A New Goodness-of-Fit Test: Free Chi-Square (FCS)

Ozge TEZEL [iD] , Bugra Kaan TIRYAKI [iD] , Eda OZKUL* [iD] , Orhan KESEMEN [iD]

Department of Statistics and Computer Sciences, Faculty of Science, Karadeniz Technical University, Trabzon, Turkey

Highlights
• This paper presents a new goodness-of-fit test.
• The proposed method is binning free and distribution free test.
• It is more sensitivity, easy to use and fast.

Article Info	Abstract
	This paper presents a new goodness-of-fit technique for testing the assumption of univariate distributions which is based on the theoretical distribution function of the hypothesized distribution. The existing methods are examined in two different categories: binning and binning-free. The most widely known binning test is the Chi-square test. The Kolmogorov-Smirnov, the Cramer-von Mises and the Anderson-Darling goodness-of-fit tests come to the forefront as the binning-free tests. When tests are evaluated in terms of distributions, it is examined in two different classes: the not distribution-free tests and the distribution-free tests. The desired goodness-of-fit test method for a researcher should be binning-free, distribution-free, more sensitivity, easy to use and fast. In this study, a test method is proposed which provides almost all the options that a researcher would want. The Monte-Carlo simulation methods are used to demonstrate the success of the proposed method. In these simulations, the normality test was applied for symmetric distributions whereas the lognormality test was applied for non-symmetric distributions. The proposed test method has demonstrated superiority in many aspects compared to other selected test methods on both simulations and three different real-life datasets.

1. INTRODUCTION

The methods that test whether a randomly selected sample comes from the claimed theoretical distribution, are called goodness-of-fit tests. The goodness-of-fit test is one of the most important issues of the inferential statistics. The normality test is considered as an assumption in many statistical procedures [1,2]. Many goodness-of-fit tests have been developed for the uniform, the exponential, the Laplace and the Cauchy distributions apart from the normality tests [3-6]. There are discussed many popular techniques in the literature for goodness-of-fit. These techniques can be listed as the binning test, tests based on the empirical distribution function, tests based on the characteristic function, tests based on the moment generating function, tests based on the regression, correlation and moments [7].

The goodness-of-fit tests were originally developed as the chi-square (χ^2) test by Pearson in 1900 [8]. The χ^2 goodness-of-fit test has become a widely used classical technique because it is simple, and its distribution is known and still maintains its popularity. On the other hand, the disadvantages of the χ^2 test are that it does not have enough sample numbers, the number of binning is chosen by the researcher and the expected frequencies in each bin should be at least 5. In this paper, the χ^2 test was called binning chi-square (BCS) to avoid the confusions.

There are many goodness-of-fit tests based on the empirical distribution function are developed. The basic principle of these tests is to determine how well the cumulative distribution of the sample fits the theoretical cumulative distribution in the hypothesis. The goodness-of-fit tests based on the empirical distribution

*Corresponding author, e-mail: eda.ozkul.gs@gmail.com.tr

function give more consistent results than χ^2 goodness-of-fit test. This is because they directly use the observed values [9]. Kolmogorov-Smirnov, Cramér-Von Mises and Anderson Darling goodness-of-fit tests are included in this category.

Kolmogorov-Smirnov (KS) test is the most popular goodness-of-fit test. It was developed by Kolmogorov and Smirnov in 1930s [10]. The KS test statistic is based on the difference between the experimental distribution function using observational values and the theoretical distribution function in the hypothesis. The KS test is preferred to the chi-square test in the case of small sample size.

Cramer-von Mises (CVM) test was developed by Cramer and von-Mises [11,12]. This test method gives better results than the KS test when different alternative hypotheses are used. It is based on the square of the difference between the empirical distribution function and the theoretical distribution function in the hypothesis.

Anderson and Darling [13] proposed the Anderson-Darling goodness-of-fit test (AD) by adapting the CVM test. Although the AD test is a stronger test in some cases than the KS test, it calculates the critical value depending on the distribution.

The most important binning test and the empirical distribution-based tests were given in Section 2. A new goodness-of-fit test: Free Chi-Square (FCS) was introduced in Section 3. These tests were applied to some alternative hypotheses and their performances were compared in Section 4. In Section 5, the proposed method was performed on the three real-life examples.

2. THE GOODNESS-OF-FIT TESTS

The goodness-of-fit tests are used to determine whether a given sample comes from a population with a defined theoretical distribution. In other words, let the observed $x_1, x_2, \ldots, x_n$ be a random sample come from a population with a continuous $F(x)$ distribution. This unknown $F(x)$ distribution function should be confirmed with the help of hypothesis by using a goodness-of-fit test. The null hypothesis to be used for this reason is given as follows,

$H_0: F(x) = F_0(x; \theta)$ or $H_0: The\ data\ follow\ a\ specified\ distribution.$

Where $F_0(x, \theta)$ is the specified distribution function with θ parameter. The alternative hypothesis is given as follows,

$H_1: F(x) \neq F_0(x; \theta)$ or $H_1: The\ data\ do\ not\ follow\ the\ specified\ distribution.$

In order to show whether the data come from the specified distribution, it is first necessary to calculate the test statistics, which its distribution and the critical value are known, according to the null hypothesis. The most basic feature that distinguishes the goodness-of-fit test from each other is the calculation of the test statistics in a different way. Despite the existence of many goodness-of-fit tests in the literature, the BCS, the CVM, the KS and the AD goodness-of-fit tests were used to compare the performance of the tests in this study.

2.1. The Binning Chi-Square (BCS) Goodness-of-Fit Test

The BCS test is one of the most used goodness-of-fit tests. The chi-square goodness-of-fit test which is binning-based method, has taken part in almost all the basic statistical books because it is easy to apply and understand. If the examined data is not binning data, this data can be obtained by calculating a histogram. The BCS test is based on the inconsistency between observed and expected frequencies.

The BCS investigates whether there is a statistically significant difference between observed and expected frequencies. This test statistic is obtained as,

$$\chi^2 = \sum_{j=1}^{m} \frac{\left(O_j - E_j\right)^2}{E_j}$$

(1)

where O_j is the observed frequency for bin j and E_j is the expected frequency for bin j. The obtained test statistic follows, approximately, a chi-square distribution with $(m-1)$ degrees of freedom. The calculated test statistic is compared to the critical value from the chi-square critical value with $(m-1)$ degrees of freedom and $(1-\alpha)$ confidence level so that it can be decided as a result of the test. If the calculated $\chi^2 <$ critical $\chi^2_{m-1,1-\alpha}$, then we cannot reject the null hypothesis.

In order to apply the BCS goodness-of-fit test properly, the expected frequencies in each class should be at least 5. If it is not, the classes must be combined with other classes until assumption is satisfied. The disadvantage of the BCS test is that it requires enough sample size. Another disadvantage is that the number of bins is determined by the researcher.

2.2. The Cramer-von Mises (CVM) Goodness-of-Fit Test

The CVM test is based on the square of the difference between the empirical distribution function and the theoretical distribution function in the hypothesis [14]. The CVM test statistic can be given as follows as a result of many updates [15],

$$W^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[F_0\left(x_{(i)}\right) - \frac{2i-1}{2n}\right]^2$$

(2)

where $F_0(x)$ is the specified distribution function in the hypothesis, $(2i-1)/2n$ is the empirical distribution function and $x_{(i)}$ denotes the ordered sample.

2.3. The Kolmogorov-Smirnov (KS) Goodness-of-Fit Test

The KS test is based on investigation of the two cumulative distribution functions [15]. The first of this is the cumulative distribution function $F_0(x)$ given in null hypothesis. The second is the empirical distribution function $F_n(x)$ calculated from the sample. The KS test statistic is defined by

$$D = \max_{1 \le i \le n} \left|F_0\left(x_{(i)}\right) - F_n(x_{(i)})\right|.$$

(3)

The empirical distribution function can also be calculated in the form of $F_n\left(x_{(i)}\right) = i/n$ as well as $F_n\left(x_{(i)}\right) = (i-1)/n$. Or it can be calculated by taking the absolute maximum of them [16].

2.4. The Anderson-Darling (AD) Goodness-of-Fit Test

Another the empirical distribution-based goodness-of-fit test is the Anderson-Darling (AD) test. The AD test pays more attention to the tails of distribution than the CVM and KS tests [17]. Anderson and Darling [18] have proposed the AD goodness-of-fit test by adapting the Cramer-von Mises and Kolmogorov-Smirnov tests. The AD test uses the specified hypothesis distribution in the calculation of the critical values. Although this test is a more sensitive method, the critical values for each distribution must be calculated. The AD test statistic is obtained as,

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\log_e\left(F_0(x_{(i)})\right) + \log_e\left(1 - F_0\left(x_{n-(i)+1}\right)\right)\right]$$

(4)

where $x_{(i)}$ denotes the ordered sample and $F_0(x)$ is the specified distribution function in the hypothesis.

3. THE FREE CHI-SQUARE (FCS) GOODNESS-OF-FIT TEST

In statistics, a random variable from any given continuous distribution can be transformed into a random variable from the uniform distribution defined on the interval $[0,1)$. Let, X is the continuous random variable and we will assume its cumulative distribution function is F_X $(X \sim F_X(x))$. In the circumstances, U obtained which is a random variable with uniform distribution from the equation $U = F_X(X)$ [19].

In the goodness-of-fit tests, the null hypothesis and the alternative hypothesis are given as follows.

H_0: $F(x) = F_0(x)$
H_1: $F(x) \neq F_0(x)$

Suppose that x random values come from a population with the $F_0(x)$ distribution function, the independent U_i values of the theoretical distribution function in Equation (5) will be defined as $U_i \sim Uniform(0,1)$.

$$U_i = F_0(X_i) \tag{5}$$

The variable V is transformed to the uniform distribution in $(0, 2\pi)$ range by using Equation (6). In other words, it can be defined as $V_i \sim Uniform(0, 2\pi)$.

$$V_i = 2\pi U_i \tag{6}$$

If uniformly distributed random values are multiplied by 2π and scattered over the unit circle, it is assumed that each data is a unit vector.

The proposed test statistic is obtained by the resultant vector calculated by the vertical and horizontal components of the unit vectors. If we want to calculate the resultant length, the mean horizontal component of it can be computed by using Equation (7),

$$\bar{C} = \frac{1}{n} \sum_{i=1}^{n} \cos(V_i) \tag{7}$$

and the mean vertical component of it can be computed by using Equation (8),

$$\bar{S} = \frac{1}{n} \sum_{i=1}^{n} \sin(V_i). \tag{8}$$

Here $(V_1, V_2, \dots, V_n) \sim U(0, 2\pi)$, $c = \cos V$ and $s = \sin V$. Then we have

$$E(c) = \int_0^{2\pi} \cos V \frac{1}{2\pi} \, dV = 0,$$
$$E(s) = \int_0^{2\pi} \sin V \frac{1}{2\pi} \, dV = 0$$

and similarly,

$$E(c^2) = E(s^2) = \frac{1}{2},$$
$$E(c \cdot s) = 0.$$

In this case, the average horizontal and vertical components $(\bar{C}, \bar{S})$ calculated by taking the average of the sine and cosine of the random variable converted to uniform distribution in the $(0, 2\pi)$ converges to $N(\mu, \Sigma)$ with

$$\vec{\mu} = 0 \quad \text{and} \quad \Sigma = \begin{bmatrix} \dfrac{1}{2n} & 0 \\ 0 & \dfrac{1}{2n} \end{bmatrix}.$$

If the random variables $(\bar{C}, \bar{S})$ are standardized, $(\sqrt{2n}\bar{C})$ and $(\sqrt{2n}\bar{S})$ converges to $N(0,1)$ [20]. Then we have

$$\sqrt{2n}\,\bar{C} = \sqrt{2n}\frac{1}{n}\sum_{i=1}^{n}\cos(V_i)$$
$$= \sqrt{\frac{2}{n}}\sum_{i=1}^{n}\cos\big(2\pi F_0(x_i)\big),$$

and similarly,

$$\sqrt{2n}\,\bar{S} = \sqrt{2n}\frac{1}{n}\sum_{i=1}^{n}\sin(V_i)$$
$$= \sqrt{\frac{2}{n}}\sum_{i=1}^{n}\sin\big(2\pi F_0(x_i)\big).$$

If the sample size is enough large, each horizontal and vertical component random variables has standard normal distribution according to the central limit theorem. In the simulation study for the determination of the sample size, random values $(\bar{C}, \bar{S})$ of the random samples with different size generated from uniform distribution in the range of $[0,2\pi]$ are computed. A dataset was obtained by generating 100 of each $\bar{C}$ and $\bar{S}$ random values and tested by the Kolmogorov-Smirnov goodness-of-fit test whether or not it came from the standard normal distribution. Type I error was obtained by 10,000 runs and it is shown in Figure 1.
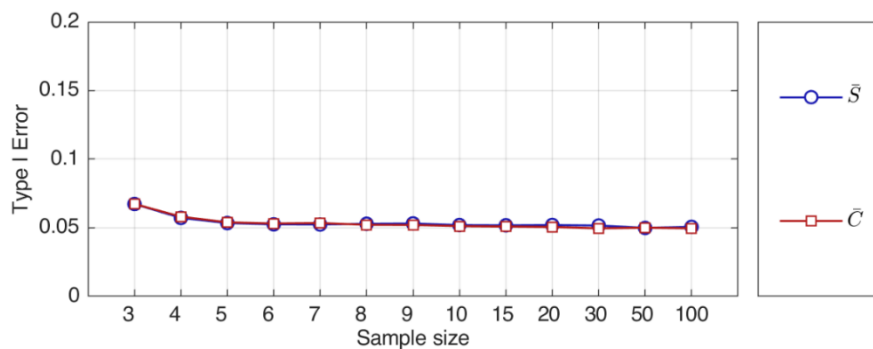
Figure 1. *Type I error of the random variables $\bar{C}$ and $\bar{S}$ obtained by KS Normality test with $\alpha=0.05$*

In Figure 1, the increase in Type I error at the small sample size of 4 indicates that the normality assumption is invalid. This assumption indicates that the proposed method is available for at least a sample size of 4. The standard mean resultant-squared vector length of the circular data shown in vector form on the unit circle, is calculated as follows

$$\bar{R}^2 = \bar{C}^2 + \bar{S}^2. \tag{9}$$

Since $(\sqrt{2n}\bar{C})$ and $(\sqrt{2n}\bar{S})$ converge in distribution to a $N(0,1)$, $(\sqrt{2n}\bar{C})^2$ and $(\sqrt{2n}\bar{S})^2$ converge to χ_1^2 each and their sum,

$$R^2 = \frac{2}{n}\left[\left(\sum_{i=1}^{n}\cos\big(2\pi F_0(X_i)\big)\right)^2 + \left(\sum_{i=1}^{n}\sin(2\pi F_0(X_i)\big)\right)^2\right] \sim \chi_2^2.$$

(10)

The resultant vector is accepted as the test statistic of the proposed method. The resultant vector is the sum of squares of 2 random variables from standard normal distribution, and thus it has the chi-square distribution with the 2 degrees of freedom.

The graphical representation of the proposed method is given in Figure 2.
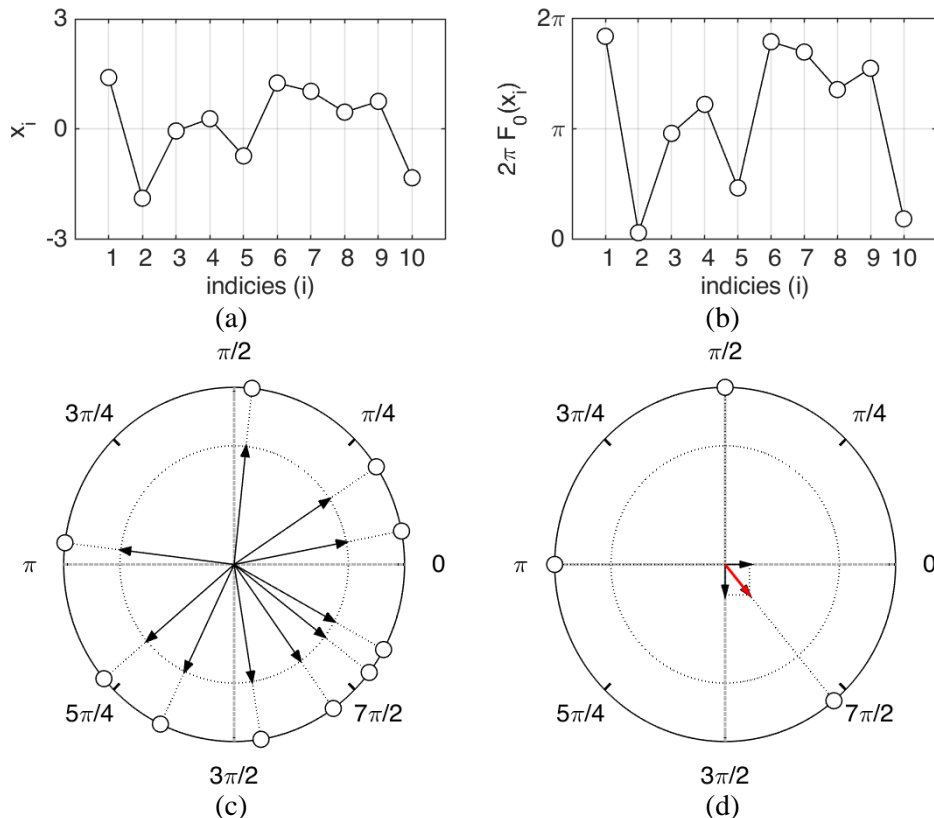
Figure 2. *Conversion of the linear data to the circular data and determination of the resultant vector; (a) Linear data (b) Conversion to the circular data (c) Unit vectors (d) The resultant vector*

Figure 2(a) shows 10 data generated from an arbitrary distribution and the circular form of these data $\big(Uniform(0,2\pi)\big)$ is given in Figure 2(b). Figure 2(c)-(d) shows the unit vector of the circular data and the resultant vector of them, respectively.

As the test statistic approaches zero, it can be said that the data are uniformly distributed, and thus the hypothesized distribution is the real distribution of the data. This assumption is weak as the resultant vector diverges from zero.

The algorithm of the proposed method is given in the Algorithm 1.

Algorithm 1. *FCS* goodness-of-fit test

Step 1. Define the distribution function and parameters for the null hypothesis,
Step 2. Calculate the R^2 test statistic,
Step 3. Determine the critical value from the chi-square table with 2 degrees of freedom and the chosen α,
Step 4. If the R^2 test statistic is greater than the critical value, reject the null hypothesis.

The critical values of the proposed method can be given in a single-line table (Table 1) without the need for a large table value. This table is a chi-square distribution table with 2 degrees of freedom.

Table 1. *The chi-square distribution table with 2 degrees of freedom*

α	0.2	0.1	0.05	0.025	0.01	0.001
$\chi^2_{v=2,1-\alpha}$	3.2189	4.6052	5.9915	7.3778	9.2103	13.8155

Exercise. The calculation of the test statistic in Table 2 is given to show the applicability of the proposed method in a simple way. The null hypothesis for the sample in Table 2, which its sample size is 10, is given as having a normal distribution with ($\mu = 10, \sigma^2 = 4$) parameters and $\alpha = 0.05$ significance level. The result is that the null hypothesis cannot be rejected.

Table 2. *Calculation of the proposed test statistic*

i	x_i	$\theta_i = 2\pi F_0(x_i)$	$\cos(\theta_i)$	$\sin(\theta_i)$
1	13.5142	6.0353	0.9694	-0.2453
2	10.2069	3.4004	-0.9667	-0.2559
3	14.9308	6.2402	0.9991	-0.0430
4	11.5484	4.9046	0.1910	-0.9816
5	11.1112	4.4658	-0.2441	-0.9698
6	11.8232	5.1460	0.4202	-0.9074
7	9.6953	2.7612	-0.9285	0.3713
8	8.8668	1.7938	-0.2212	0.9752
9	11.9410	5.2408	0.5041	-0.8636
10	8.0430	1.0299	0.5149	0.8573
		Σ	1.2383	-2.0628
		$(\Sigma)^2$	1.5334	4.2553
		R^2	**1.1577**	
		Result ($\alpha = 0.05$)	**$1.1577 < 5.9915 : H_0$ is not rejected**	

4. THE TYPE I ERROR AND THE POWER COMPARISONS

In this section, the Free Chi-Square (FCS) test was compared with the Binning Chi-Square (BCS), the Cramer von Mises (CVM), the Kolmogorov-Smirnov (KS) and the Anderson-Darling (AD) tests in terms of the type I error and the test powers by using Monte-Carlo simulation. In all comparisons, the significance coefficient was taken as $\alpha = 0.05$.

The simulation was performed using MATLAB® 2017b software. The Statistics and Machine Learning Toolbox was used for BCS (chi2gof), KS (kstest), AD (adTest) commands. The command developed by Ben Sada [21] was used for CVM (cmTest) command. The number of binning in the BCS (chi2gof) command was chosen to 10 by default. However, the software optimizes the bins by combining with neighbor bins, when the expected frequency is less than 5 in a bin.

4.1. Comparison of Power Values in Symmetrical Distributions

In this section, each simulation was run 10,000 times for $n = \{10, 20, 30, 50, 100, 200, 300, 400, 500, 1000, 2000\}$. For the simulation study, 4 symmetrical and one non-symmetrical distributions are selected. These are normal, uniform, Laplace, Student's t, and exponential distributions, respectively.

Samples with different size are selected from a normal distribution with the parameters ($\mu = 0, \sigma^2 = 1$). Five goodness-of-fit tests was used to test whether the samples fit the normal distribution with $\mu = 0$ and

$\sigma^2 = 1$. Type I errors of the goodness-of-fit (normality) tests for the normal distribution are given in the Table 3.

Table 3. *Type I errors of the goodness-of-fit tests for the normal distribution ($\mu = 0, \sigma^2 = 1$)*

Method	Sample Size										
	10	20	30	50	100	200	300	400	500	1000	2000
BCS	0.0447	0.0453	0.0466	0.0457	0.0524	0.0477	0.0527	0.0481	0.044	0.0537	0.0514
CVM	0.0468	0.0478	0.0459	0.0478	0.0498	0.0472	0.0481	0.0487	0.0519	0.0486	0.0524
KS	0.0492	0.0506	0.0459	0.0498	0.0497	0.0464	0.0463	0.0477	0.0529	0.0489	0.0524
AD	0.0493	0.0485	0.0467	0.0457	0.049	0.0486	0.0487	0.0486	0.0504	0.0493	0.0522
FCS	0.0455	0.0456	0.0473	0.051	0.0497	0.0477	0.0508	0.0476	0.054	0.0524	0.0521

It is shown that Type I errors are scattered around 0.05 for all sample sizes. This result indicates that *FCS* is also consistent method as the other goodness-of-fit tests. Furthermore, the changing in the Table 3 is clearly shown in the Figure 3.
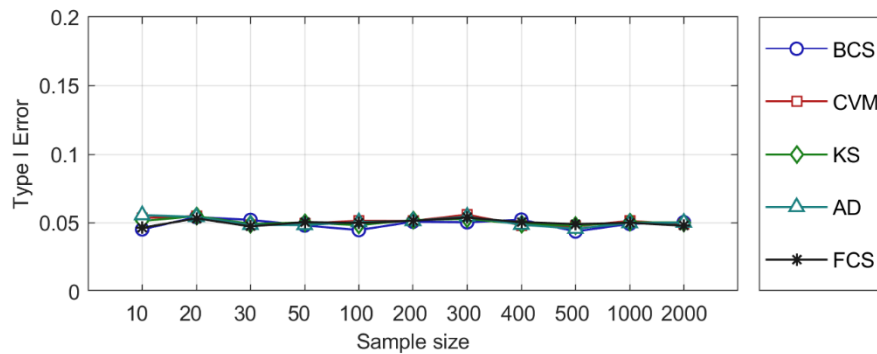
Figure 3. *Type I errors obtained as a result of the goodness-of-fit (normality) tests of samples generated from a normal distribution ($\mu = 0, \sigma^2 = 1$)*

Samples generated at different sizes from the uniform distribution ($a = -\sqrt{3}, b = \sqrt{3}$) are tested with five different goodness-of-fit methods to determine whether samples fit the normal distribution ($\mu = 0, \ \sigma^2 = 1$) or not. The power comparisons of the goodness-of-fit tests for the uniform distribution are given in the Table 4.

Table 4. *Power comparisons of the goodness-of-fit tests for the uniform distribution*

Method	Sample Size										
	10	20	30	50	100	200	300	400	500	1000	2000
BCS	0.0464	0.0495	0.0486	0.1412	0.389	0.7111	0.8982	0.9685	0.9925	1	1
CVM	0.073	0.0836	0.0934	0.1196	0.2093	0.4626	0.727	0.8907	0.9651	1	1
KS	0.0835	0.0961	0.1178	0.1506	0.2617	0.4941	0.7065	0.851	0.9344	0.9999	1
AD	0.07	0.0868	0.1027	0.1464	0.2914	0.6638	0.9135	0.986	0.9997	1	1
FCS	0.0809	0.1287	0.1775	0.2827	0.5288	0.8427	0.9613	0.9906	0.9987	1	1

As the number of sample size increases, the powers of five methods increase for the uniform distribution. The *BCS* method has the lower power than the other for $n < 50$. The *FCS* method has generally the highest power in almost all sample sizes. As the number of sample size increases, the powers of all methods increase to '1'. Also, the changing in the Table 4 is clearly shown in the Figure 4.
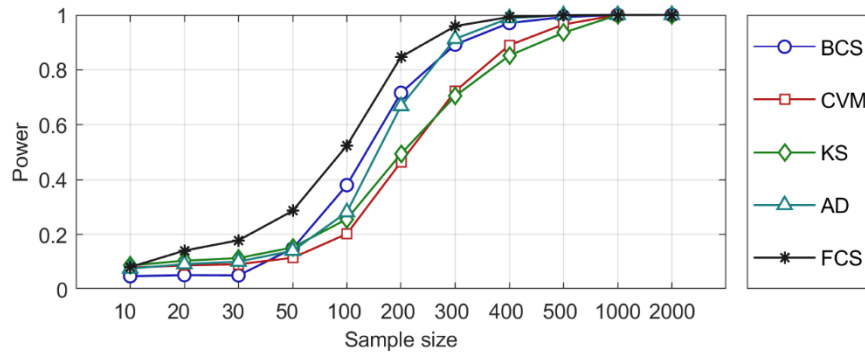
Figure 4. *Test powers obtained as a result of the goodness-of-fit (normality) tests of samples generated from a uniform distribution* $(a = -\sqrt{3}, b = \sqrt{3})$

According to the Figure 4, while the FCS shows the best sensitivity, the CVM and the KS show the lowest sensitivity according to the test powers. Although the BCS and the AD show similar sensitivities, it is observed that the BCS shows low sensitivity to $n < 50$ values.

Samples generated at different sizes from the Laplace distribution $(\mu = 0, \sigma^2 = 1)$ are tested with five different goodness-of-fit methods to determine whether samples fit the normal distribution $(\mu = 0, \sigma^2 = 1)$ or not. In the Table 5, the power comparisons of the goodness-of-fit tests for the Laplace distribution are shown.

Table 5. *Power comparisons of the goodness-of-fit tests for the Laplace distribution*

Method	Sample Size										
	10	20	30	50	100	200	300	400	500	1000	2000
BCS	0.0398	0.054	0.1156	0.2434	0.4189	0.7701	0.9266	0.9785	0.9922	0.9986	0.9993
CVM	0.032	0.0452	0.0516	0.0842	0.1967	0.5384	0.8004	0.9399	0.9838	1	1
KS	0.0395	0.0629	0.0857	0.1332	0.2648	0.5715	0.7873	0.9207	0.9713	1	1
AD	0.0423	0.0503	0.0605	0.099	0.2339	0.6129	0.8597	0.9668	0.9952	1	1
FCS	0.0945	0.1575	0.2166	0.3471	0.6163	0.9056	0.9816	0.9967	0.9998	1	1

As the number of sample size increases, the powers of five methods increase for the Laplace distribution. The CVM method has the lower power than the others for $n < 300$. The FCS has the higher power in all sample sizes than the others. As the number of sample size increases, the powers of all methods increase, but the powers of all methods except the BCS increase to '1'. In addition, the changing in the Table 5 is clearly shown in the Figure 5.
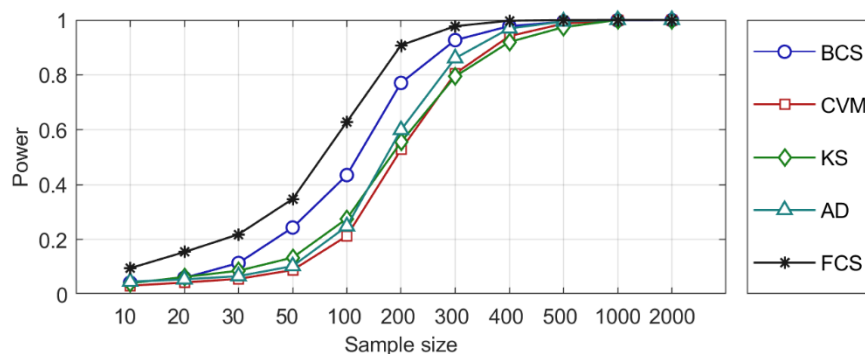

Figure 5. *Test powers obtained as a result of the goodness-of-fit (normality) tests of samples generated from a Laplace distribution* $(\mu = 0, \sigma^2 = 1)$
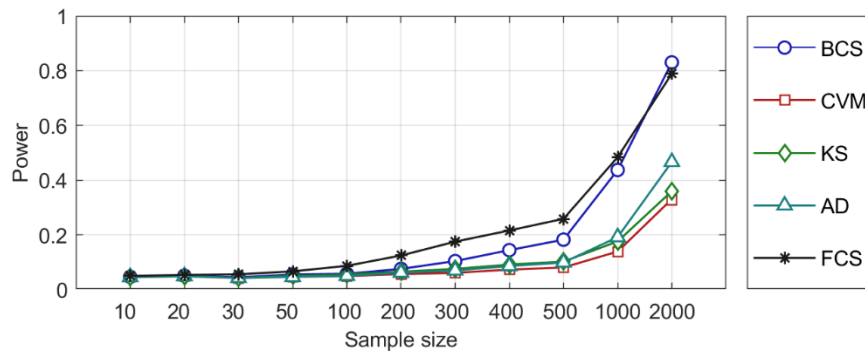
While the FCS shows the best sensitivity, the CVM and the KS show the lowest sensitivity according to the test powers. Moreover, the BCS shows higher sensitivities than the AD at higher sample sizes especially in the Figure 5.

Samples generated at different sizes from the Student's t-distribution ($\nu = 10$) are tested with five different goodness-of-fit methods to determine whether samples fit the normal distribution ($\mu = 0, \sigma^2 = 10/8$) or not. The power values of the goodness-of-fit tests for Student's t-distribution performed at the different sample sizes are given in the Table 6.

Table 6. *Power comparisons of the goodness-of-fit tests for the Student's t-distribution*

Method	Sample Size										
	10	20	30	50	100	200	300	400	500	1000	2000
BCS	0.0477	0.0488	0.0508	0.0555	0.0545	0.0795	0.1063	0.139	0.1867	0.4323	0.8299
CVM	0.0428	0.0439	0.0433	0.0474	0.0507	0.0578	0.0596	0.064	0.0734	0.1364	0.3354
KS	0.0417	0.0443	0.0473	0.0495	0.0556	0.0664	0.0778	0.0855	0.0992	0.1764	0.3684
AD	0.047	0.0461	0.045	0.0487	0.052	0.0639	0.0721	0.0799	0.0963	0.1866	0.4738
FCS	0.0466	0.0536	0.0585	0.0668	0.0832	0.1353	0.1713	0.2135	0.2577	0.4774	0.793

While the number of sample size increases, the powers of five methods increase for the Student's t-distribution. The CVM method has the lower power than the others for $n > 10$. The FCS generally has the higher power in all sample sizes than the others. Only, the powers of the KS and the FCS methods increase for all sample sizes. The power values of the other methods are not monotone increasing. In addition, the changing in the Table 6 is clearly given in the Figure 6.

Figure 6. *Test powers obtained as a result of the goodness-of-fit (normality) tests of samples generated from a Student's t-distribution ($\nu = 10$)*
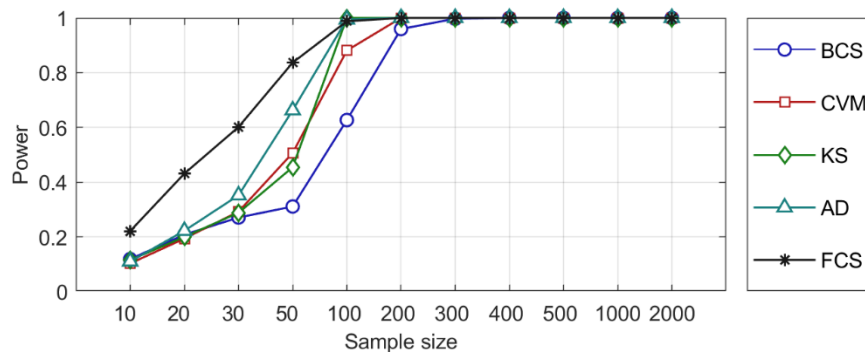
While the FCS shows the best sensitivity, the CVM and the KS show the lowest sensitivity according to the test powers. Furthermore, the BCS shows higher sensitivities than the AD at higher sample sizes especially in the Figure 6.

Samples generated at different sizes from the exponential distribution ($\mu = 1$) are tested with five different goodness-of-fit methods to determine whether samples fit the normal distribution ($\mu = 1, \sigma^2 = 1$) or not. The power comparisons of the goodness-of-fit tests for the exponential distribution are given in the Table 7.

Table 7. *Power comparisons of the goodness-of-fit tests for the exponential distribution*

	Sample Size										
Method	**10**	**20**	**30**	**50**	**100**	**200**	**300**	**400**	**500**	**1000**	**2000**
BCS	0.1196	0.2042	0.2684	0.3015	0.6281	0.9637	0.998	1	1	1	1
CVM	0.1035	0.1879	0.292	0.4829	0.8765	1	1	1	1	1	1
KS	0.1156	0.1926	0.2793	0.4332	1	1	1	1	1	1	1
AD	0.1168	0.2115	0.353	0.6457	0.9937	1	1	1	1	1	1
FCS	0.2259	0.4329	0.6062	0.833	0.9882	1	1	1	1	1	1

While the number of sample size increases, the powers of five methods increase for the exponential distribution. The FCS has the lower power within $n = 100$ sample size than the KS and AD methods, but the FCS method has the higher power in other sample sizes than the others. While the number of sample size increases, the powers of all methods increase to '1' greater than $n = 400$. Moreover, the changing in the Table 7 is clearly given in the Figure 7.

Figure 7. *Test powers obtained as a result of the goodness-of-fit (normality) tests of samples generated from an exponential distribution* ($\mu = 1$)

According to the Figure 7, while the FCS shows the best sensitivity, the BCS show the lowest sensitivity according to the test powers. Also, the AD shows higher sensitivities than the KS and the CVM at some sample sizes.

**4.2. Comparison of Power Values in Non-Symmetrical Distributions**

In this section, each simulation was run 10,000 times for $n = \{10, 20, 30, 50, 100, 200, 300, 400, 500, 1000, 2000\}$. Log-normal, normal, Weibull, gamma and chi-square distributions are selected for these simulations. Despite normal distribution is a symmetric distribution, the other four distributions are chosen as non-symmetrical distributions.

Samples generated at different sizes from the log-normal distribution ($\mu = 0.8432, \sigma^2 = 0.7147$) are tested with five different goodness-of-fit methods to determine whether samples fit the log-normal distribution ($\mu = 0.8432, \sigma^2 = 0.7147$) or not. Type I errors of the goodness-of-fit (LogNormality) tests for the log-normal distribution are given in the Table 8.

***Table 8.** Type I errors of the goodness-of-fit tests for the log-normal distribution*

	Sample Size										
Method	**10**	**20**	**30**	**50**	**100**	**200**	**300**	**400**	**500**	**1000**	**2000**
BCS	0.0494	0.0476	0.047	0.0494	0.0512	0.0518	0.0508	0.0487	0.0506	0.0514	0.0462
CVM	0.0542	0.0467	0.0497	0.0468	0.0493	0.0494	0.0524	0.048	0.0463	0.0535	0.049
KS	0.0543	0.047	0.05	0.0473	0.0507	0.0495	0.0549	0.0467	0.0499	0.0512	0.0496
AD	0.0546	0.0476	0.0495	0.0461	0.0473	0.0485	0.0537	0.0492	0.0474	0.0539	0.0499
FCS	0.0489	0.0464	0.0501	0.0492	0.0519	0.0481	0.054	0.0521	0.0514	0.0519	0.0453

Type I errors appear to distribute around 0.05. According to this result, the proposed FCS method is also work consistently as the other methods for non-symmetric distributions. Also, the changing in the Table 8 is clearly shown in the Figure 8.
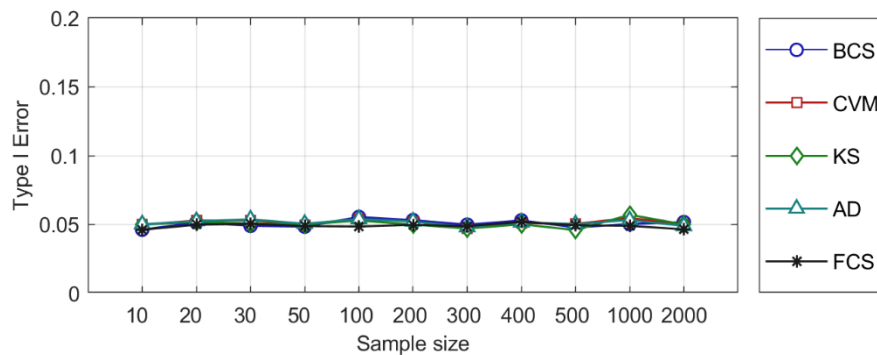
***Figure 8.** Type I errors obtained as a result of the goodness-of-fit (LogNormality) tests of samples generated from a log-normal distribution ($\mu = 0.8432, \sigma^2 = 0.7147$)*

Samples generated at different sizes from the normal distribution ($\mu = 3, \sigma^2 = 6$) are tested with five different goodness-of-fit methods to determine whether samples fit the log-normal distribution ($\mu = 0.8432, \sigma^2 = 0.7147$) or not. The power comparisons of the goodness-of-fit tests for the normal distribution are given in the Table 9.

***Table 9.** Power comparisons of the goodness-of-fit tests for the normal distribution*

	Sample Size										
Method	**10**	**20**	**30**	**50**	**100**	**200**	**300**	**400**	**500**	**1000**	**2000**
BCS	0.0994	0.1652	0.2919	0.5292	0.8849	0.9966	0.9998	1	1	1	1
CVM	0.1554	0.2496	0.3787	0.636	0.9588	1	1	1	1	1	1
KS	0.166	0.2925	0.4218	0.6724	0.9711	1	1	1	1	1	1
AD	0.776	0.952	0.9918	0.9998	1	1	1	1	1	1	1
FCS	0.264	0.5266	0.7259	0.9268	0.9982	1	1	1	1	1	1

While the number of sample size increases, the powers of five methods increase for the normal distribution. The BCS method has the lowest power and the AD has the highest power in all sample sizes. Although the FCS test shows lower power than the AD, it has more high power than the others. As the number of sample size increases, the powers of all goodness-of-fit methods increase to '1'. In addition, the changing in the Table 9 is clearly shown in the Figure 9.
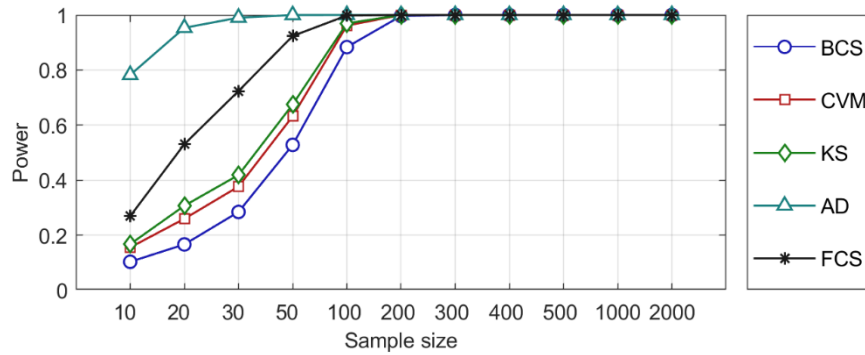
Figure 9. *Test powers obtained as a result of the goodness-of-fit (LogNormality) tests of samples generated from a normal distribution ($\mu = 3, \sigma^2 = 6$)*

According to the Figure 9, while the AD test shows the highest sensitivity, the BCS test show the lowest sensitivity according to the test powers. Although the FCS test shows lower sensitivity than the AD, it has more sensitivity than the others.

Samples generated at different sizes from the Weibull distribution ($a = 3.21, b = 1.23$) are tested with five different goodness-of-fit methods to determine whether samples fit the log-normal distribution ($\mu = 0.8432, \sigma^2 = 0.7147$) or not. The power values of the goodness-of-fit tests for Weibull distribution performed at the different sample sizes are given in the Table 10.

Table 10. *Power comparisons of the goodness-of-fit tests for the Weibull distribution*

	Sample Size										
Method	**10**	**20**	**30**	**50**	**100**	**200**	**300**	**400**	**500**	**1000**	**2000**
BCS	0.0473	0.0588	0.0987	0.2201	0.5181	0.7455	0.8501	0.904	0.9431	0.9908	0.9998
CVM	0.0968	0.1147	0.138	0.1928	0.4002	0.8063	0.968	0.996	0.9997	1	1
KS	0.0979	0.1216	0.1571	0.2216	0.4527	0.841	0.9736	0.998	1	1	1
AD	0.262	0.3639	0.469	0.6426	0.9112	0.9987	1	1	1	1	1
FCS	0.1103	0.1927	0.2814	0.4364	0.7454	0.9656	0.9964	0.9997	1	1	1

While the number of sample size increases, the powers of five methods increase for the Weibull distribution. Although the FCS test shows lower power than the AD, it has more high power than the others. As the number of sample size increases, the powers of all goodness-of-fit methods increase to '1'except the BCS method. Furthermore, the changing in the Table 10 is clearly shown in the Figure 10.
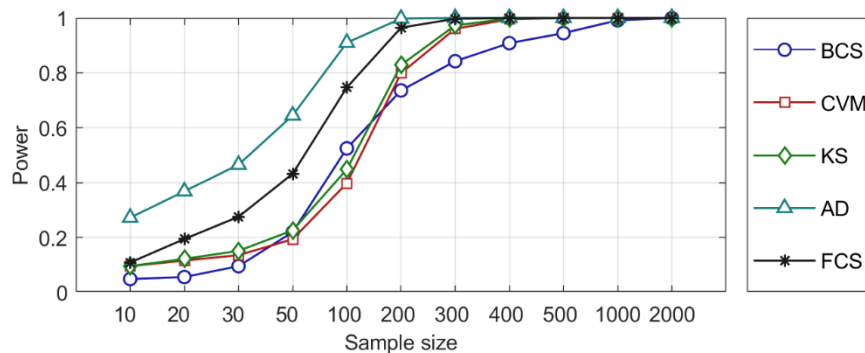

Figure 10. *Test powers obtained as a result of the goodness-of-fit (LogNormality) tests of samples generated from a Weibull distribution ($a = 3.21, b = 1.23$)*
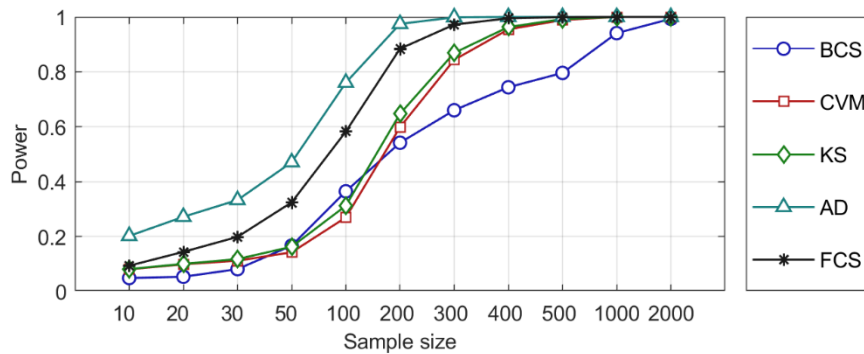
According to the Figure 10, while the BCS test has generally the lowest sensitivity according to the test powers, the AD test has the highest sensitivity. Although the FCS test shows lower sensitivity than the AD, it has more sensitivity than the others.

Samples generated at different sizes from the gamma distribution ($a = 1.5, b = 2$) are tested with five different goodness-of-fit methods to determine whether samples fit the log-normal distribution ($\mu = 0.8432, \sigma^2 = 0.7147$) or not. The power comparisons of the goodness-of-fit tests for the gamma distribution are given in the Table 11.

Table 11. *Power comparisons of the goodness-of-fit tests for the gamma distribution*

Method	Sample Size
	10	20	30	50	100	200	300	400	500	1000	2000
BCS	0.0489	0.0523	0.0851	0.166	0.368	0.5406	0.6537	0.736	0.8027	0.9362	0.9918
CVM	0.0894	0.0987	0.1121	0.1483	0.2681	0.6002	0.847	0.9541	0.9909	1	1
KS	0.0878	0.1007	0.1208	0.169	0.3114	0.649	0.8762	0.9675	0.9943	1	1
AD	0.2073	0.2711	0.338	0.4785	0.7683	0.9721	0.9985	1	1	1	1
FCS	0.0941	0.1513	0.2109	0.3276	0.5918	0.8847	0.9757	0.9953	0.9991	1	1

While the number of sample size increases, the powers of five methods increase for the gamma distribution. The BCS has the lower power in all sample sizes except within '50' and '100' sample sizes than the other methods. Although the FCS test shows lower power than the AD, it has more high power than the others. As the number of sample size increases, the powers of all goodness-of-fit methods increase to '1'except the BCS method. Also, the changing in the Table 11 is clearly shown in the Figure 11.

Figure 11. *Test powers obtained as a result of the goodness-of-fit (LogNormality) tests of samples generated from a gamma distribution ($a = 1.5, b = 2$)*
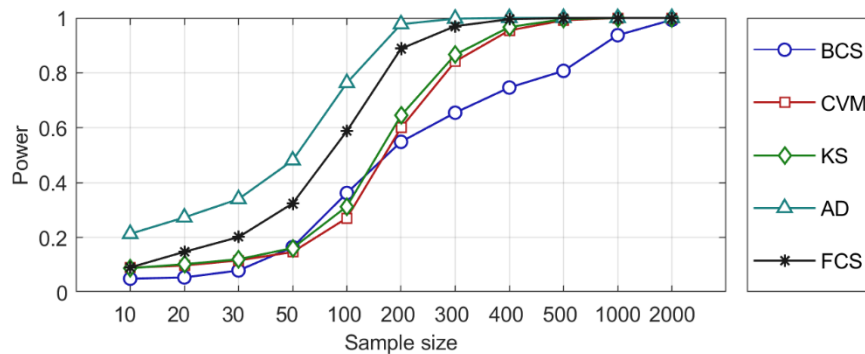
In the Figure 11, while the BCS test has generally the lowest sensitivity according to the test powers, the AD test has the highest sensitivity. Although the FCS test shows lower sensitivity than the AD, it has more sensitivity than the others.

Samples generated at different sizes from the chi-square distribution ($\nu = 3$) are tested with five different goodness-of-fit methods to determine whether samples fit the log-normal distribution ($\mu = 0.8432, \sigma^2 = 0.7147$) or not. The power values of the goodness-of-fit tests for chi-square distribution performed at the different sample sizes are given in the Table 12.

Table 12. *Power comparisons of the goodness-of-fit tests for the chi-square distribution*

Method	Sample Size										
	10	20	30	50	100	200	300	400	500	1000	2000
BCS	0.0514	0.056	0.0832	0.1644	0.3668	0.5508	0.665	0.7418	0.8077	0.9375	0.994
CVM	0.0917	0.0996	0.1124	0.1447	0.2705	0.599	0.8459	0.955	0.9908	1	1
KS	0.0901	0.1073	0.123	0.1663	0.3076	0.6534	0.8731	0.9655	0.9938	1	1
AD	0.2084	0.2742	0.3463	0.4795	0.7664	0.9769	0.9981	0.9999	1	1	1
FCS	0.0925	0.156	0.2089	0.3225	0.5917	0.8903	0.9766	0.9953	0.9991	1	1

While the number of sample size increases, the powers of five methods increase for the chi-square distribution. The BCS has the lower power in all sample sizes except within '50' and '100' sample sizes than the other methods. Although the FCS test shows lower power than the AD, it has more high power than the others. As the number of sample size increases, the powers of all goodness-of-fit methods increase to '1'except the BCS method. Furthermore, the changing in the Table 12 is clearly shown in the Figure 12.

Figure 12. *Test powers obtained as a result of the goodness-of-fit (LogNormality) tests of samples generated from a chi-square distribution* $(\nu = 3)$

In the Figure 12, while the BCS test has generally the lowest sensitivity according to the test powers, the AD test has the highest sensitivity. Although the FCS test shows lower sensitivity than the AD, it has more sensitivity than the others.

5. REAL LIFE EXAMPLES

Three different popular real-world applications are chosen to demonstrate the performance of the proposed method. These are sleep time data [22], Hurricane Katrina data and global warming data [23].

5.1. Sleep Time Data

Students were surveyed about various aspects of sleep behavior during a two-week period [24]. The average sleep time per day (h) for 100 of the students were observed [22]. It is thought that the data on the average sleep time shows the normal distribution. The distribution parameters were estimated by the maximum likelihood estimation (MLE). The null and the alternative hypothesis are given as follows,

H_0: The average sleep time data follow the normal pdf.
H_1: The average sleep time data do not follow the normal pdf.

In order to test these hypothesis, Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Binning Chi-Square (BCS) and the proposed Free Chi-Square (FCS) goodness-of-fit tests were applied at $\alpha = 0.05$ significance level. The obtained results, the p-values and the test statistics of goodness-of-fit tests are given in Table 13.

Table 13. *Goodness-of-fit test results of the sleep time data*

		Test Statistic	*p*-value	Result
Kolmogorov-Smirnov (KS)	:	0.0618	0.8166	Cannot Reject
Anderson-Darling (AD)	:	0.3969	0.8515	Cannot Reject
Binning Chi-Square (BCS)	:	5.5484	0.4756	Cannot Reject
Free Chi-Square (FCS)	:	1.0827	0.7090	Cannot Reject

All these tests support the acceptance of the null hypothesis. Thus, the average sleep time follows the normal distribution.

5.2. Hurricane Katrina Data

Hurricane Katrina is one of the most devastating hurricanes in the United States in the last 100 years. One of the most important variables of Hurricane Katrina is the pressure wind velocity. There are 63 observations of the wind velocity in this dataset [23]. It is thought that the data on pressure wind velocity shows the Weibull distribution. The distribution parameters were estimated by MLE. The null and the alternative hypothesis are given as follows,

H_0: The wind velocity data of Hurricane Katrina follow the Weibull pdf.
H_1: The wind velocity data of Hurricane Katrina do not follow the Weibull pdf.

In order to test these hypothesis, Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Binning Chi-Square (BCS) and the proposed Free Chi-Square (FCS) goodness-of-fit tests were applied at $\alpha = 0.05$ significance level. The obtained results, the p-values and the test statistics of goodness-of-fit tests are given in the Table 14.

Table 14. *Goodness-of-fit test results of Hurricane Katrina data*

		Test Statistic	*p*-value	Result
Kolmogorov-Smirnov (KS)	:	0.0884	0.9510	Cannot Reject
Anderson-Darling (AD)	:	0.3632	0.8835	Cannot Reject
Binning Chi-Square (BCS)	:	0.9582	0.811	Cannot Reject
Free Chi-Square (FCS)	:	1.2701	0.7351	Cannot Reject

All these tests support the acceptance of the null hypothesis. Thus, the wind velocity measurement of the Hurricane Katrina follows the Weibull distribution.

5.3. Global Warming Data

Global warming is affected by two interacting elements, such as atmospheric temperature and the amount of carbon dioxide in the atmosphere. In this application, CO_2 data related to global warming were used. This data set was obtained from: http://booksite.elsevier.com/9780124171138. CO_2 data were collected in Point Barrow in Alaska from 1974 to 2004. It is thought that the data on the amount of CO_2 shows the

gamma distribution. The distribution parameters were estimated by MLE. The null and the alternative hypothesis are given as follows,

H_0: The CO_2 data follow the gamma pdf.
H_1: The CO_2 do not follow the gamma pdf.

In order to test these hypothesis, Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Binning Chi-Square (BCS) and the proposed Free Chi-Square (FCS) goodness-of-fit tests were applied at $\alpha = 0.05$ significance level. The obtained results, the p-values and the test statistics of goodness-of-fit tests are given in the Table 15.

Table 15. *Goodness-of-fit test results of the CO_2 data*

		Test Statistic	p-value	Result
Kolmogorov-Smirnov (KS)	:	0.0884	0.9510	Cannot Reject
Anderson-Darling (AD)	:	0.3632	0.8835	Cannot Reject
Binning Chi-Square (BCS)	:	0.9582	0.811	Cannot Reject
Free Chi-Square (FCS)	:	1.2701	0.7351	Cannot Reject

All these tests support the acceptance of the null hypothesis. Thus, the amount of CO_2 of the Point Barrow follows the gamma distribution.

6. CONCLUSION

In this paper, a free chi-square test was proposed. It has many advantages over the existing tests such as, binning-free, distribution-free, more sensitivity, easy to use and fast. The BCS is a test with disadvantages due to binning selection and it requires a sufficient sample size for the chi-square approximation to be valid. On the other hand, the CVM and KS tests can be preferred because they are binning free goodness-of-fit tests. But the KS test has the disadvantage that it is more sensitive to deviations near the center of the distribution than at the tails and the KS table values are used in the interpretation of the results. The AD test, which is a more sensitive test, gives good results for some distributions but cannot be applied to all distributions. Moreover, the calculation of the different critical values table for each distribution indicates that the AD test is not distribution-free approach. On the other hand, the proposed FCS method is not only a sensitive method, but also a binning-free and distribution-free test. In addition, this test method does not require grouped data or ordered data, nor does it require an empirical distribution function. It is independent of the degree of freedom since it uses only a chi-square table with 2 degrees of freedom. Exercise in Section 3 shows that it is a simple and easy to calculate. Therefore, it can be preferred by many researchers. Simulation studies for the test statistic calculated using the trigonometric approach of the proposed method are given in Section 4. It is observed that the proposed method is more sensitive than the other methods for symmetric distributions, also it gives the correct results according to Type I error. Although simulation studies of non-symmetric distributions show that the AD test is more sensitive, the proposed method gives more sensitive results than the rest of the methods. But the AD cannot be applied to all distributions and it is not distribution-free approach. In these simulations, studies were made for different significance levels except $\alpha = 0.05$, and similar results were observed. So, they were not put into this study.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1]   Yazici, B., Yolacan, S., "A comparison of various tests of normality", Journal of Statistical Computation and Simulation, 77(2): 175-183, (2007).

[2]   Romao, X., Delgado, R., Costa, A., "An empirical power comparison of univariate goodness-of-fit tests for normality", Journal of Statistical Computation and Simulation, 80(5): 545-591, (2010).

[3]   Dudewicz, E., Van Der Meulen, E., "Entropy-based tests of uniformity", Journal of the American Statistical Association, 76(376): 967-974, (1981).

[4]   Crzcgorzewski, P., Wirczorkowski, R., "Entropy-based goodness-of-fit test for exponentiality", Communications in Statistics-Theory and Methods, 28(5): 1183-1202, (1999).

[5]   Choi, B., Kim, K., "Testing goodness-of-fit for Laplace distribution based on maximum entropy", Statistics, 40(6): 517-531, (2006).

[6]   Mahdizadeh, M., Zamanzade, E., "New goodness of fit tests for the Cauchy distribution", Journal of Applied Statistics, 44(6): 1106-1121, (2017).

[7]   Dhumal, B., Shirke, D., "A modified one-sample test for goodness-of-fit", Journal of Statistical Computation and Simulation, 85(2): 422-429, (2015).

[8]   Gibbons, J., Chakraborti, S., Nonparametric statistical inference 5th ed., CRC Press, Boca Raton, (2011).

[9]   Bain, L., Engelhardt, M., Introduction to probability and mathematical statistics 2nd ed., Brooks/Cole, USA, (1992).

[10]   Stephens, M., "EDF statistics for goodness of fit and some comparisons", Journal of the American statistical Association, 69(347): 730-737, (1974).

[11]   Cramér, H., "On the composition of elementary errors", Scandinavian Actuarial Journal, 11(1): 13-74, (1928).

[12]   Von Mises, R., "Wahrscheinlichkeitsrechnung und Ihre Anwendung in der Statistik und Theoretischen Physik", F. Deuticke, 6: 13-74, (1931).

[13]   Anderson, T., Darling, D., "Asymptotic theory of certain goodness of fit criteria based on stochastic processes", The Annals of Mathematical Statistics, 23(2): 193-212, (1952).

[14]   Darling, D., "The kolmogorov-smirnov, cramer-von mises tests", The Annals of Mathematical Statistics, 28(4): 823-838, (1957).

[15]   Stephens, M., "Use of the Kolmogorov-Smirnov, Cramér-Von Mises and related statistics without extensive tables", Journal of the Royal Statistical Society. Series B (Methodological), 32(1): 115-122, (1970).

[16]   Noughabi, H., Arghami, N., "Monte Carlo comparison of seven normality tests", Journal of Statistical Computation and Simulation, 81(8): 965-972, (2011).

[17]   Farrell, P., Rogers-Stewart, K., "Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test", Journal of Statistical Computation and Simulation, 76(9): 803-816, (2006).

[18]   Anderson, T., Darling, D., "A test of goodness of fit", Journal of the American Statistical Association, 49(268): 765-769, (1954).

[19]   Dodge, Y., The Oxford dictionary of statistical terms, Oxford University Press on Demand, Great Britain, (2006).

[20]   Jammalamadaka, S. R., Gupta, A. S., Topics in Circular Statistics, World Scientific Publishing Co. Pte. Ltd., Singapore, (2001).

[21]   Ben Sada, A. "Cramer-von Mises test for goodness-of-fit of a single sample". Retrieved Apr 18, 2018, from MATLAB Central File Exchange, (2015).

[22]   Devore, J., Probability and Statistics for Engineering and the Sciences 9th ed., Cengage Learning, USA, (2016).

[23]   Ramachandran, K., Tsokos, C., Mathematical statistics with applications in R 2nd ed., Academic Press, (2015).

[24]   Onyper, S., Thacher, P., Gilbert, J., Gradess, S., "Class start times, sleep, and academic performance in college: A path analysis", Chronobiology International, 29(3): 318-335, (2012).