

NRR statistic for the extension Weibull distribution

W. Treidi and N. Seddik-Ameur

*Laboratory of probability and statistics,
Badji Mokhtar university-Annaba-Algeria.*

Abstract

The extension Weibull distribution is a new model generated from Weibull distribution to model the bathtub failure rate life time data. Characterized by three parameters, this model has many advantages in applications. In this work, we propose the construction of a modified chi-squared goodness-of-fit test based on the Nikulin-Rao-Robson (NRR) statistic for this distribution when the parameters are unknown. This test is based on maximum likelihood estimators on non-grouped data and follows chi-square distribution. Simulations and real data sets from reliability and survival analysis are proposed to show the performances of the results obtained through this study.

AMS subject classification: 62F03-62G05-62G10.

Keywords: Chi-squared test, maximum likelihood estimation, NRR statistic.

1. Introduction

Weibull distribution still attracts a great deal of researchers and reliability engineers. Depending on its shape parameter values, the failure rate can be decreasing describing early failures, monotone indicative of useful life or increasing describing aging or wear-out failures. Despite its flexibility, this distribution can not modelize the reliability of some real systems, so many extended models have been discussed in the statistical literature during recent years. Among these, we can mention the first generalization called the exponentiated Weibull (EW) distribution, introduced by Mudholkar and Srivastava (1993), for more details one can see Nadarajah *et al.* (2013), and the recent one, the gamma exponentiated Weibull model proposed by Castellares and Lemonte (2015).

In this work we are mainly interested in the construction of a modified chi-squared goodness-of-fit test for the extension Weibull model introduced by Xie *et al.* (2002). This one is a new distribution generated from Weibull distribution to model the bathtub failure rate life time data. Characterized by three parameters, this model has many

advantages in applications. It can be considered as generalizations of the model proposed by Chen (2000) and the exponential power distribution studied by Smith and Bain (1975). Statistical properties were detailed in Xie *et al.* (2004). Later and using Markov chain Monte Carlo simulations, Gupta *et al.* (2008) provided bayesian analysis and used graphical and numerical methods to fit real data to this distribution. They showed that the extension Weibull distribution is more capable to model diverse problems than the Weibull model does. Testing the validity of this model has not been investigated yet what motivated us to propose a modified chi-squared goodness-of-fit test for this distribution when the parameters are unknown. We shall adapt the Nikulin-Rao-Robson (NRR) statistic proposed separately by Nikulin (1973) and Roa and Robson (1974). The NRR statistic Y^2 based on maximum likelihood estimation on initial data, is a natural modification of the wellknown Pearson statistic X^2 and follows chi-square distribution. When the parameters of the model are unknown, It was demonstrated that the Pearson statistic X^2 does not follow chi-square distribution and depends on the estimation method used which makes the test inapplicable. So, different techniques were proposed to assess the adequacy of models used in the analysis such as graphical methods, Hsuan–Robson–Mirvaliev (HRM) statistic based on moment-type estimators, the Dzhaparidze–Nikulin (DN) statistic (Dzhaparidze and Nikulin, 1974), the McCulloch test (McCulloch, 1985).

The paper is organized as follow. The extension Weibull distribution and the theory of the NRR statistic are presented respectively in section 2 and 3. The section 4 is devoted to the construction of a modified chi-squared goodness-of-fit test based on the NRR statistic for the extension Weibull distribution and all the elements of the statistic Y^2 are given. To evaluate our results, we conducted an intensive simulation study (10, 000 samples of each size $n = 30$; $n = 50$; $n = 100$; $n = 250$; $n = 500$) and two examples of real data from reliability and survival analysis are proposed in section 5. Anderson-Darling and Kolmogorov-Smirnov statistics are also calculated.

2. Extension Weibull model

The cumulative density function of the extension Weibull model, characterized by $\alpha > 0$ the scale parameter and $\beta, \lambda > 0$ shape parameters, is

$$F(t) = 1 - \exp \left\{ -\lambda \alpha \left(e^{\left(\frac{t}{\alpha}\right)^\beta} - 1 \right) \right\}$$

Its probability distribution function and hazard rate are defined by

$$f(t) = \lambda \beta \left(\frac{t}{\alpha} \right)^{\beta-1} \exp \left\{ \left(\frac{t}{\alpha} \right)^\beta + \lambda \alpha (1 - e^{\left(\frac{t}{\alpha}\right)^\beta}) \right\}$$

$$h(t) = \lambda \beta \left(\frac{t}{\alpha} \right)^{\beta-1} \exp \left\{ \left(\frac{t}{\alpha} \right)^\beta \right\}$$

The hazard rate $h(t)$ can be increasing when $\beta \geq 1$, and has a bathtub shape when $0 < \beta < 1$, which enable it to be used in reliability applications and study and survival analysis.

3. Nikulin-Rao-Robson statistic test

Suppose that T_1, T_2, \dots, T_n is a n -sample from a parametric family $F(t, \theta)$ and consider the problem of testing the hypothesis H_0

$$H_0 : P \{T_i \leq t \mid H_0\} = F(t, \theta), \quad t \geq 0,$$

where $\theta = (\theta_1, \theta_2, \dots, \theta_s)^T \in \Theta \subset \mathbb{R}^s$ represents the parameters vector.

The description of the Nikulin-Rao-Robson (NRR) test is given as follows.

sample data T_1, T_2, \dots, T_n are grouped in r sub-intervals I_1, I_2, \dots, I_r mutually disjoint $I_j =]a_{j-1}, a_j]$, where $j = \overline{1, r}$.

The limit intervals a_j are obtained such as

$$p_j(\theta) = \int_{a_{j-1}}^{a_j} f(t, \theta) dt = \frac{1}{r}, \quad j = 1, 2, \dots, r$$

So

$$a_j = F^{-1} \left(\frac{j}{r} \right), \quad j = \overline{1, r-1}.$$

If $v = (v_1, v_2, \dots, v_r)^T$ is the vector of the frequencies obtained by grouping data into these intervals I_j

$$v_j = \text{card} \{i : T_i \in I_j, i = 1, 2, 3, \dots, n\}$$

the NRR statistic is defined by

$$Y_n^2(\hat{\theta}_n) = X_n^2(\hat{\theta}_n) + \frac{1}{n} L^T(\hat{\theta}_n) (I_n(\hat{\theta}_n) - J(\hat{\theta}_n))^{-1} L(\hat{\theta}_n)$$

where

$$X_n^2(\theta) = X_n^T(\theta) X_n(\theta)$$

with

$$X_n(\theta) = \left(\frac{v_1 - np_1(\theta)}{\sqrt{np_1(\theta)}}, \dots, \frac{v_r - np_r(\theta)}{\sqrt{np_r(\theta)}} \right)^T$$

$J(\theta)$ is the information matrix for grouped data

$$J = J(\theta) = B(\theta)^T B(\theta)$$

with

$$B(\theta) = \left[\frac{1}{\sqrt{p_j(\theta)}} \frac{\partial p_j(\theta)}{\partial \theta_j} \right]_{r \times s}$$

$$L(\hat{\theta}_n) = \left(L_1(\hat{\theta}_n), L_2(\hat{\theta}_n), \dots, L_s(\hat{\theta}_n) \right)^T$$

with

$$L_k(\hat{\theta}_n) = \sum_{j=1}^r \frac{v_j}{p_j} \frac{\partial p_j(\hat{\theta}_n)}{\partial \theta_k}, \quad k = 1, \dots, s$$

$I_n(\hat{\theta}_n)$ represents the Fisher information matrix and $\hat{\theta}_n$ the maximum likelihood estimator of the unknown parameter vector θ . The statistic Y_n^2 follows asymptotically chi-squared distribution χ_{r-1}^2 with $(r - 1)$ degrees of freedom. Based on maximum likelihood estimators on non-grouped data, the NRR statistic Y^2 recover information lost while grouping data. An overview on chi-squared tests applications is given in Voinov et al. (2013).

4. Nikulin-Rao-Robson statistic for extension Weibull distribution

Consider a sample $T = (T_1, T_2, \dots, T_n)^T$. To verify if these data fit the extension Weibull distribution, $P(T_i \leq t) = F_{EW}(t, \theta)$, with unknown parameters $\theta = (\alpha, \beta, \lambda)^T$, we construct a modified chi-squared goodness-of-fit test by adapting the NRR statistic developed in the previous section.

Data are grouped into I_j intervals with limits a_j such as

$$\hat{a}_j = F_{EW}^{-1}(j)$$

so,

$$\hat{a}_j = \left(\alpha^\beta \ln \left(1 - \frac{\ln \left(1 - \frac{j}{r} \right)}{\lambda \alpha} \right) \right)^{\frac{1}{\beta}}$$

In order to provide the formula of the statistic Y^2

$$Y_n^2(\hat{\theta}_n) = X_n^2(\hat{\theta}_n) + \frac{1}{n} L^T(\hat{\theta}_n) (I_n(\hat{\theta}_n) - J(\hat{\theta}_n))^{-1} L(\hat{\theta}_n)$$

we calculate the statistic $X_n^2(\theta_n)$ for the extension Weibull distribution which is deduced from its reliability function $S(t)$ because

$$p_j = F_{EW}(a_j) - F_{EW}(a_{j-1}) = S(a_{j-1}) - S(a_j)$$

therefore

$$X_n^2(\theta_n) = \frac{1}{n} \sum_{j=1}^r \frac{(v_j - np_j)^2}{p_j} = \frac{1}{n} \sum_{j=1}^r \frac{(v_j - n[S(a_{j-1}) - S(a_j)])^2}{S(a_{j-1}) - S(a_j)}$$

with $v = (v_j)_r$ is the frequencies of grouped data into the I_j intervals chosen.

4.1. Calculation of the information matrix $J(\theta)$

The components of the estimated symmetric matrix

$$J(\hat{\theta}) = B(\hat{\theta})^T B(\hat{\theta})$$

are obtained as follows.

$$J_{11} = \sum_j \frac{1}{p_j} \left(\frac{\partial p_j}{\partial \alpha} \right)^2,$$

$$J_{22} = \sum_j \frac{1}{p_j} \left(\frac{\partial p_j}{\partial \beta} \right)^2,$$

$$J_{33} = \sum_j \frac{1}{p_j} \left(\frac{\partial p_j}{\partial \lambda} \right)^2,$$

$$J_{12} = J_{21} = \sum_j \frac{1}{p_j} \left(\frac{\partial p_j}{\partial \alpha} \right) \left(\frac{\partial p_j}{\partial \beta} \right),$$

$$J_{13} = J_{31} = \sum_j \frac{1}{p_j} \left(\frac{\partial p_j}{\partial \alpha} \right) \left(\frac{\partial p_j}{\partial \lambda} \right),$$

$$J_{23} = J_{32} = \sum_j \frac{1}{p_j} \left(\frac{\partial p_j}{\partial \beta} \right) \left(\frac{\partial p_j}{\partial \lambda} \right),$$

and the vector

$$L(\hat{\theta}_n) = \left(L_1(\hat{\theta}_n) = \sum_{j=1}^r \frac{v_j}{p_j} u_{j1}, L_2(\hat{\theta}_n) = \sum_{j=1}^r \frac{v_j}{p_j} u_{j2}, L_3(\hat{\theta}_n) = \sum_{j=1}^r \frac{v_j}{p_j} u_{j3} \right)^T$$

$$J(\theta) = B(\theta)^T B(\theta),$$

$$B(\theta) = (b_{jk}(\theta))_{r \times s}$$

$$b_{jk} = \frac{1}{\sqrt{p_j}} \frac{\partial p_j(\theta)}{\partial \theta_i}, \quad i = 1, \dots, r, \quad k = 1, \dots, s$$

Where the derivatives $(u_{jk})_{r \times s}$ of the cumulative distribution function $p_j(\theta)$ are given in the simple form

$$u_{j1} = \frac{\partial p_j(\theta)}{\partial \alpha} = S(a_{j-1}) \left[-\lambda \left(\exp \left\{ \left(\frac{a_{j-1}}{\alpha} \right)^\beta \right\} - 1 \right) + \left(\frac{a_{j-1}}{\alpha} \right) h(a_{j-1}) \right] \\ - S(a_j) \left[-\lambda \left(\exp \left\{ \left(\frac{a_j}{\alpha} \right)^\beta \right\} - 1 \right) + \left(\frac{a_j}{\alpha} \right) h(a_j) \right]$$

$$u_{j2} = \frac{\partial p_j(\theta)}{\partial \beta} = - \left(\frac{a_{j-1}}{\alpha} \right) \frac{\alpha}{\beta} h(a_{j-1}) S(a_{j-1}) \ln \left(\frac{a_{j-1}}{\alpha} \right) + \left(\frac{a_j}{\alpha} \right) \frac{\alpha}{\beta} h(a_j) S(a_j) \ln \left(\frac{a_j}{\alpha} \right)$$

$$u_{j3} = \frac{\partial p_j(\theta)}{\partial \lambda} = -\alpha \left(\exp \left\{ \left(\frac{a_{j-1}}{\alpha} \right)^\beta \right\} - 1 \right) S(a_{j-1}) + \alpha \left(\exp \left\{ \left(\frac{a_j}{\alpha} \right)^\beta \right\} - 1 \right) S(a_j)$$

where

$$S(a_j) = \exp \left\{ -\lambda \alpha \left(e^{\left(\frac{a_j}{\alpha} \right)^\beta} - 1 \right) \right\}$$

is the reliability function and $h(a_j)$ the hazard rate function of the EW distribution.

4.2. Calculation of Fisher information matrix $I(\theta)$

The elements of Fisher information matrix defined by

$$I_n(\theta)_{i,j} = -E \left[\frac{\partial^2 \ln f(t, \theta)}{\partial \theta_i \partial \theta_j} \right]$$

are necessary for the construction of Y^2 the NRR statistic. After several simplifications, we obtain all the elements of $I(\theta)$:

$$\frac{\partial^2 l(\theta)}{\partial \alpha^2} = \frac{n(\beta - 1)}{\alpha^2} + \sum_{i=1}^n \left[\frac{\left(\frac{t_i}{\alpha} \right) h(t_i) (-\beta - \beta \left(\frac{t_i}{\alpha} \right)^\beta + 1)}{\alpha} + \frac{\left(\frac{t_i}{\alpha} \right)^\beta \beta (\beta + 1)}{\alpha^2} \right]$$

$$\begin{aligned} \frac{\partial^2 l(\theta)}{\partial \beta^2} &= \left(-\frac{n}{\beta^2} \right) \\ &+ \sum_{i=1}^n \left[\left(1 + \left(\frac{t_i}{\alpha} \right)^\beta \right) \left(-\lambda \alpha \left(\frac{t_i}{\alpha} \right)^\beta \ln \left(\frac{t_i}{\alpha} \right)^2 \exp \left(\frac{t_i}{\alpha} \right)^\beta \right) + \left(\frac{t_i}{\alpha} \right)^\beta \ln \left(\frac{t_i}{\alpha} \right)^2 \right] \end{aligned}$$

$$\frac{\partial^2 l(\theta)}{\partial \lambda^2} = -\frac{n}{\lambda^2}$$

$$\frac{\partial^2 l(\theta)}{\partial \beta \partial \lambda} = \frac{\partial^2 l(\theta)}{\partial \lambda \partial \beta} = \sum_{i=1}^n \left[-\alpha \left(\frac{t_i}{\alpha} \right)^\beta \ln \left(\frac{t_i}{\alpha} \right) \exp \left(\frac{t_i}{\alpha} \right)^\beta \right]$$

$$\begin{aligned} \frac{\partial^2 l(\theta)}{\partial \alpha \partial \beta} &= \frac{\partial^2 l(\theta)}{\partial \beta \partial \alpha} = -\frac{n}{\alpha} + \sum_{i=1}^n \left[\left(1 + \left(\frac{t_i}{\alpha} \right)^\beta - \frac{1}{\beta} \right) \lambda \beta \left(\frac{t_i}{\alpha} \right)^\beta \ln \left(\frac{t_i}{\alpha} \right) \exp \left(\frac{t_i}{\alpha} \right)^\beta \right. \\ &\quad \left. + \lambda \left(\frac{t_i}{\alpha} \right)^\beta \exp \left(\frac{t_i}{\alpha} \right)^\beta - \frac{\left(\frac{t_i}{\alpha} \right)^\beta (1 + \ln \left(\frac{t_i}{\alpha} \right) \beta)}{\alpha} \right] \end{aligned}$$

$$\frac{\partial^2 l(\theta)}{\partial \alpha \partial \lambda} = \frac{\partial^2 l(\theta)}{\partial \lambda \partial \alpha} = n + \sum_{i=1}^n \left[\frac{\left(\frac{t_i}{\alpha} \right) h(t_i)}{\lambda} - \exp \left(\frac{t_i}{\alpha} \right)^\beta \right]$$

The statistic Y^2 does't depend on the parameters, so we can use the estimated information matrix $I_n(\hat{\theta}_n)$. As all the components of Y^2 the NRR statistic for extension Weibull model when the parameters are unknown, are provided, therefore Y^2 can be deduced easily.

5. Simulations and application

To evaluate the results obtained in this work, we conducted an intensive simulation study (10, 000 samples of different sizes). In order to show the effectiveness of the test proposed in this work, we applied theses results to a real data set from reliability. In this case, Anderson-Darling and Kolmogorov-Smirnov statistics are also calculated.

5.1. Maximum likelihood estimators

To demonstrate the performances of the maximum likelihood estimators, we generate from extension Weibull distribution with the parameter values $\alpha = 0.8, \beta = 2.5, \lambda = 3.5, N = 10, 000$ simulated samples (with sizes $n = 30; n = 100; n = 250; n = 500$). Using the *R* software and Barzilai-Borwein algorithm (BB) (Ravi. V 2009), average simulated values of the maximum likelihood estimators $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$ parameters and their mean square errors (MSE) are calculated and presented in Table 1.

Table 1: Maximum likelihood estimators of parameters and their mean square errors

$N = 10, 000$	$n = 30$	$n = 50$	$n = 100$	$n = 250$	$n = 500$
$\hat{\alpha}$	0.8647	0.7731	0.8153	0.7939	0.7922
MSE	0.0041	0.0007	0.0002	$3.60.e^{-05}$	$6.03.e^{-05}$
$\hat{\beta}$	2.5697	2.4659	2.3876	2.4770	2.5510
MSE	0.0048	0.0011	0.0126	0.0005	0.0026
$\hat{\lambda}$	3.8013	3.7790	3.7300	3.5985	3.5018
MSE	0.0908	0.0778	0.0529	0.01214	$3.25.e^{-06}$

The values of the MLEs are very close to the supposed values of the parameters.

5.2. NRR statistic Y_n^2

For testing the null hypothesis H_0 that a sample belongs to extension Weibull distribution, we calculate Y^2 the NRR statistic for 10, 000 simulated samples with sizes $n = 30; n = 100; n = 250; n = 500$, respectively. For different theoretical level significance ($\epsilon = 0.02, 0.05, 0.1$), we compute means of the number of non-rejection of the null hypothesis, when $Y_n^2 \leq \chi_{\epsilon}^2(r - 1)$ then, we report the results of empirical and the corresponding theoretical values in Table 2.

As It can be seen, the values of the empirical level calculated are very close with

Table 2: Empirical levels and corresponding theoretical levels ($\epsilon = 0.02, 0.05, 0.1$)

$N = 10,000$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.1$
$n = 30$	0.9870	0.9517	0.9258
$n = 50$	0.9854	0.9576	0.9054
$n = 100$	0.9803	0.9522	0.9028
$n = 250$	0.9789	0.9501	0.9002
$n = 500$	0.9756	0.9499	0.8898

those of their corresponding theoretical level. So, we conclude that the proposed test is well suited to the extension Weibull distribution.

Table 3: Life times of 50 devices from Aarset (1987).

0.1	0.2	1	1	1	1	1	2	3	6
7	11	12	18	18	18	18	18	21	32
36	40	45	46	47	50	55	60	63	63
67	67	67	67	72	75	79	82	82	83
84	84	84	85	85	85	85	85	86	86

5.3. Application

We consider the lifetimes of 50 devices provided by Aarset (1987) and studied by other authors. For testing the null hypothesis H_0 that these data (given in Table 3) are fitted by an extension Weibull distribution, we use the NRR statistic obtained in this work. Using *R* software and Barzilai-Borwein algorithm (BB) (Ravi. V 2009), we firstly calculate the maximum likelihood estimators $\hat{\alpha} = 101.0917$, $\hat{\beta} = 0.8402$, $\hat{\lambda} = 0.01388$ of the unknown parameters α, β, λ ; and the estimated Fisher information matrix

$$I(\hat{\theta}) = \begin{pmatrix} 0.0002546459 & -0.0255712941 & 6.2368342619 \\ -0.0255712941 & -94.14084510 & 2220.92706251 \\ 6.2368342619 & 2220.92706251 & -2.592642e + 05 \end{pmatrix}$$

The quantities W and $X_n^2(\hat{\theta})$ are obtained

$$W = \frac{1}{n} L^T(\theta_n) (I_n(\hat{\theta}_n) - J(\hat{\theta}_n))^{-1} L(\hat{\theta}_n) = -0.001128842 \quad ; \quad X_n^2(\hat{\theta}) = 8.8$$

We can then find the value of $Y_n^2 = 8.798871$. For level of significance $\epsilon = 0.02$, the critical value is $\chi_{0.02}^2(5 - 1) = 11.6678$ and for $\epsilon = 0.05$ the critical value is

Table 4: Results of goodness-of-fit testing.

	kolmogorov-Smirnov Test		Anderson-Darling test	
	KS	p-value	AD	p-value
Extension Weibull	0.1796	0.0794	3.0008	0.02752

$\chi_{0.05}^2(5 - 1) = 9.4877$. So, the NRR statistic Y_n^2 is inferior to critical values, this allows us to say that these data fit suitably an extension Weibull model.

On the other hand, we calculate the values of the Kolmogorov–Smirnov (KS) statistic and the Anderson–Darling (AD) statistic. The results are provided below

References

- [1] MV. Aarset, (1987), How to identify bathtub hazard rate. *IEEE Trans Reliab*, 36:106–108.
- [2] F. Castellares, A. J. Lemonte, (2015), A new generalized Weibull distribution generated by gamma random variables, *Journal of the Egyptian Mathematical Society*, Volume 23, Issue 2, Pages 382–390.
- [3] Chen Z. (2000), A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function, *Statistics and Probability Letters*, 49:155–161.
- [4] Dzhaparidze, K. O., Nikulin, M. S. (1974). On a modification of the standard statistic of Pearson, *Theory of Probability and Its Applications*, 19:851–853.
- [5] A. Gupta, B. Mukherjee, S.K. Upadhyay (2008) Weibull extension model: A Bayes study using Markov chain Monte Carlo simulation, *Reliability Engineering & System Safety*, 93, 1434–1443.
- [6] S. Nadarajah, G.M. Cordeiro, E.M.M. Ortega, (2013), The exponentiated Weibull distribution: a survey, *Stat. Papers*, 54, pp. 839–877.
- [7] McCulloch, C. E., (1985). Relationships among some chi-squared goodness of fit statistics. *Communications in Statistics—Theory and Methods*, 14:593–603.
- [8] Mirvaliev, M., (2001). An investigation of generalized chi-squared type statistics, Ph.D.Thesis, Academy of Science, Republic of Uzbekistan, Tashkent.
- [9] G.S. Mudholkar, D.K. Srivastava, (1993), Exponentiated Weibull family for analyzing bathtub failure-rate data, *IEEE Trans. Reliab.*, 42, pp. 299–302.
- [10] Rao, K. C., Robson, D. S., (1974). A Chi-square statistic for goodness-of-fit tests within the exponential family, *Communication in Statistics*, 3:1139–1153.
- [11] Ravi, V., Gilbert, P. D. (2009), BB: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *J. Statist. Software*, 32(4).

- [12] R.M., Smith, L.J., Bain, (1975) , An exponential power life-testing distribution. *Communication in Statistics*, 4:469–481.
- [13] V. Voinov, M. Nikulin, N. Balakrishnan, (2013), *Chi-Squared Goodness of Fit Tests with Applications*, Academic Press, Elsevier.
- [14] M. Xie, T.N. Goh, Y. Tang, (2004), On changing points of mean residual life and failure rate function for some generalized Weibull distributions, *Reliability Engineering and System Safety*, 84, 293–299.