Thailand Statistician
January 2025; 23(1): 217-227
http://statassoc.or.th
Contributed paper

Discriminating Between New Weibull Pareto Distribution and Weibull Distribution

K.N.V.R. Lakshmi*[a] and Boyapati Srinivasa Rao [b]
[a] Andhra Loyola College, Vijayawada, Andhra Pradesh, India.
[b] Department of Mathematics and Humanities, R.V.R & J.C College of Engineering, Chowdavaram, Guntur, Andhra Pradesh, India.
*Corresponding author; e-mail: knvrlakshmi@gmail.com

Received: 30 April 2021
Revised: 8 October 2022
Accepted: 26 November 2022

Abstract

In this paper, distinguision between the new Weibull Pareto distribution and the Weibull distribution, test statistics based on likelihood functions, order statistics, and population quantiles are suggested. The percentiles of the proposed test statistics are tabulated with the aid of simulated sampling distributions of the test statistics due to the non-tractability of their exact sampling distributions. The test statistics power is also tabulated, and a comparison between the powers for a given sample and the level of significance is calculated.

Keywords: New Weibull Pareto distribution (NWPD), Weibull distribution (WD), likelihood function, order statistics, quantiles, power of the test.

1. Introduction

Let X be a random variable from a Pareto distribution with its cumulative distribution function (cdf) for $x \geq \theta$ given by

$$F_1(x;\theta,k) = 1 - \left(\frac{\theta}{x}\right)^k, \tag{1}$$

where $\theta > 0$ and $k > 0$ are the scale and shape parameters respectively. The probability density function (pdf) associated to (1) is

$$f_1(x;\theta,k) = \frac{k\theta^k}{x^{k+1}}. \tag{2}$$

The form of cdf of new Weibull-Pareto distribution (NWPD) is

$$F(x) = \int_0^{\frac{1}{R(x)}} f_2(x)dx, \tag{3}$$

where $R(x)$ is that the survival function of the Pareto distribution and is given by $R(x)=1-F_1(x;\theta,k)$ while $f_2(x)$ is that the pdf of a Weibull distribution and is given by

$$f_2(x)=\alpha\lambda(\lambda x)^{\alpha-1}e^{-(\lambda x)^{\alpha}} \tag{4}$$

where $x>0$, $\alpha>0$ and $\lambda>0$. Using (4) and $R(x)=\left(\frac{\theta}{x}\right)^{k}$ in (3) the cdf of the NWPD is given by

$$F(X)=\int_{0}^{\frac{1}{\left(\frac{\theta}{x}\right)^{k}}}\alpha\lambda(\lambda x)^{\alpha-1}e^{-(\lambda x)^{\alpha}}dx \tag{5}$$

$$F(x)=1-e^{-\lambda^{\alpha}\left(\frac{x}{\theta}\right)^{\alpha k}}. \tag{6}$$

If we let $\delta=\lambda^{\alpha}$ and $\beta=\alpha k$, then the cdf of the new Weibull-Pareto distribution may be written as

$$F(x)=1-e^{-\delta\left(\frac{x}{\theta}\right)^{\beta}}. \tag{7}$$

The pdf is given by

$$g(x)=\frac{\beta\delta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1}e^{-\delta\left(\frac{x}{\theta}\right)^{\beta}}, \tag{8}$$

where $x>0,\ \beta>0,\ \theta>0$ and $\delta>0$. The pdf the r^{th} order statistic $X_{r:n}$, is given by

$$\alpha_{(r)}=\frac{n!}{(r-1)!(n-r)!}\times\left[1-e^{-\delta\left(\frac{x}{\theta}\right)^{\beta}}\right]^{r-1}\times\left[e^{-\delta\left(\frac{x}{\theta}\right)^{\beta}}\right]^{n-r}\times\frac{\beta\delta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1}e^{-\delta\left(\frac{x}{\theta}\right)^{\beta}}. \tag{9}$$

The other distributional properties are thoroughly discussed by Nasiru and Luguterah (2015). The cdf of Weibull distribution is given by

$$F(x)=1-e^{-(\lambda x)^{\alpha}}. \tag{10}$$

Graph of the pdf of NWPD and WD for selected parameter values are plotted graphically.

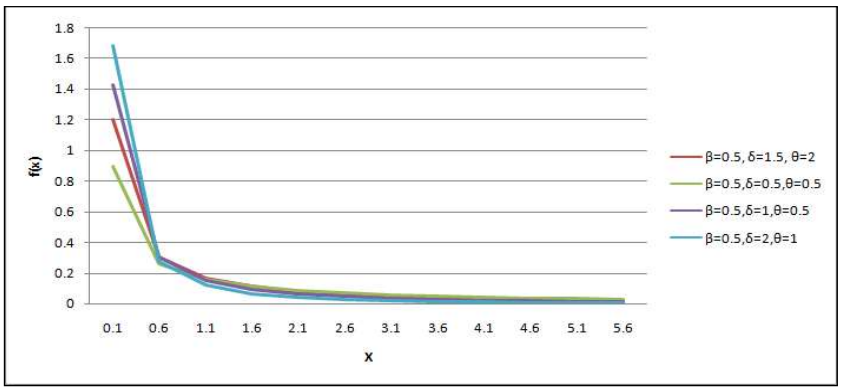


Figure 1. PDFs of NWPD

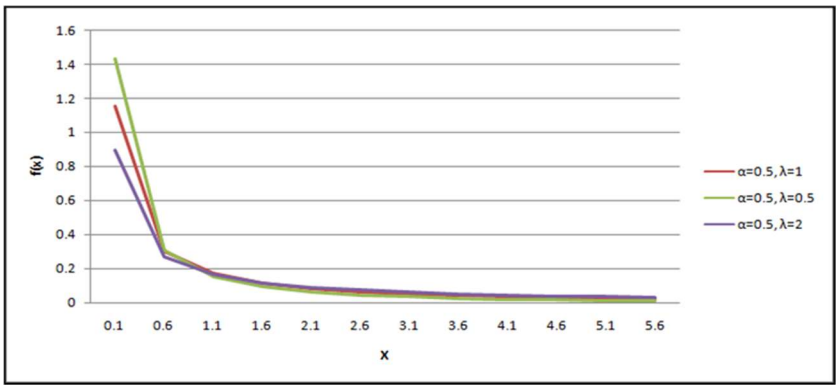


Figure 2. PDFs of WD

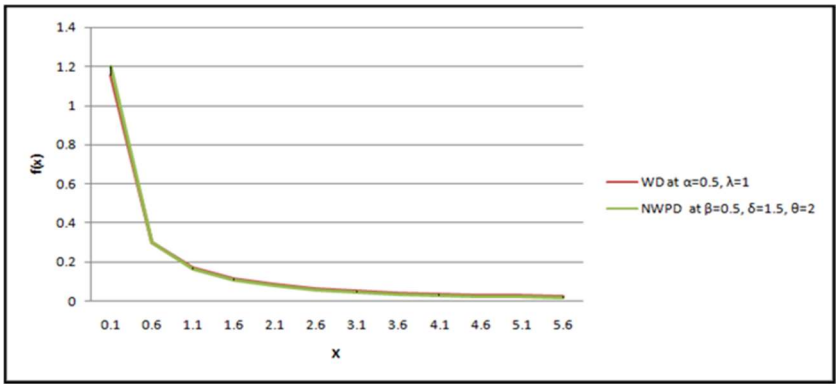


Figure 3. Comparison of NWPD and WD

The significance of these two distributions as existence testing models has been examined by many people. In fact, Weibull distribution inferential properties are simple, efficient, and mathematically tractable, while the new Weibull Pareto distribution is not. We are thus inspired to investigate whether the Weibull distribution is a viable alternative to the new Weibull Pareto distribution, if only to adopt the analytical and efficient inferential characteristics of the Weibull distribution for data following the new Weibull Pareto distribution. To distinguish between the new Weibull Pareto distribution (NWPD) and the Weibull distribution (WD), we proposed three separate test statistics: likelihood ratio criterion, order statistics approach, and quantile criterion method. The standard graphs in Section 1 reveal that the frequency curve of NWPD with parametric values $\beta = 0.5,\ \delta = 1.5$ and $\theta = 2$ shows that there is a similarity with WD for the parametric values $\lambda = 1$ and $\beta = 0.5$. The null and alternative populations were chosen because the graphs of frequency curves for these distributions on the same scale look similar, which motivated the researchers to see if one distribution could be used instead of the other. Hence, we can develop a test criterion to identify whether WD can be an alternative to NWPD.

The research works on discriminating between the probability models include, Gupta and Kundu (2003), Gupta and Kundu (2004), Kundu and Manglick (2004), Gupta et. al. (2005), Kundu et. al. (2005), Kundu and Raqab (2007), Rao and Kantam (2011), Rao et al. (2013), Rao and Kantam

(2013), Bromideh and Valizadeh (2014), Rao and Kantam (2014), Ravi Kumar and Kantam (2016), Bodhisuwan and Sangsanit (2017), Qaffou and Zoglat (2017), Raqab et al. (2017), Rao and Sricharani (2018). The remaining paper is laid out as follows. Sections 2, 3, and 4 describe the proposed test statistics based on likelihood function, order statistics, and population quantiles, as well as their percentiles and power values. In Section 5, a comparison of the three statistics in terms of their power is discussed.

2. Test Statistics Based on Likelihood Ratio (LR) Criterion

2.1. Procedure

Let a random sample $x_1, x_2, ..., x_n$ is drawn from a null population P_0. Let $\hat{P}_0$ denote the value of the likelihood of the sample $x_1, x_2, ..., x_n$ w.r.t. the null population, at the values of the parameters using $x_1, x_2, ..., x_n$. Let $\hat{P}_1$ denote the value of likelihood function w.r.t. to alternative population, using the sample $x_1, x_2, ..., x_n$. The ratio L_1 / L_0 in a way represents the ratio of the likelihood of P_1 to that of P_0 with a sample drawn from P_0. Therefore, P_1 / P_0 is the ratio of a smallest probability to a larger probability and hence is expected to be small. The null hypothesis H_0: The samples are drawn from P_0.

Can be tested using the percentiles of the likelihood ratio L_1 / L_0. Therefore $T_1 = L_1 / L_0$ can be taken as the test statistic to test the above null hypothesis. Since the distribution of L_1 / L_0 is not analytical, we compute the percentiles of empirical sampling distributions of $T_1 = L_1 / L_0$ with the help of 10,000 simulation runs of sample size $n = 2,3,...,10$.

2.2. Power of the test

The likelihood ratio (LR) type statistic suggested by us would be meaningful when it is able to distinguish between the null and alternative populations. Conveniently fix the levels of significance as 0.1%, 2.5% and 5%. The critical value of L_1 / L_0 corresponding to the level of significance is identified from the percentile table, say C_0. 10,000 random samples of size each $n = 2,3,...,10$ from the alternative population are generated. The individual sample values are substituted in L_1. Using the same sample the value of L_0 is also computed in order to get 10,000 values of L_1 / L_0 for a given sample size. The proportion of the values of L_1 / L_0 that exceeded the critical value out of 10,000 is computed and considered as the power of the test statistic at the chosen level of significance. A large value of the power shows that the test statistic is able to distinguish between the null and alternative populations.

2.3. Application of LR criterion for discriminating NWPD and WD models

Let us consider NWPD as the null population, say (P_0) and WD is considered as the alternative population say (P_1). A random sample of size n is generated from NWPD (P_0). The generated sample is thus substituted in the alternative (P_1). The likelihood functions L_0 and L_1 of the populations P_0 and P_1, respectively are calculated. Thus in all, we get as many values of L_1 / L_0 as the number of repetitions are done, generally called the simulation runs. We repeat this procedure 10,000 times for accuracy and precision. Among these 10,000 values various specified cut off points

(percentiles) would form the critical values useful for testing method. The estimates yielded from the chosen method of estimation are given in Table 1.

Table 1. Percentiles of sampling distributions of $T_1 = L_1 / L_0$.

n	0.999	0.99	0.975	0.95	0.05	0.025	0.01	0.001
2	1.3083	1.1532	1.0955	1.0520	0.9053	0.9035	0.9020	0.9003
3	1.3401	1.1661	1.0991	1.0505	0.8645	0.8620	0.8597	0.8565
4	1.3033	1.1617	1.0874	1.0325	0.8268	0.8239	0.8205	0.8151
5	1.3347	1.1544	1.0775	1.0262	0.7907	0.7869	0.7835	0.7769
6	1.3708	1.1620	1.0789	1.0176	0.7569	0.7526	0.7483	0.7407
7	1.3170	1.1344	1.0540	0.9949	0.7245	0.7203	0.7162	0.7056
8	1.3304	1.1303	1.0399	0.9812	0.6942	0.6890	0.6845	0.6746
9	1.3370	1.1165	1.0282	0.9633	0.6665	0.6609	0.6550	0.6459
10	1.3018	1.0964	1.0032	0.9374	0.6388	0.6324	0.6263	0.6167

The power of the test statistics T_1 is also tabulated for three levels of significance (0.1%, 2.5% and 5%) at sample sizes $n = 2,3,...,10$ by simulating samples from P_1 and using the values of T_1. The count of T_1 value that fall beyond the values of Table 1 shall speak of the power of T_1. These are given in Table 2.

Table 2 Power of T_1

n	0.999	0.975	0.95
2	0.9866	0.9694	0.9417
3	0.9865	0.9673	0.9391
4	0.9865	0.9661	0.9379
5	0.9860	0.9663	0.9374
6	0.9862	0.9666	0.9358
7	0.9852	0.9660	0.9344
8	0.9849	0.9638	0.9337
9	0.9848	0.9650	0.9346
10	0.9847	0.9639	0.9333

The values in the table show that even with a small sample size of 2, the power is still greater than 90%. As a result, it can be assumed that the test statistic T_1 proposed in this section has high power to differentiate between the null and alternative populations.

3. Test Statistics Based on Order Statistics

3.1. Procedure

Consider the two continuous probability distributions P_0 and P_1 as null and alternative populations respectively. Let $x_{(1)}, x_{(2)}, ..., x_{(n)}$ denote an ordered sample of size n from P_0. Here we test the null hypothesis, H_0 : The sample has come from the population P_0, against the alternative hypothesis, H_1 : The sample has come from the population P_1. The test statistic which is the correlation-type goodness-of-fit test using the order statistics is given by

$$T_2 = \frac{\sum_{i=1}^{n} x_{(i)} \alpha_i}{\sqrt{\sum_{i=1}^{n} x_{(i)}^2 \sum_{i=1}^{n} \alpha_i^2}}, \tag{11}$$

where $x_{(i)}$ is the i^{th} order sample observation and α_i is expected value of i^{th} order statistic in a sample of size n drawn from P_0. If $x_{(1)}, x_{(2)}, ..., x_{(n)}$ is truly a sample from the null population, the formula T_2 can be used as a test statistic to distinguish between the null and alternative populations using critical values. As a result, in order to use the test statistic T_2, the sampling distribution of T_2 and its percentiles are needed. Since the sampling distribution of T_2 requires non-linear function of order statistics, it is not tractable analytically in all populations, necessitating the use of empirical sampling distributions and empirical percentiles of T_2 obtained by simulation.

3.2. Power of the test

The power of the test refers to the probability of not rejecting a wrong null hypothesis, which is determined by the alternative distribution assumed. The test statistic T_2 is determined by randomly selecting a sample of a given size from each of the alternative populations, with α_i in the T_2 formula representing the null population. The critical values are compared to the corresponding calculated value of T_2. This process is repeated 10,000 times, and the number of null hypothesis rejections out of 10,000 is used to determine the test's power.

3.3. Application of OS approach for discriminating NWPD and WD models

The statistics T_2 defined is based on moments of order statistics in samples from null population which may not be available to probability models; the same is the case with new Weibull Pareto distribution. Here we have calculated the moments of order statistics α_i for NWPD and the procedure is as follows:

Let $x_{(1)}, x_{(2)}, ..., x_{(n)}$ be a complete ordered random sample of size *n* from the new Weibull Pareto distribution. Here we test the null hypothesis H_0 : The sample has come from the new Weibull Pareto distribution, against the alternative hypothesis, H_1 : The sample has come from the Weibull distribution. Proceeding in the same lines above, our proposed test statistic,

$$T_2 = \frac{\sum_{i=1}^{n} x_{(i)}\alpha_i}{\sqrt{\sum_{i=1}^{n} x_{(i)}^2 \sum_{i=1}^{n} \alpha_i^2}}.$$

Here we consider the pdf of r^{th} order statistics $X_{r:n}$, $1 \le r \le n$ as

$$\alpha_{(r)} = \frac{n!}{(r-1)!(n-r)!}\left[F(x)\right]^{r-1}\left[1-F(x)\right]^{n-r} f(x), -\infty < x < \infty, \tag{12}$$

where $f(x)$ and $F(x)$ are pdf and cdf of the NWPD and is given by

$$\alpha_{(r)} = \frac{n!}{(r-1)!(n-r)!} \times \left[1-e^{-\delta\left(\frac{x}{\theta}\right)^{\beta}}\right]^{r-1} \times \left[e^{-\delta\left(\frac{x}{\theta}\right)^{\beta}}\right]^{n-r} \times \frac{\beta\delta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1} e^{-\delta\left(\frac{x}{\theta}\right)^{\beta}}. \tag{13}$$

Now, based on 10,000 simulations, the statistic T_2 is simulated using the Monte Carlo method. T_2 percentiles for sample sizes $n = 2,3,...,10$ and different levels of significance are shown in Table 3. The T_2 percentiles will be used to evaluate the null hypothesis that a given sample comes from a new Weibull Pareto distribution. Therefore, a large value of T_2 implies a strong linear relation between $x_{(i)}$ and $\alpha_{(i)}$, which means that the sample is taken from new Weibull Pareto distribution.

Table 3. Percentiles of T_2 based on order statistics

n	0.999	0.99	0.975	0.95	0.05	0.025	0.01	0.001
2	1.0000	0.9999	0.9995	0.9982	0.0661	0.0310	0.0113	0.0020
3	0.9997	0.9975	0.9946	0.9889	0.1051	0.0534	0.0233	0.0039
4	0.9996	0.9987	0.9980	0.9972	0.2137	0.1102	0.0465	0.0081
5	1.0000	0.9999	0.9998	0.9997	0.3207	0.1637	0.0767	0.0125
6	1.0000	0.9999	0.9997	0.9995	0.3762	0.2046	0.0993	0.0166
7	1.0000	0.9999	0.9998	0.9996	0.4453	0.2536	0.1361	0.0388
8	1.0000	0.9999	0.9998	0.9995	0.4450	0.2791	0.1541	0.0432
9	1.0000	0.9999	0.9998	0.9994	0.4825	0.2928	0.1528	0.0423
10	1.0000	0.9999	0.9997	0.9992	0.5217	0.3138	0.1779	0.0520

The power of the test statistics T_2 is tabulated for two levels of significance (0.1%, 2.5% and 5%) at sample sizes $n = 2,3,...,10$ by simulating samples from P_1 and using the values of T_2. The calculations of the power were performed for the above mentioned alternative with various sample sizes and various levels of significance for complete samples presented in Table 4.

Table 4 Power of T_2

n	0.999	0.975	0.95
2	1.0000	0.9744	0.9510
3	0.9988	0.9764	0.9961
4	0.9990	0.9809	0.9620
5	0.9999	0.9548	0.9151
6	0.9996	0.9782	0.9574
7	0.9998	0.9833	0.9591
8	0.9997	0.9756	0.9507
9	0.9999	0.9778	0.9530
10	0.9995	0.9772	0.9523

4. Test Statistic Based on Quantiles

Let P_0 and P_1 be two continuous probability distributions which can be regarded as null and alternative populations respectively. Let $x_{(1)}, x_{(2)}, ..., x_{(n)}$ be an ordered random sample of size n (ordering is mandatory as the method indicates) from the null population P_0. Here, we test the null hypothesis, H_0: The sample has come from the population P_0, against the alternative hypothesis, H_1: The sample has come from the population P_1. The test statistic T_3 defined is calculated based on the population quantiles which are available analytically for all the populations with invertible cumulative distribution functions.

Let $p_i = i/(n+1)$ be the i^{th} population quantile from H_0 and let $v_i = F^{-1}(p)$ then the proposed test statistic is correlation-type goodness-of-fit test using percentiles is given by

$$T_3 = \frac{\sum_{i=1}^{n} x_{(i)} v_i}{\sqrt{\sum_{i=1}^{n} x_{(i)}^2 \sum_{i=1}^{n} v_i^2}}. \tag{14}$$

If $x_{(1)}, x_{(2)}, ..., x_{(n)}$ is truly a sample from the null population, the formula T_3 would serve as a test statistic to discriminate a null population and the corresponding alternative population with the help of critical values. This statistic represents the correlation between x_i and v_i, $i = 1, 2, ..., n$ where v_i is the i^{th} quantile of the null population. The statistic T_3 is simulated through Monte Carlo method based on 10,000 simulations.

4.1. Power of the test

The test statistic T_3 determined by randomly selecting a sample of a given size from the alternative population, with v_i in the formula for $x_{(i)}$ representing the null population. The critical values are compared to the corresponding calculated value of T_3. This process is repeated 10,000

times, and the number of null hypothesis rejections out of 10,000 is used to determine the test's power.

4.2. Application of quantiles method for discriminating NWPD and WD

Let $x_{(1)}, x_{(2)}, ..., x_{(n)}$ be a complete ordered random sample of size n from NWPD. Here, we test the null hypothesis, H_0: The sample has come from NWPD, against the alternative hypothesis, H_1: The sample has come from WD. Calculate the population quantiles p_i for NWPD. Let $v_i = F^{-1}(p)$. Using these v_i's the proposed test statistic T_3 is calculated. Table 5 shows the percentiles of the empirical sampling distribution of T_3 for $n = 2,3,...,10$ based on 10,000 Monte Carlo simulation cycles. The percentiles calculated in this way serve as critical values for testing the null hypothesis that a given sample is from NWPD.

Table 5. Percentiles of T_3 based on quantiles

n	0.999	0.99	0.975	0.95	0.05	0.025	0.01	0.001
2	1.0000	1.0000	1.0000	0.9998	0.8260	0.8111	0.8021	0.7966
3	1.0000	0.9996	0.9989	0.9976	0.8290	0.8004	0.7755	0.7416
4	0.9997	0.9986	0.9975	0.9960	0.8315	0.8084	0.7833	0.7382
5	0.9994	0.9979	0.9961	0.9939	0.8359	0.8098	0.7820	0.7341
6	0.9990	0.9972	0.9948	0.9921	0.8481	0.8184	0.7899	0.7397
7	0.9988	0.9965	0.9944	0.9915	0.8470	0.8171	0.7890	0.7345
8	0.9982	0.9958	0.9933	0.9903	0.8620	0.8319	0.8015	0.7401
9	0.9979	0.9953	0.9929	0.9897	0.8648	0.8343	0.8014	0.7463
10	0.9980	0.9947	0.9924	0.9891	0.8570	0.8424	0.8092	0.7476

The power of the test statistic with WD as alternative in succession is evaluated as follows. 10,000 random samples of size $n = 2,3,...,10$ from WD are generated and the values T_3 are calculated at each sample. While calculating T_3 for each sample drawn from WD and the v_{ijs} in the formula for T_3 are retained as those of the null population. The proportion values of T_3 that fall above the percentiles out of 10,000 samples are noted down, which represent the power of the statistic T_3 corresponding the three levels of significance (0.1%, 2.5% and 5%) at sample sizes $n = 2,3,...,10$. T_3's power is measured by the number of T_3 values that fall outside of table 5's range. Table 6 lists them all.

Table 6. Power of T_3

n	0.999	0.975	0.95
2	1.0000	1.0000	0.9483
3	0.9895	0.9754	0.9497
4	0.9904	0.9743	0.9505
5	0.9897	0.9750	0.9499
6	0.9898	0.9750	0.9499
7	0.9900	0.9749	0.9497
8	0.9902	0.9748	0.9501
9	0.9899	0.9751	0.9507
10	0.9901	0.9754	0.9754

5. Comparative Study

The test statistics T_1, T_2 and T_3 which have powers in Tables 2, 4, and 6, respectively, indicating that all three statistics discriminate between new Weibull Pareto and Weibull distributions very well. This holds true regardless of sample size. However, the statistics based on the formulae T_1, T_2 require the use of a reliable estimator or specially tabulated expected values of order statistics of samples drawn from new Weibull Pareto distribution. The statistic T_3, on the other hand, is based on the population quantiles of the new Weibull Pareto distribution corresponding to a given sample of size n, which can be easily determined by inverting the analytical expression of the new Weibull Pareto distribution's population distribution function. This inversion does not actually require any special tables, making it much easier. We can see that quantile-based coverage probabilities are more effective than likelihood ratios and order statistics-based coverage probabilities.

References

Balakrishnan N, Malik HJ. Moments of order statistics from truncated log-logistic distribution. J Stat Plan Inference. 1987; 17: 251–267.

Bodhisuwan W, Sangsanit Y. Discriminating between the Weibull and Power Lindley Distributions Utilizing Fisher Information. Thail Stat. 2017; 15: 55-68.

Bromideh AA, Valizadeh R. Discrimination between Gamma and Log-Normal Distributions by Ratio of Minimized Kullback-Leibler Divergence. Pak J Stat Oper Res. 2014; 9: 443-453.

Gupta DR, Kundu D. Discriminating between Weibull and generalized exponential distributions. Comput Stat Data Anal. 2003; 43: 179-196.

Gupta DR, Kundu D. Discriminating between gamma and generalized exponential distributions. J Stat Comput Simul. 2004; 74(2): 107–121.

Gupta DR, Kundu D, Manglic A. Discriminating between log normal and generalized exponential distributions. J Stat Plan Inference. 2005; 127: 213-227.

Kundu D, Manglick A. Discriminating between Weibull and log-normal distributions. Nav Res Logist. 2004; 51(6): 893-905.

Kundu D, Raqab MZ. Discriminating between the generalized Rayleigh and log-normal distribution. J Theor Appl Stat. 2007; 41(6): 505-515.

Nasiru S, Luguterah A. The new Weibull-Pareto distribution. Pak J Stat Oper Res. 2015; 11(1): 103-114.

Qaffou A, Zoglat A. Discriminating between normal and Gumbel distribution. Revstat Stat J. 2017; 15(4): 523-536.

Rao BS, Kantam RRL. Significance test for half logistic distribution. Stoch Qual Control. 2011; 26: 53-62.

Rao BS, Kumar SCh, Kantam RRL, Rosaiah K. Discriminating between half normal and half logistic distribution. J Comput Inf Syst 2013; 7(4): 59- 63.

Rao BS, Kantam RRL. Discriminating between generalized exponential distribution and some life test models based on population quantiles. J Mod Appl Stat Methods. 2013; 12(2): 336-343.

Rao BS, Kantam RRL. Discriminating between log-logistic and Rayleigh distributions. Pak J Stat Oper. 2014; 10(1): 1-7.

Rao BS, Sricharani P. Discriminating between Dagum and Burr-III distributions. Int J Sci Res Math Stat Sci. 2018; 5(5): 104-107.

Raqab MZ, Al-Jarallah RA, Al-Mutairi DK. Closeness of Lindley distribution to Weibull and gamma distributions. Commun Stat Appl Methods. 2017; 24: 129-142.

Ravi KMS, Kantam RRL. Discrimination between Gamma and Rayleigh models. Int J Math Eng Manag Sci. 2016; 1(2): 89-92.