BENIN JOURNAL OF STATISTICS
ISSN 2682-5767
Vol. 3, pp. 82– 90 (2020)

On the Parameter Estimation of Johnson's System of Distribution

A.T. Soyinka[1*], A.A. Olosunde[2], F.S. Apantaku[3], A.A. Akintunde[4] and O.A. Wale-Orojo[5]

[1,3,4,5] *Department of Statistics, Federal University of Agriculture Abeokuta, PMB 2022, Nigeria*; [2]*Department of Mathematics, Obafemi Awolowo University Ile-ife 220005, Nigeria*

Abstract. The estimation of the two shape parameters of the Johnson's System of Distributions (JSD) via the Maximum Likelihood Estimation (MLE) method has received considerable attention in the literature. This paper is a research expounded in this direction, albeit via the methods of least squares and moment. The location and the scale parameters of the JSD are obtained in closed form under certain regularity conditions. In order to circumvent the computational agony in the estimation procedure, a third shape parameter is introduced by means of the pivotal quantity method. The utility of the method is illustrated using both simulated and real-life data.

Keywords: Johnson distribution, pivotal quantity, moments, confidence interval, least squares.

Published by: Department of Statistics, University of Benin, Nigeria

1. Introduction

Fitting distributions to data has a long history and many different procedures have been advocated (George et al., 2009). Johnson in 1949, derived a system of curves with various shapes flexibility to represent a set of empirical data (George and Ramachandran, 2011). The shape parameter in the Johnson system aimed at accommodating for the various flexibility from normal random variates was however unable to influence the behavioural pattern of the underlying normal distribution. So in order to make the behavioural pattern of the underlying distribution flexible, Soyinka et al. (2019) introduced the third shape parameter β. The introduction of the new shape parameter was achieved by employing the exponential power distribution as the underlying distribution. Exponential power distribution is a family distribution that accommodates normal distribution when $\beta = 1$, laplace distribution when $\beta = 0.5$ and many other distributions depending on the value of the newly introduced shape parameter which is data dependent (Olosunde and Soyinka, 2019). The advantage of JSD to the modeling of empirical observations cannot be over emphasized because of its practical relevance to real life processes. Some of which include the ability to model non-normal heteroscedastic data with/without lower and/or upper reference points and generate different behavioural patterns from varying distributional shapes depending on random combination of mean-variance-skewness-kurtosis properties (Andrea, 2016). Also, according to Rennolls and Wang (2005), the four parameter JSD was first introduced into forest literature by Hafley and Schreuder (1977), and since then it has been widely used in forest diameter (and height) distribution modelling (Hafley and Buford, 1985; Knoebel and Burkhart, 1991; Zhou and McTague, 1996; Kamziah et al., 1999; Li et al., 2002; Scolforo et al., 2003; Zhang et al., 2003). Recently JSD was used in prediction of life birth (Soyinka et al., 2019) and in the modeling of diameter distributions of nauclea diderrichii stands (Ogundipe et al., 2018). So due to empirical importance of JSD, the need to obtain the estimate of all its parameters in closed form via exact solution to ensure accuracy in modeling, and in case that is impossible, to ensure that the numerical solution converged is a necessity. Considering Lindsey (1999) and George and Ramachandran (2011), the uni-dimensional random variable (rv) X for every $x \in X$ is said to be a member of the Johnson families of distribution with different normalizing transformations to define lower bounded Johnson system S_L, bounded Johnson system S_B and unbounded Johnson system S_B respectively if it has the probability density functions (pdf's):

*Corresponding author. Email: soyinkataiwo@yahoo.ie

Case 1

$$f(z) = \frac{\delta}{\lambda z 2^{1+\frac{1}{2\beta}}\Gamma(1+\frac{1}{2\beta})}\exp\left[-\frac{1}{2}\left(\alpha+\delta\ln z\right)^{2\beta}\right]; x > a. \quad (1.1)$$

Case 2

$$f(z) = \frac{\lambda\delta}{z(1-z)2^{1+\frac{1}{2\beta}}\Gamma(1+\frac{1}{2\beta})}\exp\left[-\frac{1}{2}\left(\alpha+\delta\ln\frac{z}{1-z}\right)^{2\beta}\right]; a \le x \le \lambda + a. \quad (1.2)$$

Case 3

$$f(z) = \frac{\delta}{\lambda\sqrt{1+z^2}2^{1+\frac{1}{2\beta}}\Gamma(1+\frac{1}{2\beta})}\exp\left[-\frac{1}{2}\left(\alpha+\delta\ln z+\sqrt{1+z^2}\right)^{2\beta}\right]; -\infty < x < \infty. \quad (1.3)$$

where $z = \frac{x-a}{\lambda}$. The parameters α, $\beta > 0$ and $\delta > 0$ are the shape parameters that can be obtained through maximum likelihood or moments approach, while a and $\lambda > 0$ are the location and scale parameters, respectively.

Note:

(i) pdf (1.1),(1.2) and (1.3) reduce to the normal distribution related JSD when $\beta = 1$.
(ii) Each of the three cases reduces to the standardized exponential power distribution (SEPD $[0, 1, \beta]$) when the exponent in pdf's (1.1)-(1.3) is defined as a random variable.
(iii) The log-likelihood function is non-regular with respect to the location 'a' and scale 'λ' parameters.

Though the current form of the pdf (1.1), (1.2) and (1.3) cannot give a closed form solution of the location 'a' and scale parameter 'λ' because of the violation of the regularity condition, this study is aimed at establishing propositions via pivotal approach and shape dependent quantile limits to ease the estimation of all JSD parameters via least square and moments approach while maximizing the obtained parameter estimates through its likelihood function using r package 'bbmle' (Bolker, 2020). Finally, the deviance statistic was used to determine the effectiveness of the moment and the MLE approach.

2. Materials and method

2.1 *Pivotal quantity for the scale and location parameters*

DEFINITION 2.1 *(Pivotal Quantity). A pivotal quantity (P) for a parameter θ is a random variable $P(X|\theta = [a, \lambda])$ whose value depends on both (the data) X and on the value of the unknown parameter θ but whose distribution is known to be independent of θ. For the case of the normal distribution $N(a, \lambda^2)$, the pivotal quantity $p_1 = \frac{x-a}{\lambda}$ and $p_2 = \frac{(x-a)^2}{\lambda^2}$ has distributions $N(0, 1)$ and χ^2_1 that are independent of a and λ^2. Therefore, the quantities p_1 and p_2 are pivotal quantities for a and λ^2 respectively (Toulis, 2017; Mood, 1974).*

PROPOSITION 2.2 *If X be defined for the pdf (1.1), (1.2) and (1.3) then a rv P, for each of the transformation function $p \in P$ has the gamma distribution*

$$f(p) = \frac{\delta}{2^{1+\frac{1}{2\beta}}\Gamma(1+\frac{1}{2\beta})}\exp\left[-\frac{1}{2}\left(\alpha+\delta p\right)^{2\beta}\right] \sim \Gamma(\tfrac{1}{2\beta}, 2); -\infty < p < \infty, \beta > 0. \quad (2.1)$$

Proof. Note that $z = \frac{x-a}{\lambda}$ in (1.1), (1.2) and (1.3). The transformation and its derivatives can be obtained as $p(x|a, \lambda) = \ln\left(\frac{x-a}{\lambda}\right) \Rightarrow x = a + \lambda e^p$ which implies $dx = \lambda e^p dp$. $p(x|a, \lambda) = \ln\left(\frac{x-a}{\lambda+a-x}\right) \Rightarrow x = a + \frac{\lambda e^p}{(1+e^p)}$ with $dx = \frac{\lambda e^p}{(1+e^p)^2}dp$ and finally we have $p(x|a, \lambda) = \sinh^{-1}\left(\frac{x-a}{\lambda}\right) \Rightarrow x = a + \lambda sinhp$ with $dx = \lambda\cosh p dp$. Substituting for p and the derivatives in (1.1), (1.2) and (1.3) and simplifying the equation $f(p) = f(x)\frac{dx}{dp}$ we have (2.1). ■

Note: Since (2.1) is independent of location parameter 'a' and scale parameter 'λ' then each of the transformation functions is a pivotal quantity for the location 'a' and scale 'λ' parameters.

2.2 *rth moment estimation of JSD*

PROPOSITION 2.3 *The rth moment is*

$$E(p^r) = \frac{1}{\Gamma(\frac{1}{2\beta})} \left(\frac{2^{\frac{1}{2\beta}}}{\delta} \right)^r \sum_{i=0}^{r} {}^rC_i(-1)^i \left(\frac{\alpha}{2^{\frac{1}{2\beta}}} \right)^i \int_0^p h^{\frac{r+1-i}{2\beta}-1} \exp(-h)dh. \quad (2.2)$$

Proof. From (2.1) using the change of variable technique with the transformation $h = \frac{1}{2}(\alpha + \delta p)^{2\beta}$ and evaluating $\int_{-h}^{h} p^r f(p)dp$ we obtain (2.2). Implying that for a complete gamma function the first $E(p)$ and second $E(p^2)$ moments for the rv (p) are $\frac{1}{\delta}\left(\frac{2^{\frac{1}{2\beta}}\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{2\beta})} - \alpha\right)$ and $\frac{1}{\delta^2}\left(\frac{2^{\frac{1}{\beta}}\Gamma(\frac{3}{2\beta})}{\Gamma(\frac{1}{2\beta})} - \frac{2^{1+\frac{1}{2\beta}}\alpha\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{2\beta})} + \alpha^2\right)$ respectively. Other higher moments of rv (p) can be obtained in the same way along side its variance, skewness and kurtosis. ■

Substituting the various pivotal transformation functions and simplifying further, we obtain the first moment and variance for the initial rv $x \in X$ as follow:

Table 1: First moment and variance of JSD rv (x) in S_U

JSD	Mean	Variance
S_L	$a + \lambda \exp\left(\frac{1}{\delta}\left(\frac{2^{\frac{1}{2\beta}}\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{2\beta})} - \alpha\right)\right)$	$\lambda^2\left(\exp(H) - \exp\frac{2}{\delta}\left(\frac{2^{\frac{1}{2\beta}}\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{2\beta})} - \alpha\right)\right)$
S_B	$a + \lambda\frac{\exp\left(\frac{1}{\delta}\left(\frac{2^{\frac{1}{2\beta}}\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{2\beta})}-\alpha\right)\right)}{1+\exp\left(\frac{1}{\delta}\left(\frac{2^{\frac{1}{2\beta}}\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{2\beta})}-\alpha\right)\right)}$	$\lambda^2\left[\exp(H) - \left(\frac{\exp\frac{1}{\delta}\left(\frac{2^{\frac{1}{2\beta}}\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{2\beta})}-\alpha\right)}{1+\exp\frac{1}{\delta}\left(\frac{2^{\frac{1}{2\beta}}\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{2\beta})}-\alpha\right)}\right)^2\right]$
S_U	$a + \lambda \sinh\left(\frac{1}{\delta}\left(\frac{2^{\frac{1}{2\beta}}\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{2\beta})} - \alpha\right)\right)$	$\lambda^2\left(\exp(H) - \sinh^2\left(\frac{1}{\delta}\left(\frac{2^{\frac{1}{2\beta}}\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{2\beta})} - \alpha\right)\right)\right)$

where H is the obtained second moment of rv p. Subsequently the skewness and kurtosis of JSD can be derived as

$$\lambda^3\left(\exp(I) - 3\exp(J+H) + 2\exp(3J)\right) \quad (2.3)$$

$$\lambda^4\left(\exp(K) - 4\exp(J+I) + 6\exp(2J+H) - 3\exp(4J)\right) \quad (2.4)$$

where H=second moment of p, I=skewness of p, J=first moment, and K=kurtosis of p. So putting H, I, J and K in (2.3) and (2.4), we obtain the estimate of the skewness and kurtosis of JSD for the rv X for each of the three cases under study. Finally, since the cumulative distribution function (cdf) of (2.1) tends to the standardized exponential power distribution (SEPD), either as complete or incomplete gamma function as the case may be, then the maximum likelihood estimate of 'β' though not in closed form can be obtained using SEPD nomalp codes in r environment (Mineo and Ruggerri, 2003b; Olosunde and Soyinka, 2018). So, having obtained 'β' via normalp, its value can be substituted into $E(p)$ and $E(p^2)$ to obtain '$\hat{\alpha}$' and '$\hat{\delta}$' provided the estimate of '$\hat{a}$' and '$\hat{\lambda}$' have been earlier obtained. Having established the Johnson transformation function as a pivotal quantity for the location and scale parameters and its rth moment expressions, we proceed to compute the probability that an unknown random interval of the transformation function contains the estimate of the parameter a and λ. Hence, our next step is to make a proposition that will serve as a foundation to the estimate of the location a and the scale λ parameters.

PROPOSITION 2.4 *Let* $p \sim \Gamma(\frac{1}{2\beta}, 2)$ *from (2.1) and let V be a random variable distributed as standardized exponential power distribution SEPD* $(0, 1, \beta)$ *then a new random variable* T *defined as* $T = \frac{\alpha+\delta p}{v}$ *has the pdf*

$$f(t) = \frac{\beta}{Beta\left(\frac{1}{2\beta}, \frac{1}{2\beta}\right)} \left(1 + t^{2\beta}\right)^{-\frac{1}{\beta}} \tag{2.5}$$

Proof. Supposing p and v are independent $f(p, v) = \frac{\delta}{k^2} \exp(-\frac{1}{2}\left[v^{2\beta} + (\alpha + \delta p)^{2\beta}\right])$, then using change of variable technique with Jacobian $\frac{w}{\delta}$ and $v = w$, we obtain the joint density $f(p, v) = \frac{w}{k^2} \exp(-\frac{w^{2\beta}}{2}\left[1 + t^{2\beta}\right])$ which upon further integration $f(t) = \int f(t, w)dw$ yields the marginal (2.5), where $k = 2^{1+\frac{1}{2\beta}}\Gamma(1 + \frac{1}{2\beta})$. ■

COROLLARY 2.5 *The cdf of (2.5) is*

$$F(t) = beta\left(t, \frac{1}{2\beta}, \frac{1}{2\beta}\right) \tag{2.6}$$

Proof. From (2.5) using the transformation $\tan\theta = T^{\beta}$, the integral $\int_0^T (\sin\theta)^{\frac{1}{\beta}-1}(\cos\theta)^{\frac{1}{\beta}-1}d\theta$ is the regularized incomplete beta function (2.6). ■

DEFINITION 2.6 *(Interval Estimation) Let* $x_1, x_2, .., x_n$ *be a random sample from the density* $f(x|\theta = [\alpha, \beta, \delta, a, \lambda])$ *(1.1)-(1.3). Let* $p(x|a, \lambda)$ *be a pivotal quantity whose distribution is independent of* (a, λ)*; then the two statistics* $T_1 = t(x_1, x_2, .., x_n)$ *and* $T_1 = t(x_1, x_2, .., x_n)$ *defined over a percentile interval* $Pr(t_1 < p(x|a, \lambda) < t_2) \equiv \gamma$ *such that* $t_1 < t_2$*; is said to be the lower* (t_1) *and upper* (t_2) *confidence limits respectively for the pivot* $p(x|a, \lambda)$ *within* 100γ *percent confidence interval, where* γ *is the confidence coefficient and does not depend on* (a, λ) *(Olosunde and Soyinka, 2018).*

From (2.5) the area spanned by the entire X-domain within the interval $t_1 < t_2$ is $\int_{t_1}^{t_2} f(t)dt = F(t) = F(t_2) - F(t_1) \equiv \gamma\%$, $0 \leq \gamma \leq 1$. Suppose that the probability values for the lower and upper confidence limit are $F(t_1) = \frac{1-\gamma}{2}$ and $F(t_2) = \frac{1+\gamma}{2}$ respectively then from (2.6) we can deduce

$$(t_1; t_2) = \left[beta^{-1}\left[\left(\frac{1-\gamma}{2}, \frac{1}{2\beta}, \frac{1}{2\beta}\right)\right]; beta^{-1}\left[\left(\frac{1+\gamma}{2}, \frac{1}{2\beta}, \frac{1}{2\beta}\right)\right]\right] \tag{2.7}$$

where (2.7) is the interval estimate of the pivot value (t_1, t_2), which is the regularized inverse incomplete beta function. Likewise the point estimate of the pivot value t for any arbitrary probability value γ is

$$t = beta^{-1}\left(\gamma, \frac{1}{2\beta}, \frac{1}{2\beta}\right) \tag{2.8}$$

Note: Using codes in R environment, we obtain the pivot, interval or point estimate for various confidence coefficient $0 < \gamma < 1$ from (2.7) and (2.8).

2.3 *Estimation of JSD parameters*

Estimation of the location a and the scale parameter λ via least square approach

If for every observed values $x_1, x_2, .., x_n$ there exist a random interval $(x < [a, \lambda], x \geq [a, \lambda])$ which spans an area γ over a percentile scale; then there exist a corresponding pivot interval $(t_1(x|a, \lambda); t_2(x|a, \lambda))$ matching each observed value which spans the same area.

DEFINITION 2.7 *Let* $x_1, x_2, .., x_n$ *be a random sample from the density* $f(x|\theta = [\alpha, \beta, \delta, a, \lambda])$*, if* $x_1 \leq x_2 \leq, .., \leq x_n$ *represents the relative standing of each observed value after ordering in ascending order, then the ordered pair* x_j, t_j *for* $j = 1, 2, .., n$ *signifies the ordered observed values and its corresponding pivot values which occupy the same theoretical probability area* j/n *(George and Ramachandran, 2011; Olosunde and Soyinka, 2019).*

Substituting $\gamma = j/n$ in (11), we obtain the desired pivot values t_j for each ordered observed value assuming equal cumulative increase that reflects relativity within observations. Next we obtain the least square estimates of the location 'a' and scale 'λ' parameters from the linear regression of ordered pair x_j, t_j.

Table 2: The transformation function, the regression model and the estimator for the location and scale parameters scale (λ) parameters of JSD

Tansformation function $p(x\|a,\lambda)$	Linear regression model (t_i, x_i)	Location parameter 'a' estimate	Scale parameter 'λ' estimate
$\ln\left(\frac{x-a}{\lambda}\right)$	$x_i = a + \lambda e^{t_i} + E$	$\bar{x} - \frac{\lambda}{n}\sum_{i=1}^{n} e^{(t_i)}$	$\frac{n\sum_{i=1}^{n} x_i e^{(t_i)} - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} e^{t_i})}{(n\sum_{i=1}^{n} e^{2t_i}) - (\sum_{i=1}^{n} e^{t_i})^2}$
$\ln\left(\frac{x-a}{\lambda+a-x}\right)$	$x_i = a + \lambda\left(\frac{e^{(t_i)}}{1+e^{(t_i)}}\right) + E$	$\bar{x} - \frac{\lambda}{n}\sum_{i=1}^{n}\left(\frac{e^{(t_i)}}{1+e^{(t_i)}}\right)$	$\frac{n\sum_{i=1}^{n} x_i \frac{e^{(t_i)}}{1+e^{(t_i)}} - (\sum_{i=1}^{n} x_i)\left(\sum_{i=1}^{n}\frac{e^{(t_i)}}{1+e^{(t_i)}}\right)}{\left(n\sum_{i=1}^{n}\frac{e^{(t_i)}}{1+e^{(t_i)}}\right)^2 - \left(\sum_{i=1}^{n}\frac{e^{(t_i)}}{1+e^{(t_i)}}\right)^2}$
$\sinh^{-1}\left(\frac{x-a}{\lambda}\right)$	$x_i = a + \lambda\sinh(t_i) + E$	$\bar{x} - \frac{\lambda}{n}\sum_{i=1}^{n}\sinh(t_i)$	$\frac{n\sum_{i=1}^{n} x_i \sinh(t_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n}\sinh(t_i))}{n\sum_{i=1}^{n}\sinh^2 t_i - (\sum_{i=1}^{n}\sinh t_i)^2}$

E-Error

2.4 *Maximum Likelihood Estimation (MLE) of the shape parameters (α and δ)*

From the log-likelihood function of (2.1)

$$\ln L(p) = n\ln\left(\frac{\delta}{2^{1+\frac{1}{2\beta}}}\Gamma(1+\frac{1}{2\beta})\right) - \frac{1}{2}\sum_{i=1}^{n}(\alpha+\delta p_i)^{2\beta} \tag{2.9}$$

the derivative of (2.9) with respect to α and δ are $\frac{\partial \ln L(p)}{\partial\alpha} = -\beta\sum_{i=1}^{n}(\alpha+\delta p_i)^{2\beta-1}$ and $\frac{\partial \ln L(p)}{\partial\delta} = \frac{n}{\delta} - \beta\sum_{i=1}^{n} p_i(\alpha+\delta p_i)^{2\beta-1}$; $\forall\, \beta \in \Re$. Likewise the Fisher information matrix $I(\alpha,\delta)$ of JSD can be obtained by arranging the following expectations of the second derivatives log likelihood function into a 2x2 matrix $I(\alpha,\delta) = -\begin{pmatrix} E(\frac{\partial^2 \ln L(p)}{\partial\alpha^2}), E(\frac{\partial^2 \ln L(p)}{\partial\alpha\partial\delta}) \\ E(\frac{\partial^2 \ln L(p)}{\partial\alpha\partial\delta}), E(\frac{\partial^2 \ln L(p)}{\partial\delta^2}) \end{pmatrix}$. Note: The standard error of the Fisher estimates in the leading diagonals tends to zero as $n \longrightarrow \infty\, \forall\, \beta \geq 1$. This implies that as n increases, $I(\alpha,\delta)$ becomes large, variance of and becomes small and the sampling of different initial values of $\hat{\alpha}$ and $\hat{\delta}$ from the iterative procedure $\begin{pmatrix}\hat{\alpha}_n \\ \hat{\delta}_n\end{pmatrix} = \begin{pmatrix}\hat{\alpha}_{n-1} \\ \hat{\delta}_{n-1}\end{pmatrix} + \left(I(\hat{\alpha}_{n-1},\hat{\delta}_{n-1})^{-1} S(\hat{\alpha}_{n-1},\hat{\delta}_{n-1})\right)$ will lead to mle estimates of $\hat{\alpha}$ and $\hat{\delta}$ within the close neighbourhood of the initial values of $\hat{\alpha}_o$ and $\hat{\delta}_o$ where $S(\hat{\alpha}_{n-1},\hat{\delta}_{n-1}) = (\frac{\partial \ln L(p)}{\partial\alpha}, \frac{\partial \ln L(p)}{\partial\delta})^T$. Owing that the estimation of $\hat{\alpha}$ and $\hat{\delta}$ from the polynomial obtained via the log likelihood first derivatives cannot be obtained in closed form due to polynomial power that is not an integer; we estimate the initial values of $\hat{\alpha}_o$ and $\hat{\delta}_o$ from the first and second moments of rv(p) and eventually obtain the parameter value from the iterative procedure that maximizes (2.9).

PROPOSITION 2.8 *The estimate $(\hat{\alpha}_o, \hat{\delta}_o)$ of the shape parameter $(\hat{\alpha}, \hat{\delta})$ are*

$$\hat{\delta}_o^2 = \frac{2^{\frac{1}{\beta}}}{E(p^2)-[E(p)]^2}\left[\frac{\Gamma(\frac{3}{2\beta})}{\Gamma(\frac{1}{2\beta})} - \Gamma^2(\frac{1}{\beta})\Gamma^2(\frac{1}{2\beta})\right] and\ \hat{\alpha}_o = \frac{2^{\frac{1}{\beta}}\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{2\beta})} - \hat{\delta}_o E(p) \tag{2.10}$$

Proof. Evaluate the variance of rv(p) $E(p^2) - [E(p)]^2$ and make $\hat{\delta^2}_o$ the subject of the formula. Afterwards, put the obtained value of $\hat{\delta}_o$ into the first moment and solve for $\hat{\alpha}_o$. Starting with $(\hat{\alpha}_o, \hat{\delta}_o)$ as the initial values, we obtain the values $(\hat{\alpha}, \hat{\delta})$ that maximizes the log-likelihood function of (2.9) using 'bbmle' package in r environment. ■

2.5 *Properties of JSD*

Soyinka et al. 2019, established that the incomplete gamma characteristic function of the lower S_L and bounded S_B JSD in series form is

$$\phi(t) = \frac{\exp\left(-\frac{it\alpha}{\delta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)} \sum_{n=0}^{\infty} \frac{\left(\frac{it}{\delta}2^{\frac{1}{2\beta}}\right)^n \left[\gamma\left(\frac{n+1}{2\beta},u\right)-\gamma\left(\frac{n+1}{2\beta},s\right)\right]}{\Gamma(n+1)}, s < p < u \tag{2.11}$$

Note: Equation (2.11) converges to a finite value via ratio test of convergence. Likewise, the characteristics function of the unbounded S_U JSD is

$$\phi(t) = \frac{\exp\left(-\frac{it\alpha}{\delta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)} \sum_{n=0}^{\infty} \frac{\left(\frac{it}{\delta}2^{\frac{1}{2\beta}}\right)^n \Gamma\left(\frac{n+1}{2\beta}\right)}{\Gamma(n+1)}. \tag{2.12}$$

In addition, the Rényi entropy of order J for JSD is

$$H_J(p) = \ln\left(\frac{k}{\delta}\right) + \frac{1}{2\beta(J-1)} \ln J. \tag{2.13}$$

This suggests that the Rényi entropy is dependent on the shape parameter(β) and thus pre-fixing the shape parameter will lead to unreliable result.
The rth moments of the rv(p) was obtained as

$$E(p^r) = \frac{(-1)^r\alpha^r}{\Gamma\left(\frac{1}{2\beta}\right)\delta^r} \sum_{n=0}^{r} {}^rC_n(-1)^n\alpha^{-n}2^{\frac{n}{2\beta}} \left[\gamma\left(\frac{n+1}{2\beta},u\right) - \gamma\left(\frac{n+1}{2\beta},s\right)\right] \tag{2.14}$$

for S_L and S_B whose domain is finite $(s < p < u) \in \Re^+$, and for the S_U whose domain spans the entire real line the rth moment is

$$E(p^r) = \frac{(-1)^r\alpha^r}{\Gamma\left(\frac{1}{2\beta}\right)\delta^r} \sum_{n=0}^{r} {}^rC_n(-1)^n\alpha^{-n}2^{\frac{n}{2\beta}}\Gamma\left(\frac{n+1}{2\beta}\right). \tag{2.15}$$

Using the ratio test of convergence, the series solution in the rth moment (2.14) converges to a finite value provided $r - n < n$.

3. Results and discussion

3.1 *Applications*

A real life data and a simulated data was used to demonstrate all the results obtained. First a practical data on time taken by pilot to make a known corrective action during flight as obtained by Lehmann and Romano (2005) was first considered. The second and third examples are from simulated data with sizes $n = 500$ and $n = 1000000$ generated in R software using normalp package. Codes to compute various estimates are written in R environment.

EXAMPLE 3.1 *Twenty pilots were tested in a flight simulator and the time for each of them to complete a certain corrective action was measured in seconds. The results are as follow: 5.2, 5.6, 7.6, 6.8, 4.8, 5.7, 9.0, 6.0, 4.9, 7.4, 6.5, 7.9, 6.8, 4.3, 8.5, 3.6, 6.1, 5.8, 6.4, 4.0.*

The data from each example were evaluated assuming S_L, S_B and S_U JSD. Kenneth and David (2013) on shifting negative AIC values.

Table 3 showed the estimates of the JSD parameters obtained via the moment and the MLE approach. Table 4 revealed the Kolmogorov Smirnov test of the goodness of fitted model to the samples. The $p-$value for the KS test are all > 0.05 indicating that the JSD distribution fits the data. Finally in Table 5, following

Table 3: Parameters estimates and standard error of JSD for the different examples with respects to the various cases in the study

Example	Parameter	Case 1	Case 2	Case 3	β
Example 1	a	2.5130(0.3328)	-3.1317 (0.8368)	4.4005(0.1859)	1.5434
	λ	1.9968 (0.1716)	14.8665 (1.3273)	2.9533	
Moment	δ	1.0974 (0.0022)	0.3758 (0.002)	2.2498 (0.0022)	
	α	0.0742 (0.0273)	-0.1766 (0.0061)	0.8155 (0.0047)	
MLE	δ	1.6309 (1.176×10^{-7})	0.4874 (6.127×10^{-6})	0.7786 (1.021×10^{-6})	
	α	-0.7073 (1.2736×10^{-6})	-0.1044 (3.016×10^{-6})	0.2129 (3.251×10^{-7})	
Example 2	a	0.7795(0.1756)	-4.8955 (0.4138)	2.7237(0.0981)	2.5009
	λ	2.02534 (0.0902)	15.0901 (0.6607)	2.9879 (0.1293)	
Moment	δ	1.82768 (0.0005)	0.5174 (0.0001)	2.78495 (0.0348)	
	α	-0.2609 (0.006)	-0.5572 (0.0126)	0.7187 (0.002)	
MLE	δ	0.7934 (6.2×10^{-6})	0.346 (1.204×10^{-5})	0.5827 (9.109×10^{-8})	
	α	0.0144 (1.358×10^{-5})	-0.7323 (4.339×10^{-6})	0.3204 (5.585×10^{-8})	
Example 3	a	0.6638(0.0023)	-6.7178 (0.0055)	3.2112(0.00131)	2.6727
	λ	2.6466 (0.0012)	19.6779 (0.0088)	3.8995 (0.0017)	
Moment	δ	1.9141 (9.12×10^{-7})	0.5334 (6.47×10^{-9})	3.1986 (7.33×10^{-9})	
	α	-0.2999 (1.4573×10^{-7})	-0.7463 (2.077×10^{-7})	0.9193 (2.15×10^{-7})	
MLE	δ	0.5075 (0.0576)	0.5784 (0.211)	0.4252 (0.9311)	
	α	0.4894 (0.0146)	-0.4651 (0.0065)	0.4377 (0.03455)	

Table 4: Kolmogorov-Smirnov test on Example 1

F_X	Case 1	Case 2	Case 3
0.1453	0.0014	0.9802	0.2393
0.3643	0.055	0.1983	0.6612
0.2129	0.1501	5.44×10^{-9}	0.0928
0.2079	0.352	8.17×10^{-15}	0.0067
0.0522	0.2498	1.036×10^{-29}	9.788×10^{-7}
0.01	0.094	1.867×10^{-48}	1.03×10^{-10}
0.0073	0.098	9.82×10^{-61}	1.14×10^{-12}
$D_{cal}(p-value)$	0.2867 (0.9627)	0.7143 (0.053)	0.5714 (0.2121)

Table 5: Deviance statistics for moment and ML estimates with AIC, BIC estimate of $\beta > 1$

	Deviance for moment estimate	Deviance for MLE	AIC $\beta > 1$	AIC $\beta = 1$	BIC $\beta > 1$	BIC $\beta = 1$
Example 1 S_L	9.098	1.2805	7.527	2.6064	4.711	4.859
S_B	5.373	0.7277	9.4478	5.0321	5.2638	3.9672
S_U	1.955	4.693	12.18	8.4493	1.2986	2.2722
Example 2 S_L	124620	1735.85	174.7	1793.58	32.244	1723.42
S_B	246.53	604.28	182.45	899.33	1.0558	591.85
S_U	74822.1	186.4	205.3	514.13	15.577	173.97
Example 3 S_L	2.6×10^8	6.3×10^6	1.5×10^5	1.9×10^6	237.1	6.3×10^6
S_B	1.2×10^5	2.4×10^5	4.2×10^5	1.1×10^6	3.9×10^4	2.4×10^5
S_U	4.0×10^8	7.3×10^5	5.5×10^5	1.9×10^6	4.7×10^4	7.3×10^5

the estimates of $(\hat{\alpha}, \hat{\delta})$ from the moment and the MLE approach, we determine the estimate with the least null deviance, since residual deviance is a difference of two null deviance (Helie, 2006). For each examples in the row of Table 5, three different analysis were evaluated assumming S_L,S_B and S_U JSD making nine analysis in all. The deviance statistics from MLE (column 2) indicates that MLE performed better than moment estimate for small sizes $n > 30$. Though we have a different result in S_U JSD (case 3) in example 1, this result was however not significant $\chi^2_{cal} = 4.693 < \chi^2_{15,0.05,tab} = 7.261$. For large sample sizes, both the moment and the MLE approach performed better in different cases based on their deviant statistics, but overall MLE perfomed twice better than moments approach based on the empirical evidence in this study. In addition, column 4 and 5 presented the estimate of the Akaike information criteria, to justify the inclusion of the fifth parameter β. Column 4 explains the result of AIC when the new parameter is introduced $\beta > 1$, while column 5 signifies the situation when the new parameter is removed $\beta = 1$. The AIC statistics however justified the significance of the introduced new parameter in the accurate modeling of large samples. On the other hand, unlike AIC, the Bayesian information criteria (BIC) demonstrated the superiority of 5-parameter JSD on the 4-parameter JSD irrespective of the sample sizes.

4. Conclusion

This study introduces the 5th parameter into the JSD to influence its underlying behavioural pattern. Some properties of JSD, dependent on the introduced parameter, were derived. These properties however suggested that fixing the introduced parameter as we have in previous studies will lead to unreliable analysis. The significance of the introduced 5th parameter is further justified using AIC and BIC. In addition, the study applied the pivotal quantity approach to make the non-regular JSD pdf become regular. This eases the estimation of its parameters via moments and least square approach over a standardized quantile scale. The least square approach to estimating the location and scale parameter of the different JSD cases is a good estimator, though not uniformly the most powerful, but it gives a minimum variance on likelihood. The results via the moments and least square approach was then maximized to obtain the parameter estimates with the highest likelihood value. Though the two approach were effective for the different assumptions, the empirical evidence via deviance statistics showed the dominance of MLE estimates over the moment estimate. The parameter estimation procedure engaged in this study were proven, encouraging and aside that it is easy to understand; it also has a standard error that tends to zero as sample sizes increase to infinity going by the convergence of the maximization technique. Further research on some properties of JSD which includes its kullback leibler information, its inverse distribution function as well as its confidence estimates can be carried out for the purpose of establishing hypothesis testing for one sample and two sample JSD. The codes to reproduce the results in this study is available in R for easy usage by practitioners.

Acknowledgement

We would like to appreciate the useful suggestions of the anonymous reviewers. Those suggestions have not only improved the paper, it has also enriched our understanding of the work.

References

Andrea, C. (2016). Log-concave functions. Partially supported by FIR Project 2013.

Bolker, B. and R Development Core Team (2020). Bbmle: Tools for general maximum likelihood estimation. R package version 1.0.23.1.https://CRAN.R-project.org/package=bbmle.

George, F. and Ramachandran, K.M. (2011). Estimation of parameters of Johnson's system of distributions. Journal of Modern Applied Statistical Methods, **10**: 494-504.

George, F., Ramachandran, K. M. and Lihua, L. (2009). Gene selection with Johnson distribution. Journal of Statistical Research, **43**: 117-125.

Gupta, R. and Balakrishnan, N. (2012). Log-concativity and monotonicity of hazard and reversed hazard functions of univariate and multivariate skew-normal distributions. Metrika, **75**: 181-191.

Helie, S. (2006). An introduction to model section: tools and algorithms. Tutorials in Quantitative Methods for Psychology, **2**: 1-10.

Kamziah, A. K., Ahmad, M. I. and Jaffirin, L. (1999). Nonlinear regresion approach to estimating Johnson SB parameters for diameter data. Canadian Journal of Forestry Resources, **29**: 310-314.

Kenneth, P.B. and David, R.A. (2013). Model selection and multimodel inference: A practical information theoretic approach. Springer.

Lehmann, E.L. and Romano, J. P. (2005). Testing statistical hypotheses (3rd ed.). Springer-Verlag, New York.

Lindsey, J.K. (1999). Multivariate elliptically contoured distributions for repeated measurements. Biometrics, **55**: 1277-1280.

Mineo, A.M. and Rugerri, M. (2003). On the estimation of the structure parameter of a normal distribution of order p. Statistica, **63**: 109–122.

Mood, A. (1974). Introduction to the theory of statistics. McGraw-Hi1lseries in Probability and Statistics.

Ogundipe, O.C., Osho, J.S.A. and Olakunle, O.D. (2018). Modelling diameter distributions of nauclea diderrichii stands in Omo Forest Reserve Nigeria with Johnson S_B. Journal of Agricultural Research and Technology, **18**: 149-154.

Olosunde, A.A., and Soyinka, A.T. (2018). On interval estimation for exponential power distribution parameters. Journal of Data Science, **10**: 93–104.

Olosunde, A.A., and Soyinka, A.T. (2019). Interval estimation for symmetric and asymmetric exponential power distribution parameters. Journal of the Iranian Statistical Society, 18: 237-252.

Rennolls, K. and Wang, M. (2005). A new parameterization of Johnson's distribution with application to fitting forest tree diameter data. DOI:10.1139/x05-006.

Soyinka, A.T., Olosunde, A.A., Wale-Orojo, O.A. and Atanda, O.D. (2019). Properties of generalized Johnson systems of distribution with applications in modeling and prediction of life birth. 15th International Conference on Electronics Computer and Computation (ICECCO).

Toulis, P. (2017). A useful pivotal quantity. Journal of American Statistician, **71**: 272-274.

Zacks, S. (1981). Parametric statistical inference basic theory and modern approaches. Pergamon Press State University of New York, Binghamton.