UtilitasMathematica

ISSN 0315-3681 Volume 122, 2025

Modified statistics for fitting a new bounded Distribution with Applications

[1]Rochdi Nouadri and [2] Nacira Seddik-Ameur

[1,2]Laboratoire of probability and Statistics LaPS

Badji Mokhtar University- Annaba
E-mail : rochdi.nouadri@univ-annaba.dz

naciraseddik@yahoo.fr

Abstract:

Bounded probability functions are very important in life test experiments relating to data on fractions, percentages, survival times from a deadly diseas and other fields, so researchers tried to generate new models with finite support capable to describe such phenomenon. In this work, we propose a new distribution so-called the Unit Generalized Exponential distribution (UGE) defined on (0,1). Unless the increasing and bathtub shaped hazard function, its probability density function can be unimodal, decreasing and left-skewed which enable it to be used in several fields of application. Using different approaches, we propose new powerful test statistics, based on any efficient estimator, capable to fit datasets to this model without regarding any alternative as used in classical selection models. Having the lowest variance and the highest rate of convergence to the limit, they also recover the information lost while grouping data. Thousands of simulated samples of varying sizes were utilized to demonstrate the practicability of the proposed tests. Two real-world data illustrated the flexibility of this model particularly for skewed data, compared to existing distributions, including the Kumaraswamy, Unit Weibull, Unit Generalized Inverse Weibull, Unit Burr III, and Unit Gompertz distributions.

Classification AMS: 62F12- 62G10- 62N05

Keys words: Estimation methods, Modified chi-squared tests, Dzhaparidze-Nikulin statistic, McCulloch statistic, Nikulin-Rao-Robson statistic, Singh statistic.

1 Introduction

Checking the validity of the model used in statistical analysis is the first problem for practitionars, particularly in some sensitive areas such the electronic, aerospace, medical and industrial studies, because the model chosen have an important impact on the obtained results. Thus statisticians have always tried to product criteria for fitting models. The most common goodness-of-fit tests are those based on the empirical distribution function (EDF) and the Pearson statistic X^2; but when the parameters are unknown, most of them cannot be applicable. Over that, the existing goodness-of-fit are not sufficiently powerful. This fact has led many researchers to propose free distribution modifications of Pearson statistic. Different procedures were given where the power of the tests depends on the estimation method, or on the use of grouped or non-grouped data. Among the proposed statistics, Dzhaparidze and Nikulin (1974) gave a new form U^2 of the statistic X^2 for the exponential distribution. The advantages of this statistic is the use of any $\sqrt{n}$-consistent estimates on grouped or non-grouped data, and Its insensitivity to the alternative hypothesis (Voinov et al., 2009). On the other hand, Nikulin (1973), and Rao and Robson (1974) introduced separately a modified chi-square statistic Y^2 based on maximum likelihood estimation and equiprobable grouping classes for the Gaussian and the exponential distribution respectively. After some years, McCulloch (1985) derived from Y^2, a very interesting powerful statistic S^2, called the McCu-test, with the lowest variance and the highest rate of convergence to the limit. Later, Singh (1987) proposed a new statistic, noted $Q^2(\theta)$, which can also use any $\sqrt{n}$-consistent estimates. However and despite of their effectiveness, these tests have not sufficiently investigated because they are not given in their explicit form and must be adapted for the hypothezised model. Recently the statistic Y^2 was developed for some distributions, such as a competing risk model (Chouia and Seddik-Ameur, 2014), the generalized Rayleigh distribution (Tilbi and Seddik-Ameur, 2017), the Toppe-Leone Extended exponential distribution (Aidi et al., 2022), the New Weibull-G family (Meribout et al., 2023), but for the other statistics, the applications are very limited. The statistics U^2 and S^2 were developed for the Gaussian and the power generalized Weibull distributions (Bagdonavicius et al., 2006) but Q^2 has not been considered yet. Based on these approaches, we adapte the different statistics for the unit generalized exponential distribution (UGE) introduced in this work. The goodness-of-fit tests proposed recover the information lost while grouping data and don't require competitor models for the comparison as used in classical model selection methods.

This paper is organized as follows. The unit generalized exponential distribution is introduced in section 2. General moments and Fisher information matrix are calculated in section 3. Based on the techniques cited above, different modified statistics were constructed for fitting this new model in section 4. The section 5 is devoted to the performance, the power and the practicability of the proposed statistics by an intensive simulation study and the usefulness of the proposed distribution by two applications.

2 Unit generalized exponential distribution

The generalized exponential distribution introduced by Gupta and Kundu (1999) is defined by:

$$f(t) = \alpha\lambda(1 - e^{-\lambda t})^{\alpha-1}e^{-\lambda t}$$

where α and λ are respectively the shape and scale parameters.

By using the transformation $t = \frac{x}{1-x}$, we derive the corresponding bounded model on $(0,1)$ called the unit generalized exponential distribution (UGE). Therefore, the probability density function (pdf) of this new model is obtained by:

$$f(x) = \alpha\lambda(1 - x)^{-2}exp(-\lambda(\frac{x}{1-x}))[1 - exp(-\lambda(\frac{x}{1-x}))]^{\alpha-1} \qquad \alpha > 0 \text{and} \lambda > 0.$$

The corresponding cumulative distribution $F(x)$ and the hazard rate $h(x)$ functions have the forms:

$$F(x) = [1 - exp(-\lambda(\frac{x}{1-x}))]^{\alpha} \quad \alpha > 0, \lambda > 0 \text{ and } x \in [0,1].$$

$$h(x) = \frac{\alpha\lambda(1-x)^{-2} exp(-\lambda(\frac{x}{1-x}))[1-exp(-\lambda(\frac{x}{1-x}))]^{\alpha-1}}{1-[1-exp(-\lambda(\frac{x}{1-x}))]^{\alpha}}$$

The quantile function is obtained by:

$$Q(u) = \frac{-log(1-u^{1/\alpha})}{\lambda - log(1-u^{1/\alpha})}$$

where u is uniformely distributed on the intervall $(0,1)$.

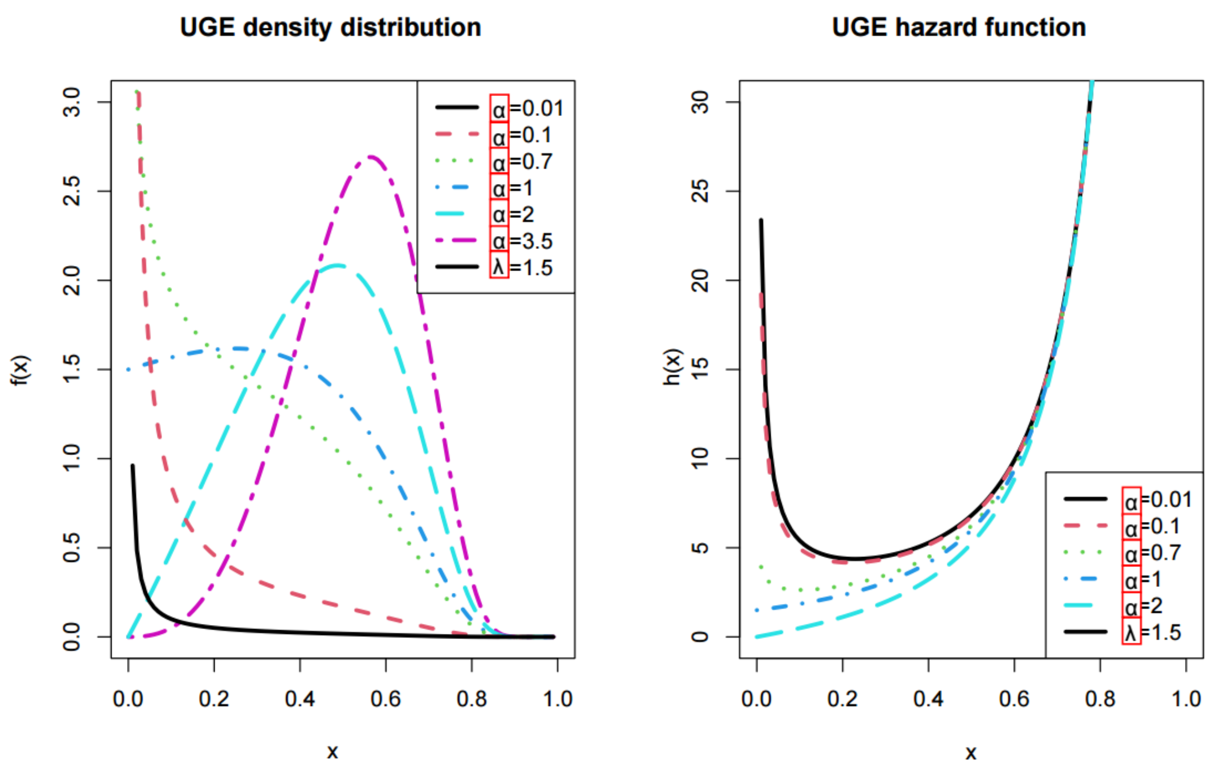

Fig. 1: **pdf and hazard rate plots of the UGE distribution**

Depending on the parameter values, the UGE pdf and the hazard rate functions displayed in Figure 1 show the flexibility of this new model.

For small values of α, the pdf and the rate failure functions shape are decreasing like which describe early failures. Noted that for larger values of α, the pdf gradually becomes unimodal with hazard function bathtub and increasing shaped which characterize many real-world systems where failures occur frequently in early life (due to defects), then stabilize, and later increase again due to aging or wear-out.

3 Moments and information matrix

3.1 Moments:

The skeweness and the kurtosis of any propobility distribution give an idea on the comportement of the phenomenon, for example on the presence of outliers or on the asymetry of the distribution. We provide in this section the moments of the UGE distribution necesssary for the computation of these coefficients.

The general moment of order s is given by:

$$E[X^s] = \int_0^\infty x^s \alpha\lambda(1-x)^{-2} exp(-\lambda(\frac{x}{1-x}))[1-exp(-\lambda(\frac{x}{1-x}))]^{\alpha-1} dx$$

After some simplifications, we obtain:

$$E[X^s] = 1 - \frac{s(-1)^\alpha}{\alpha^{\alpha+1}\lambda k^\alpha}[(s+1)\Gamma(\alpha+1)\beta(2,s) - (s-1)\Gamma(\alpha+1)\beta(2,s-1)]$$

Then, we deduce the coefficients of skewness sk and kurtosis kur:

$$sk = \frac{1-\frac{3(-1)^\alpha}{\alpha^{\alpha+1}\lambda k^\alpha}[(4)\Gamma(\alpha+1)\beta(2,3)-(2)\Gamma(\alpha+1)\beta(2,2)]}{Var[x]^{\frac{3}{2}}}$$

$$kur = \frac{1-\frac{4(-1)^\alpha}{\alpha^{\alpha+1}\lambda k^\alpha}[(5)\Gamma(\alpha+1)\beta(2,4)-(3)\Gamma(\alpha+1)\beta(2,3)]}{(E[x^2]-E[x]^2)^2}$$

3.2 Fisher information matrix

We consider a sample of n independant and identically distributed variables X_i from the UGE distribution with α and λ parameters, the log-likelihood function is:

$$\ln l(x;\alpha,\lambda) = n(\ln(\lambda)+\ln(\alpha)) - 2\sum_{i=1}^n \ln(1-x_i) +$$

$$(\alpha-1)\sum_{i=1}^n \ln[1-exp(-\lambda(\frac{x_i}{1-x_i}))] - \lambda\sum_{i=1}^n (\frac{x_i}{1-x_i})$$

The maximum likelihood estimators $\hat{\alpha}_{MLE}$, $\hat{\lambda}_{MLE}$ of the unknown parameters are obtained by cancelling the first derivative functions of the $\ln l(x;\alpha,\lambda)$ with respect to the unknown parameters and the second derivatives give the elements of the estimated Fisher information matrix I_n as follow:

$$I_{11} = (\frac{-n}{\alpha^2})$$

$$I_{22} = (\frac{-n}{\lambda^2}) - \sum_{i=1}^{n} \left((\frac{x_i}{1-x_i}) \frac{exp(-\lambda(\frac{x_i}{1-x_i}))}{(1-exp(-\lambda(\frac{x_i}{1-x_i})))} \right)$$

$$I_{12} = \sum_{i=1}^{n} \left((\frac{x_i}{1-x_i}) \frac{exp(-\lambda(\frac{x_i}{1-x_i}))}{(1-exp(-\lambda(\frac{x_i}{1-x_i})))} - 1 \right)$$

$$I_{21} = I_{12} = \sum_{i=1}^{n} \left((\frac{x_i}{1-x_i}) \frac{exp(-\lambda(\frac{x_i}{1-x_i}))}{(1-exp(-\lambda(\frac{x_i}{1-x_i})))} - 1 \right)$$

4 Modified statistics for fitting the UGE distribution

Let us consider an $n-$sample $(X_1, X_2, \ldots, X_n)$, for testing the null hypothesis H_0, that this sample belongs to the UGE distribution with the unknown parameter vector $\theta = (\alpha, \lambda)^T$, we propose the construction of some powerful test statistics using efficient estimators of the unknown parameters. In addition of recovering all the information lost while grouping data, these tests don't require the use of alternatives as in classical model selection methods.

4.1 Dzhaparidze-Nikulin statistic

The first goodness-of-fit test considered in this work is that introduced by Dzhaparidze and Nikulin (1974), named Dzhaparidze-Nikulin statistic. This one doesn't depend on a specific method of estimation as the maximum likelihood estimators which may not exist sometimes. Despite its usefulness, this statistic has not been developed in the statistical literature. Defined by:

$$U^2(\theta) = Z_n^T(\theta)[I - B(\theta)(B^T(\theta)B(\theta))^{-1}B^T(\theta)]Z_n(\theta)$$

This statistic follows a chi-square distribution with $r - 1 - s$ degrees of freedom, where r represents the number of the grouped classes chosen, and s the number of the estimated parameters.

First, we calculate any $\sqrt{n}-$consistent estimator $\tilde{\theta}$ of the unknown parameter. Then, we grouped the observations $(X_1, X_2, \ldots, X_n)$ into r non-intersecting cells $\Delta_j = [a_{j-1}; a_j[$, with $p_j(\theta)$:

$$p_j(\theta) = \int_{\Delta_j} dF(x;\theta), \quad j = 1,\ldots r.$$

If $v = (v_1, v_2, \ldots, v_r)^T$ represents the vector of the frequencies of these classes Δ_j, so:

$$p_j(\theta) = F(a_j) - F(a_{j-1})$$
$$p_j(\theta) = [1 - exp(-\lambda(\frac{a_j}{1-a_j}))]^\alpha - [1 - exp(-\lambda(\frac{a_{j-1}}{1-a_{j-1}}))]^\alpha$$

Then, we can calculate the vector $Z(\theta)$:

$$Z(\theta) = (\frac{v_1 - np_1(\theta)}{\sqrt{np_1(\theta)}}, \ldots, \frac{v_r - np_r(\theta)}{\sqrt{np_r(\theta)}})^T$$

The elements of the $(r \times 2)$ matrix $B(\theta) = (b_{jk})_{r\times 2}$, necessary for the calculation of the statistic U^2, are defined by:

$$b_{j1} = \frac{1}{\sqrt{p_j(\theta)}} \int_{\Delta_j} \frac{\partial f(x,\theta)}{\partial \alpha}, \quad j = 1,\ldots r.$$

$$b_{j2} = \frac{1}{\sqrt{p_j(\theta)}} \int_{\Delta_j} \frac{\partial f(x,\theta)}{\partial \lambda}, \quad j = 1,\ldots r.$$

For the $UGE(\alpha, \lambda)$ distribution, we obtain:

$$b_{j1} = \frac{1}{\sqrt{p_j(\theta)}} [[1 - exp(-\lambda(\frac{a_j}{1-a_j}))]^\alpha log([1 - exp(-\lambda(\frac{a_j}{1-a_j}))])$$
$$-[1 - exp(-\lambda(\frac{a_{j-1}}{1-a_{j-1}}))]^\alpha log([1 - exp(-\lambda(\frac{a_{j-1}}{1-a_{j-1}}))])]$$

$$b_{j2} = \frac{1}{\sqrt{p_j(\theta)}} [\frac{\alpha a_j}{1-a_j} exp(-\lambda(\frac{a_j}{1-a_j})[1 - exp(-\lambda(\frac{a_j}{1-a_j}))]^\alpha$$
$$-\frac{\alpha a_{j-1}}{1-a_{j-1}} exp(-\lambda(\frac{a_{j-1}}{1-a_{j-1}})[1 - exp(-\lambda(\frac{a_{j-1}}{1-a_{j-1}}))]^\alpha]$$

and I is the unit matrix.

By replacing the unknown parameter vector $\theta = (\alpha, \lambda)^T$ by its estimator $\tilde{\theta}$, we can deduce the value of the criteria $U^2(\tilde{\theta})$ for the UGE distribution.

In this work, an extensive simulation study showed that the maximum likelihhod (MLE), the Anderson-Darling (AD), the Cramer-Von-Mises (CVM) and the maximum product spacing (MPS) estimators are the most efficient ones for our model, the results are not given here but can be consulted. The null hypothesis H_0 is rejected for $U^2 \succ \chi^2$ with $r - 2 - 1 = r - 3$ degrees of freedom.

4.2 Nikulin-Rao-Robson statistic

Among the modified chi-square test statistic used till now is that introduced separatly by Nikulin (1973) and Rao and Robson (1974), named the NRR statistic and noted Y^2. Based on maximum likelihood estimators on non-grouped data, this test recovers all the information given by the sample. Many researchers were interested by this method which modifieded the classical chi-square test by givening a correction term when the distribution is not specified and the parameters must be estimated. This statistic Y^2 follows a chi-square distribution with $r-1$ degrees of freedom. In this work, we use the following computational form deduced by Van Der Vaart (1998):

$$Y^2(\hat{\theta}) = X^2(\hat{\theta}) + \frac{1}{n} L^T(\hat{\theta})\left(I_n(\hat{\theta}_n) - J(\hat{\theta}_n)\right)^{-1} L(\hat{\theta}_n)$$

where $\hat{\theta}$ is the maximum likelihood estimator and $X^2(\hat{\theta})$ is the Pearson statistic.
As this statistic is based on equiprobable grouping data Δ_j', so we must have:

$$p_j(\theta) = \int_{\Delta_j'} dF(x;\theta) = \frac{1}{r} \quad \text{forany} \quad j = 1, \dots r.$$

therefore, the limit intervalls a_j' of the r grouped equiprobable cells $\Delta_j' = [a_{j-1}', a_j'[$, are obtained by

$$a_j' = F^{-1}(j/r)$$

which gives for the UGE distribution:

$$a_j' = F^{-1}(j/r) = \frac{-log(1-u^{1/\alpha})}{\lambda - log(1-u^{1/\alpha})}$$

where u is uniformely distributed on $(0;1)$.
The components of the estimated Fisher information on grouped data $J(\hat{\theta})$, calculated on the equiprobable cells Δ_j' are defined by:

$$J_{11} = \sum_j \frac{1}{p_j}\left(\frac{\partial p_j(\theta)}{\partial \alpha}\right)^2; \; J_{22} = \sum_j \frac{1}{p_j}\left(\frac{\partial p_j(\theta)}{\partial \lambda}\right)^2; \; J_{12} = J_{21} = \sum_j \frac{1}{p_j}\left(\frac{\partial p_j(\theta)}{\partial \alpha}\right)\left(\frac{\partial p_j(\theta)}{\partial \lambda}\right)$$

The vector L is:

$$L = (L_1, L_2)^T \quad L_1 = \sum_{j=1}^{r} \frac{v_j}{p_j(\theta)} \frac{dp_j(\theta)}{d\alpha}; \qquad L_2 = \sum_{j=1}^{r} \frac{v_j}{p_j(\theta)} \frac{dp_j(\theta)}{d\lambda}$$

The elements $\frac{\partial p_j(\theta)}{d\alpha}$ and $\frac{\partial p_j(\theta)}{d\lambda}$ calculated in the precedent subsection are adapted to the equiprobable classes Δ_j' and I_n represents the Fisher information matrix given in subsection 3.2.

The value of Y^2 can be deduced and H_0 is rejected for $Y^2 \succ \chi^2$ with $r-1$degrees of freedom.

4.3 McCulloch statistic

The NRR statistic have attracted the attention of many researchers particularly McCulloch (1985) who gave a very interesting decomposition of this one. He derived a new powerful statistic S^2, which can be used separately from Y^2. The author showed that this test, named the $McCu-test$, has the lowest variance and the highest rate of convergence to the limit. The power gain compared to Y^2 and U^2 is showed for normality against logistic distribution (Voinov et al., 2013) which makes it a very good alternative to existing goodness-of-fit tests. Despite its power, this statistic hasn't been sufficiently exploited for fitting models.

Based on the maximum likelihood estimators on initial data, It's defined by:

$$S_n^2 = Z_n^T(\hat{\theta})B[(I_n(\theta) - J_n(\theta))^{-1} + (B^TB)^{-1}]B^TZ_n(\hat{\theta})$$

The elements of this criteria test are defined in the precedent subsections. The vector $Z(\hat{\theta})$, the Fisher information matrix on initial data $I_n(\hat{\theta})$, the matrix $J_n(\hat{\theta})$ and the corresponding matrix B are calculated for the chosen grouped data Δ_j'. Therefore, we obtain the $McCu-test$ statistic value for the UGE distribution. The null hypothesis H_0 is rejected for $S^2 \succ \chi^2$ with 2 degrees of freedom.

4.4 Singh statistic

On an other hand, Singh (1987) introduced a new chi-square goodness-of-fit test called Singh statistic and noted Q^2. The particularity of this one, is its used in more general conditions than Y^2 and S^2. The author showed that this criteria does'nt depend on the estimation method used for calculating the unknwon parameters or on the grouping data cells. Note that despite its nice properties, this test has not been used till now. The general formula of Singh statistic is given by:

$$Q^2(\tilde{\theta}) = Z_*^T(\tilde{\theta})(I - BI_n^{-1}B^T)Z_*(\tilde{\theta})$$

The distribution of this statistic is a chi-square with $r-1$ degrees of freedom, where r is the number of grouped classes chosen. All its elements are provided above and the vector $Z_*(\tilde{\theta})$ is defined by

$$Z_*(\tilde{\theta}) = Z(\tilde{\theta}) - BI_n^{-1}W(\tilde{\theta})$$

where the vector $W(\tilde{\theta})$ is:

$$W_1(\tilde{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\partial \ln f(x_i,\theta)}{\partial \alpha}|_{\alpha=\tilde{\alpha}}$$

$$W_2(\tilde{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\partial \ln f(x,\theta)}{\partial \lambda}|_{\lambda=\tilde{\lambda}}$$

Based on this technique, we construct a new flexible statistic test Q^2 for fitting the new model UGE. In this case, we obtain:

$$W_1(\tilde{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left[\frac{1}{\tilde{\alpha}} + \ln[1 - exp(-\tilde{\lambda}(\frac{x_i}{1-x_i}))]\right]$$

$$W_2(\tilde{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left[\frac{1}{\tilde{\lambda}} + (\frac{x_i}{1-x_i})\left[\frac{(\tilde{\alpha}-1)exp(-\tilde{\lambda}(\frac{x_i}{1-x_i}))}{1-exp(-\tilde{\lambda}(\frac{x_i}{1-x_i}))} - 1\right]\right]$$

Because of their efficiency, the MLE, CVM , AD and MPS estimators are used in this study to calculate the value of $Q_n^2(\tilde{\theta})$. H_0 is not accepted for $Q^2 \succ \chi^2$ with $r-1$ degrees of freedom.

5 Simulation study and Applications

An extensive simulation study is conducted to show the practicability of the different tests proposed in this study for fitting the UGE distrbution and their power compared to the estimation methods used and grouping data cells. At this end, $N = 10{,}000$ samples with different sizes $(n = 20{,}30{,}50{,}100$ and $n = 150)$ are generated from the new model. First, we estimated the unknown parameters by different methods. The results are not reported here but they showed that the MLE, AD, CvM and MPS estimates are more efficient than those obtained by other methods. The values of the statistics U^2 and Q^2 with the cited estimates are computed for all the simulated samples for equiprobability and non-equiprobability grouping classes (see Tables 2,3) and they are compared to the corresponding theoretical chi-square distributions with levels of significance $(\epsilon = 0.025$, 0.05 and 0.1); while Y^2 and S^2 are calculated for MLE and equiprobable cells (see Table 1).

n	$\epsilon = 0.025$	$\epsilon = 0.05$	$\epsilon = 0.1$
$Y^2(\theta_{MLE})$			
20	(0.0238)	(0.0439)	(0.1131)
30	(0.0243)	(0.0441)	(0.1092)

50	(0.0247)	(0.0447)	(0.1081)
100	(0.0251)	(0.0516)	(0.1080)
150	(0.0249)	(0.0506)	(0.1056)
$S^2(\theta_{MLE})$			
20	(0.0259)	(0.0410)	(0.0857)
30	(0.0244)	(0.0416)	(0.0904)
50	(0.0246)	(0.0470)	(0.0923)
100	(0.0254)	(0.0476)	(0.1069)
150	(0.0251)	(0.0511)	(0.1000)

Table 1 values of Y^2 and S^2 for equiprobability cells

n	$\epsilon = 0.025$	$\epsilon = 0.05$	$\epsilon = 0.1$	n	$\epsilon = 0.025$	$\epsilon = 0.05$	$\epsilon = 0.1$
		$Q^2(\theta_{MLE})$				$U^2(\theta_{MLE})$	
20	(0.0257)	(0.0550)	(0.0924)		(0.0340)	(0.0864)	(0.1530)
30	(0.0254)	(0.0521)	(0.0933)		(0.0313)	(0.0554)	(0.1164)
50	(0.0253)	(0.0520)	(0.0943)		(0.0309)	(0.0485)	(0.1137)
100	(0.02480)	(0.0511)	(0.0988)		(0.0290)	(0.0486)	(0.1090)
150	(0.02520)	(0.0494)	(0.1009)		(0.0246)	(0.0533)	(0.0991)
		$Q^2(\theta_{AD})$				$U^2(\theta_{AD})$	
20	(0.0262)	(0.0520)	(0.0932)		(0.02560)	(0.0598)	(0.1063)
30	(0.0255)	(0.0516)	(0.0956)		(0.02426)	(0.0527)	(0.1057)
50	(0.0245)	(0.0509)	(0.0967)		(0.02540)	(0.0511)	(0.1030)
100	(0.0253)	(0.0494)	(0.0982)		(0.02517)	(0.0505)	(0.1016)
150	(0.02490)	(0.0498)	(0.0984)		(0.02502)	(0.0503)	(0.1003)
		$Q^2(\theta_{MPS})$				$U^2(\theta_{MPS})$	
20	(0.0229)	(0.0573)	(0.0914)		(0.0262)	(0.0544)	(0.1042)
30	(0.0239)	(0.0564)	(0.0960)		(0.0259)	(0.0534)	(0.1040)
50	(0.0256)	(0.0559)	(0.1022)		(0.0244)	(0.0514)	(0.1035)
100	(0.0255)	(0.0496)	(0.0981)		(0.0247)	(0.0510)	(0.1019)
150	(0.0248)	(0.0499)	(0.1010)		(0.0252)	(0.0501)	(0.1012)
		$Q^2(\theta_{CM})$				$U^2(\theta_{CM})$	
20	(0.0231)	(0.0588)	(0.0954)		(0.0259)	(0.0552)	(0.1070)
30	(0.0265)	(0.0546)	(0.0955)		(0.0258)	(0.0550)	(0.1060)
50	(0.0261)	(0.0523)	(0.0958)		(0.0243)	(0.0544)	(0.0952)
100	(0.0243)	(0.0492)	(0.1018)		(0.0255)	(0.0484)	(0.1048)

150	(0.0252)	(0.0474)	(0.0984)		(0.0246)	(0.0494)	(0.1004)

Table 2. :values of Q^2 and U^2 for MLE, AD, MPS and CM parameter estimators for equiprobability cells

The obtained results indicate that for Y^2 and S^2, the empirical and theoretical values of level of significance are very close to each other which confirm the use of these statistics to fit data to the UGE distribution in a satisfactory manner (Table1). We can also say that S^2 gave best results than Y^2 for large samples. On an another hand, when the MLE estimators cannot be found, we consider the alternative statistics Q^2 and $U^2$to fit data to the UGE model. The results summarized in Table 2 demonstrate the relative power under different estimation methods ($\hat{\theta}_{MLE}$, $\tilde{\theta}_{AD}$, $\tilde{\theta}_{MPS}$, and $\tilde{\theta}_{CM}$). We constate that for larger samples, both statistics yield similar performance, indicating their asymptotic equivalence as sample size increases and U^2 appears to be a more powerful and adaptable test, especially for small sample sizes.

n	$\epsilon = 0.025$	$\epsilon = 0.05$	$\epsilon = 0.1$		$\epsilon = 0.025$	$\epsilon = 0.05$	$\epsilon = 0.1$
		$Q^2(\theta_{MLE})$				$U^2(\theta_{MLE})$	
20	(0.0269)	(0.0459)	(0.0908)		(0.02370)	(0.0443)	(0.1084)
30	(0.0241)	(0.0464)	(0.0931)		(0.02460)	(0.0567)	(0.1064)
50	(0.0257)	(0.0469)	(0.0945)		(0.02470)	(0.0535)	(0.1034)
100	(0.0252)	(0.0476)	(0.0952)		(0.02520)	(0.0533)	(0.1027)
150	(0.0249)	(0.0491)	(0.0965)		(0.02490)	(0.0522)	(0.0991)
		$Q^2(\theta_{AD})$				$U^2(\theta_{AD})$	
20	(0.0263)	(0.0446)	(0.0904)		(0.02630)	(0.0566)	(0.0948)
30	(0.0241)	(0.0454)	(0.0907)		(0.02620)	(0.0534)	(0.0966)
50	(0.0258)	(0.0460)	(0.0911)		(0.02470)	(0.0531)	(0.1054)
100	(0.0246)	(0.0474)	(0.0939)		(0.02570)	(0.0515)	(0.1018)

150	(0.0250)	(0.0477)	(0.0952)		(0.02530)	(0.0510)	(0.1011)
		$Q^2(\theta_{MPS})$				$U^2(\theta_{MPS})$	
20	(0.0263)	(0.0532)	(0.0919)		(0.0225)	(0.0554)	(0.1075)
30	(0.0258)	(0.0516)	(0.0938)		(0.0263)	(0.0542)	(0.1067)
50	(0.0257)	(0.0493)	(0.1054)		(0.0240)	(0.0461)	(0.1040)
100	(0.0248)	(0.0494)	(0.0957)		(0.0246)	(0.0507)	(0.0961)
150	(0.0249)	(0.0497)	(0.0966)		(0.0249)	(0.0503)	(0.1033)
		$Q^2(\theta_{CM})$				$U^2(\theta_{CM})$	
20	(0.0264)	(0.0451)	(0.0918)		(0.0237)	(0.0530)	(0.1075)
30	(0.0253)	(0.0471)	(0.0958)		(0.0262)	(0.0512)	(0.1067)
50	(0.0248)	(0.0482)	(0.0964)		(0.0242)	(0.0491)	(0.1040)
100	(0.0249)	(0.0510)	(0.0975)		(0.0249)	(0.0494)	(0.0961)
150	(0.0251)	(0.0504)	(0.0983)		(0.0250)	(0.0499)	(0.1033)

Table 3. values of Q^2 and U^2 for MLE, AD, MPS and CM parameter estimators for non-equiprobability cells

The simulated results (Table 3) of Q^2 and U^2 under non-equiprobable grouping classes, considering different estimation methods, enhances the power of both tests. U^2 appears to be more stable and powerful across different estimation methods, while Q^2 remains a reliable alternative with lower variability. We note that the Anderson-Darling estimation method gave the best results for the Q^2 statistic and that the MPS and CvM estimators gave better results than the MLE for the U^2 statistic. Also, the use of U^2 is more suitable for small and moderate sample sizes. Furthermore, non-equiprobable grouping cells appears to improve test performance, making it a preferable approach when applying these goodness-of-fit tests.

5.1 Applications

The flexibility of the proposed model UGE is illustrated by two real data sets. After using classical methods for testing the fit of these observations to the UGE

distribution, the constructed goodness-of-fit test statistics are provided to confirm the results. Note that these statistics are very practical because they conclude or not the fit and they also permit to distinguish between the studied model and its competing bounded distributions like:

Unit generalized inverse Weibull distribution: $F_{IGUW}(x) = e^{-A(\frac{C(1-x)}{x})^B}$

Gompertz distribution: $F_{UG}(x) = e^{-C(x^{-D}-1)}$

Unit $Burr - III$ **distribution:** $F_{UBurr-III}(x) = (1 + (\frac{1}{x} - 1)^d)^{-k}$

Kumaraswamy distribution: $F_{KUM}(x) = 1 - (1 - x^n)^s$

Example 1:

The observations are the values of the force exerted on the glass of the airplane window reported by Fuller et al. (1994) which are divided by 100 to get data between [0,1]:

18.83,20.80,21.657,23.03,23.23,24.05,24.321,25.50,25.52,25.80,26.69,26.77,26.78,27.05,27.67,29.90,31.1

The empirical and theoretical pdf and cdf of these data are reported in Figure 2 as well as the PP-plots and the QQ-plots.

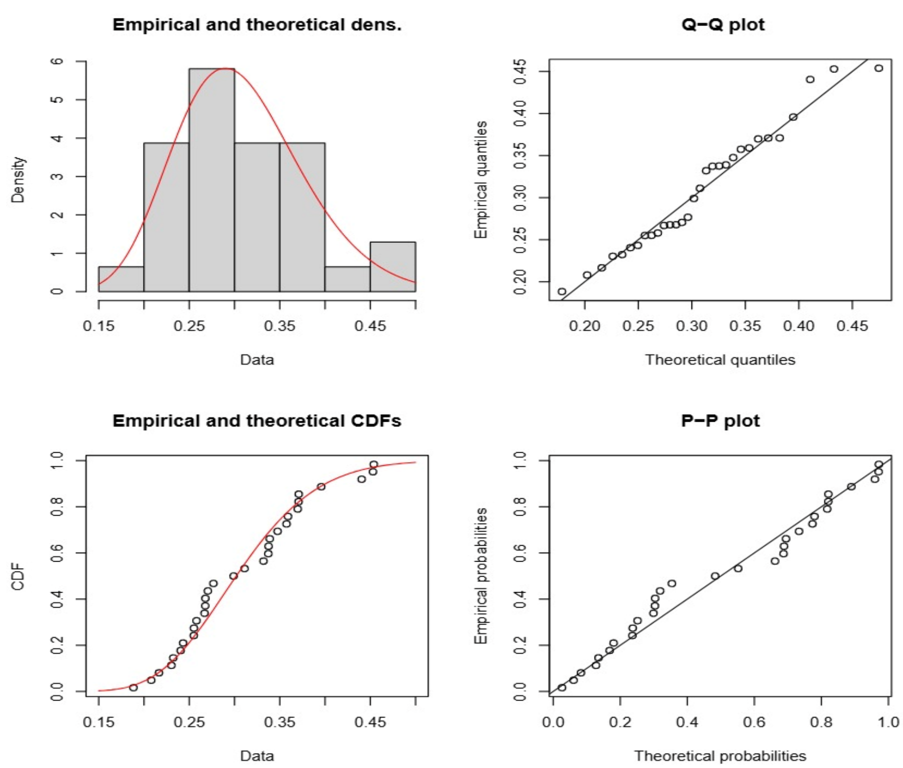

Fig. 2: pdf, cdf, pp-plot and qq-plot for data of The force exerted on the glass of the airplane window

Without lost of generality, we calculate the MLE and AD estimates, we use firstly classical model selection criteria to fit these data to the cited models. The obtained results (Table 4), the pdf plots and cdf plots for the different models (Figure 3, Figure 4) proved that this new model can be used instead of its alternatives to describe this dataset.

Distributions	$estimators$	LL	AIC	BIC
UGE (α,λ)	$\alpha_{MLE} =$ 21.0305 $\lambda_{MLE} =$ 7.9277	38.7757	-73.5515	-70.6835
	$\alpha_{AD} =$ 15.2618 $\lambda_{AD} =$ 7.2007	38.5395	-73.0791	-70.2111
$UGIW(B,A,C)$	$B_{MLE} =$ 3.3441 $A_{MLE} =$ 0.3805 $C_{MLE} =$ 0.4938	38.2989	-70.5978	-66.2958
	$B_{AD} =$ 6.0710^{-10} $A_{AD} =$ 0.6930 $C_{AD} =$ 0.8426	38.2989	-70.5978	-66.2958
$Kum(s,n)$	$s_{MLE} =$ 152.05 $n_{MLE} =$ 4.6213	37.2578	-70.5157	-67.6477
	$s_{AD} =$ 152.05 $n_{AD} =$ 4.6213	37.2578	-70.5157	-67.6477
$UBurr-III(k,d)$	$k_{MLE} =$ 0.0170			

	$d_{MLE}=$ 70.8399	23.6859	-43.3718	-40.5038
	$k_{AD}=$ 0.0173			
	$d_{AD}=$ 54.0221	22.8039	-41.6078	-38.7398
$UW(b,a)$	$b_{MLE}=$ 5.8993			
	$a_{MLE}=$ 0.2133	38.7082	-73.4164	-70.5485
	$a_{AD}=$ 5.5019			
	$b_{AD}=$ 0.2376	38.5797	-73.1594	-70.2914
$UG(D,C)$	$D_{MLE}=$ 4.6205			
	$C_{MLE}=$ 0.0022	37.5053	-71.0106	-68.1426
	$D_{AD}=$ 4.4654			
	$C_{AD}=$ 0.0028	37.4722	-70.9444	-68.0764

Table 4. Model selection criteria for Fuller data（1994）for UGE against the alternatives

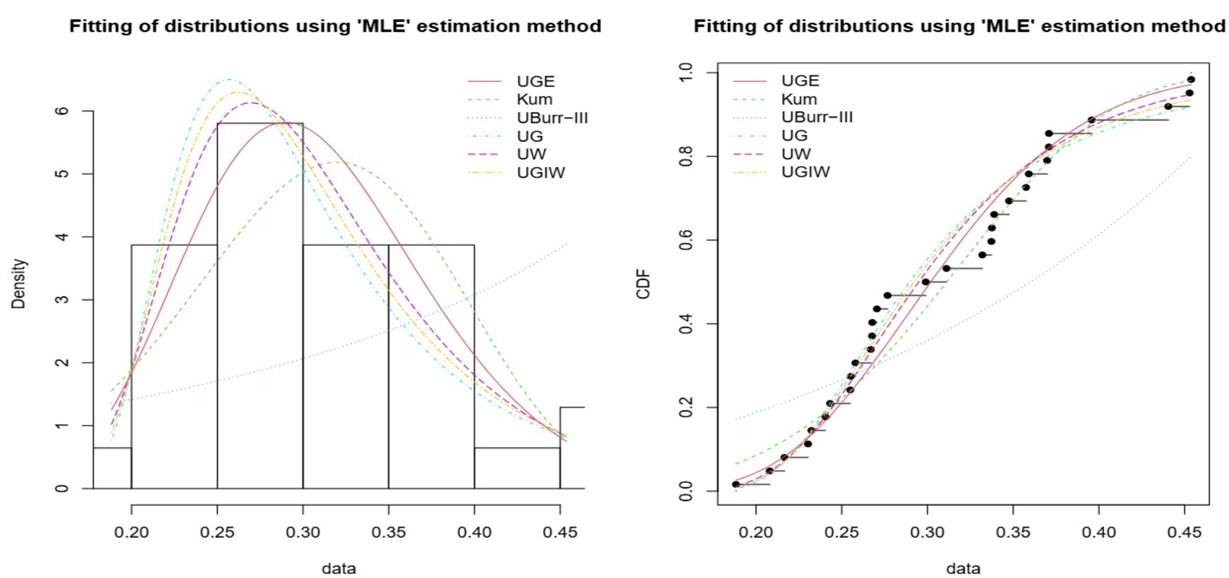

Fig. 3: Fitting of distributions with 'MLE' method estimation for the

values of The force exerted on the glass of the airplane window

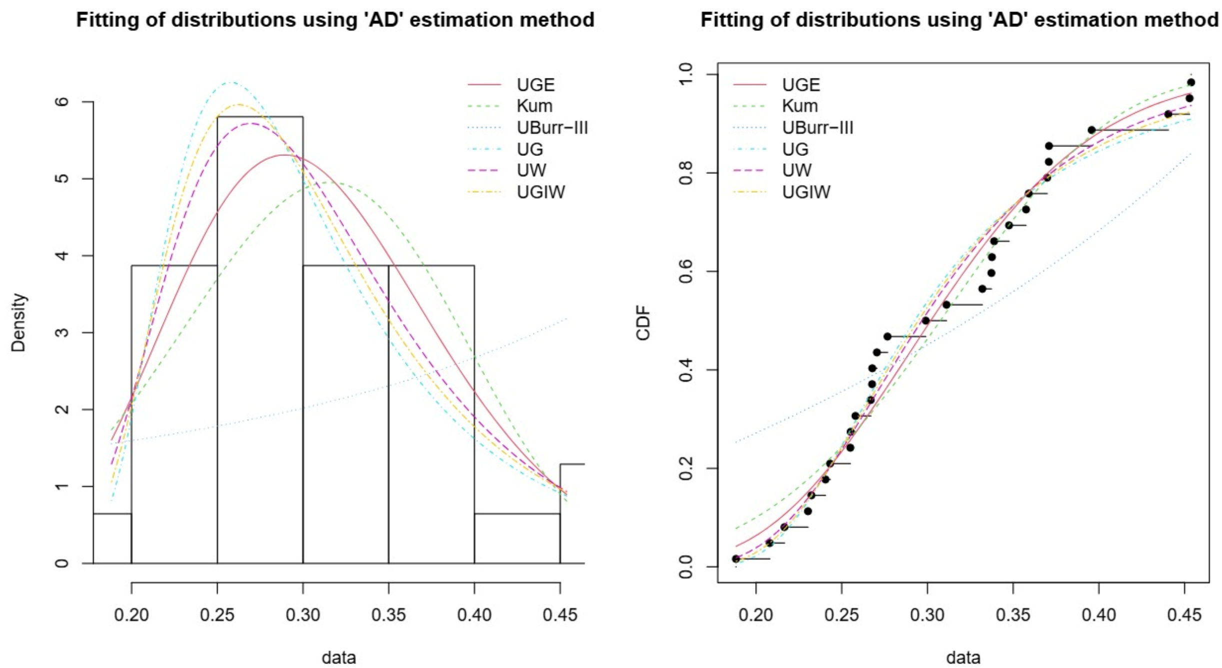

Fig. 4: Fitting of distributions with 'AD' method estimation for the values of The force exerted on the glass of the airplane window

As shown in this work, we propose the use of the statistics Y^2, S^2, U^2 and Q^2 for testing the null hypothesis H_0 that this sample belongs to the UGE distribution. The computation is based on the MLestimates and the AD estimates. If we choose, for example $r = 7$ grouping classes, the intermediate calculations are given below:

$$B(\hat{\theta}_{MLE}) = \begin{bmatrix} -0.02295 & -0.01005 \\ -0.00065 & 0.00545 \\ 0.00999 & 0.01361 \\ 0.01307 & 0.17182 \\ 0.11499 & 0.04882 \\ -0.00599 & -0.05866 \\ -0.11726 & -0.15573 \end{bmatrix}$$

$$L = (L_1 = 0.01006, \qquad L_2 = 0.37676)$$

$$J(\hat{\theta}_{MLE}) = B^T(\hat{\theta}_{MLE})B(\hat{\theta}_{MLE}) = \begin{bmatrix} 0.02781 & 0.02684 \\ 0.02684 & 0.05992 \end{bmatrix}$$

$$I_n(\hat{\theta}_{MLE}) = \begin{bmatrix} 2.02610^{-10} & -3.66010^{-5} \\ -3.66010^{-5} & 6.610947 \end{bmatrix}$$

$$Z = (0.7282006, 0.2715377, 0.2715377, -1.1540352, -0.2036533, 0.7467287, 0.2432196)$$

$$W = \frac{1}{\sqrt{31}}(-2.1310^{-5}, 14.31)$$

	Y^2	S^2	U^2	Q^2
statistic values	2.9251	1.2952	1.1152	2.5793
critical values	$X^2_{r-1} =$ 12.59	$X^2_s =$ 5.99	$X^2_{r-s-1} =$ 9.49	$X^2_{r-1} =$ 12.59

Table 5. statistic values for MLE estimators

For a level of significance $\varepsilon = 0.05$, the values of the criteria tests (Table 5) are less than those of the critical corresponding chi-square distributions. So, we conclude that all the goodness-of-fit tests confirmed the null hypothesis H_0.

As the explicit form of the maximum likelihood estimators doesn't exist always, so we can use the statistics U^2 and Q^2 for testing H_0. At this end, we proceed as above but the values of the statistics are based, for example, on the AD estimates. The components of the corresponding matrices are:

$$B(\hat{\theta}_{AD}) = \begin{bmatrix} -0.02528 & -0.01385 \\ -0.00090 & 0.00751 \\ 0.01377 & 0.01875 \\ 0.01647 & 0.14108 \\ 0.11700 & 0.05218 \\ -0.00261 & -0.05624 \\ -0.11711 & -0.14285 \end{bmatrix}$$

$$L = (L_1 = 0.16886, \qquad L_2 = -0.08795)$$

$$J(\hat{\theta}_{AD}) = B^T(\hat{\theta}_{AD})B(\hat{\theta}_{AD}) = \begin{bmatrix} 0.02851 & 0.02591 \\ 0.02591 & 0.04679 \end{bmatrix}$$

$$I_n(\hat{\theta}_{AD}) = \begin{bmatrix} 0.00046 & 0.05505 \\ 0.05505 & 6.46184 \end{bmatrix}$$

$$Z = (-0.07968924, 1.22191962, 0.27153769, -1.15403520, -0.20365327, 0.74672866, 0.41184056)$$

$$W = \frac{1}{\sqrt{31}}(0.1206, 14.15)$$

Then, we obtain :

$$Q^2(\tilde{\theta}_{AD}) = 3.5605 \quad \text{and} \quad U^2(\tilde{\theta}_{AD}) = 2.0282$$

For a level of significance $\varepsilon = 0.05$, the corresponding chi-square critical values are $X^2_{r-1} = 12.59$ and $X^2_{r-s-1} = 9.49$. So, the null hypothesis H_0 can not be rejected which affirm that the observed data can be modeled by the UGE distribution.

Example 2

The usefulness of the new model appear also in modelling left skewed data. We suppose that the observations of time to failure of turbocharger of one type of engine studied by Xu et al., (2003) follow the UGE distribution. The observed failure times given below are divided by 10 to get values between [0,1]:

4.8,6.5,7.0,7.3,7.7,8.0,8.4,1.6,3.5,3.0,4.6,2.0,3.9,5.0,5.6,6.1,6.5,7.1,7.3,7.8,8.1,8.4,2.6,4.5,5.1,5.8,6.3,6.7,7.3,7.7,7.9,8.3,8.5,5.3,6.0,8.7,8.8,5.4,6.0,9.0.

For testing this hypothesis, we proceed as in example 1. First, we represent the pp-plots and the QQ-plots of these observations (Figure 5). The classical methods of model selection with different estimates showed that our model fit these failure time observations better than the alternatives (Table 6, Table 7).

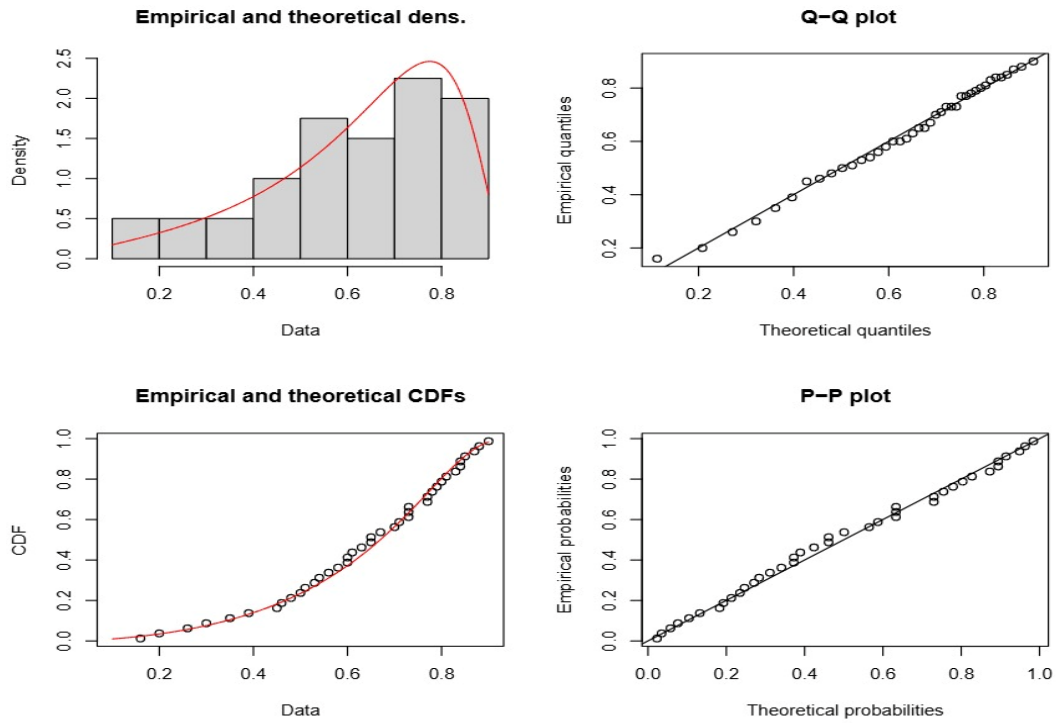

Fig. 5: pdf, cdf, qq-plot and pp-plot failure time of turbocharger

Distributions	Estimators	LL	AIC	BIC
$UGE(\alpha, \lambda)$	$\alpha_{MLE} =$ 1.5853 $\lambda_{MLE} =$ 0.5122	13.8294	-23.6588	-20.2811
	$\alpha_{AD} =$ 1.4531 $\lambda_{AD} = 0.4825$	13.7543	-23.5087	-20.1309
$UGIW(B, A, C)$	$B_{MLE} =$ 1.0286 $A_{MLE} =$ 0.7545 $C_{MLE} =$ 1.4651	7.7719	-9.5439	-4.4773
	$B_{AD} =$ 2.910^{-09} $A_{AD} =$ 0.6931 $C_{AD} =$ 0.8347	-761.72	1529.44	1534.51

$Kum(s,n)$	$s_{MLE} =$ 2.4924 $n_{MLE} =$ 3.1564	12.3428	-20.6857	-17.3079
	$s_{AD} = 2.2269$ $n_{AD} =$ 3.0377	12.2268	-20.4536	-17.0758
$UBurr-III(k,d)$	$k_{MLE} =$ 1.8753 $d_{MLE} =$ 1.5006	11.1585	-18.3171	-14.9394
	$k_{AD} =$ 1.9567 $d_{AD} =$ 1.4651	11.1054	-18.2109	-14.8331
$UW(b,a)$	$b_{MLE} =$ 1.4611 $a_{MLE} =$ 2.1273	10.6883	-17.3767	-13.9989
	$b_{AD} = 1.4614$ $a_{AD} =$ 2.2538	10.6205	-17.2410	-13.8632
$UG(D,C)$	$D_{MLE} =$ 0.6951 $C_{MLE} =$ 1.9321	8.1274	-12.2549	-8.8771
	$D_{AD} =$ 0.9971 $C_{AD} =$ 1.2052	7.6749	-11.3498	-7.9721

Table 6. classical model selection of the UGE against alternative distributions

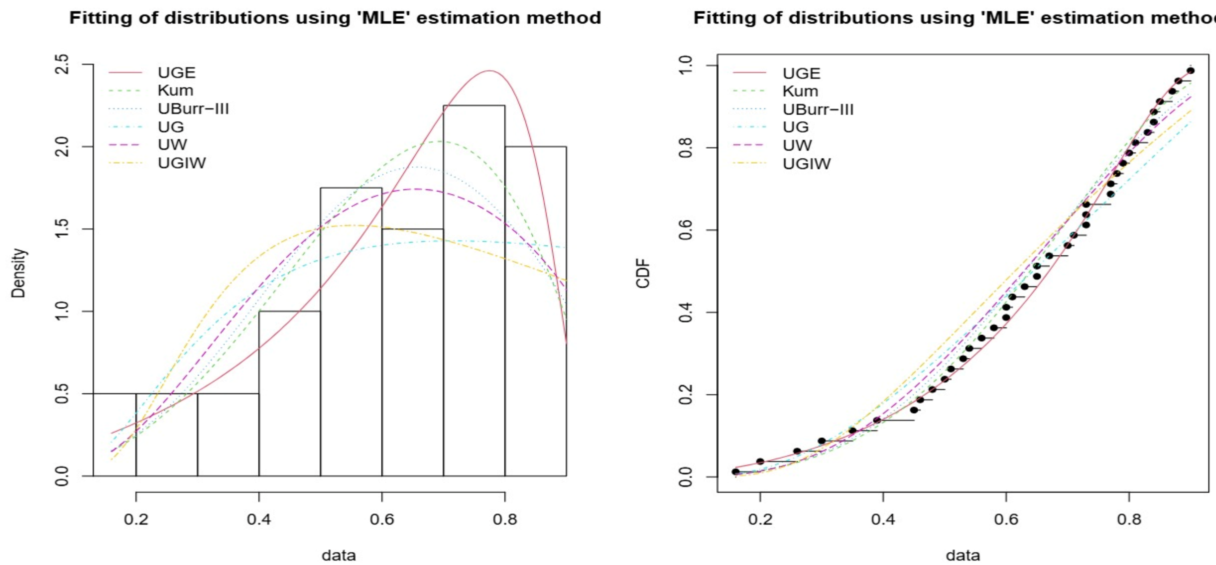

Fig. 6: Fitting of distributions with 'MLE' method estimation for time to failure (103h) of turbocharger of one type of engine

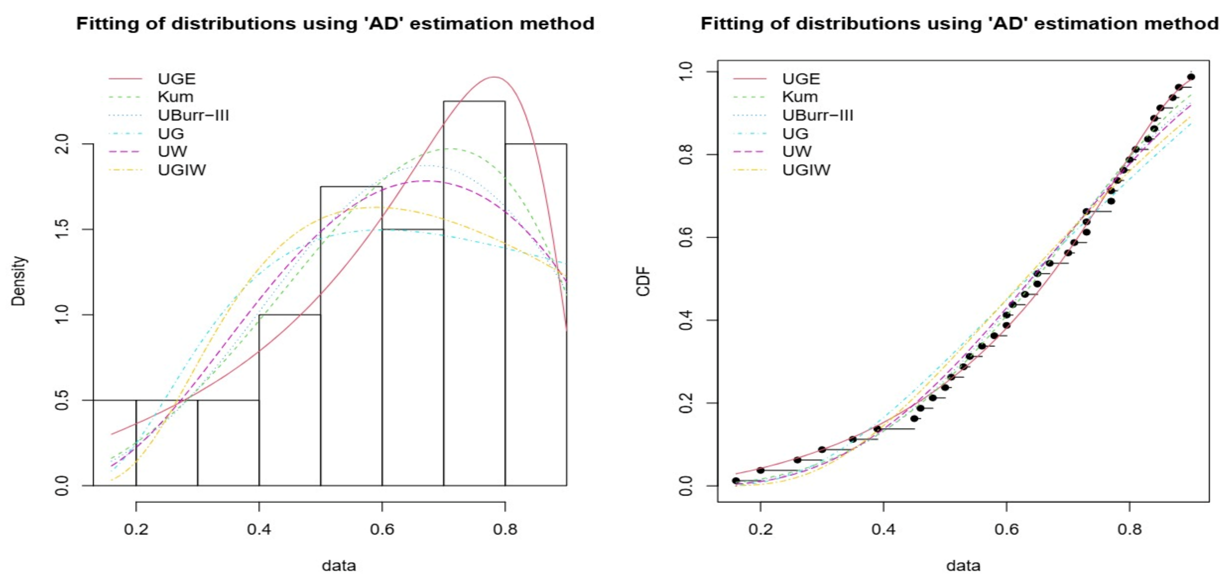

Fig. 7: Fitting of distributions with 'AD' method estimation for time to failure (103h) of turbocharger of one type of engine

To confirm this fact, we compute the corresponding test statistics Y^2, S^2, U^2 and Q^2. If we choose $r = 6$ grouping classes, the necessary calculations are given for the MLE and AD estimation methods respectively:

$$B(\hat{\theta}) = \begin{bmatrix} -0.32672 & -0.10441 \\ 0.03033 & 0.11783 \\ 0.18290 & 0.21065 \\ 0.86594 & 0.71828 \\ 0.40619 & 0.03509 \\ -0.48356 & -1.36337 \end{bmatrix}$$

$$L = (L_1 = 1.305785, \qquad L_2 = 3.416131)$$

$$J(\hat{\theta}) = B^T(\hat{\theta})B(\hat{\theta}) = \begin{bmatrix} 1.2898 & 1.3717 \\ 1.3717 & 2.4451 \end{bmatrix}$$

$$I_n(\hat{\theta}) = \begin{bmatrix} 8.7010^{-09} & -0.00152 \\ -0.00152 & 266.422 \end{bmatrix}$$

$$Z = (0.1081749, 0.5163978, 0.1290994, -0.2581989, -0.2581989, 0.3916068)$$

$$W = \frac{1}{\sqrt{40}}(-5.9010^{-4}, 103.23)$$

$$B(\hat{\theta}_{AD}) = \begin{bmatrix} -0.32961 & -0.11391 \\ 0.03309 & 0.12855 \\ 0.19954 & 0.22465 \\ 0.82445 & 0.73146 \\ 0.42394 & 0.05082 \\ -0.47927 & -1.34270 \end{bmatrix}$$

$$L = (L_1 = 1.6284, \qquad L_2 = 4.0360)$$

$$J(\hat{\theta}_{AD}) = B^T(\hat{\theta}_{AD})B(\hat{\theta}_{AD}) = \begin{bmatrix} 1.2387 & 1.3547 \\ 0.5093 & 0.5733 \end{bmatrix}$$

$$I_n(\hat{\theta}_{AD}) = \begin{bmatrix} 0.0228 & 2.4923 \\ 2.4923 & 271.470 \end{bmatrix}$$

$$Z = (0.2131557, 0.5163978, 0.1290994, -0.2581989, -0.2581989, 0.4469937)$$

$$W = \frac{1}{\sqrt{40}}(-2.2024, 53.5212)$$

We obtain, for a level of a significance $\epsilon = 0.05$:

	Y^2	S^2	U^2	Q^2
statistic values (MLE)	0.7864	0.0112	0.3657	0.3936
statistic values (AD)			0.3912	0.4242
critical values	$X^2_{r-1} =$ 11.07	$X^2_s = 5.99$	$X^2_{r-s-1} = 7.81$	$X^2_{r-1} = 11.07$

Table 7. criteria test values for MLE and AD estimators

For a level of a significance $\varepsilon = 0.05$, as shown in Table 7, the null hypothesis H_0 is not rejected, we can deduce that this new distribution can also model asymetric data.

Conclusion

Using different approaches, we have constructed new powerful goodness-of-fit test statistics. Based on efficient estimators, these statitics are capable to fit datasets to a new model (UGE) without regarding any alternative as used in classical selection models. Having the lowest variance and the highest rate of convergence to the limit, they also recover the information lost while grouping data and don't require a specific estimation method. Thousands of simulated samples of varying sizes were utilized to demonstrate the practicability of the proposed tests. Two real-world data illustrated the flexibility of this model particularly for skewed data, compared to existing distributions, including the Kumaraswamy, Unit Weibull, Unit Generalized Inverse Weibull, Unit Burr III and Unit-Gompertz Distributions. The findings from this study provide a robust foundation for improved model validation techniques, paving the way for more reliable statistical analyses in diverse application areas.

Statements and declarations

The authors confirm that this work is not funded by any institution or any organisation and attest that there is no conflict of interest.

References

Aidi, K., Seddik-Ameur, N., Abd EL-Baset A. Ahmad, Mundher A. Khaleel, 2022. The Topp-Leone Extended Exponential Distribution: Estimation Methods and Applications, Pakistan Journal of Statistics Operation Research 18 (4).

BagdonaviË‡ cius, V., Clerjaud, L., Nikulin, M., 2006. Power generalized Weibull in accelerated life testing, Preprint 0602, Statistique MathÃ© matique et ses Applications, I.F.R. 99, UniversitÃ© Victor Segalen Bordeaux 2, p. 28.

Chouia, S., Seddik-Ameur, N., 2014. A Goodness-of-fit Test for a Competing Risk Model International Journal of Applied Mathematics and Statistics 52 (2), 34-41.

Dzhaparidze, K.O., Nikulin, M.S., 1974. On a modification of the standard statistic of Pearson.Theory of Probability and its Applications 19, 851â “853.

Fuller E.J., Frieman S, Quinn J, et al. Fracture mechanics approach to the design of glass aircraft windows: A case study. SPIE Proc. 1994; 2286:419â “430.

Gupta, R. D. and Kundu, D. (1999). Theory & methods: Generalized exponential distributions. Australian & New Zealand Journal of Statistics, 41(2):173{188.

Kumaraswamy, P. (1980). Generalized probability density-function for double-bounded random processes. Journal of Hydrology. 462:79-88.

McCulloch, C.E., 1985. Relationships among some chi-squared goodness of fit statistics.Communications in Statistics â “ Theory and Methods 14, 593â “603.

Mazucheli, J., 2018. The unit weibull distribution and associated inference, journal of Applied Probability and Statistics, Vol. 13, No. 2, pp. 1-22.

Mazucheli, J., Menezes, A. F., Dey, S., 2019.Unit-Gompertz Distribution with Applications, Statistica, 79, issue 1, p. 25-43, https://EconPapers.repec.org/RePEc:bot:rivsta:v:79:y:2019:i:1:p:25-43.

Meribout, K. K.,Seddik Ameur, N., Goual, H., 2023.Goodness-of-fit tests for the new weibull-G family of distribution, Advances in Mathematics: Scientific Journal.

Nikulin, M.S., 1973. Chi-square test for continuous distributions, Theory of Probability and its Applications 18, 638â “639.

Rao K. C., Robson D.S .(1974). "A Chi-square statistic for goodness-of-fit for tests with in the exponential family".Communications in statistics 3:1139-53.

Singh, A.C., 1987. On the optimality and a generalization of Rao-Robsonâ s statistic. Communications in Statistics - Theory and Methods 16, 3255â "3273.Tilbi, D., Seddik-Ameur, N. 2018. Chi-squared goodness-of-fit tests for the generalized Rayleigh distribution, Journal of Statistical Theory and Practice, Volume 11, Issue 4, pp. 594-603, (2017), http://dx.doi.org/10.1080/15598608.2017.1286276.

Tilbi, D. and Seddik-Ameur, N. 2017. Chi-squared goodness-of-fit tests for the generalized Rayleigh distribution, Journal of Statistical Theory and Practice, 11:4, 594-603, DOI:10.1080/15598608.2017.1286276.

Xu et al., 2003.Application of Neural Networks in Forecasting Engine Systems Reliability. Applied Soft Computing, 2, pp. 255-268.