Neutrosophic Sets and Systems, Vol. 49, 2022
University of New Mexico

A new Cramèr–von Mises Goodness-of-fit test under Uncertainty

Muhammad Ahsan-ul-Haq[1,*]

[1] "College of Statistical & Actuarial Sciences, University of the Punjab, Pakistan; ahsanshani36@gmail.com"

* Correspondence: ahsanshani36@gmail.com

Abstract: The Cramer-von Mises test is commonly used to determine how well-observed sample data fits a given model. The existing Cramer-von Mises test under traditional statistics is commonly used when sample data in reliability work are resolute and precise. In this paper, we introduced a Neutrosophic Cramer-von Mises (NCVM) test under neutrosophic statistics. The necessary measures and procedures are presented to perform the test. For the application purpose, we consider the real-life data sets of failure time batteries and ball bearings. It is inferred that the NCVM test is more instructive than the classical CVM test under indeterminacy.

Keywords: Cramer-von Mises; Neutrosophic Weibull; Neutrosophic Rayleigh; Goodness of fit

1. Introduction

The statistical techniques have been utilized in every practical field for modeling data sets, prediction, and forecasting purposes. The application of these modeling statistical techniques/tests is made under specific suppositions, and infringement of these assumptions could prompt deluding interpretation and dependable outcomes [1, 2]. One of the fundamental presumptions that various statistical techniques are associated with the distribution of observed data follows a specified distribution. Typically, it is expected that the obtained information follows the normal distribution. In some viable situations, data sets don't need to be normally distributed. Therefore, researchers planned a few tests to valuation some hypotheses about the distribution of the information being scrutinized. Various tests, for the most part, known as "goodness-of-fit" are employed to evaluate whether an example of observations can be considered as a sample from a given distribution. The frequently utilized goodness-of-fit tests are; Kolmogorov–Smirnov [3, 4], Anderson–Darling [5, 6], Pearson's chi-square [7], Cramèr–von Mises [8, 9], Shapiro–Wilk [10], Jarque–Bera [11, 12], D'Agostino–Pearson [13] and Lilliefors [14].

The Cramèr–von Mises (CVM) test is a criterion utilized for the evaluation of the goodness of fit. The CVM test is the generalization of the Anderson-Darling test. The CVM test is the assessment of the minimum distance between hypothetical and sample probability distribution. Stephens [16] utilized the CVM goodness-of-fit test based on the experimental distribution function considering normal and exponential distributions. It was found that the CVM test appears more powerful test than chi-square. Al-zahrani [17] introduced the CVM goodness of fit test for Topp-Leone distribution.

The classical Cramèr–von Mises test can't be applied when the sample observations are neutrosophic numbers. So the principle motivation behind this study is to present another Cramèr–von Mises goodness-of-fit test within the sight of indeterminacy. We will introduce the technique to fit the neutrosophic Weibull and Rayleigh distributions on the lifetime of batteries and ball-bearings data sets.

2. Preliminaries

Suppose that $X_N = X_L + X_U I_N; I_N \in [I_L, I_U]$ denotes the neutrosophic number (NN) that follows the neutrosophic Weibull distribution with neutrosophic shape parameter $\beta_N = \beta_L + \beta_U I_N; I_N \in [I_L, I_U]$ and the neutrosophic scale parameter $\alpha_N = \alpha_L + \alpha_U I_N; I_N \in [I_L, I_U]$. Here X_L is the determinate part and $X_U I_N$ is the indeterminate part with an indeterminacy constant $I_N \in [I_L, I_U]$. Note that neutrosophic Weibull random variable X_N reduces to classical Weibull distribution when $I_N = 0$.

Neutrosophic statistics is the augmentation of classical statistics. This field acquires significance because of dealing with the data sets of values more specifically an interval, for more detail reader can consult the following references [18-20]. For the presentation of the neutrosophic environment, normally a subsequent "N" is utilized such as X_N.

3. Neutrosophic Weibull distribution

The neutrosophic Weibull (NW) distribution was introduced by [21]. The cumulative distribution function of NW distribution is

$$F(X_N) = 1 - \exp\left\{ -\left(\frac{X_N}{\alpha_N} \right)^{\beta_N} \right\}, \quad X_N > 0. \tag{1}$$

where $\alpha_N \in [\alpha_L, \alpha_U], \beta_N \in [\beta_L, \beta_U]$

4. Neutrosophic Rayleigh distribution

The Neutrosophic Rayleigh (NR) distribution was introduced by [22]. The cumulative distribution function of NR distribution is

$$F(X_N) = 1 - \exp\left\{ -\frac{1}{2}\left(\frac{X_N}{\alpha_N} \right)^{2} \right\}, \quad X_N > 0. \tag{2}$$

where $\alpha_N \in [\alpha_L, \alpha_U]$

5. Neutrosophic Cramèr-von-Mises

The CVM test is a non-parametric test of the hypothesis. It is utilized to test whether an example comes from a particular distribution when the observed data set is precise or determined. When the data set is imprecise then the exiting CVM test cannot be used to test the goodness of fit due to indeterminacy in the data. We modify the classical CVM test and proposed the neutrosophic Cramer-von-Mises (NCVM) test for the data having neutrosophic numbers. The proposed test will bring about terms of indeterminacy interval which will be more successful when compared to the classical CVM test. The assumption for the NCVM test are

- The data consists of imprecise observations.
- The observations in the interval are mutually independent.

"Suppose $X_{1N}, X_{2N}, X_{3N}, ..., X_{nN}$ is a neutrosophic random sample from a neutrosophic population having a neutrosophic cumulative distribution function, say $F(X_N)$". Then the NCVM is given by

$$CVM_N = \frac{1}{12n_N} + \sum_{i=1}^{n_N} \left\{ \frac{2i-1}{2n_N} - \left[1 - \exp\{-M_N(i)\} \right] \right\}^2; \; CVM_N \in [CVM_L, CVM_U] \tag{3}$$

where $n_N \in [n_L, n_U]$ are the neutrosophic random samples and $M_N(i) = \left(\frac{X_N(i)}{\alpha_N}\right)^{\beta_N}$.

6. Applications of Neutrosophic Cramèr-von-Mises Test

This section examines the use of the newly introduced test. For the application purposed we consider two real-life data sets.

6.1. Application on data Set I (lifetime in 100 h of 23 batteries

The first data set is regarding the lifetime of batteries also utilized by [21]. The lifetime in 100 h of 23 batteries is given in Table 1.

Table 1. The lifetime of batteries

Sr. No	X_N	Sr. No	X_N	Sr. No	X_N	Sr. No	X_N
1	[2.9,3.99]	7	[12.65,17.4]	13	[17.4,23.93]	19	[26.07,35.84]
2	[5.24,7.2]	8	[13.24,18.21]	14	[17.8,24.48]	20	[30.29,41.65]
3	[6.56,9.02]	9	[13.67,18.79]	15	[19.01,26.14]	21	[43.97,60.46]
4	[7.14,9.82]	10	[13.88,19.09]	16	[19.34,26.59]	22	[48.09,66.13]
5	[11.6,15.96]	11	[15.64,21.51]	17	[23.13,31.81]	23	[73.48,98.04]
6	[12.14,16.69]	12	[17.05,23.45]	18	[23.34,32.09]		

The mechanical investigators are intrigued to test either the given informational collection follows Weibull distribution or not. It is not difficult to take note that the data observations are given in indeterminacy intervals instead of the specific observation. So the classical CVM test is not appropriate. Therefore, we will utilize the option NCVM test proposed in section 5 is used for these neutrosophic numbers.

Assume that we need to test the following hypothesis:

H_0=The sample observation follows to neutrosophic Weibull distribution.

H_1=The distribution of sample observation is not neutrosophic Weibull distribution.

The numerical computations are listed in Table 2. The parameters of Neutrosophic Weibull distribution are estimated using the maximum likelihood estimation method. The estimated values are $\hat{\alpha}_N \in [22.936, 31.427]$ and $\hat{\beta}_N \in [1.465, 1.481]$. The test statistic values of the proposed CVM_N test for the considered lifetime of batteries data are shown as

$$CVM_N = \frac{1}{12[n_L, n_U]} + \sum_{i=1}^{n_N} \left\{ \frac{2i-1}{2[n_L, n_U]} - \left[1 - \exp\left\{-M_N(i)\right\}\right] \right\}^2$$

$$CVM_N = \frac{1}{12[23, 23]} + \sum_{i=1}^{n_N} \left\{ \frac{2i-1}{2[23, 23]} - \left[1 - \exp\left\{-M_N(i)\right\}\right] \right\}^2$$

$CVM_N \in [0.1112, 0.3617]$

Table 2. The necessary calculation of the NCVM test for the first data

ith	X_N	$M_N(i)$	$F_N(X_N)$	ith-term
1	[2.9,3.99]	[0.0484, 0.0772]	[0.0472, 0.0743]	[0.0006, 0.0028]
2	[5.24,7.2]	[0.1150, 0.1832]	[0.1087, 0.1674]	[0.0019, 0.0104]
3	[6.56,9.02]	[0.1599, 0.2549]	[0.1477, 0.2250]	[0.0015, 0.0135]
4	[7.14,9.82]	[0.1810, 0.2887]	[0.1656, 0.2507]	[0.0002, 0.0097]
5	[11.6,15.96]	[0.3684, 0.5879]	[0.3082, 0.4445]	[0.0127, 0.0619]
6	[12.14,16.69]	[0.3938, 0.6277]	[0.3255, 0.4662]	[0.0075, 0.0516]
7	[12.65,17.4]	[0.4183, 0.6672]	[0.3418, 0.4869]	[0.0035, 0.0417]
8	[13.24,18.21]	[0.4472, 0.7132]	[0.3606, 0.5099]	[0.0012, 0.0338]
9	[13.67,18.79]	[0.4686, 0.7467]	[0.3741, 0.5261]	[0.0000, 0.0245]
10	[13.88,19.09]	[0.4792, 0.7643]	[0.3807, 0.5343]	[0.0010, 0.0147]
11	[15.64,21.51]	[0.5708, 0.9103]	[0.4349, 0.5976]	[0.0005, 0.0199]
12	[17.05,23.45]	[0.6477, 1.0330]	[0.4767, 0.6441]	[0.0005, 0.0208]
13	[17.4,23.93]	[0.6672, 1.0641]	[0.4869, 0.6550]	[0.0032, 0.0124]
14	[17.8,24.48]	[0.6898, 1.1001]	[0.4983, 0.6672]	[0.0079, 0.0064]
15	[19.01,26.14]	[0.7596, 1.2111]	[0.5321, 0.7021]	[0.0097, 0.0051]
16	[19.34,26.59]	[0.7790, 1.2418]	[0.5411, 0.7111]	[0.0176, 0.0014]
17	[23.13,31.81]	[1.0124, 1.6146]	[0.6367, 0.8010]	[0.0065, 0.0070]
18	[23.34,32.09]	[1.0259, 1.6354]	[0.6415, 0.8051]	[0.0142, 0.0020]
19	[26.07,35.84]	[1.2063, 1.9228]	[0.7007, 0.8538]	[0.0107, 0.0024]
20	[30.29,41.65]	[1.5028, 2.3961]	[0.7775, 0.9089]	[0.0049, 0.0037]
21	[43.97,60.46]	[2.5941, 4.1359]	[0.9253, 0.9840]	[0.0012, 0.0086]
22	[48.09,66.13]	[2.9577, 4.7161]	[0.9481, 0.9911]	[0.0002, 0.0032]
23	[73.48,98.04]	[5.5034, 8.3957]	[0.9959, 0.9998]	[0.0003, 0.0005]

6.2. Application on data Set I (lifetime of ball-bearings)

The second data set is about the service life of ball-bearing data [23]. The second data set is listed in below Table 3.

Table 3. The failure life of 21 ball bearings

Sr. No	X_N	Sr. No	X_N	Sr. No	X_N	Sr. No	X_N
1	[0.70, 0.81]	7	[0.85, 1.03]	13	[0.23, 0.74]	19	[0.34, 1.11]
2	[0.63, 0.81]	8	[0.67, 0.73]	14	[0.76, 0.95]	20	[0.07, 1.17]
3	[0.35, 0.41]	9	[0.96, 1.04]	15	[0.80, 0.86]	21	[0.41, 0.44]
4	[0.70, 0.72]	10	[1.07, 1.26]	16	[1.06, 1.21]		
5	[1.12, 1.43]	11	[0.95, 1.35]	17	[0.60, 0.70]		
6	[0.47, 1.39]	12	[0.82, 1.02]	18	[0.85, 1.01]		

For the second application, we utilized the ball-bearing failure time data test either it follows Neutrosophic Raleigh or not.

Assume that we need to test the following hypothesis:

H_0 =The sample observation follows to neutrosophic Raleigh distribution.
H_1 =The distribution of sample observation is not neutrosophic Raleigh distribution.

The maximum likelihood estimates for NR distribution are $\hat{\alpha}_N \in [0.52447, 0.70812]$. The numerical computation of NCVM is presented in Table 4. The test statistic values of the proposed CVM_N test for the considered lifetime of batteries data are shown as

$$CVM_N = \frac{1}{12[21,21]} + \sum_{i=1}^{n_N} \left\{ \frac{2i-1}{2[21,21]} - \left[1 - \exp\left\{-M_N(i)\right\}\right] \right\}^2$$

$CVM_N \in [0.1564, 0.3550]$

Table 4. The necessary calculation of the NCVM test for second data

ith	X_N	$M_N(i)$	$F_N(X_N)$	ith-term
1	[0.70, 0.81]	[0.0178, 0.3352]	[0.0089, 0.1543]	[0.0002, 0.0170]
2	[0.63, 0.81]	[0.1923, 03861]	[0.0917, 0.1756]	[0.0004, 0.0108]
3	[0.35, 0.41]	[0.4203, 0.9772]	[0.1895, 0.3865]	[0.0050, 0.0715]
4	[0.70, 0.72]	[0.4453, 1.0338]	[0.1996, 0.4036]	[0.0011, 0.0562]
5	[1.12, 1.43]	[0.6111, 1.0628]	[0.2633, 0.4122]	[0.0024, 0.0392]
6	[0.47, 1.39]	[0.8031, 1.0921]	[0.3307, 0.4208]	[0.0047, 0.0252]
7	[0.85, 1.03]	[1.3088, 1.3084]	[0.4802, 0.4802]	[0.0291, 0.0291]
8	[0.67, 0.73]	[1.4429, 1.3084]	[0.5140, 0.4802]	[0.0246, 0.0151]
9	[0.96, 1.04]	[1.6320, 1.4750]	[0.5578, 0.5217]	[0.0234, 0.0137]
10	[1.07, 1.26]	[1.7814, 1.7998]	[0.5896, 0.5934]	[0.0188, 0.0199]
11	[0.95, 1.35]	[1.7814, 2.0344]	[0.5896, 0.6384]	[0.0080, 0.0192]
12	[0.82, 1.02]	[2.0998, 2.0748]	[0.6500, 0.6456]	[0.0105, 0.0096]
13	[0.23, 0.74]	[2.3267, 2.1157]	[0.6876, 0.6528]	[0.0085, 0.0033]
14	[0.76, 0.95]	[2.4445, 2.1570]	[0.7054, 0.6599]	[0.0039, 0.0003]
15	[0.80, 0.86]	[2.6266, 2.4572]	[0.7311, 0.7073]	[0.0016, 0.0003]
16	[1.06, 1.21]	[2.6266, 2.7300]	[0.7311, 0.7446]	[0.0000, 0.0000]
17	[0.60, 0.70]	[3.2810, 2.9198]	[0.8061, 0.7677]	[0.0004, 0.0003]
18	[0.85, 1.01]	[3.3504, 3.1661]	[0.8127, 0.7947]	[0.0004, 0.0015]
19	[0.34, 1.11]	[4.0848, 3.6346]	[0.8703, 0.8375]	[0.0001, 0.0019]
20	[0.07, 1.17]	[4.1622, 3.8531]	[0.8752, 0.8544]	[0.0028, 0.0055]
21	[0.41, 0.44]	[4.5603, 4.0781]	[0.8977, 0.8698]	[0.0062, 0.0113]

7. Discussion and Conclusion

In this section, we will compare the efficiency of the proposed NCVM test under the neutrosophic environment with the existing classical CVM test. The proposed test is more efficient when dealing with data having imprecise observation or indeterminacy as the proposed method provides results in the form of indeterminacy. For comparison purposes, we use the same data set for classical CVM. Note that the data given in Tables 1 and 3 have a determinate part as well as the indeterminate part. The determinate part will be used for the existing CVM test and the same data set is used for the NCVM test. The critical value at 1% and 5% are $CVM_{1\%,23} = 0.267$ and $CVM_{5\%,23} = 0.187$, respectively. From Table 2 it is unmistakably that the proposed test gives the results in the form of indeterminacy interval rather than determinate part only. Utilizing Equation (3) the value of statistic as indeterminacy interval can be written as $0.1112 + 0.3617I; I_N \in [0, 0.6926]$. Note that the proposed test gives a decent portion of indeterminacy. At a 1% level of significance, the probability of accepting the true null hypothesis is 0.99, the probability of rejecting the true null hypothesis is 0.01 and the probability of indeterminacy is 0.69. For instance, $CVM = 0.3617$ is the value of classical CVM and $CVM_U = 0.3617$ gives the indeterminate part under uncertainty. By contrasting with crucial values, we can see that the determinant component of the information follows the Weibull distribution, but the uncertain part does not. Similarly, for the failure of ball bearings data the value of statistics as indeterminacy interval can be written as $0.1564 + 0.3550I; I_N \in [0, 0.5994]$. By contrasting with critical values, we note that the determinant part follows the Rayleigh distribution, yet the uncertain part of the information doesn't follow the Rayleigh distribution.

It is concluded that the proposed NCVM test under neutrosophic statistics provides information about the measure of indeterminacy, but the classical CVM test does not. Furthermore, the existing test delivers accurate statistics values, which are not necessary for uncertainty. As a result, under neutrosophic statistics, the proposed NCVM goodness-of-fit test is particularly efficacious when used under uncertainty.

Funding: "This research received no external funding".

Conflicts of Interest: "The authors declare no conflict of interest."

References

1. Nimon, K. F. Statistical assumptions of substantive analyses across the general linear model: a mini-review. *Frontiers in Psychology* **2012**, *3*, 322. doi:10.3389/fpsyg.2012.00322.
2. Hoekstra, R.; Kiers, H.; Johnson, A. Are Assumptions of Well-Known Statistical Techniques Checked, and Why? *Front. Psychol.* **2012**, *3*, 137, doi:10.3389/fpsyg.2012.00137.
3. Kolmogorov, A. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari* **1933**, *4*, 83–91.
4. Smirnov, N. Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Stat.* **1948**, *19*, 279–281, doi:10.1214/aoms/1177730256.
5. Anderson, T.W.; Darling, D.A. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Ann. Math. Stat.* **1952**, *23*, 193–212, doi:10.1214/aoms/1177729437.
6. Anderson, T.W.; Darling, D.A. A Test of Goodness-of-Fit. *J. Am. Stat. Assoc.* **1954**, *49*, 765–769, doi:10.2307/2281537.
7. Pearson, K. Contribution to the mathematical theory of evolution. II. Skew variation in homogenous material. *Philos. Trans. R. Soc. Lond.* **1895**, *91*, 343–414, doi:10.1098/rsta.1895.0010.
8. Cramér, H. On the composition of elementary errors. *Scand. Actuar. J.* **1928**, *1*, 13–74, doi:10.1080/03461238.1928.10416862.
9. Von Mises, R.E. *Wahrscheinlichkeit, Statistik und Wahrheit*; Julius Springer: Berlin, Germany, 1928.
10. Shapiro, S.S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*, 591–611, doi:10.1093/biomet/52.3-4.591.
11. Jarque, C.M.; Bera, A.K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econ. Lett.* **1980**, *6*, 255–259, doi:10.1016/0165-1765(80)90024-5.

12. Jarque, C.M.; Bera, A.K. A test for normality of observations and regression residuals. *Int. Stat. Rev.* **1987**, *55*, 163–172, doi:10.2307/1403192.
13. D'Agostino, R.B.; Belanger, A.; D'Agostino, R.B., Jr. A suggestion for using powerful and informative tests of normality. *Am. Stat.* **1990**, *44*, 316–321, doi:10.2307/2684359.
14. Lilliefors, H.W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* **1967**, *62*, 399–402, doi:10.2307/2283970.
15. Van Soest, J. Some experimental results concerning tests of normality. *Stat. Neerl.* **1967**, *21*, 91–97, doi:10.1111/j.1467-9574.1967.tb00548.x.
16. Stephens, Michael A. "Use of the Kolmogorov–Smirnov, Cramer–Von Mises and related statistics without extensive tables." *Journal of the Royal Statistical Society: Series B (Methodological)* **1970**. 115-122.
17. Al-zahrani, B. *Goodness-of-Fit for the Topp-Leone Distribution*. **2012.** *6*(128), 6355–6363.
18. Smarandache, F. *Definitions derived from neutrosophic*, **2003.** Infinite Study.
19. Smarandache, F. Introduction to Neutrosophic Statistics, Sitech and Education Publisher, Craiova. *Romania-Educational Publisher, Columbus, Ohio, **2014.** USA*, *123*.
20. Zeina, M. B., & Hatip, A. Neutrosophic Random Variables. *Neutrosophic Sets and Systems*, **2021**. *39*(1), 4.
21. Aslam M. A new goodness of fit test in the presence of uncertain parameters. Complex & Intelligent Systems. **2021** Feb;7(1):359-65.
22. Khan, Z., Gulistan, M., Kausar, N. & Park, C. Neutrosophic Rayleigh Model with Some Basic Characteristics and Engineering Applications. IEEE Access. **2021**. 9, 71277–71283.
23. Lieblein, J. & Zelen, M. Statistical investigation of the fatigue life of deep-groove ball bearings. J. Res. Natl. Bur. Stand. **1956**. 57, 273–316.

Received: Dec. 3, 2021. Accepted: April 4, 2022.