African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 3, 2024 (pp. 79-94)
www.abjournals.org

ON FITTING LIVE-STREAMING DATA: A LEVEL THREE POLYNOMIAL COMPONENT PROBABILITY MODEL DEVELOPMENT

Okolo Chiemeka Nwankwor[1], Kenneth Kevin Uadiale[2],

and Emeka Emmanuel Otti[3]

[1]Department of Industrial Mathematics and Applied Statistics, Faculty of Physical Sciences, Ebonyi State University, Ebonyi, Nigeria.

[2]Department of Mathematics and Statistics, Faculty of Pure and Applied Sciences, Federal University Wukari, Taraba, Nigeria.

[3]Department of Mathematics, Faculty of Sciences, Case Western Reserve University, USA.

Cite this article:

Okolo C. N., Kenneth K. U., Emeka E. O. (2024), On Fitting Live-Streaming Data: A Level Three Polynomial Component Probability Model Development . African Journal of Mathematics and Statistics Studies 7(3), 79-94. DOI: 10.52589/AJMSS-RDFCZYVM

Manuscript History

Received: 2 May 2024

Accepted: 25 Jun 2024

Published: 8 Aug 2024

Copyright © 2024 The Author(s). This is an Open Access article distributed under the terms of Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), which permits anyone to share, use, reproduce and redistribute in any medium, provided the original author and source are credited.

ABSTRACT: *A new probability distribution termed Stream distribution is proposed and studied. This distribution is a mathematical combination of level three polynomial components and exponential distribution. The mathematical and statistical properties of the development are studied, with empirical emphasis: on the inequality relationship within the measures of central tendency, and the coefficient of variation. The model parameter was estimated using the method of maximum likelihood, where the asymptotic and consistent properties are numerically studied as well. The flexibility of Stream distribution is shown, through an application to a Live-Streaming data set and showed high efficiency in the inferential performance. The distribution is therefore recommended for forecasting needs in the light of live online audience engagement.*

KEYWORDS: Stream distribution, Coefficient of variation, Polynomial component, Live-Streaming data, Comparative analysis.

INTRODUCTION

The analysis and modeling of lifetime data is a crucial aspect of statistical research in various scientific and technological fields. The field of lifetime data analysis has experienced rapid growth and expansion in recent years, with significant advancements in methodology, theory, and application. In the context of modeling real-life phenomena, continuous probability distributions and transformation methods have been proposed. Data plays a vital role in tracking events and revealing patterns and behaviors of outcomes. Lifetime data refers to data whose events exhibit a propensity for failure or success after a measurable amount of valid cycles.

The conception of models to handle such random variables emerged in the twentieth century and developed into two main subjects: reliability theory, concerned with modeling lifetimes for components and systems in engineering and industrial fields, and survival analysis, used in biological fields. In probability distribution theory, flexibility and tractability are highly valued in modeling lifetime data. While tractable distributions are useful in theory, flexible distributions are more relevant in industrial applications.

Transforming data to satisfy assumptions is a common statistical practice, but it's preferable to use probability distributions that best fit the available data set. Recent efforts have focused on developing new distributions and their extensions to accommodate the increasing amount of data from various fields. Lindsay (1995) proposed a model for developing new distributions from parent k-distributions, which can be combined to form new distributions. However, this approach can be complex and sometimes useless, so most mixtures are taken from one family of distribution with different parameters and/or distributions that share the same support or range.

Compound distributions, which result from assuming a random variable is distributed according to a parameterized distribution with random parameters, offer an alternative approach. Sankaran (1970) introduced compounding in the Poisson-Lindley distribution, while others have explored convolutions and mixtures of various distributions. This paper aims to propose a probability distribution for modeling live-streaming data, building on the existing literature and methodologies in lifetime data analysis.

MODEL PROPOSITION

Stream distribution as a continuous category is derived mathematically following the concept of integration and normalizing constant, after a multiplicative combination of the exponential distribution and a three level polynomial component. In probability theory and statistics, the exponential distribution is the model of the time between events in a poisson process. The event is always independent and continuous at a constant average rate. It has a probability density function (pdf) defined as

$$f(x;\, \theta) = \{\, \theta e^{-\theta x} \qquad x \geq 0 \quad 0, \qquad x < 0$$

where the rate parameter $\theta > 0$, and the cumulative distribution function is given as

$$F(x;\, \theta) = \{\, 1 - e^{-\theta x} \qquad x \geq 0 \quad 0, \qquad x < 0$$

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

More so, the three level polynomial components is subjectively defined thus:

$$y = \quad 1 + x + x^3$$

Stream Distribution

The Stream distribution denoted as $S_D(x)$ is derived, recalling that $\int_{-\infty}^{\infty} f(x)dx = 1$. This implies that continuous probability models can be derived from functions or combination of functions $f(x)$; so far as their integration equals one. This is also termed normalizing constant method. Now, let C_n be normalizing constant, then

$$S_D(x, \theta) = C_n f(x)$$

Where $C_n = \left[\int_0^{\infty} f(x)dx\right]^{-1}$ and $f(x) = \theta e^{-\theta x}\,(1 + x + x\text{^}3)$

$$\rightarrow \qquad C_n = \frac{\theta^3}{6+\theta^2+\theta^3}$$

$$\rightarrow \qquad S_D(x, \theta) = \frac{\theta^3}{6+\theta^2+\theta^3}\left[\theta e^{-\theta x}\,(1 + x + x^3)\right]$$

$$= \frac{\theta^4(\,1+x+\,x^3\,)}{e^{\theta x}(\theta^3+\theta^2+6)}\ ,\ x > 0,\ \theta > \ 0$$

PROPERTIES OF STREAM DISTRIBUTION

Cumulative Distribution Function (CDF)

The Cumulative Distribution Function (CDF) for Stream distribution is an integral derivation of the proposed pdf in equation (1),

$$F(x) = \int_0^x f(t, \theta)dt = \int_0^x \frac{\theta^4(\,1+t+\,t^3\,)}{e^{\theta t}(\theta^3+\theta^2+6)}\,dt$$

$$= \frac{\theta^4}{\theta^3+\theta^2+6}\left[\int_0^x e^{-\theta t}dt \ + \int_0^x t\,e^{-\theta t}dt \ + \int_0^x t^3\,e^{-\theta t}dt\ \right]$$

$$= \frac{\theta^4}{\theta^3+\theta^2+6}\left(\frac{\left[\theta^3+\theta^2+6-(\theta^3+\theta^2+6)e^{-\theta x}-\theta x\left(6+3\theta x+\theta^2 x^2\right)e^{-\theta x}\right]}{\theta^4}\right)$$

$$F(x, \theta) = 1 - \left(1 + \frac{\theta x\,[\,\theta^2 + \theta^2\,x^2 + 3\theta x + 6\,]}{\theta^3 + \theta^2 + 6}\right)e^{-\theta x} \qquad (2)$$

Moment Generating Function

The moment generating function of Stream Distribution is derived as:

$$M_x(\mathrm{t}) \ = E\,(\,e^{tx}\,) \ = \int_0^{\infty} e^{tx}\,f(x)\ dx$$

$$= \int_0^{\infty} \mathrm{e}^{\mathrm{tx}}\,\frac{\theta^4}{\theta^3+\theta^2+6}\,(1 + \ x + \ x^3\,)e^{-\theta x}\,\mathrm{dx}$$

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 3, 2024 (pp. 79-94)

$$= \int_0^{\infty} \frac{\theta^4}{\theta^3+\theta^2+6} \left(1 + x + x^3 \right) e^{-x(\theta-t)} \, dx$$

Given that $\int_0^{\infty} x^m e^{-\theta x} \, dx = \frac{\Gamma(m+1)}{\theta^{m+1}}$ and $(u + v)^{-n} = \sum_{r=0}^{\infty} \binom{r+a}{r} u^{-n-r} v^r$

$$= \frac{\theta^4}{\theta^3+\theta^2+6} \left[\frac{1}{\theta-t} + \frac{1}{(\theta-t)^2} + \frac{6}{(\theta-t)^4} \right]$$

$$= \frac{\theta^4}{\theta^3+\theta^2+6} [\frac{1}{\theta} \sum_{r=0}^{\infty} \frac{t^r}{\theta^r} + \frac{1}{\theta^2} \sum_{r=0}^{\infty} \binom{r+1}{r} \frac{t^r}{\theta^r} + \frac{6}{\theta^4} \sum_{r=0}^{\infty} \binom{r+6}{r} \frac{t^r}{\theta^r}$$

$$M_x(t) = \sum_{r=0}^{\infty} \left(\frac{\theta^3 + \theta^2 (r+1)! + (r+3)!}{\theta^r (\theta^3+\theta^2+6)} \right) \frac{t^r}{r!}$$

Moment

The r^{th} moment of the Stream distribution is obtained

$$E(x^r) = \frac{r! \left[\theta^3 + \theta^2 (r+1) + (r+1)(r+2)(r+3) \right]}{\theta^r (\theta^3+\theta^2+6)} \tag{3}$$

Therefore, the first-four moments about origin of Stream Distribution are given as:

$$\mu_1' = \frac{\theta^3 + 2\theta^2 + 24}{\theta (\theta^3+\theta^2+6)} = \mu \qquad \mu_2' = \frac{2(\theta^3 + 3\theta^2 + 60)}{\theta^2 (\theta^3+\theta^2+6)}$$

$$\mu_3' = \frac{6(\theta^3 + 4\theta^2 + 120)}{\theta^3 (\theta^3+\theta^2+6)} \qquad \mu_4' = \frac{24(\theta^3 + 5\theta^2 + 210)}{\theta^4 (\theta^3+\theta^2+6)}$$

The central moment about the mean of Stream distribution is:

$$\mu_n = E[(X - E[X])^n] = \sum_{j=0}^{n} \binom{n}{j} (-1)^{n-j} \mu_j' \mu^{n-j}$$

$$\mu_2 = \mu_2' - \mu^2$$

$$= \frac{\theta^6 + 4\theta^5 + 2\theta^4 + 84\theta^3 \ 60\theta^2 + 144}{\theta^2 (\theta^3+\theta^2+6)^2} = \sigma^2$$

$$\mu_3 = \mu_3' - 3\mu_2' \mu + 2\mu^3$$

$$= \frac{2(\theta^9 + 6\theta^8 + 6\theta^7 + 200\theta^6 + 270\theta^5 + 108\theta^4 + 324\theta^3 + 432\theta^2 + 864)}{\theta^3 (\theta^3+\theta^2+6)^3}$$

$$\mu_4 = \mu_4' - 4\mu_3' \mu + 6\mu_2' \mu^2 - 3\mu^4$$

$$= \frac{3(3\theta^{12} + 24\theta^{11} + 44\theta^{10} + 968\theta^9 + 2336\theta^8 + 2016\theta^7 + 7488\theta^6 + 13248\theta^5 + 5760\theta^4 + 31104\theta^3 + 24192\theta^2 + 31104)}{\theta^4 (\theta^3+\theta^2+6)^4}$$

Coefficient of Variation, Skewness, Kurtosis, and Index of Dispersion

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

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

$$CV = \frac{\sigma}{\mu_1'} = \frac{\sqrt{(\theta^6 + 4\theta^5 + 2\theta^4 + 84\theta^3 \ 60\theta^2 + 144)}}{(\theta^3 + 2\theta^2 + 24)}$$

$$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} = \frac{2(\theta^9 + 6\theta^8 + 6\theta^7 + 200\theta^6 + 270\theta^5 + 108\theta^4 + 324\theta^3 + 432\theta^2 + 864)}{(\theta^6 + 4\theta^5 + 2\theta^4 + 84\theta^3 \ 60\theta^2 + 144)^{3/2}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} =$$

$$\frac{3(3\theta^{12} + 24\theta^{11} + 44\theta^{10} + 968\theta^9 + 2336\theta^8 + 2016\theta^7 + 7488\theta^6 + 13248\theta^5 + 5760\theta^4 + 31104\theta^3 + 24192\theta^2 + 31104)}{(\theta^6 + 4\theta^5 + 2\theta^4 + 84\theta^3 \ 60\theta^2 + 144)^2}$$

$$\gamma = \frac{\sigma^2}{\mu_1'} = \frac{(\theta^6 + 4\theta^5 + 2\theta^4 + 84\theta^3 \ 60\theta^2 + 144)}{\theta(\theta^3 + 2\theta^2 + 24)(\theta^3 + \theta^2 + 6)}$$

OTHER PROPERTIES OF THE STREAM DISTRIBUTION

Mean Residual Life Function (MRL)

In reliability studies, Mean Residual Life Function (MRL) is the expected additional lifetime, given that a component has survived until time t. This is defined as:

$$m(x) = E[X - x \mid X > x] = \frac{1}{1 - \mathrm{F(x)}} \int_x^\infty [1 - \mathrm{F(t)}]\, dt$$

Where we consider $\mathrm{A} = \frac{1}{1 - \mathrm{F(x)}}$

The mean residual life function of Stream distribution is given as

$$m(x) = \mathrm{A} \int_x^\infty \left[1 - 1 - \left(1 + \frac{\theta t\,[\theta^2 + \theta^2 t^2 + 3\theta t + 6]}{\theta^3 + \theta^2 + 6}\right) e^{-\theta t} \right] dt$$

With $$A = \frac{1}{1 - \left\{1 - \left(1 + \frac{\theta x\,[\theta^2 + \theta^2 x^2 + 3\theta x + 6]}{\theta^3 + \theta^2 + 6}\right) e^{-\theta x}\right\}}$$

$$m(x) = \frac{\theta^3 + 2\theta^2 + \theta^3 x + \theta^3 x^2 + 6\theta^2 x^2 + 18\theta x + 24}{\theta(\theta^3 + \theta^2 + 6) + \theta x(\theta^2 + \theta^2 x^2 + 3x\theta + 6)}$$

Thus at $x = 0$, $m(0) = \frac{(\theta^3 + 2\theta^2 + 24)}{\theta(\theta^3 + \theta^2 + 6)} = \mu$

Hazard Function

According to Gross and Clark [1975], hazard function accounts for the risk of failure of a system at varying times x. At the other hand Survival Function is the probability that a system survives beyond a given time $x, x \geq 0$. The Hazard Function of Stream distribution is given as

$$H(x, \theta) = \frac{f(x, \theta)}{1 - F(x, \theta)} = \frac{f(x, \theta)}{S(x, \theta)}$$

where $S(x, \theta)$ is the survival function

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

$$= \frac{\frac{\theta^4}{\theta^3+\theta^2+6}(1+x+x^3)\,e^{-\theta x}}{\left[1+\frac{\theta x[\theta^2+\theta^2x^2+3\theta x+6]}{\theta^3+\theta^2+6}\right]e^{-\theta x}}$$

$$H(x,\theta) = \frac{\theta^4(1+x+x^3)}{(\theta^3+\theta^2+6)+\theta x(\theta^2+\theta^2x^2+3x\theta+6)}$$

Bonferroni and Lorenz Curve

Bonferroni [1930] gave a curve that measures for the conditional mean of a distribution; whereas, Dagum [1985] referred to the Lorenz curve as the measure of inequality of the variability of X. Let X be a non-negative continuous random variable, with positive and finite expected value μ, and distribution F; then Bonferroni curve is obtained as

$$B(p) = \frac{1}{p\mu}\int_0^q x\,f(x)dx$$

$$B(p) = \frac{1}{p\mu}\left[\int_0^\infty x\,f(x)dx - \int_q^\infty xf(x)dx\right] = \frac{1}{p\mu}\left[\mu - \int_q^\infty x\,f(x)dx\right]$$

While the Lorenz curve is obtained as

$$L(p) = \frac{1}{\mu}\int_0^q x\,f(x)dx$$

$$L(p) = \frac{1}{\mu}\left[\int_0^\infty x\,f(x)dx - \int_q^\infty xf(x)dx\right] = \frac{1}{\mu}\left[\mu - \int_q^\infty x\,f(x)dx\right]$$

The relationship between the Boneferroni curve and Lorenz curve is given as

$$B(p) = \frac{1}{\mu}\int_0^p F^{-1}(x)dx = \frac{L(p)}{p}$$

$$Where\, \mu = E(X),\ q = F^{-1}(p)\ and\, p \in [0,1]$$

Thus, when $X \sim Juchez(\theta)$, *the* $B(p)$ *and* $L(p)$ *of Stream distribution are defined as*

$$B(p) = \frac{1}{p}\left(1 - \left[\frac{(\theta^3+2\theta^2+24) - q(\theta^4+2\theta^3+24) - q^2(\theta^4+12\theta^2) - 4\theta^3q^3 - \theta^4q^4)}{(\theta^3+2\theta^2+24)}\right]\right)$$

$$L(p) = \left(1 - \left[\frac{(\theta^3+2\theta^2+24) - q(\theta^4+2\theta^3+24) - q^2(\theta^4+12\theta^2) - 4\theta^3q^3 - \theta^4q^4)}{(\theta^3+2\theta^2+24)}\right]\right)$$

Stochastic Ordering

Given that $X \sim Juchez(\theta_1)$ and $Y \sim Juchez(\theta_2)$, and if $\theta_1 > \theta_2$ then $X \leq_{lr} Y$ and hence $X \leq_{hr} Y$, $X \leq_{mrl} Y$ and $X \leq_{st} Y$. Where lr, hr, mrl and st represent the likelihood ratio order, hazard rate order, mean residual life order and stochastic order respectively. Thus,

$$\frac{f_X(x)}{f_Y(x)} = \frac{\theta_1^4(\theta_2^3+\theta_2^2+6)}{\theta_2^4(\theta_1^3+\theta_1^2+6)}e^{x(\theta_2-\theta_1)}, \quad x > 0. \tag{4}$$

If, for $\theta_2 > \theta_1$,

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

$$\frac{d}{dx}\frac{f_x(x)}{f_y(y)} = (\theta_2 - \theta_1)\frac{f_x(x)}{f_y(x)} < 0, \quad (5)$$

From equations (4) and (5), $\frac{f_x(x)}{f_y(x)}$ is decreasing in x.That implies $X \leq_{lr} Y$.

Remark:

- $X \leq_{st} Y$ if $F_x(x) \geq F_y(x)\ \forall x$;
- $X \leq_{hr} Y$ if $h_x(x) \geq h_y(x)\ \forall x$;
- $X \leq_{mrl} Y$ if $m_x(x) \geq m_y(x)\ \forall x$

These conditions hold if a random variable X is said to be lesser than a random variable Y. These implications are well known [Shaked and Shanthikumar, 1994]:

$$X \leq_{lr} Y \Rightarrow X \leq_{hr} Y \Rightarrow X \leq_{mrl} Y \text{ and } X \leq_{hr} Y \Rightarrow X \leq_{st} Y$$

Entropy Measure

Entropy measures the uncertainty, or randomness of a system, say probability distribution. The Rényi [1961] entropy of a random variable X, following the Stream distribution is given by:

$$T_R(s) = \frac{1}{1-s}\log\left(\int f^s(x)\,dx\right) \quad \text{where } s > 0 \text{ and } s \neq 1$$

$$= \frac{1}{1-s}\log\left(\int_0^\infty \left(\frac{\theta^4}{\theta^3+\theta^2+6}\right)^s (1+x+x^3)^s e^{-\theta s x}dx\right)$$

$$\text{But } (1+a)^m = \sum_{i=0}^{\infty}\binom{m}{i} a^i$$

$$= \frac{1}{1-s}\log\left(\int_0^\infty \left(\frac{\theta^4}{\theta^3+\theta^2+6}\right)^s \sum_{i=0}^{\infty}\binom{s}{i} (x+x^3)^i e^{-\theta s x}dx\right)$$

$$= \frac{1}{1-s}\left(\log \sum_{i=0}^{\infty}\binom{s}{i}\left(\frac{\theta^4}{\theta^3+\theta^2+6}\right)^s \int_0^\infty (1+x^2)^i x^i e^{-\theta s x}dx\right)$$

$$= \frac{1}{1-s}\log\left(\sum_{i=0}^{\infty}\binom{s}{i}\sum_{j=0}^{\infty}\binom{i}{j}\left(\frac{\theta^4}{\theta^3+\theta^2+6}\right)^s \int_0^\infty x^{2j+i} e^{-\theta s x}dx\right)$$

$$\text{We have that } \int_0^\infty x^m e^{-\theta x}\,dx = \frac{\Gamma(m+1)}{\theta^{m+1}} = \frac{m!}{\theta^{m+1}}$$

$$T_R(s) = \frac{1}{1-s}\log\left(\sum_{i=0}^{\infty}\binom{s}{i}\sum_{j=0}^{\infty}\binom{i}{j}\left(\frac{\theta^4}{\theta^3+\theta^2+6}\right)^s \frac{\Gamma(2j+i+1)}{(\theta s)^{2j+i+1}}\right)$$

$$T_R(s) = \frac{1}{1-s}\log\left(\sum_{i=0}^{\infty}\binom{s}{i}\sum_{j=0}^{\infty}\binom{i}{j}\frac{\theta^{4s-2j-i-1}}{(\theta^3+\theta^2+6)^s}\frac{(2j+i)!}{s^{2j+i+1}}\right)$$

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 3, 2024 (pp. 79-94) www.abjournals.org

Order Statistics

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from Stream Distribution. Let $X_1 < X_2 < , \ldots, < X_n$ denote the corresponding order statistics. The pdf and the cdf of the kth order statistics say $Y = X_k$ is given by:

$$f_Y(\mathrm{y}) = \frac{\mathrm{n!}}{(k-1)!(n-k)!} F^{k-1}(y)\{1 - F(y)\}^{n-k} f(y)$$

$$f_Y(\mathrm{y}) = \frac{\mathrm{n!}}{(k-1)!(n-k)!} \sum_{l=0}^{n-k} \binom{n-k}{l} (-1)^l F^{k+l-1}(y) f(y)$$

$$F_Y(\mathrm{y}) = \sum_{j=k}^{n} \binom{n}{j} F^j(y)\{1 - F(y)\}^{n-j}$$

$$F_Y(\mathrm{y}) = \sum_{j=k}^{n} \sum_{l=0}^{n-j} \binom{n}{j} \binom{n-j}{l} (-1)^l F^{j+l}(y)$$

Thus the pdf and the cdf of kth order statistics of Stream distribution are given by

$$f_Y(\mathrm{y}) = \frac{\mathrm{n!}\, \theta^3 (1 + x + x^3)\, e^{-\theta x}}{(\theta^3 + \theta^2 + 6)(j-1)!(n-j)!} \sum_{l=0}^{n-j} \binom{n-j}{l} (-1)^l \left[1 + \frac{\theta x\, [\theta^2 + \theta^2 x^2 + 3\theta x + 6]}{\theta^3 + \theta^2 + 6}\right] e^{-\theta x(i+j+1)} \quad (6)$$

$$F_Y(\mathrm{y}) = \sum_{j=k}^{n} \binom{n}{j} \sum_{l=0}^{n-j} \binom{n-j}{l} \sum_{k}^{j} \binom{j}{k} \sum_{m=0}^{k} \binom{k}{m} (-1)^{j+l} \left[1 + \frac{\theta x\, [\theta^2 + \theta^2 x^2 + 3\theta x + 6]}{\theta^3 + \theta^2 + 6}\right] e^{-\theta x(j+l)}$$

That implies that the pdf of minimum order statistics is obtained by substituting $j = k = 1$ in equation (6) to have:

$$f_{1:n} = \frac{\mathrm{n}[\theta^3 (1 + x + x^3) e^{-\theta x}]}{(\theta^3 + \theta^2 + 6)} \sum_{l=0}^{n-1} \binom{n-1}{l} (-1)^l \left[1 + \frac{\theta x\, [\theta^2 + \theta^2 x^2 + 3\theta x + 6]}{\theta^3 + \theta^2 + 6}\right] e^{-\theta x(i+2)}$$

While the corresponding pdf of maximum order statistics is obtained by making the substitution of $j = k = n$ in equation (6)

$$f_{n:n} = \frac{\mathrm{n}[\theta^3 (1 + x + x^3) e^{-\theta x}]}{(\theta^3 + \theta^2 + 6)} \left[1 + \frac{\theta x\, [\theta^2 + \theta^2 x^2 + 3\theta x + 6]}{\theta^3 + \theta^2 + 6}\right] e^{-\theta x(i+n+1)}$$

Limiting Distribution

If $X_1, .., X_n$ is a random sample, and if $\underline{X} = \frac{X_1 + \ldots + X_n}{n}$ denotes the sample mean then by the usual central limit theorem, $\frac{X_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ approaches the standard normal distribution $N(0,1)$ as $n \to \infty$.

There could be an interest in deriving the asymptotic of the extreme values $X_{n;n} = max(X_1, .., X_n)$ and $X_{1;n} = min(X_1, .., X_n)$. Bensid (2017) gave many examples on the Lindley family distribution.

The limiting distribution of sample minima and maxima of Stream distribution is

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 3, 2024 (pp. 79-94)

$$\frac{F(tx)}{F(t)} = x\frac{f(tx)}{f(t)}$$

$$= x\frac{\theta^4(1+ tx +t^3x^3)\, e^{-\theta tx}}{\theta^4(1+ t +t^3)e^{-\theta t}}$$

$\frac{F(tx)}{F(t)} = x$, for $X_{1;n}$ minima

$$\frac{1-F(tx)}{1-F(t)} = lim_{t\to\infty}\left[\frac{\left(1+\frac{\theta^3(t+x)+\theta^3(t+x)^3 + 3\theta^2(t+x)^2 + 6\theta(t+x)}{\theta^3+\theta^2+6}\right)e^{-\theta(t+x)}}{\left(1+\frac{\theta^3 t+\theta^3t^3 + 3\theta^2t^2 + 6\theta t}{\theta^3+\theta^2+6}\right)e^{-\theta t}}\right]$$

$= e^{-\theta x}$, for $X_{n;n}$ maxima

Maximum Likelihood Estimator

Let $X_i, i = 1,2,3, \dots, n$, be a random variable from Stream Distribution, the maximum likelihood estimator (MLE) is obtained thus:

$$Lf(x,\theta) = \left(\frac{\theta^4}{(\theta^3+\theta^2+6)}\right)^n \prod_{i=1}^{n} (1+x + x^3)\, e^{-\theta\sum_{i=1}^{n} x_i}$$

$$lnLf(x,\theta) = 4nln\theta - nln(\theta^3+\theta^2+6) + \sum_{i=1}^{n} ln(1+x+x^3) - \theta\sum_{i=1}^{n} x_i$$

In estimation of MLE, the estimator is maximized at $\frac{\partial\, lnL}{\partial\theta} = 0$*, then*

$$\frac{\partial\, lnLf(x,\theta)}{\partial\theta} = \frac{4n}{\theta} - \frac{n(3\theta^2+2\theta)}{\theta^3+\theta^2+6} + 0 - \sum_{i=1}^{n} x_i = 0$$

$$\frac{4n(\theta^3+\theta^2+6) - n\theta(3\theta^2+2\theta)}{\theta(\theta^3+\theta^2+6)} = \sum_{i=1}^{n} x_i$$

$$\frac{4(\theta^3+\theta^2+6) - \theta(3\theta^2+2\theta)}{\theta(\theta^3+\theta^2+6)} = \frac{\sum_{i=1}^{n} x_i}{n} = 0$$

$$(\theta^3+2\theta^2 + 24) - (\theta^4+\theta^3+6\theta)\bar{x} = 0$$

MLE has the following properties:

- The estimator $\hat{\theta}_n$ of θ is consistent if $\hat{\theta}_n \xrightarrow{p} \theta$ as $n \to \infty$. This also implies that

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| > \epsilon\right) = 0$$

- The estimator $\hat{\theta}_n$ of θ is asymptotically normal:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{D} N\left(0, \frac{1}{I(\theta)}\right)$$

 Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

Simulation Study

Using numerical approach, the quantile function for the Stream distribution can be obtained from this expression $x = F^{-1}(u)$, which is derived from $F(x) = u$; where $F(x)$ is the distribution function given by equation (2); where 0 < u < 1. And it provides for the generating of "n" random Stream samples.

$$u = 1 - \left(1 + \frac{\theta x\,[\,\theta^2 + \theta^2 x^2 + 3\theta x + 6\,]}{\theta^3 + \theta^2 + 6}\right) e^{-\theta x}$$

$$1 - u = \left(1 + \frac{\theta x\,[\,\theta^2 + \theta^2 x^2 + 3\theta x + 6\,]}{\theta^3 + \theta^2 + 6}\right) e^{-\theta x}$$

$$ln\left(1 + \frac{\theta x\,[\,\theta^2 + \theta^2 x^2 + 3\theta x + 6\,]}{\theta^3 + \theta^2 + 6}\right) - ln\, ln\,(1-u) - \theta x = 0$$

$$(\theta^3 + \theta^2 + 6)(1-u) - [\,(\theta^3 + \theta^2 + 6\,) + \theta x(\theta^2 + \theta^2 x^2 + 3\theta x + 6)\,]\, e^{-\theta x} = 0$$

EMPIRICAL ANALYSIS

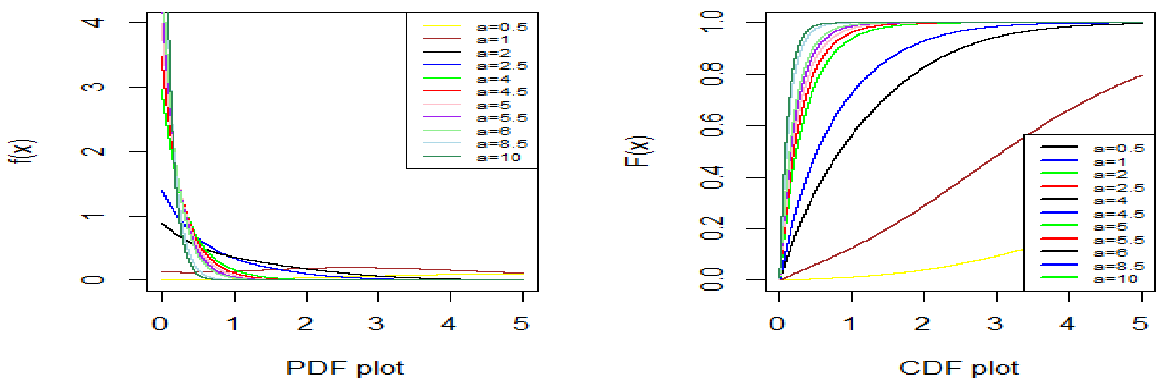

Figure 1: Pdf and CDF plots for Stream Distribution for Different Levels of Parameters

In Figure 1, the plot for pdf of the Stream distribution for selected values of θ shows that the distribution is positively skewed and unimodal; whereas the cdf is an increasing function at various parameter values and converges at $F(x) = 1$ as supposed.

Table 1: Mean, Median and Mode Comparison for Random Simulated Sample Data Using Stream Distribution

Statistics	Data1 $(n = 10)$	Data2 $(n = 50)$	Data3 $(n = 100)$	Data4 $(n = 500)$
Mode	2.000	1.000	4.000	2.000
Median	2.000	2.000	3.000	3.000
Mean	3.500	2.720	3.440	2.870

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

In Table 1, it is observed in data1 for $n = 10$, that *mode (M) = median (m) < mean* (μ). In data2 for $n = 50$, *mode (M) < median (m) < mean* (μ); whereas, in data3 for $n = 100$ *mode (M) > median (m) < mean* (μ). Finally, *mode (M) < median (m) > mean* (μ) as seen in data4 for $n = 500$. We could deduce clearly that as n increases the inequality tends not to show any patterned consistency.

As given by Lindley distribution**,** *mode (M) < median (m) < mean* (μ) under certain conditions; data 2 is seen to adhere to this**.** Abadir (2005), however, stated that "for a unimodal and positively skewed distributions whose first three moments exist, the inequality mode (M) < median (m) < mean (μ) does not necessarily hold". Consequently, data 1, 3 and 4 adhere to Abadir's proposition; and it is seen in the Stream moment derivations in equation (3) that the first three moments exist.

Table 2: Coefficient of Variation (CV) Comparison of Different One-parameter Distributions, Valued at $\theta = 1$.

Distributions	$CV = \frac{\sigma}{\mu}$
Exponential	1
Lindley	0.8819
Akash	0.7693
Shanker	0.8819
Sujatha	0.7617
Ishita	0.7693
Aradhana	0.7551
Akshaya	0.6425
Stream	0.6361

In Table 2, the coefficient of variation also known as the relative standard deviation (RSD) is compared across other parameter probability distributions. The CV for Exponential Distribution equals 1, which implies that the standard deviation and mean are equal; this is different for other listed distributions. According to Everitt (1998), "higher CV of a model indicates greater dispersion around the mean of the model". By implication, lower values of CV, indicate greater precision of its model. Following the result obtained in Table 2, Stream Distribution has the lowest variance when valued at $\theta = 1$. It is worthy of note that this trend is consistent for other parameter θ values. Therefore, Stream Distribution, could be comparatively considered a more efficient model.

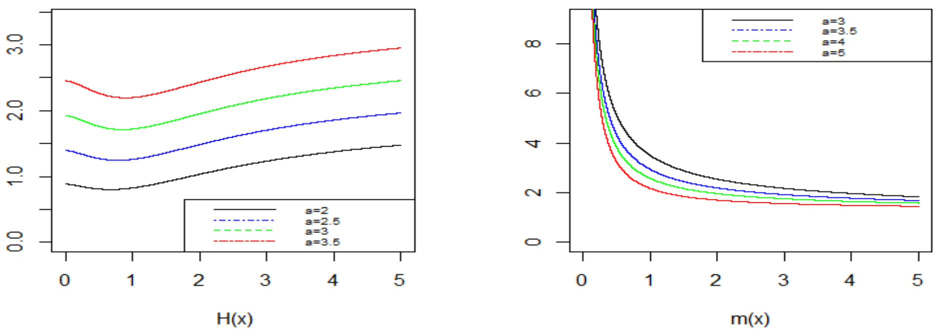

Figure 2: Hazard Plots and Mean Residual Life Plot for Stream Distribution for Different Levels of Parameters

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

In figure 2, the two plots show both increasing and decreasing trends respectively. As a result, Stream distribution is an increasing failure rate model. It is well-known that the MRL and the Hazard function have strong relationship with each other and also to the reliability function. Hence, both the MRL and the hazard functions are able to uniquely determine the distribution of the lifetime of items. In addition, these two functions usually have opposite monotonic trends and represent the aging behavior of a component from different points of view. From the graphs in figures 2, it is confirmed that an increasing failure rate function implies a decreasing MRL function.

Remark: At $x = 0$, $m(0) = \frac{(\theta^3 + 2\theta^2 + 24)}{\theta(\theta^3 + \theta^2 + 6)} = \mu$; and $H(0) = f(0) = \frac{\theta^4}{\theta^3 + \theta^2 + 6}$. This quantity refers to the component failure of the distribution.

Table 3: Biasedness and Consistency of the MLE

N	$\hat{\theta}_{mle}$ at $\theta = 10$	$\hat{\theta}_{mle}$ at $\theta = 15$
20	10.4503	15.6409
50	10.1047	15.2298
80	10.0831	15.2037
100	10.0825	15.1972
200	10.0639	15.0498
400	10.0210	15.0272
600	10.0154	15.0253
800	10.0088	15.0179
1000	10.0049	15.0040

Table 3 shows that MLE is positively bias as $E(\hat{\theta}_{mle}) - \theta > 0$. In addition, as n increases, the MLE's tend to converge to the true parameter values with high probability; which gives a confirmation note to the consistency of the estimator. More so, this convergence will never meet up to equal the true parameter as n keeps increasing, hence we ascribe the MLE to be asymptotically normal.

Computation of the Average Bias and Mean Square Error for $M = 1000$ Monte Carlo Simulations; over the selected values of (n, θ).

$$Average\ Bias = \left[\frac{1}{M}\sum_{i=1}^{M} (\hat{\theta}_i - \theta)\right] \ and\ MSE = \left[\frac{1}{M}\sum_{i=1}^{M} (\hat{\theta}_i - \theta)^2\right] \tag{71}$$

Table 4: Average Bias of the Estimator $\hat{\theta}$

N	$\theta = 8$	$\theta = 10$	$\theta = 15$
20	0.3586	0.4503	0.6409
50	0.1556	0.1047	0.2298
80	0.0994	0.0831	0.2037
100	0.0475	0.0825	0.1972
200	0.0428	0.0639	0.0498
400	0.0167	0.0210	0.0272

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

Table 5: MSE of the Estimator $\hat{\theta}$

N	$\theta = 8$	$\theta = 10$	$\theta = 15$
20	3.0782	5.1541	12.1505
50	1.1013	1.7315	4.4184
80	0.5980	1.0819	2.7699
100	0.5085	0.7982	2.0735
200	0.2535	0.4146	1.0323
400	0.1258	0.1956	0.5009

From Tables 4 and 5, we deduce that the estimates of the average bias and the mean square error decrease as the sample size n increases. In addition, MSE estimates increases as θ increases, for each of the sample sizes.

Table 6: Statistical Table for the PDF of Stream Distribution ($\theta = 0.1$ to $\theta = 0.5$)

X	$\theta = 0.1$	$\theta = 0.2$	$\theta = 0.25$	$\theta = 0.3$	$\theta = 0.35$	$\theta = 0.4$	$\theta = 0.45$	$\theta = 0.5$
1	0.000045	0.00065	0.0015	0.0029	0.0051	0.0083	0.0124	0.0178
2	0.00015	0.00195	0.0042	0.0080	0.0133	0.0203	0.0291	0.0397
3	0.00038	0.00450	0.0094	0.0167	0.0264	0.0384	0.0524	0.0678
4	0.00077	0.00820	0.0163	0.0275	0.0414	0.0573	0.0743	0.0916
5	0.00132	0.01275	0.0241	0.0387	0.0554	0.0729	0.0899	0.1054
6	0.00204	0.01777	0.0320	0.0488	0.0665	0.0832	0.0976	0.1088
7	0.00300	0.02290	0.0392	0.0569	0.0737	0.0878	0.0980	0.1039
8	0.00389	0.02783	0.0453	0.0626	0.0771	0.0874	0.0928	0.0936
9	0.00500	0.03232	0.0500	0.0658	0.0771	0.0831	0.0839	0.0805
10	0.00619	0.03619	0.0533	0.0667	0.0743	0.0762	0.0732	0.0668
11	0.00744	0.03936	0.0522	0.0656	0.0696	0.0678	0.0620	0.0538
12	0.00872	0.04178	0.0557	0.0630	0.0635	0.0589	0.0512	0.0423
13	0.0100	0.04438	0.0551	0.0593	0.0569	0.0502	0.0415	0.0326
14	0.0113	0.04466	0.0535	0.0548	0.0500	0.0419	0.0330	0.0247
15	0.0125	0.04435	0.0513	0.0499	0.0433	0.0346	0.0259	0.0184
16	0.0138	0.04476	0.0484	0.0448	0.0370	0.0281	0.0200	0.0135
17	0.0149	0.04353	0.0452	0.0398	0.0313	0.0226	0.0153	0.0098
18	0.0161	0.04229	0.0418	0.0350	0.0262	0.0180	0.0116	0.0071
19	0.0171	0.04071	0.0382	0.0305	0.0217	0.0142	0.0087	0.0050
20	0.0181	0.03886	0.0347	0.0263	0.0178	0.0111	0.0064	0.0036
21	0.0189	0.03682	0.0313	0.0226	0.0145	0.0086	0.0048	0.0025
22	0.0197	0.03406	0.0280	0.0192	0.0118	0.0066	0.0035	0.0017
23	0.0203	0.03241	0.0249	0.0163	0.0095	0.0051	0.0025	0.0012
24	0.0209	0.03015	0.0221	0.0137	0.0076	0.0039	0.0018	0.0008
25	0.0214	0.02790	0.0194	0.0115	0.0060	0.0029	0.0013	0.0006

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

In Table 6, the trend reveals that the pdf of Stream Distribution is unimodal; and that the axiom $0 < P(x) < 1$ holds across the variable and the parameter values.

A fourteen-week (daily) observation was carried out over a facebook live-streaming program. The data represents time or cycle-to-event data, which is the weekly average number of viewers before it went below 10,000, which was the target for outreach success.

11.2, 10.9, 13.2, 12.0, 11.5, 11.1, 10.8, 10.3, 13.8, 12.5, 12.3, 12.0, 11.1, 13.7, 14.3, 15.3, 13.1, 12.0, 11.8, 10.9, 13.5, 12.6, 13.4, 14.2, 11.6,13.7, 12.6, 11.6, 14.0, 11.0, 13.6, 12.0, 11.5, 11.9, 10.7, 12.6, 12.5, 13.7, 13.5, 12.4, 13.0, 13.2, 12.0, 14.3, 14.3, 12.5, 11.0, 12.0, 13.2, 12.0, 13.5, 13.2, 12.5, 11.6, 14.0, 12.9, 10.5, 13.4, 14.0, 10.5, 12.6, 13.4, 14, 12.6, 13, 12.8, 13.7, 12.7, 13.6, 14.5, 13.4, 12.9, 11.0, 15.1, 13.6, 12.4, 12.9, 11.2, 10.7, 12.3, 13.5, 12.6, 13.5, 12.3, 13.5, 12.4, 12.3, 11.2, 13.5, 10.3, 11.3 (in thousands).

Finally, we test for the flexibility of the Stream distribution, in comparison with some renowned one parameter distributions. Literature has it that two or more parameter distributions usually show superiority over one parameter distributions due to its robustness; hence the comparative choice of similar one parameter distributions.

Among many tools, we employ: lnL (Log-Likelihood), AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion), for performance comparison. The models are given by:

$$AIC = -2lnL + 2k, \; BIC = -2lnL + k \, ln * n \qquad (72)$$

where n is the number of observations, k is the number of estimated parameters and L is the value of the likelihood function evaluated at the parameter estimates

Table 7: Performance Comparison for Stream Probability Distribution:

Model	**Parameter Estimate**	lnL	**AIC**	**BIC**	Rank
Stream	0.3142	-255.36	512.73	515.24	1
Exponential	0.0795	-321.42	644.85	647.36	9
Akash	0.2342	-269.27	540.55	543.06	4
Lindley	0.1487	-292.66	587.31	589.82	8
Shanker	0.1561	-287.75	577.50	580.01	7
Sujatha	0.2270	-271.99	545.97	548.48	5
Aradhana	0.2212	-274.07	550.13	552.64	6
Amarendra	0.3074	-257.19	516.38	518.89	2
Akshaya	0.2946	-260.80	523.60	526.11	3

From Table 7, the best distribution corresponds to the smallest value in AIC, BIC statistics, and or the highest value in lnL. It can be easily seen from Table 7 that the Stream distribution outperforms other distributions in terms of the inferential measures.

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

CONCLUSION

The paper aimed at proposing a new probability distribution suitable for modeling live-streaming data. Different distributions have their niche in modeling data from various fields of life. The empirical analyses carried out in this study are sufficient to project Stream distribution as a novel one parameter distribution with respect to live-streaming data modeling, which to the best of my knowledge is unprecedented in the field of distribution. Since all the models compared have one parameter, it follows that the Stream distribution provides the better fit. A further back-up for this finding, can be obtained by observing the result in Table 2, where it is shown that the coefficient of variation for Stream distribution is least among other listed distributions. These imply that the newly proposed distribution is more efficient in modeling live-streaming data.

REFERENCES

Abadir, K. M. (2005). The mean-median-mode inequality: counterexamples. *Econometric Theory*, 477-482.

Bensid A. E. and Zeghdoudi H, 2017. On the Lindley family distribution: quantile function, limiting distribution of sample minima and maxima and entropies.

Bonferroni, C.E. (1930). *Elementi di statistica generate*, Libreria Seber, Firenze.

Deshpande, J. V., & Purohit, S. G. (2015). *Lifetime Data: Statistical Models and Methods* (Vol. 16). World Scientific Publishing Company.

Friedman, Jerome, Hastie and Robert, 2009. The elements of Statistical learning, volume 2. Springer series in statistics New York, NY, USA.

Ghitany, M. E., Al-Awadhi, F. A., & Alkhalfan, L. (2007). Marshall–Olkin extended Lomax distribution and its application to censored data. *Communications in Statistics—Theory and Methods*, *36*(10), 1855-1866.

Gross A.J., Clark VA., 1975. Survival Distributions: Reliability Applications in the Biometrical Sciences, John Wiley, New York.

Jasiulewicz, H., & Kordecki, W. (2003). Convolutions of Erlang and of Pascal distributions with applications to reliability. *Demonstratio Mathematica*, *36*(1), 231-238.

Ghitany, M. E., Atieh, B., & Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and computers in simulation*, *78*(4), 493-506.

Lindsay B. G., 1995. Mixture models: theory, geometry and applications, NSF-CBMS Regional Conference Series in Probability and Statistics, Hayward, CA, USA: Institute of Mathematical Statistics, ISBN 0-940600-32-3, JSTOR 4153184.

McNolty, F.; Doyle, J. and Hansen E. (Nov., 1980) "Properties of the Mixed Exponential Failure Process" *Technometrics, Taylor and Francis, Ltd.* Vol. 22, No. 4 , pp. 555-565

Merovci, F. (2013). Transmuted lindley distribution. *Int. J. Open Problems Compt. Math*, *6*(2), 63-72.

Oguntunde, P. E., Owoloko, E. A., & Balogun, O. S. (2016). On a new weighted exponential distribution: theory and application. Asian Journal of Applied Sciences, 9(1), 1-12.

Renyi A., 1961. On measures of entropy and information, in: Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, vol I, University of California Press, Berkeley, pp.547-561.

Sankaran, M. (1970). 275. Note: The discrete Poisson-Lindley distribution. *Biometrics*, 145-149.

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 3, 2024 (pp. 79-94)

Shanker, R. (2016). Sujatha distribution and its applications. *Statistics in Transition. New Series*, *17*(3), 391-410.

Shanker R. (2015). Shanker Distribution and its applications. *International Journal of Statistics and Application*, 5(6), 338-348.

Shanker R., (2015). Akash distribution and its applications. *International Journal of Probability and Statistics*, 4(3), 65-75.

Shanker R. (2016). Aradhana distribution and its applications. *International Journal of Statistics and Application*, 6(1), 23-34.

Shanker R. (2016). Amarendra distribution and its applications. *American Journal of Mathematics and Statistics*, 6(1), 44-56.

Shanker R. (2016). Sujatha distribution and its applications. *Statistics in Transition New Series*, 17(3), 391-410.

Shanker R. (2017). Akshaya distribution and its applications. *American Journal of Mathematics and Statistics*, 7(2), 51-59.

Article DOI: 10.52589/AJMSS-RDFCZYVM
DOI URL: https://doi.org/10.52589/AJMSS-RDFCZYVM