Int. J. Nonlinear Anal. Appl. 13 (2022) 1, 2655-2663
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.5987

Comparison of the percentile estimation method and mixture (maximum likelihood and least square) method for estimating parameters of Johnson bounded distribution

Sanaa Mohammed Naeem[a,*], A. N. Hussein[a]

[a] *College of Management and Economics, University of Basrah, Iraq*

(Communicated by Madjid Eshaghi Gordji)

Abstract

The Johnson Bonded Distribution (JSBD) is one of the distributions belonging to the Johnson Distribution family (JD) This family is considered one of the flexible distributions that enable it to represent the random behaviour of many phenomena. The method of percentile estimation and the Mixture method (Maximum Likelihood and Least Square) were used to find estimates of the distribution parameters. The simulation results showed the advantage of the Mixture method at large sample sizes, the accuracy of the two methods is close at average sample sizes.

Keywords: Johnson distribution, Johnson bounded distribution, percentile method, Mixture method.

1. Introduction

The Johnson Bounded Distribution (JBD) is one of the distributions that belong to the family of Johnson Distribution . These distributions are characterized by being flexible distributions, as through this family of probability distribution it is possible to cover a wide range of shapes for triple and quadrilateral moments $\sqrt{B_1.B_2}$. The first category of Johnson Distribution known as Johnson Unbounded distribution and it can cover shapes for undefined triple , quadruple moments, while the second category of these distributions is known as Johnson Bounded Distribution , and these

*Corresponding author

Email addresses: sunamohammed70@gmail.com (Sanaa Mohammed Naeem), ali.hussien@uobasrah.edu.iq (A. N. Hussein)

Received: September 2021 *Accepted:* December 2021

distributions cover shapes for specific (third and fourth) central moments. These distributions were developed from a standard normal distribution by making a transformation between a standard natural random variable and a random variable . JD distributions were used to represent the random behavior of many application because they are flexible probability distributions such as human pollution applications (exposure) air pollution data , rain data and other phenomena Statistical theory pays wide attention to methods of estimating parameters of probability distributions. The method of greatest possibility and method of least squares are among the most important methods of estimation . The method of estimating using the percentile is one of the methods through which the estimators of the probability distribution are found by equating the percentile values of the actual data with their theoretical value.

2. Distribution Johnson S_B

J ohnson suggested a family of continuous distributions using the transformations method for the relationship between a standard normal variable (Z) and a random variable (x) that the family of distributions proposed by Johnson consists of three probability distributions (SU,SB,SL) these differ Distributions by different conversion method .

Assuming (x) a random variable and(z)standard natural variable ,Johnson suggested the following transformation :

$$z = \beta + \delta g\left(x\right) \tag{2.1}$$

Since the transformation function $g(x), (\beta, \delta)$is the parameters shape ,SL is known as the log –normal distribution system and the general formula for the transformation function is general formula for this distribution as follows:

$$z = \beta + \delta \log\left(\frac{x-\theta}{\lambda}\right) \quad ; x > \theta \tag{2.2}$$

Since θ is the location parameter , λ is the scale parameter ,the system of log-normal distributions ,while S_B is known as the system of definite or restricted distributions whose general formula is as follows :

$$z = \beta + \delta \log\left(\frac{x-\theta}{\lambda+\theta-\lambda}\right) \qquad \theta < x < \theta+\lambda \tag{2.3}$$

This system of probability distributions covers for random distributions that are restricted to one or both sides ,such as gamma curves, beta distributions and many other distributions . as for the system of undefined or restricted S_U distributions ,the general formula is as follows.

$$\begin{aligned} z &= \beta + \delta \log\left(\left(\frac{x-\theta}{\lambda}\right) + \left(\left(\frac{x-\theta}{\lambda}\right)^2 + 1\right)^{\frac{1}{2}}\right), \quad -\infty < x < \infty \\ z &= \beta + \delta \sinh^{-1}\left(\frac{x-\theta}{\lambda}\right) \end{aligned} \tag{2.4}$$

3. The probability density function of the Johnson distribution

Applying the transformations to formula (2.3) we get the probability density function for the Johnson (S_B) distribution as follows :

$$p\left(\xi\right) = \frac{\delta}{\sqrt{2\pi}} \frac{1}{\xi/(1-\xi)} \exp\left(-\frac{1}{2}\left(\beta + \delta \log\left(\frac{\xi}{1-\xi}\right)\right)^2\right) \tag{3.1}$$

Where : $\xi = \frac{x-\theta}{\lambda}$, $\theta < x < \theta + \lambda$.
There is no explicit formula for the probability density function for this distribution , and it can be obtained using ordinary methods for calculating the following integral

$$F(\xi) = \int_{\theta}^{u} \frac{\delta}{\sqrt{2\pi}} \frac{1}{\xi/(1-\xi)} \exp\left(-\frac{1}{2}\left(\beta + \delta \log\left(\frac{\xi}{1-\xi}\right)\right)^2\right) dx \tag{3.2}$$

4. Some characteristics of the distribution

In this section , some properties of the Johnson S_B distribution will be derived , as follows:

Arithmetic mean . The arithmetic mean of a distribution is obtained by finding the following integral .

$$E(x) = \int_{\theta}^{\theta+\lambda} \frac{\delta}{\sqrt{2\pi}} \frac{1}{\xi/(1-\xi)} \exp\left(-\frac{1}{2}\left(\beta + \delta \log\left(\frac{\xi}{1-\xi}\right)\right)^2\right) \mathrm{dx} \tag{4.1}$$

[5] developed the arithmetic mean formula for the Johnson S_B distribution

$$E(x) = \theta + \lambda \frac{\exp\left(\frac{1}{\delta}\left(\frac{2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)} - \theta\right)\right)}{1 + exp\left(\frac{1}{\delta}\left(\frac{2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)} - \theta\right)\right)} \tag{4.2}$$

The variance of the Johnson distribution is obtained from the following formula :

$$V(x) = \lambda^2 \left(\exp(H) - \left(\frac{2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)} - \theta\right)^2\right) \tag{4.3}$$

5. Estimation methods

Two methods were used to estimate the parameters of the Johnson S_B distribution, the first method is the percentile method ,and the second method is a mixture method of the maximum Likelihood method and the Least Squares method.

5.1. Method of percentile estimation

The term percentile plays an important role in descriptive statistics , and percentiles can be used with the cumulative distribution function . By solving the resulting equations simultaneously, we get estimates for the unknown parameters.

To estimate the parameters of the Johnson S_B distribution , we find an equation (k) where (k) represents the number of parameters for the distribution . And that is by equating (k) of the percentiles or percentages of the standard normal distribution with the corresponding percentages or percentiles for the community , assuming that a represents the percentage or percentile where $\{a_j : 1 \leq j \leq k\}$ and on the assumption that z_{α_j} represents the percentile or the percentage of the standard normal distribution as that x_{α_j} represents the percentile or percentage of the population where :

$$z_{\alpha_j} = \phi^{-1}(\alpha_j)$$
$$x_{\alpha_j} = F^{-1}(\alpha_j)$$

Since (ϕ^{-1}) is the inverse of the standard normal distribution function. (F^{-1}) is the invers of the distribution function.
To find estimates of the parameters of the probability density function of the Johnson distribution , the steps of the percentage method include finding the solution to the following system of equations.

$$z_{\alpha_j} = \gamma + \eta ln\left(\frac{\widehat{x}_{\alpha_j} - \xi}{\lambda + \xi - \widehat{x}_{\alpha_j}}\right) \tag{5.1}$$

Assuming that

$$m = x_{3z} - x_z \tag{5.2}$$
$$n = x_{-z} - x_{-3z} \tag{5.3}$$
$$p = x_z - x_{-z} \tag{5.4}$$
$$x = \xi + \frac{\lambda}{1 + e^{\frac{\gamma - z}{\eta}}} \tag{5.5}$$

Using the relationship between hyperbolic function and exponential functions produces the following

$$m = \frac{\lambda \sinh\left(\frac{z}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma - 2z}{\eta}\right)} \tag{5.6}$$
$$n = \frac{\lambda \sinh\left(\frac{z}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma - 2z}{\eta}\right)} \tag{5.7}$$
$$p = \frac{\lambda \sinh\left(\frac{z}{\eta}\right)}{\cosh\left(\frac{z}{\eta}\right) + \cosh\left(\frac{\gamma}{\eta}\right)} \tag{5.8}$$

By performing som algebraic operations , we get the estimation of the following parameters .

$$\eta = \frac{z}{\cosh^{-1}\left(\frac{1}{2}\left(\left(1 + \frac{p}{m}\right)\left(1 + \frac{p}{n}\right)\right)^{\frac{1}{2}}\right)} \tag{5.9}$$
$$\gamma = \eta \sinh^{-1}\left(\frac{\left(\frac{p}{n} - \frac{p}{m}\right)\left(\left(1 + \frac{p}{m}\right)\left(1 + \frac{p}{n}\right) - 4\right)^{\frac{1}{2}}}{2\left(\frac{p}{m}\frac{p}{n} - 1\right)}\right) \tag{5.10}$$

5.2. Maximum Likelihood & Least Square Method

This method was suggested by [4]. In this method ,the least squares method and the greatest possibility method are combined to obtain an estimate of the parameters of the probability density function of the Johnson S_B distribution. The objective of combining the two methods to obtain a system of nonlinear equations is the weighting function of the JSBD as follows :

$$L(x) = \frac{\delta^n}{(2\pi)^{\frac{n}{2}}} \Pi_{i=1}^{n} \left(\frac{\lambda}{(\lambda + \Theta - xi)(xi - \Theta)}\right) e^{-\frac{1}{2}\sum\left[\beta + \delta ln\left(\frac{x_i - \Theta}{\lambda + \Theta - x_i}\right)\right]^2} \tag{5.11}$$

To get the estimators values using the combined method (Maximum Likelihood &Least Squares) it is first done by finding estimates for the shape parameters (δ, β) using the greatest possibility method, as follows:

Finding the logarithm of the weighting function (5.11) we get the following.

$$\ln L = nlog\,(\delta) + nlog\,(\lambda) + \frac{n}{2}\log(2\pi) - \sum_{i}^{n}\log(\lambda+\theta-x) - \sum_{i}^{n}\log(x_i-\theta) - \frac{n\gamma}{2} + \delta\sum_{i}^{n}\log\left(\frac{x_i-\theta}{\lambda+\theta-x_i}\right) \tag{5.12}$$

Taking the first derivative of the parameter δ and being equal to zero we get.

$$\delta^2\sum_{i=1}^{n}\ln\left(\frac{xi-\Theta}{\lambda+\Theta-xi}\right)^2 + \beta\delta\sum_{i=1}^{n}\ln\left(\frac{xi-\Theta}{\lambda+\Theta-xi}\right) - n = 0 \tag{5.13}$$

Performing some algebraic and simplifying operations, we get the following.

$$\widehat{\delta} = \left(\frac{n-\beta\delta\sum_{i=1}^{n}\ln\left(\frac{xi-\Theta}{\lambda+\Theta-xi}\right)}{\sum_{i=1}^{n}\ln\left(\frac{xi-\Theta}{\lambda+\Theta-xi}\right)^2}\right)^{\frac{1}{2}} \tag{5.14}$$

In order to obtain an estimate for the shape parameter β , we find the partial derivative of the weighting equation (5.11) with respect to the parameter β and then equalize it to zero, we get the following

$$\widehat{\beta} = \frac{-1}{n}\sum_{i=1}^{n}\delta\ln\left(\frac{x_i-\theta}{\lambda+\theta-x_i}\right) \tag{5.15}$$

The value of parameter β can be substituted into equation (5.13) as follows:

$$\begin{aligned}
&\delta^2\sum ln\left(\frac{xi-\Theta}{\lambda+\Theta-xi}\right)^2 + \delta\sum\ln\left(\frac{xi-\Theta}{\lambda+\Theta-xi}\right) - n = 0\\
&\delta^2\sum\ln\left(\frac{xi-\Theta}{\lambda+\Theta-xi}\right)^2 - \frac{\delta^2\sum\ln\left(\frac{xi-\Theta}{\lambda+\Theta-xi}\right)}{n} - \sum\ln\left(\frac{xi-\Theta}{\lambda+\Theta-xi}\right) - n = 0\\
&\delta^2\sum ln\left(\frac{xi-\Theta}{\lambda+\Theta-xi}\right)^2 - \frac{\delta^2}{n}\left[\sum\ln\left(\frac{xi-\Theta}{\lambda+\Theta-xi}\right)\right]^2 - n = 0\\
&\delta = \left(\frac{n}{\sum ln\left(\frac{xi-\Theta}{\lambda+\Theta-xi}\right)^2 - \frac{1}{n}\left[\sum ln(\frac{xi-\Theta}{\lambda+\Theta-xi})\right]^2}\right)^{\frac{1}{2}}
\end{aligned} \tag{5.16}$$

It has been shown that finding numerical solutions to the parameters (θ,λ) from partial derivatives of the weighting function is not an easy task , so we will apply the least squares method to get estimates for the parameters ,(θ,λ) as follows

From the relationship in equation (2.3) we can the following function :

$$f\,(\Theta.\lambda) = \sum_{i=1}^{n}\left(xi-\Theta-\lambda e^{\left(\frac{z-\widehat{\beta}}{\widehat{\delta}}\right)}\right)^2 \tag{5.17}$$

Finding the derivative with respect to the parameter (θ) and equality by zero and simplification we get the following .

$$\sum xi = n\widehat{\Theta} + \lambda\sum_{i=1}^{n}e^{\frac{z_i-\widehat{\beta}}{\delta}} \tag{5.18}$$

Finding the derivative with respect to the parameter (λ) and equality by zero and simplification we get the following

$$\sum \mathrm{xi} e^{\frac{zi-\beta}{\lambda}} = \Theta \sum e^{\frac{zi-\beta}{\lambda}} + \lambda \sum \left(e^{\frac{zi-\beta}{\delta}}\right)^2 \tag{5.19}$$

Note that the variable z_i is a random variable that follows the standard normal distribution , so we apply the Least Squares method to solve the natural equations (5.18), (5.19) above that produces .

$$\widehat{\lambda} = \frac{n \sum xi\ \mathrm{e}^{\left(\frac{zi-\beta}{\delta}\right)} - \sum xi\ \sum e^{\left(\frac{zi-\beta}{\delta}\right)}}{n \sum \left[e^{\left(\frac{zi-\beta}{\delta}\right)}\right]^2 - \left[\sum e^{\frac{zi-\beta}{\delta}}\right]^2} \tag{5.20}$$

$$\widehat{\Theta} = \widehat{x} - \lambda \left(\sum e^{\frac{z-\beta}{\delta}}\right) \tag{5.21}$$

6. Simulation

Simulation was used between method of percentile and the combined method of percentile and the combined method of Maximum Likelihood- Least Squares . Monte Carlo simulation was used to estimate parameters of the Johnson SB distribution because it is a technique that allows us to choose different sizes of samples and to generate data from the Johnson SB distribution the following transformation was used :

$$y = (1 + e^{-\frac{z-\beta}{\delta}})^{-1} \tag{6.1}$$

$$x = \lambda y + \theta \tag{6.2}$$

Where

$$Z \sim N(0,1)$$

Simulation was carried out using Matlab program and was replicated (1000) times to obtain the required homogeneity. The following assumptions were taken .
Four different sizes of samples (10,25,35,75, 100 and 500)
Number of parameters were suggested according to the relationships among them as in the Table 1

Table 1: Suggested parameters

Parameter	Case 1	Case 2	Case 3	Case 4
β	1	0.5	1	0.5
δ	1	0.5	0.5	1
θ	10	10	10	10
λ	10	10	1‘0	10

7. Comparison between Estimation Methods

After the data are estimated using the methods mentioned in the theoretical side of the research , the comparison between the estimates is made using the mean squares criterion, where this standard measures the average squared distance between the real value and the estimated value, and the estimation method is better whenever the scale is less ,and the formula for this scale is as follows:

$$\mathrm{MSE}\left(\widehat{\theta}_i\right) = \frac{1}{R}\sum_{i=1}^{R}\left(\widehat{\theta}_i - \theta\right)^2 \tag{7.1}$$

8. Analysis of the result

The simulation results were presented to reach the accuracy of the previously mentioned estimation methods using the MSE statistical standard , arranged according to the relationship between the parameters in the probability density function .

Table 2: Simulation result for estimating the pdf parameters of the JSBD when the sample size is different

n	Case1	Value	MSE		Best
			PM	ML	
15	β	1	0.218624	0.28402	PM
	δ	1	1.703642	1.84821	PM
	θ	10	0.509346	0.63291	PM
	λ	10	0.862131	0.88252	PM
25	β	1	0.562858	0.6573	PM
	δ	1	1.677574	1.69321	PM
	θ	10	0.494598	0.5021	PM
	λ	10	0.732131	0.76421	PM
35	β	1	0.45321	0.43621	ML
	δ	1	0.94372	0.99431	PM
	θ	10	0.49831	0.4282	ML
	λ	10	0.69321	0.73281	PM
75	β	1	0.29631	0.20321	ML
	δ	1	0.68224	0.49431	ML
	θ	10	0.097521	0.09621	ML
	λ	10	0.38219	0.29937	ML
100	β	1	0.09721	0.06329	ML
	δ	1	0.10216	0.04739	ML
	θ	10	0.07821	0.06821	ML
	λ	10	0.08431	0.07732	ML
500	β	1	0.05321	0.04732	ML
	δ	1	0.09752	0.08942	ML
	θ	10	0.07531	0.06652	ML
	λ	10	0.07648	0.06943	ML

Table 3: Simulation result for estimating the pdf parameters of the JSBD when the sample size is different

n	Case1	Value	MSE		Best
			PM	ML	
15	β	0.5	0.98492	0.99782	PM
	δ	0.5	1.85931	1.99212	PM
	θ	10	0.67438	0.68421	PM
	λ	10	1.8775	1.8837	PM
25	β	0.5	0.74026	0.77201	PM
	δ	0.5	1.36913	1.7321	PM
	θ	10	0.83721	0.89931	PM
	λ	10	0.88321	0.99461	PM
35	β	0.5	0.5532	0.58321	ML
	δ	0.5	1.75829	1.9 431	PM
	θ	10	0.43733	0.5682	ML
	λ	10	0.65421	0.76621	PM
75	β	0.5	0.70049	0.8153	ML
	δ	0.5	0.65327	0.5932	ML
	θ	10	0.07521	0.06211	ML
	λ	10	0.3919	0.2332	ML
100	β	0.5	0.0852	0.07721	ML
	δ	0.5	0.9421	0.8321	ML
	θ	10	0.09431	0.07532	ML
	λ	10	0. 5321	0. 5151	ML
500	β	0.5	0.06221	0.06173	ML
	δ	0.5	0.07832	0.07046	ML
	θ	10	0.09431	0.06328	ML
	λ	10	0.06831	0.04931	ML

Table 4: Simulation result for estimating the pdf parameters of the JSBD when the sample size is different

n	Case1	Value	MSE		Best
			PM	ML	
15	β	0.5	0.98492	0.99782	PM
	δ	0.5	1.85931	1.99212	PM
	θ	10	0.67438	0.68421	PM
	λ	10	1.8775	1.8837	PM
25	β	0.5	0.74026	0.77201	PM
	δ	0.5	1.36913	1.7321	PM
	θ	10	0.83721	0.89931	PM
	λ	10	0.88321	0.99461	PM
35	β	0.5	0.5532	0.58321	ML
	δ	0.5	1.75829	1.9 431	PM
	θ	10	0.43733	0.5682	ML
	λ	10	0.65421	0.76621	PM
75	β	0.5	0.70049	0.8153	ML
	δ	0.5	0.65327	0.5932	ML
	θ	10	0.07521	0.06211	ML
	λ	10	0.3919	0.2332	ML
100	β	0.5	0.0852	0.07721	ML
	δ	0.5	0.9421	0.8321	ML
	θ	10	0.09431	0.07532	ML
	λ	10	0. 5321	0. 5151	ML
500	β	0.5	0.06221	0.06173	ML
	δ	0.5	0.07832	0.07046	ML
	θ	10	0.09431	0.06328	ML
	λ	10	0.06831	0.04931	ML

Table 5: Simulation result for estimating the pdf parameters of the JSBD when the sample size is different

n	Case1	Value	MSE		Best
			PM	ML	
15	β	1	1.88731	1.89631	PM
	δ	0.5	1.67941	1.8462	PM
	θ	10	0.7632	0.9321	PM
	λ	1'0	1.78341	1.9932	PM
25	β	1	0.8641	0.9672	PM
	δ	0.5	1.8632	1.9957	PM
	θ	10	0.9573	0.8347	PM
	λ	1'0	0.65221	0.6994	PM
35	β	1	0.53672	0.5293	ML
	δ	0.5	0.9284	0.9783	PM
	θ	10	0.7634	0.6932	ML
	λ	1'0	0.6875	0.6982	PM
75	β	1	0.48441	0.4212	ML
	δ	0.5	0.6281	0.6164	ML
	θ	10	0.08431	0.0636	ML
	λ	1'0	0.4217	0.3065	ML
100	β	1	0.08332	0.07321	ML
	δ	0.5	0.73381	0.8064	ML
	θ	10	0.05621	0.05073	ML
	λ	1'0	0. 0956	0. 0648	ML
500	β	1	0.07787	0.05252	ML
	δ	0.5	0.06372	0.04942	ML
	θ	10	0.08154	0.06421	ML
	λ	1'0	0.08752	0.04982	ML

Table 6: Simulation result for estimating the pdf parameters of the JSBD when the sample size is different

n	Case1	Value	MSE		Best
			PM	ML	
15	β	0.5	0.877804	0.96210	PM
	δ	1	1.03475	1.572475	PM
	θ	10	0.98061	0.99090	PM
	λ	10	0.963611	0.98271	PM
25	β	0.5	0.725213	0.816096	PM
	δ	1	0.844529	0.881553	PM
	θ	10	0.821667	0.847427	PM
	λ	10	0.93958	0.953368	PM
35	β	0.5	0. 5967	0.405744	ML
	δ	1	0.64739	0.60150	ML
	θ	10	0.68671	0.62267	ML
	λ	10	0. 5925	0.69378	PM
75	β	0.5	0. 25053	0. 21418	ML
	δ	1	0. 3092	0. 18898	ML
	θ	10	0. 4234	0.32565	ML
	λ	10	0. 2248	0.2119	ML
100	β	0.5	0. 09937	0.09892	ML
	δ	1	0.07014	0.02922	ML
	θ	10	0.05206	0.05069	ML
	λ	10	0. 0956	0. 0648	ML
500	β	0.5	0.01258	0.010390	ML
	δ	1	0.01659	0.01337	ML
	θ	10	0.011833	0.00828	ML
	λ	10	0.039494	0.01546	ML

9. Conclusions

1. The simulation results showed the advantage of the PM method over the ML method in small sample sizes (15,25) and the results showed closeness in the estimates at medium sample sizes (35) while the results showed a relative advantage of the ML method over the PM method at large sample sizes (75,500) .
2. The simulation results showed an accuracy in the estimation that is directly proportional to the increase in the sample size, and this is in line with the statistical theory

References

[1] M.R. Flynn, *Fitting human exposure data with the Johnson SB distribution*, J. Exposure Sci. Environ. Epidemio. (16) 2006 56–62.

[2] F. George, *Johnson's System of Distributions and Microarray Data Analysis*, Theses, University of South Florida, 2007.

[3] D. T. Mage, *An explicit solution for SB, parameters using four percentile points*, Technomet. 22(2) (1980) 247–251.

[4] P. Palimitesta and C. Provasi, *Estimating distribution functions in Johnson translation system by the starship procedure with simulated annealing*, Int. J. Stat. 0 (1-2) (2000) 97–110.

[5] A. T. Soyink, *On the parameter estimation of Johnson's system of distribution*, Benin J. Stat. 3(28) (2020) 82–90.