k	μ'_1 (or μ)	μ'_2	μ'_3	μ'_4	σ^2	S	K
(0.9)	4.7776	197.2291	12580.12	926941.7	13.2061	4.3293	23.4075
(0.5)	2.4811	26.9505	694.0732	29216.75	4.5600	5.5261	53.6763
(0.1)	1.1146	2.6602	9.9786	51.6491	1.1907	2.2818	11.1218
(-0.1)	0.9111	1.5747	3.9396	12.8005	0.8628	1.7872	7.6106
(-0.5)	0.6901	0.7918	1.2081	2.2492	0.5617	1.2757	4.9837
(-0.9)	0.5683	0.4995	0.5588	0.7389	0.4202	1.0004	3.9662
(-5)	0.2419	0.0744	0.0261	0.01009	0.1262	0.2237	2.4846
(-10)	0.1569	0.0296	0.0062	0.0014	0.0709	-0.0223	2.4316

The interest of Proposition 6 is to provide an exact formula based on sums and manageable coefficients, from which one can derive the following approximation:

$$\mu'_r \approx \frac{r}{1-k} \sum_{m=0}^{M} \sum_{\ell=0}^{r-1} \binom{r-1}{\ell} (-1)^{r-1-\ell} \frac{1}{m!} \Gamma\left(\frac{1}{1-k}(\ell-(m-1)k)+1, 1\right),$$

where M denotes a large integer, such as $M = 40$ for instance.

Based on μ'_1, μ'_2, μ'_3 and μ'_4, we get standard moment measures such as the mean, variance, coefficient of variation, skewness, and kurtosis of X, given by $\mu = \mu'_1$, $\sigma^2 = \mu'_2 - \mu^2$,

$$S = \frac{1}{\sigma^3} E[(X-\mu)^3] = \frac{\mu'_3 - 3\mu\mu'_2 + 2\mu^3}{\sigma^3}$$

and

$$K = \frac{1}{\sigma^4} E[(X-\mu)^4] = \frac{\mu'_4 - 4\mu\mu'_3 + 6\mu^2\mu'_2 - 3\mu^4}{\sigma^4},$$

respectively.

A numerical work on these moment measures is performed by considering an arbitrary panel of values for k satisfying $k < 1$, with a focus on $k \in (-1, 1)$, and the results are collected in Table 2.

From Table 2, we observe a wide range of possible values for the mean and variance. For the considered values of k, there is a decreasing trend in these measures as k increases. Also, we confirm that the JCA distribution can be right skewed since positive values of S are observed and left skewed since $S < 0$ for $k = -10$. The JCA distribution may be of all the "kurtic states", since K may be lower, equal or greater than the reference value of 3. For the considered values of k, when k increases, we observe that S and K increase. All of these facts are proof of the versatility of the moments of the JCA distribution.

3.3 Order Statistics

Order statistics are essential tools for modeling various random systems and are the basis of useful statistical methods. We may refer to the book of [11]. Here some distributional properties of the order statistics for the JCA distribution are given. First, given n independent and identically random variables $X_1, \ldots, X_n$ with the JCA distribution, the order statistics are random variables defined by sorting the realizations of $X_1, \ldots, X_n$ in increasing order. They are commonly denoted by $X_{(1)}, \ldots, X_{(n)}$. Then, for $i = 1, \ldots, n$, the cdf of $X_{(i)}$ is given by the following formula:

$$F_{(i)}(x; k) = \sum_{j=i}^{n} \binom{n}{j} F(x; k)^j S(x; k)^{n-j},$$

that is, for $x \geq 0$,

$$F_{(i)}(x;k) = \sum_{j=i}^{n} \binom{n}{j} \left[1 - e^{-x(1+x)^{-k}}\right]^{j} e^{-(n-j)x(1+x)^{-k}},$$

and $F_{(i)}(x;k) = 0$ for $x < 0$. The order statistics $X_{(1)}$ and $X_{(n)}$ are of particular interest in many areas related to extremes. In particular, they appear naturally to define the spread of the sample $(X_1, \dots, X_n)$, that is $R_n = X_{(n)} - X_{(1)}$. Here, by taking $i = 1$, the cdf of $X_{(1)} = \inf(X_1, \dots, X_n)$ is

$$F_{(1)}(x;k) = 1 - e^{-nx(1+x)^{-k}}, \quad x \geq 0 \tag{4}$$

and $F_{(1)}(x;k) = 0$ for $x < 0$, and , for $i = n$, the cdf of $X_{(n)} = \sup(X_1, \dots, X_n)$ is

$$F_{(n)}(x;k) = \left(1 - e^{-x(1+x)^{-k}}\right)^{n}, \quad x \geq 0$$

and $F_{(n)}(x;k) = 0$ for $x < 0$. We also have the following simple formula:

$$P(X_{(1)} > x, X_{(n)} \leq y) = \left(e^{-x(1+x)^{-k}} - e^{-y(1+y)^{-k}}\right)^{n}, \quad y > x \geq 0.$$

Based on these functions, further developments on $X_{(1)}$ or $X_{(n)}$ are possible. For instance, the rth moment of $X_{(1)}$ is given as

$$\mu_{(1),r} = E(X_{(1)}^{r}) = r \int_{0}^{+\infty} x^{r-1}(1 - F_{(1)}(x;k))dx = r \int_{0}^{+\infty} x^{r-1} e^{-nx(1+x)^{-k}} dx.$$

This integral can be calculated numerically or developed analytically in the same way as in Proposition 6. The asymptotic distribution of $nX_{(1)}$ is examined below.

Proposition 7 *The sequence of random variable $(nX_{(1)})_{n\in\mathbb{N}^*}$ converges in distribution to the $\mathcal{E}(1)$ distribution.*

Proof. Based on (4), for any $x \geq 0$, we have

$$\lim_{n\to+\infty} P(nX_{(1)} \leq x) = \lim_{n\to+\infty} F_{(1)}\left(\frac{x}{n};k\right) = \lim_{n\to+\infty} \left(1 - e^{-x(1+x/n)^{-k}}\right) = 1 - e^{-x}.$$

Also, when $x < 0$, we have $P(nX_{(1)} \leq x) = 0$. Therefore, fo any $x \in \mathbb{R}$, we have $\lim_{n\to+\infty} P(nX_{(1)} \leq x) = P(X_* \leq x)$, where X_* denotes a random variable with the $\mathcal{E}(1)$ distribution. This ends the proof. ■

4 Applications

The JCA distribution is now being discussed for data fitting purposes.

4.1 Estimation

When data are observed, one can assume that, a priori, they come from the JCA(k) distribution. Based on this assumption, the parameter k is unknown, and must be estimated accurately. In this regard, one can use the so-called maximum likelihood method as developed below. We can refer to [5] for the general context of this method.

Let $x_1, \dots, x_n$ be n independent observations of a random variable X with the JCA(k) distribution. At the basis of the method, there is the log-likelihood function for k specified by

$$L(k) = \prod_{i=1}^{n} f(x_i;k) = \left[\prod_{i=1}^{n}(1+x_i)\right]^{-k-1} \left[\prod_{i=1}^{n}(1+(1-k)x_i)\right] e^{-\sum_{i=1}^{n} x_i(1+x_i)^{-k}}.$$

Then the maximum likelihood estimate (MLE) of k is given as $\hat{k} = \text{argmax}_{k\in(-\infty,1)} L(k)$ which, by the continuity of the logarithmic function over $(0,+\infty)$, also corresponds to $\hat{k} = \text{argmax}_{k\in(-\infty,1)} \log[L(k)]$, where

$$\log[L(k)] = -(k+1)\sum_{i=1}^{n}\log(1+x_i) + \sum_{i=1}^{n}\log(1+(1-k)x_i) - \sum_{i=1}^{n} x_i(1+x_i)^{-k}.$$

Thus, $\hat{k}$ is the solution of the following equation: $\partial \log[L(k)]/\partial k \mid_{k=\hat{k}} = 0$, with

$$\frac{\partial}{\partial k}\log[L(k)] = -\sum_{i=1}^{n}\log(1+x_i) - \sum_{i=1}^{n}\frac{x_i}{1+(1-k)x_i} + \sum_{i=1}^{n} x_i(1+x_i)^{-k}\log(1+x_i).$$

A correct analytical expression of $\hat{k}$ remains impossible, but practice only requires a numerical evaluation of it which can be easily obtained using specific tools in statistical software. By applying the existing theory on MLEs, the random version of $\hat{k}$ is asymptotically normal with mean k and variance $V = \{-\partial^2 \log[L(k)]/\partial k^2 \mid_{k=\hat{k}}\}^{-1}$, where

$$\frac{\partial^2}{\partial k^2}\log[L(k)] = -\sum_{i=1}^{n}\frac{x_i^2}{(1+(1-k)x_i)^2} - \sum_{i=1}^{n} x_i(1+x_i)^{-k}(\log(1+x_i))^2.$$

In particular, the (asymptotic) estimated standard error (SE) of $\hat{k}$ is obtained by $s = \sqrt{V}$. The asymptotic distribution can be used to develop diverse statistical tests or confidence intervals at different levels. Also, based on $\hat{k}$, estimates for the unknown functions of the JCA distribution can be obtained by the plug-in technique: $F(x;\hat{k})$ is the estimated cdf of $F(x;k)$, $f(x;\hat{k})$ is the estimated pdf of $f(x;k)$, and so on. These estimated functions are essential for data fitting, among other things.

4.2 Data fitting

Based on the estimation method described above, we now provide concrete applications on real data to show how the JCA distribution can be used in practice. Two data sets in the field of engineering are taken from [10]. The data are measures coming from the perforation operation on sheet metal; they are the length of the burr formed circularly around the hole on the sheet of metal on the other side of the hole. The unit is the millimeter (mm).

The first data set, named data set A, contains 50 such observations when the hole diameter and sheet thickness are 12 mm and 3.15 mm, respectively. The data are: 0.04, 0.02, 0.06, 0.12, 0.14, 0.08, 0.22, 0.12, 0.08, 0.26, 0.24, 0.04, 0.14, 0.16, 0.08, 0.26, 0.32, 0.28, 0.14, 0.16, 0.24, 0.22, 0.12, 0.18, 0.24, 0.32, 0.16, 0.14, 0.08, 0.16, 0.24, 0.16, 0.32, 0.18, 0.24, 0.22, 0.16, 0.12, 0.24, 0.06, 0.02, 0.18, 0.22, 0.14, 0.06, 0.04, 0.14, 0.26, 0.18, 0.16

The second data set, named data set B, is obtained in similar circumstances to data set A, but with a hole diameter and sheet thickness of 9 mm and 2 mm, respectively. The data are: 0.06, 0.12, 0.14, 0.04, 0.14, 0.16, 0.08, 0.26, 0.32, 0.22, 0.16, 0.12, 0.24, 0.06, 0.02, 0.18, 0.22, 0.14, 0.22, 0.16, 0.12, 0.24, 0.06, 0.02, 0.18, 0.22, 0.14, 0.02, 0.18, 0.22, 0.14, 0.06, 0.04, 0.14, 0.22, 0.14, 0.06, 0.04, 0.16, 0.24, 0.16, 0.32, 0.18, 0.24, 0.22, 0.04, 0.14, 0.26, 0.18, 0.16.

All the details regarding the context and experiments behind these data are given in [10]. Histogram analysis shows that these data sets are unimodal in nature, with skewed shapes trending to the left. In view of the shape possibilities of the pdf of the JCA distribution (see Figure 1), the JCA distribution is clearly a good option. Thus, in this statistical framework, we aim to study the goodness of fit of the JCA distribution with those of the distributions presented in Table 3.

Exponential and Lindley distributions are common competitors to one-parameter survival distributions. Other distributions are considered as they appear in the sections above in one form or another. For all the distributions, we employ the maximum likelihood method to estimate the parameters. The R software

Table 3: One-parameter survival distributions competing with the JCA distribution

Name	Abbreviation	Parameter	Cdf ($x \geq 0$)	Pdf ($x \geq 0$)
Exponential	EX	α	$1 - e^{-\alpha x}$	$\alpha e^{-\alpha x}$
One-parameter Weibull	WEB	θ	$1 - e^{-x^{\theta}}$	$\theta x^{\theta-1} e^{-x^{\theta}}$
One-parameter linear failure rate	LF	β	$1 - e^{-x-\beta x^2}$	$(1+2\beta x) e^{-x-\beta x^2}$
Lindley	LIN	η	$1 - \frac{1+\eta+\eta x}{1+\eta} e^{-\eta x}$	$\frac{\eta^2}{1+\eta}(1+x)e^{-\eta x}$

Table 4: MLEs of the distribution parameters and their SEs for data set A.

Distribution	k	α	θ	β	η
JCA	-9.0242	-	-	-	-
	(0.6321)	-	-	-	-
EX	-	6.1275	-	-	-
	-	(0.8665)	-	-	-
WEB	-	-	5.2292	-	-
	-	-	(0.0795)	-	-
LF	-	-	-	25.9488	-
	-	-	-	(4.2479)	-
LIN	-	-	-	-	6.9028
	-	-	-	-	(0.8777)

(see [16]) is used to determine the MLEs of these parameters through the so-called L-BFGS-B algorithm. The following statistical measures are considered: AIC, CAIC, BIC and HQIC all based on the estimated log-likelihood denoted by $\hat{\ell}$, CVM, AD and KS statistics with the p-value of the KS test. These measures are quite standard and their exact formula can be found in [13], for instance. As usual, the lower the values of these criteria (except for the p-value), the better the fit of the distribution.

For the considered distributions, the MLEs and SEs of the parameters for data sets A and B are given in Tables 4 and 5, respectively. If we focus on the JCA distribution, we have $\hat{k} = -9.0242$ and $\hat{k} = -9.724$ for data sets A and B, respectively. Then, the estimated pdf and cdf of the JCA distribution are given as

$$\hat{F}(x) = F(x; \hat{k}) = 1 - e^{-x(1+x)^{-\hat{k}}}$$

and

$$\hat{f}(x) = f(x; \hat{k}) = (1+x)^{-\hat{k}-1}(1 + (1-\hat{k})x)e^{-x(1+x)^{-\hat{k}}},$$

respectively.

Values of $-\hat{\ell}$ (minus $\hat{\ell}$), AIC, CAIC, BIC, HQIC, CVM, AD and KS with its p-value for data sets A and B are given in Tables 6 and 7, respectively.

From the results of Tables 6 and 7, we see that the smaller values for AIC, CAIC, BIC, HQIC, CVM, AD and KS, and the largest KS p-value are attributed to the JCA distribution. In particular, for the JCA distribution, we have AIC $= -111.8222$ and AIC $= -115.4444$, for data sets A and B, respectively, and the best competitor is the LF distribution with AIC $= -108.2193$ and AIC $= -112.4696$, for data sets A and B, respectively. Since the considered numerical indicators are quite favorable, we can say that the JCA distribution is the best among the distributions involved. In order of merit, it is followed by the LF, LIN, EX and WEB distributions.

We complete this application study by plotting the estimated pdfs and cdfs over the corresponding histograms and empirical cdfs of the data. The plots related to data sets A and B are given in Figures 3 and

Table 5: MLEs of the distribution parameters and their SEs for data set B.

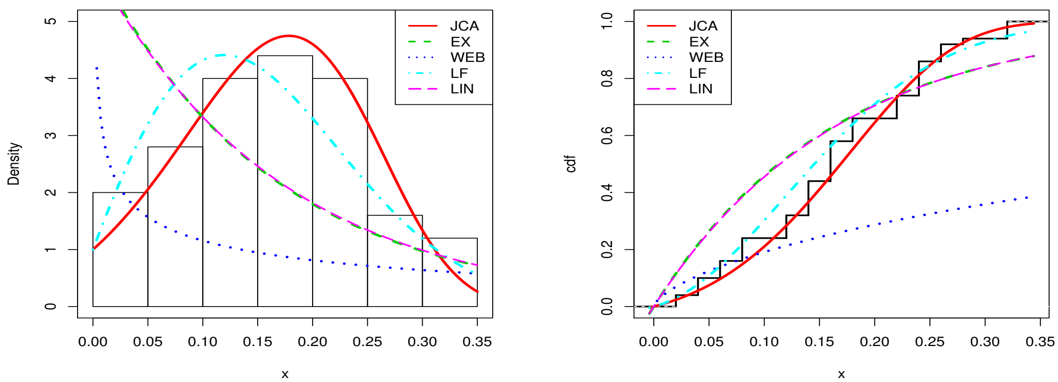
Distribution	k	α	θ	β	η
JCA	-9.724	-	-	-	-
	(0.6696)	-	-	-	-
EX	-	6.5789	-	-	-
	-	(0.9304)	-	-	-
WEB	-	-	0.6418	-	-
	-	-	(0.0759)	-	-
LF	-	-	-	29.4636	-
	-	-	-	(4.8175)	-
LIN	-	-	-	-	7.3653
	-	-	-	-	(0.9411)

Table 6: Statistical measures related to the distributions for data set A.

Distribution	$-\hat{\ell}$	AIC	CAIC	BIC	HQIC	CVM	AD	KS(p-value)
JC	-56.9110	-111.8222	-111.7388	-109.9101	-111.0941	0.0756	0.4567	0.1229(0.4360)
EX	-40.6389	-79.2778	-79.1945	-77.3658	-78.5497	0.8665	1.0985	0.2806(0.0007)
WEB	1.6146	5.2292	5.3125	7.1412	5.9573	0.2440	1.4415	0.6285(2.2×10^{-16})
LF	-55.1096	-108.2193	-108.1360	-106.3073	-107.4912	0.1016	0.6274	0.1572(0.1688)
LIN	-40.9445	-79.8891	-79.8058	-77.9771	-79.1610	0.1803	1.0828	0.2774(0.0009)

Table 7: Statistical measures related to the distributions for data set B.

Distribution	$-\hat{\ell}$	AIC	CAIC	BIC	HQIC	CVM	AD	KS(p-value)
JC	-58.7222	-115.4444	-115.3611	-113.5324	-114.7163	0.1035	0.6604	0.1396(0.2836)
EX	-44.1937	-86.3874	-86.3041	-84.4754	-85.6593	0.3240	1.7859	0.2859(0.0005)
WEB	-0.7572	0.4854	0.5688	2.3975	1.2135	0.4064	2.2152	0.6180(2.2×10^{-16})
LF	-57.2347	-112.4696	-112.3862	-110.5575	-111.7415	0.1965	1.1315	0.1920(0.0500)
LIN	-44.4578	-86.9157	-86.8324	-85.0037	-86.1876	0.3206	1.7678	0.2831(0.0007)



(i) (ii)

Figure 3: Estimated (i) pdfs and (ii) cdfs for data set A.

4, respectively.

From these figures, it is clear that the red curves related to the JCA distribution are closer to the shapes of the corresponding empirical objects. We explain this success by the nice compromise between the left and right tails of the JCA distribution, which is not achieved by the competitors. This visually confirms the high level of precision in the overall quality of the fit of the JCA distribution.

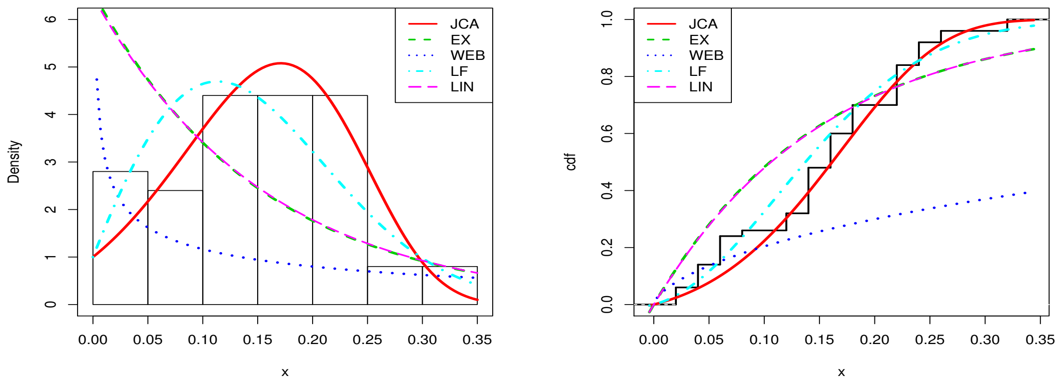
5 Conclusion and Perspectives

A new flexible one-parameter survival distribution, called JCA distribution, has been introduced. The shape behavior of the corresponding pdf and hrf, first-order stochastic dominance, moments, asymmetry and kurtosis measures have been studied in detail, revealing a particularly flexible and precise distribution for statistical modeling. In particular,

- as shown in Figure 1; the pdf has some decreasing, unimodal curves with a dominant right-tail, or unimodal with a "non-negligible left tail".
- as previously shown in Figure 2; the hrf can increase, decrease, and be constant; it benefits from flexible monotonic properties. For comparison, the shapes of this hrf are more diverse than those of the exponential distribution, which is only constant, and the Lindley distribution, which is only increasing.
- as shown in Table 2, the moment measures varied from small to large, and the coefficient skewness can be negative or positive, indicating a skew feature to the left or right.

The maximum likelihood estimate has been used to estimate the single parameter. Subsequently, the goodness of fit of the JCA distribution has been investigated with two real data sets. Adjustments obtained were found quite satisfactory over those of the exponential, Lindley, one-parameter Weibull and one-parameter linear failure rate distributions. These applied results were illustrated numerically with the help of standard statistical criteria in Tables 6 and 7, and graphically in Figures 3 and 4, for two different real-life data sets, respectively. In particular, the smaller values for AIC, CAIC, BIC, HQIC, CVM, AD and KS, and the largest KS p-value are attributed to the JCA distribution.

From the JCA distribution, one can derive multiple parameter survival distributions. Two ideas of extensions are:



(i) (ii)

Figure 4: Estimated (i) pdfs and (ii) cdfs for data set B.

- the two-parameter JCA distribution defined by the following cdf:

$$F(x;k,\alpha) = F(\alpha x;k) = 1 - e^{-\alpha x(1+\alpha x)^{-k}}, \quad x \geq 0,$$

  with $k < 1$ and $\alpha > 0$, and $F(x;k,\alpha) = 0$ for $x < 0$,

- the three-parameter JCA distribution defined by the following cdf:

$$F(x;k,\alpha,\beta) = F(\alpha x^{\beta};k) = 1 - e^{-\alpha x^{\beta}(1+\alpha x^{\beta})^{-k}}, \quad x \geq 0,$$

  with $k < 1$, $\alpha > 0$ and $\beta > 0$, and $F(x;k,\alpha,\beta) = 0$ for $x < 0$.

Both of these extensions are more complicated than the JCA distribution from a mathematical point of view, but may have interesting characteristics requiring further research, which we will leave for further study.

Acknowledgment. We thank the referee for thorough and constructive comments.

References

[1] M. V. Aarset, How to identify bathtub hazard rate, IEEE Trans. Reliab., 36(1987), 106–108.

[2] A. Algarni, On a new generalized Lindley distribution: Properties, estimation and applications, PLoS One, 16(2021), 1–19.

[3] A. Alzaatreh, F. Famoye and C. Lee, A new method for generating families of continuous distributions, METRON, 71(2013), 63–79.

[4] C. C. R. Brito, L. C. Rêgo, W. R. Oliveira and F. Gomes-Silva, Method for generating distributions and classes of probability distributions: the univariate case, Hacet. J. Math. Stat., 48(2019), 897–930.

[5] G. Casella and R. L. Berger, Statistical Inference, Brooks/Cole Publishing Company: Bel Air, CA, USA, 1990.

[6] C. Chesneau, L. Tomy and J. Gillariose, On a sum and difference of two Lindley distributions: theory and applications, REVSTAT, 18(2020), 673–695.

[7] C. Chesneau, L. Tomy and J. Gillariose, A new modified Lindley distribution with properties and applications, J. Stat. Manag. Syst., 24(2021), 1383–1403.

[8] C. Chesneau, L. Tomy, J. Gillariose and F. Jamal, The inverted modified Lindley distribution, J. Stat. Theory Pract., 14(2020), 1–17.

[9] S. T. Dara and M. Ahmad, Recent Advances in Moment Distribution and Their Hazard Rates, Lap Lambert Academic Publishing, GmbH, KG, 2012.

[10] R. Dasgupta, On the distribution of burr with applications, Sankhya B, 73(2011), 1–19.

[11] H. A. David and H. N. Nagaraja, Order Statistics, 3rd edn. Wiley, Hoboken, NJ, 2003.

[12] F. Jamal, C. Chesneau, K. Aidi and A. Ali, Theory and application of the power Ailamujia distribution, J. Math. Model., 9(2021), 391–413.

[13] S. Konishi and G. Kitagawa, Information Criteria and Statistical Modeling, Springer, New York, 2007.

[14] P. Marthin and G. S. Rao, Generalized Weibull-Lindley (GWL) distribution in modeling lifetime data, J. Math., 2020(2020), 1–15.

[15] C. C. Odom and M. A. Ijomah, Odoma distribution and its applications, Asian J. Probab. Stat., 4(2019), 1–11.

[16] R. Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL: http://www.R-project.org, 2005.

[17] A. M. Sarhan and D. Kundu, Generalized linear failure rate distribution, Comm. Statist. Theory Methods, 38(2009), 642–660.

[18] M. Shaked and J. G. Shanthikumar, Stochastic Orders and their Applications, Academic Press, New York, 1994.

[19] R. Shanker, Akash distribution and its applications, Int. J. Probab. Stat., 4(2015), 65–75.

[20] R. Shanker, Sujatha distribution and its applications, Stat. Transit.-New Series, 17(2016), 391–410.

[21] R. Shanker and K. K. Shukla, Ishita distribution and its applications, Biom. Biostat. Int. J., 5(2017), 1–9.

[22] K. K. Shukla, Prakaamy distribution with properties and applications, J. Appl. Quant. Methods., 13(2018), 30–38.

[23] K. K. Shukla, Pranav distribution and its applications, Biom. Biostat. Int. J., 7(2019), 244–254.