Journal of Statistical Research *ISSN 0256 - 422 X*
2005, Vol. 39, No. 2, pp. 103–125
Bangladesh

GOODNESS-OF-FIT TESTS OF A PARAMETRIC DENSITY FUNCTIONS: MONTE CARLO SIMULATION STUDIES

Ying Li

Department of mathematics and Statistics
University of Guelph, Ontario, Canada N1G2W1

Email: yli02@uoguelph.ca

Radhey S. Singh

Department of Mathematics and Statistics
University of Guelph, Ontario, Canada N1G2W1

Email: rssingh@uoguelph.ca

Yiguo Sun

Department of Economics
University of Guelph, Ontario, Canada N1G2W1

Email: yisun@uoguelph.ca

summary

The purpose of this paper is to use Monte Carlo simulations to evaluate the performance of six most popular statistics for testing the goodness of fit of a parametric density function. The first three tests in this study are based on the empirical distribution function which are simple and widely used. The other three are based on directed and non-directional divergence measures and derived from minimum relative entropy (MinxEnt) principle, m-spacing method and kernel method. This study aims to evaluate the behavior of these tests by examining the rejection rates under the hypothesis. It is shown that the tests based on the directed divergence measure give a good approximation to the given significance levels and are more powerful than other tests against the given alternative distributions. It also suggests that the statistics based on the MinxEnt estimator detect the distribution with higher kurtosis better than others.

Keywords and phrases: Goodness of fit tests; MinxEnt principle; m-spacing method; information measure; Monte Carlo simulation

AMS Classification: 62G10

1 Introduction

The goodness of fit tests are developed with the aim of testing the hypothesis that experimental data have come from a random variable with a theoretical distribution. How well the data are modeled by that distribution is known as goodness of fit and can be measured by several different types of test statistics.

Suppose that we have observed a sequence of independent samples $\{X_i\}_{i=1}^n$ which are drawn from a common cumulative distribution function (CDF), $F(x)$, with a probability density function (PDF), $f(x)$, $x \in R$. Consider the following goodness-of-fit test problem:

$$H_0 : F(x) = F_0(x, \theta), \text{ or } f(x) = f_0(x, \theta) \text{ for all } x \in R \text{ and some } \theta \in \Theta,$$

where the parameter vector, θ, is unspecified, $F_0(x, \theta)$ is a CDF from a parametric family, with PDF $f_0(x, \theta)$ that are measurable in x for every $\theta \in \Theta$, an open subset of a d-dimensional Euclidean space R^d, and continuous in θ for every $x \in R$.

In particular, this study considers the case of testing for normality with unknown parameters estimated by the maximum likelihood estimators (MLE). That is, the null hypothesis is

$$H_0 : f(x) = f_0(x, \theta) = \left(\sqrt{2\pi}\sigma\right)^{-1} \exp\left(-(x-\mu)^2 / 2\sigma^2\right) \text{ for all } x \in R, \text{ some } \theta \in \Theta$$

where $\theta = (\mu, \sigma^2)$, μ and σ^2 are not specified and are replaced by the sample mean $\bar{x}$ and the sample variance $\hat{\sigma}^2$, where $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ and $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$. The use of testing for normality could be applied in two areas. The first is to test the distribution of the statistics which are normally distributed due to the applicability of large sample theorems such as the central limit theorem. The second application for testing normality is related to the situations where the normal distribution is assumed to be the appropriate model for the phenomenon under investigation. The examples of use of normal distribution and lognormal distribution are given in many literatures. For example, in the analysis of cadmium and lead levels in the blood of children (Smith, Temple, and Reading, 1976), body discomfort and transmissibility scores (Griffin and Whitham, 1978), and weights of mammary tumors in rats (Fredholm, Gummarsson, Jensen, and Muntzing, 1978), applictions of normal and lognormal distributions are made.

When testing for normality for a given sample, test statistics may have different performance. This study aims to evaluate the finite-sample performances of several statistics for testing the goodness of fit of a normal distribution function. This problem is very important both in theoretical and in experimental analysis.

Many tests have been developed for testing the goodness of fit of normality. In D'gostino and Stephens (1986), five categories of them are introduced, chi-square tests, empirical distribution function (EDF) tests, moment tests, regression tests, and miscellaneous tests.

The chi-square type tests follow a chi-square distribution under H_0 and are first developed by Karl Pearson in 1900. They are based on a comparison of the frequencies from a model with the frequencies observed from an experiment. The well known Pearson chi-square test and others related to it such as the G test (Cressie and Read, 1984), the Freeman-Tukey test (Freeman and Tukey, 1950) and the Rao-Robson test (Rao and Robson, 1974) remain among the most used statistical procedures.

The EDF type tests constitute another class of goodness of fit statistics. These are based on a comparison of $F_0(x, \theta)$, the CDF of the normal distribution and the empirical distribution function $F_n(x)$ of the sample. These tests are used most often with continuous data. In the 1930s, Kolmogorov and Smirnov invented their EDF test for continuous data which is known as the KS test. Today it is still one of the best known and most commonly used goodness of fit tests because of its simplicity and intuitive nature. Crámer-von Mises (CVM) test and Anderson Darling (AD) test are defined as the integrated squared difference between the EDF and the theoretical distribution function, invoking a weight function ψ (Ozturk and Hettmansperger, 1997). Other EDF type tests include the Kuiper test, the Watson test, etc. Stephens (1974) examined these five statistics in three situations: when the hypothesized distribution $F_0(x)$ was completely specified and where $F_0(x, \theta)$ represented the uniform or normal distribution with one or more parameters to be estimated from the data. Throughout the experiments studied by Stephens, the AD test and the CVM test appeared to be the best pair of EDF statistics and the chi-squared test was not as powerful as EDF statistics. In this study, we concentrate on the KS test, the CVM test and the AD test which have attracted most attention. For moment tests, regression tests, miscellaneous tests and other tests for testing normality, see D'agostino and Stephens (1986).

Moreover, in the literature of statistic, another type of goodness-of-fit test statistics, divergence measure tests, have been developed based on the idea of information-theoretic entropy which was first introduced in communication theory by Shannon (1948), see Ullah (1996) for a good survey. The concept of divergence is related to the distance between two probability distributions.

An asymmetric divergence measure of f from f_0 is discussed in Ullah (1996). It is introduced by Kullback and Leibler (1951) and is known as relative entropy which is one of the directed divergence measures. The minimum relative entropy principle can be used in econometric estimation and hypothesis testing. Parzen (1985) has applied it to the test for goodness of fit. Based on the Kullback-Leibler relative entropy, Song (2002) developed an asymptotically distribution-free goodness-of-fit test using the m-spacing method. This method is shown to provide an extremely simple and potentially much better alternative to the classical EDF based test procedures.

Another class of divergence measure which is symmetric is referred to as the non-directional divergence measure. Test based on $\int (f(x) - f_0(x, \theta))^2 \, dx$, the integrated squared difference between a kernel estimate of $f(x)$ and the quasi-maximum likelihood estimate of $f_0(x, \theta)$ is described in Fan (1994), Aït-Sahalia (1996) and Pagan and Ullah (1999). Fan (1994) showed this test was more powerful than the KS test for the local alternatives introduced by Rosenblatt (1975).

This paper is organized as follows. Six tests including three EDF tests and three divergence measure tests are summarized in Section 2. The EDF tests, which are the KS test, the CVM test and the AD test, are first discussed in Section 2.1. A directed divergence measure, which is Kullback-Leibler divergence measure, is introduced in Section 2.2. It is applied for testing the goodness of fit by using two methods. One is the classical minimum relative entropy principle. The other is m-spacing method. Section 2.3 introduces another test of goodness of fit which is based on the non-directional divergence measure. In Section 3, a Monte Carlo simulation study is given to evaluate the performance of these statistics. After introducing the experimental design, the simulation results are given for a comparison. A conclusion and suggestions for future research are given in the last section.

2 Summary of Test Statistics

2.1 Tests Based on Empirical Distribution Function (EDF)

The empirical distribution function is used for estimating the population cumulative distribution function for a given observed sample. It is defined as

$$F_n(x) = n^{-1}\left(\text{number of } x_j \leq x\right) = n^{-1}\sum_{i=1}^{n} I\left(x_j \leq x\right),$$

where n is the sample size, and $I(\cdot)$ is the indicator function. By Glivenko-Cantelli lemma, the EDF converges uniformly to the CDF with probability one.

Conveniently, the EDF statistics are divided into two classes, the supremum class and the quadratic class (Kapur and Kesavan, 1992). In this study, we consider three EDF statistics which are Kolmogorov-Smirnov (KS) statistic, Crámer-von Mises (CVM) statistic and Anderson-Darling (AD) statistic.

KS test is one of the earliest nonparametric tests. The KS statistic belongs to the supremum class of EDF statistics which is based on the largest vertical difference between $F_0(x)$ and $F_n(x)$. As defined in Stephens (1974), without loss of generality, we assume that the values in the given sample are in ascending order, $x_1 \leq x_2 \leq \cdots \leq x_n$. The KS statistic (D) is

$$\begin{aligned} D^+ &= \max_{1\leq i\leq n}\left(i/n - z_i\right) \\ D^- &= \max_{1\leq i\leq n}\left(z_i - (i-1)/n\right) \\ D &= \max\left(D^+, D^-\right) \end{aligned} \tag{2.1}$$

where z_i is the theoretical distribution $F_0(x_i)$, D^+ is the largest vertical distance between the EDF and the theoretical distribution function when the EDF is greater than z_i, and D^- is the largest vertical distance when the EDF is less than z_i. If H_0 is true, the distance will be minimum; otherwise, the difference between the hypothetical distribution and the true distribution will be noticeable.

The CVM test statistic (W^2) is one of the quadratic EDF statistics. This class is based on $(F_0(x) - F(x))^2$ which has the following general form:

$$Q = n \int (F_0(x) - F(x))^2 \psi(x)\, dF(x), \tag{2.2}$$

where $\psi(x)$ is a suitable function that gives weight to the squared difference $(F_0(x) - F(x))^2$, see Kapur and Kesavan (1992).

As defined in Stephens (1974), the CVM test statistic $\left(W^2\right)$ can be computed as

$$W^2 = \sum_{i=1}^{n} [z_i - (2i-1)/2n]^2 + 1/12n, \tag{2.3}$$

with the weight function $\psi(x) = 1$, where n is sample size, and z_i is defined the same as above. For a given significance level, we reject H_0 if the CVM statistic is greater than the critical value.

AD test statistic (A^2) is another quadratic EDF statistics with the weight function $\psi(x) = [F(x)(1 - F(x))]^{-1}$. It is computed as

$$A^2 = -n - n^{-1} \sum_{i=1}^{n} (2i-1)(\log z_i + \log(1 - z_{n+1-i})), \tag{2.4}$$

where n is sample size, and z_i is defined the same as above (see Stephens, 1974). Similar to the KS test and the CVM test, it is a one-sided test and the null hypothesis will be rejected if A^2 is greater than the critical value for the given significance level.

2.2 Tests Based on Divergence Measure

First, we introduce the definition of entropy as an information measure. Given a random variable x with the probability density function $f(x)$, the measure of information content from observations in $f(x)$ is

$$\log(f(x))^{-1} = -\log(f(x)), \tag{2.5}$$

and the expected information in x is given by

$$H(x) = H(f(x)) = -E\log(f(x)) = -\int \log f(x)\, f(x)\, dx. \tag{2.6}$$

This definition is due to Shannon (1948), and it is a measure of average information and uncertainty. The larger $H(x)$ is, the less informative or more uncertain the data are.

This measure is first used in the field of thermodynamics. Since this definition of entropy associated with the concept of information, the use of it has penetrated almost all disciplines.

Shannon's entropy is the expected value of the function $g(f(x)) = -\log f(x)$, which satisfies $g(1) = 0$ and $g(0) = \infty$. As pointed out by Ullah (1996), any convex function $g(\cdot)$

with the condition that $g(1) = 0$ can be used as a measure of information content. Then, a class of g-entropies is given by

$$H_g(f(x)) = E[g(f(x))] = \int g(f(x)) f(x)\, dx. \tag{2.7}$$

A class of smooth functions g given bellow is discussed by Ullah,

$$g_\beta(f) = \begin{cases} (\beta-1)^{-1}\left(1 - f^{\beta-1}\right), & \beta \neq 1, \beta > 0, \\ -\log f, & \beta = 1, \end{cases} \tag{2.8}$$

where β is a non-stochastic constant, indexing the smooth function g.

Now we consider a divergence measure which is developed in terms of entropy. It is based on the ratio $\lambda = f_0/f_1$, where f_0 and f_1 represent two densities of models corresponding to two hypotheses. We consider the dissimilarity between two models on the basis of the divergence between these two densities f_0 and f_1. The difference in the models is large when λ is far from unity.

As an extension of the entropy function (2.7), given a convex function $g(\lambda)$ with $g(1) = 0$, the divergence measure of f_0 with respect to f_1 is then

$$H_g(f_1, f_0) = \int g(f_0(x)/f_1(x)) f_1(x)\, dx \geq 0 \tag{2.9}$$

by Jensen's inequality. This is called the relative entropy function.

According to Ullah (1996), the β-class of divergence measure can be obtained by applying the smooth function g as we discussed in (2.8). It is given as follows

$$H_\beta(f_1, f_0) = (\beta-1)^{-1} \int [1 - f_0(x)/f_1(x)]^{\beta-1} f_1(x)\, dx \text{ for } \beta \neq 1. \tag{2.10}$$

When $\beta \to 1$, we get

$$\begin{aligned} H(f_1, f_0) &= \int \log(f_1(x)/f_0(x)) f_1(x)\, dx \\ &= \int \log(f_1(x)) f_1(x)\, dx - \int \log(f_0(x)) f_1(x)\, dx \end{aligned} \tag{2.11}$$

which is the Kullback-Leibler (1951) generalization of Shannon's entropy in (2.6). When f_0 is a uniform density, (2.11) becomes the Shannon entropy. It is easier to see that (2.11) is not symmetric. Hence it is a directed divergence measure. The null hypothesis is rejected if the value of H is large.

Here, we consider two methods for estimating the Kullback-Leibler information measure. One is minimum relative entropy principle (MinxEnt) and the other is m-spacing method.

2.2.1 Minimum Relative Entropy (MinxEnt) Principle

To test $H_0 : F(x) = F_0(x)$, we consider the Kullback-Leibler information measure which measures the divergence of $f_0(x)$ with respect to $f(x)$. The MinxEnt principle is stated as follows:

Out of all probability distributions satisfying the given constraints, choose the distribution that is closest to the given distribution (Kapur and Kesavan, 1992).

Given $m+1$ constraints, by MinxEnt principle, (2.11) is minimized subject to

$$\int f(x)\,dx = 1, E(h_r(x)) = a_r, r = 1, \cdots, m, \tag{2.12}$$

where $h_r(x)$ are chosen from x^r, $(\log x)^r$, $|x - E(x)|^r$, etc. In practice, a_r is replaced by the method of moment (MM) estimator $\hat{a} = n^{-1}\sum_{i=1}^{n} h_r(x_i)$ or by the maximum likelihood estimator (MLE).

As suggested in Kapur and Kesavan (1992), to minimize (2.11), we form the Lagrangian:

$$\begin{aligned} L &= \int \log(f(x)/f_0(x)) f(x)\,dx - (\lambda_0 - 1)\left(1 - \int f(x)\,dx\right) \\ &\quad - \sum_{r=1}^{m} \lambda_r \left(a_r - \int h_r(x) f(x)\,dx\right). \end{aligned} \tag{2.13}$$

Make use of Euler-Lagrange equation,

$$F(x, f(x), f'(x)) = \log(f(x)/f_0(x)) f(x) + (\lambda_0 - 1) f(x) + \sum_{r=1}^{m} \lambda_r h_r(x) f(x) \tag{2.14}$$

and $\partial F/\partial f(x) = 0$, to obtain the function $f(x)$, which minimizes L, and hence (2.11), as

$$f(x) = f_0(x) \exp\left(-\lambda_0 - \sum_{r=1}^{m} \lambda_r h_r(x)\right), \tag{2.15}$$

where λ_r are estimated so as to satisfy all the constraints.

The goodness of fit test $H_0 : F(x) = F_0(x)$ is now equivalent to a parametric test $H_0 : \lambda_0 = \lambda_1 = \cdots = \lambda_m = 0$. Therefore, the likelihood ratio statistic

$$\begin{aligned} LR &= -2\left[\log\left(\prod_i f_0(x_i)\right) - \log\left(\prod_i f\left(x_i, \hat{\lambda}\right)\right)\right] \\ &= -2\sum_{i=1}^{n}\left[\log\left(f_0(x_i)/f\left(x_i, \hat{\lambda}\right)\right)\right] \\ &= -2\sum_{i=1}^{n}\left(\widehat{\lambda}_0 + \sum_{r=1}^{m} \widehat{\lambda}_r h_r(x_i)\right) \end{aligned} \tag{2.16}$$

can be applied for testing this hypothesis. Under H_0, LR converges to a Chi-square distribution with degree of freedom $m+1$.

2.2.2 M-Spacing Method

The methodology used in this section is introduced by Song (2002). It is a nonparametric procedure developed on the basis of the classical log-likelihood ratio test. This method is based on the difference between two sample quintiles whose indexes are $2m$ apart.

Kullback-Leibler divergence measure gives a distance between two density functions in terms of likelihood. In other words, it measures how likely the observations are from a distribution other than the posited one. As noted in Song, it is not very apparent how the log likelihood and the general problem of nonparametric goodness of fit test are connected with each other. However, as pointed out by Song, the information interpretation given for log likelihood by Akaike (1974, 1985, 1992, 1994) in connection of developing the Akaike information criterion (AIC) for model identification and the application of log likelihood as a general criterion of fit of models lead to the use of Kullback-Leibler divergence for testing goodness of fit.

To test $H_0 : F(x) = F_0(x,\theta)$, for some $\theta \in \Theta$, we consider the use of $H(f, f_0) = -H_1 - H_2$, where

$$\begin{aligned} H_1 &= -\int \log(f(x)) f(x)\, dx = \int_0^1 \log\left(\frac{d}{dp} F^{-1}(p)\right) dp, \\ H_2 &= \int \log(f_0(x,\theta)) f(x)\, dx. \end{aligned} \tag{2.17}$$

The idea by Song of estimating H_1 is to replace the differential operator by the estimated slope. He developed a m-spacing method which is based on the difference between two sample quantiles whose indexes are $2m$ apart. The entropy estimator of H_1 is then given as

$$I_1(m,n) = n^{-1} \sum_{i=1}^{n} \log\left(\frac{n}{2m}\left(x_{(i+m)} - x_{(i-m)}\right)\right), \tag{2.18}$$

and H_2 is estimated by

$$I_2 = n^{-1} \sum_{i=1}^{n} \log\left(f_0\left(x_{i,}\hat{\theta}\right)\right), \tag{2.19}$$

where $\hat{\theta}$ is the MLE of θ, n is the sample size, m is the order of spacings, $x_{(i)}$ denotes the ith order statistic of the sample, $x_{(j)} = x_{(1)}$ if $j < 1$ and $x_{(j)} = x_{(n)}$ if $j > n$. The null hypothesis is rejected when $(-I_1(m,n) - I_2)$ is large. In this study, the null hypothesized distribution $F_0(x,\theta)$ is normal distribution $N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$. It is well known that MLE of θ is $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)$, where

$$\hat{\mu} = n^{-1} \sum_{i=1}^{n} x_i \text{ and } \hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 .$$

Then the Kullback-Leibler divergence measure $H = -H_1 - H_2$ is estimated by $I_{mn} = -I_1(m,n) - I_2$. This is standardized as

$$S_{mn} = \sqrt{6mn}\left(I_{mn} - \log(2m) - \gamma + R_{2m-1}\right), \tag{2.20}$$

where $R_{2m-1} = \sum_{j=1}^{m} j^{-1}$ and $\gamma = \lim_{n\to\infty}(R_n - \log n) \approx 0.577215665$ is the Euler constant. Assume that

$$m(\log n)^{-1} \to \infty, \text{ and } mn^{-1/3}\log(n)^{2/3} \to 0, \text{ as } n \to \infty, \tag{2.21}$$

then under H_0, $S_{mn} \xrightarrow{D} N(0,1)$ as $n \to \infty$. These two assumptions are made on the smoothing parameter m which implies that the order of spacings m should tend to infinity slower than $n^{1/3}$ as $n \to \infty$.

2.3 Test Based on the Integrated Difference Estimated by Kernel Method

In the previous subsection, we talked about the test statistics based on the Kullback-Leibler information which is a directed divergence measure. Another class of divergence measures is developed in the following ways. Define

$$I_g(f_1, f_0) = H_g(f_1, f_0) + H_g(f_0, f_1), \tag{2.22}$$

$$K_g(f_1, f_0) = 2H_g\left(\frac{f_1 + f_0}{2}\right) - H_g(f_0) - H_g(f_1), \tag{2.23}$$

$$J_g(f_1, f_0) = \int (f_1 - f_0)(g(f_0) - g(f_1))\,dx, \tag{2.24}$$

where $H_g(f_1, f_0)$ is defined as (2.9). This class of divergence measures is symmetric and is referred to as the non-directional divergence measures.

A special case of $J_g(f_1, f_0)$ in (2.24) with g given by (2.8) ($\beta = 2$) is

$$J(f_1, f_0) = \int (f_1 - f_0)^2\,dx. \tag{2.25}$$

This can be used to test the goodness of fit $H_0 : F(x) = F_0(x, \theta)$ against the general alternative $H_1 : F(x) \neq F_0(x, \theta)$ for all $\theta \in \Theta$.

The test considered by Fan (1994) is based on the integrated squared difference between a kernel estimate of $f(x)$ and the quasi-maximum likelihood estimate (QMLE) of $f_0(x, \theta)$. Specifically, the estimator of $J(f_1, f_0)$ is obtained as follows,

$$J_n = \int \left(\tilde{f}(x) - f_0(x, \hat{\theta})\right)^2 dx, \tag{2.26}$$

where under H_0, $f_0(x, \theta)$ is estimated by $f_0(x, \hat{\theta})$, where $\hat{\theta}$ is the QMLE, and the true, unknown density function $f(x)$ is estimated by the kernel estimator $\tilde{f}(x) = (nh)^{-1}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$.

$K(.)$ is a kernel function. h is a smoothing parameter: the larger the h is, the smoother the estimated density which has larger bias and smaller variance; the smaller the h is, the less smooth the estimated density, and the density has smaller bias and larger variance.

And the asymptotic test statistic is given by

$$T_n = (2h)^{-1/2} \tilde{\sigma}^{-1} (nhJ_n - I(K)) . \qquad (2.27)$$

If $nh^{1/2+4} \to 0$, then, under H_0, $T_n \xrightarrow{D} N(0,1)$. The test statistic apparently depends on the choice of h. Although it is well-known that the optimal choice of h is of order $n^{-1/5}$ for density estimation, see Silverman (1986) for example, there is no simple data-driven method to select the optimal h for the test. Therefore, we take Fan's (1994) suggestion and use $h = h_0 \widehat{\sigma} n^{-\delta}$, where $\widehat{\sigma}$ is the standard deviation of the data, h_0 and δ are positive constant and $0 < \delta < 1$.

We have introduced above six statistics with the associated methods for applying to test the goodness of fit problem. We examine the finite-sample performances of these tests by a Monte Carlo study in the following section.

3 Monte Carlo Simulation Study

In this section we carry out Monte Carlo simulations to evaluate the performance of the goodness of fit tests discussed in Section 2 above. In particular, we study the behavior of various statistics by examining the estimated size and power of the tests. Throughout this section, the null hypothesis considered is

$$f(x) = f_0(x, \theta) = \left(\sqrt{2\pi}\sigma\right)^{-1} \exp\left(-(x-\mu)^2 / 2\sigma^2\right), \qquad (3.1)$$

where $\theta = \left(\mu, \sigma^2\right)$ is unspecified. To see how the six test statistics perform for different sample sizes, we set the sample size to be 60, 100, 200, and 600. The number of Monte Carlo simulations is 1,000. In addition, asymptotic distribution of a test statistic may not be a good approximation of the distribution of the test in finite samples, and this questions the use of critical values derived from the asymptotic theory when the sample size is not large. Therefore, for each experiment run below, we also construct size-corected critical values from 5,000 Monte Carlo simulations. The details are given below.

3.1 Comparison by Estimated Sizes

We first study the three EDF test statistics. Table 1 comes from Stephens (1974). It contains the formulas and critical values which are given for the use of the modified KS statistic (D), CVM statistic (W^2) and AD statistic (A^2). In practice, for a given data set, we first calculate the modified form T^* corresponding to T in column 1. If T^* exceeds a value in column 3 for level α, reject H_0 at the significance level α. For example, if the value of modified T^* for KS test exceeds 0.819, we reject H_0 at α=10%.

Table 1: Modifications to the EDF Statistics and Their Critical Values

Statistic T	Modified form T^*	α		
		10%	5%	2.5%
D	$D^* = D\left(\sqrt{n} - 0.01 + 0.85/\sqrt{n}\right)$	0.819	0.895	0.955
W^2	$W^{*2} = W^2\left(1 + 0.5/n\right)$	0.104	0.126	0.148
A^2	$A^{*2} = A^2\left(1 + 4/n - 25/n^2\right)$	0.656	0.787	0.918

By using the Fortran program, we generate 1,000 replicate samples of sizes n=60, 100, 200, 600 from the **standard normal distribution**. T^* is calculated from each sample. The size is then estimated by the percentage rejection of the true H_0 according to the given significance level. The results are reported in Table 2.

Table 2 indicates that CVM statistic W^{*2} has the closest percentage rejection to the given significance level for all the sample sizes in the study. It is seen that KS statistic D^* gives a very good approximation to the significance level α when sample size n = 200.

The study for evaluating the performance of LR statistic is carried out based on 1,000 Monte Carlo simulations for sample size n=60, 100, 200, 600 with the critical values from Chi-square table. If LR exceeds the upper $100\,(1-\alpha)$ percentile point of a Chi-square table with 5 degree of freedom (four moment restrictions are imposed as restrictions), H_0 is rejected at significant level α. The percentage rejections of the true H_0 according to the given significance level are shown in the second column of Table 3.

Using asymptotic critical values from the Chi-square distribution with a degree of freedom 5, Table 3 indicates that at each significant level, the rejection rates increases with the increasing of sample size. Small sample size leads to under-rejection of the null hypothesis and large sample size results in over-rejection of the null hypothesis.

Except the use of critical values from the Chi-square table, we can calculate them by Monte Carlo simulations. The Monte Carlo simulated critical values is also called size-corrected critical values, because they are different for different sample sizes. To determine the critical values of the test statistic, we generate 5,000 replicate samples at each sample size. From each sample, LR is calculated. The upper tail percentage points LR^* of the distribution of LR are then estimated by the $(1-\alpha)^{th}$ percentiles of the empirical distribution function of LR based on the observed samples. Once these critical values have been determined, the sizes of LR are estimated by Monte Carlo simulations. That is, at each sample size, we generate 1,000 samples under the null hypothesis and the size is then estimated by the proportion of the samples falling into the critical region $\{LR \geq LR^*\}$. These estimated sizes of the test are reported in the third column of Table 3 under the title of using Monte Carlo simulated critical values.

Comparing the estimated sizes using asymptotic critical values with those using Monte Carlo simulated critical values, we see that with critical values based on Monte Carlo simula-

Table 2: Estimated Sizes of Three EDF Statistics

T^*	n	10%	5%	2.5%
	60	9.2	4.6	2.7
D^*	100	9.2	4.8	3.1
	200	10.4	5.1	2.2
	600	11.1	5.8	2.9
	60	8.9	5.0	2.4
W^{*2}	100	9.3	4.2	2.1
	200	9.1	4.7	2.6
	600	9.3	5.0	1.8
	60	8.8	4.8	2.7
A^{*2}	100	9.0	3.9	1.5
	200	8.9	4.8	2.0
	600	8.5	3.3	1.5

Table 3: Estimated Size of LR Statistic

n	Using Asymptotic Critical Values			Using Monte Carlo Simulated Critical Values		
	10%	5%	2.5%	10%	5%	2.5%
60	4.9	3.0	2.2	10.9	4.1	1.8
100	6.2	4.4	2.8	10.6	4.2	1.8
200	8.2	5.1	3.8	9.6	4.5	2.4
600	11.7	7.3	4.8	9.2	3.9	1.7

Table 4: Estimated Size of Song's (2002) Statistic I

n	$\widehat{m}$	Using Asymptotic Critical Values			Using Monte Carlo Simulated Critical Values		
		10%	5%	2.5%	10%	5%	2.5%
60	3	13.2	9.5	6.2	13.4	4.1	2.2
100	4	11.6	7.9	5.4	11.5	5.2	2.8
200	6	11.4	7.3	4.6	16.2	5.1	2.4
600	12	3.2	1.7	0.9	14.9	6.6	4.1

tions, the estimated sizes are closer to the given significant levels than the use of Chi-square table.

Next, we calculate Song's (2002) test statistic I. When applying his m-spacing method, we need to specify the order of spacings m. The asymptotic theory suggests that m should be chosen according to the sample size. As suggested by Song (2002), based on large values of H favoring the alternative hypothesis, we chose m which minimizes the sample estimator I_{mn} and satisfies $I_{mn} \geq 0$. By the data-driven method,

$$\hat{m} = \min \left\{ m^* : m^* = \arg\max_m \left(I_1\left(m,n\right) : I_1\left(m,n\right) \leq -n^{-1} \sum_{i=1}^{n} \log f_0\left(x_i, \hat{\theta}\right) \right) \right\}. \quad (3.2)$$

In practice, for each sample, we choose m^* from the range $[1, n/2-1]$ which maximizes the estimated Kullback-Leibler divergence when the estimated value is positive. Then, define $\hat{m}$ as the smallest m^* over 1,000 samples.

In addition, although Song's S_{mn} test statistic converges to a standard normal distribution when sample size is sufficiently large, Song (2002) pointed out that this statistic may have substantial finite-sample bias. To correct this problem in finite samples, he suggests to reject H_0 when

$$I_{mn} \geq E\left(U_{mn}\right) + \left(6mn\right)^{-1/2} Z_{1-\alpha}, \quad (3.3)$$

where

$$E\left(U_{mn}\right) = \log\left(2m\right) - \log\left(n\right) + R_n - R_{2m-1} + \frac{2m}{n} R_{2m-1} - \frac{2}{n} \sum_{i=1}^{m} R_{i+m-2} \quad (3.4)$$

and $R_m = \sum_{i=1}^{m} j^{-1}$.

To obtain the estimated size of I_{mn} statistic, we use the same procedure as we did for Table 3 with the critical values $Z_{1-\alpha}$ from the standard normal distribution table and with Monte Carlo simulated critical values. The results and the selected $\hat{m}$ are reported in Table **??**. Comparing the estimated sizes using different critical values, it seems that using the critical values obtained from Monte Carlo simulations gives a better approximation to the sizes of Song's I statistic especially at 5% and 2.5% significant levels.

The last test of goodness of fit we consider in this study is Fan's (1994) T statistic which has approximate standard normal distribution under H_0. For the standard normal kernel $K(z) = (2\pi)^{-1/2} \exp\left(-z^2/2\right)$, Fan (1994) showed that T_n had the following form:

$$T_n = (2h)^{-1/2} \tilde{\sigma}^{-1} \left(nhJ_n - I(K)\right) \tag{3.5}$$

where

$$I(K) = \int K^2(z)\,dz = \left(2\sqrt{\pi}\right)^{-1}, \tilde{\sigma} = J(K)\int \tilde{f}^2(x)\,dx,\ J(K) = \left(2\sqrt{2\pi}\right)^{-1}$$

$$J_n = \int \tilde{f}^2(x)\,dx - 2\int \tilde{f}(x) f_0\left(x,\hat{\theta}\right)dx + \int f_0^2\left(x,\hat{\theta}\right)dx = J^1 + J^2 + J^3$$

with

$$\begin{aligned} J^1 &= \int \tilde{f}^2(x)\,dx = \left(2\sqrt{\pi}h\right)^{-1} n^{-2}\sum_i\sum_j \exp\left(-\left(4h^2\right)^{-1}\left(x_i - x_j\right)^2\right) \\ J^2 &= \int \tilde{f}(x) f_0\left(x,\hat{\theta}\right)dx = n^{-1}\left[2\pi\left(h^2+\hat{\sigma}^2\right)\right]^{-1/2}\sum_i \exp\left(-\frac{\left(x_i-\hat{\mu}\right)^2}{2\left(h^2+\hat{\sigma}^2\right)}\right) \\ J^3 &= \int f_0^2\left(x,\hat{\theta}\right)dx = \left(2\sqrt{\pi}\hat{\sigma}\right)^{-1}. \end{aligned}$$

As discussed in Pagan and Ullah (1999), the test statistic T_n has a center term that may contribute to some finite sample bias. This stems from the inclusion of the "diagonals" terms $(i = j)$ in J^1. To eliminate such an effect they introduced a modified test statistic,

$$T_{1n} = \left(\sqrt{2}\tilde{\sigma}\right)^{-1/2} nh^{1/2}\left(J_n - \left(2\sqrt{\pi}nh\right)^{-1}\right). \tag{3.6}$$

It is found that under H_0, $T_{1n} \xrightarrow{D} N(0,1)$.

For a given test, different choice of the parameter h may lead to different conclusion for a given data set. We choose h according to $h = h_0\widehat{\sigma}n^{-\delta}$, where $\widehat{\sigma} = n^{-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$. In this study, we selected the same values of h_0 and δ as in Fan (1994).

The simulation results are reported in Table 5 with the critical values from the standard normal table. The Monte Carlo experiment is similar to the procedure we described before. From this table, it seems that small values of h_0 lead to under-rejection of the null hypothesis. The estimated sizes are close to the given significant levels with the sample size increasing.

The percentage rejections with the critical values from Monte Carlo simulations are summarized in Table 6. Comparing the results with those in Table 5, it can be seen that the later gives a better approximation of the significant level for large values of h_0 at each value of δ. Among the choices of parameters we considered, $\delta = 1/4, h_0 = 1.9$ gives a better approximation to the given significant levels except for the small sample size. According to

Table 5: Estimated Size of Fan's (1994) Statistic T (Using Asymptotic Critical Values)

n	10%	5%	2.5%	10%	5%	2.5%	10%	5%	2.5%
	$\delta = 2/7, h_0 = 2.25$			$\delta = 2/7, h_0 = 2.35$			$\delta = 2/7, h_0 = 2.4$		
60	1.7	0.2	0.1	4.7	0.6	0.1	7.1	1.4	0.2
100	3.3	0.7	0.2	6.9	2.2	0.5	8.8	3.0	0.7
200	4.8	2.1	0.8	7.4	4.0	1.7	8.9	4.6	2.3
600	5.9	2.8	1.3	8.2	4.4	1.7	9.8	5.6	2.8
	$\delta = 1/4, h_0 = 1.8$			$\delta = 1/4, h_0 = 1.9$			$\delta = 1/4, h_0 = 1.95$		
60	0.3	0.1	0.0	0.9	0.1	0.0	1.7	0.2	0.1
100	1.8	0.4	0.0	2.9	0.7	0.2	5.3	1.4	0.5
200	3.7	1.3	0.4	6.1	2.7	1.0	7.6	4.1	1.9
600	6.1	3.0	1.4	9.6	5.2	2.8	12.0	6.3	3.4

this result, we choose $h = 1.9\widehat{\sigma}n^{-0.25}$ as the smooth parameter when testing the normality under the alternative distribution. This confirmed Fan (1994)'s suggestion of the choice of h.

Throughout the experiments to estimate the sizes of the tests under 10%, 5% and 2.5% significant levels, the average standard errors in estimations are 0.259 at 10% significant level, 0.167 at 5% significant level and 0.113 at 2.5% significant level.

3.2 Power Comparison

For power comparison, we investigate seven cases with non-normal distributions as alternatives. They are chosen from the generalized lambda distributions (GLD) discussed in Ramberg and Schmeiser (1974) including three symmetric and four asymmetric distributions. These distribution families are based on the inverses of the cumulative distribution functions and can be easily generated:

$$F^{-1}(z) = \lambda_1 + \lambda_2^{-1}\left[z^{\lambda_3} - (1-z)^{\lambda_4}\right], \qquad (3.7)$$

where $0 < \lambda < 1$. Table 7 comes from Fan (1994) which contains the seven selected alternative distributions defined by the parameters, along with the associated mean (μ), variance (σ^2), coefficient of skewness (α_3) and coefficient of kurtosis (α_4) values.

To evaluate the six test statistics when the null hypothesis is false, we use the critical values from Table 1 for D, W^2, A^2, and from the standard normal table for T. According to the performance of I and LR under the null hypothesis, Monte Carlo simulations are considered for determining the critical values which are based on 5,000 replicate samples

Table 6: Estimated Size of Fan's (1994) Statistic T (Using Monte Carlo Critical Values)

n	10%	5%	2.5%	10%	5%	2.5%	10%	5%	2.5%
	$\delta = 2/7, h_0 = 2.25$			$\delta = 2/7, h_0 = 2.35$			$\delta = 2/7, h_0 = 2.4$		
60	4.9	0.9	0.3	4.6	0.8	0.3	4.6	0.8	0.3
100	6.1	1.5	0.4	5.9	1.1	0.3	5.9	1.1	0.3
200	7.8	3.1	1.5	7.6	3.0	1.4	7.5	3.1	1.4
600	8.6	3.6	1.7	8.0	3.5	1.4	8.1	3.4	1.4
	$\delta = 1/4, h_0 = 1.8$			$\delta = 1/4, h_0 = 1.9$			$\delta = 1/4, h_0 = 1.95$		
60	15.5	1.5	0.3	5.0	1.2	0.3	4.9	0.9	0.3
100	11.5	1.8	0.5	6.1	1.5	0.4	5.9	1.3	0.3
200	10.7	3.1	1.3	7.7	3.3	1.5	7.6	3.0	1.4
600	8.3	3.6	1.7	8.1	3.4	1.4	8.0	3.2	1.3

Table 7: Seven Alternative Distributions Used in the Simulation Study

Case	Symmetric Distributions							
	λ_1	λ_2	λ_3	λ_4	μ	σ^2	α_3	α_4
1	0	2	1	1	0	0.0833	0	1.80
2	0	-0.3970	-0.16	-0.16	0	1.0001	0	11.61
3	0	-1	-0.24	-0.24	0	0.5323	0	126.89
Case	Asymmetric Distributions							
	λ_1	λ_2	λ_3	λ_4	μ	σ^2	α_3	α_4
4	0	1	1.4	0.25	-0.3833	0.2107	0.51	2.22
5	3.5865	0.0430	0.0252	0.0949	5.0114	5.0853	0.89	4.28
6	0	-1	-0.0075	-0.03	0.0234	0.0014	1.50	13.66
7	-0.1167	-0.3516	-0.13	-0.16	0	1.0001	0.76	11.43

from normal distribution under the corresponding null hypothesis. After determining the critical values, for each alternative, 1,000 samples of size n = 60, 100, 200, 600 are generated from the alternative distribution and the power is then estimated by the proportion of the samples falling into the critical region. In the experiments, the smooth parameter of test statistic T is chosen as $h = 1.9\widehat{\sigma}n^{-0.25}$ which we have discussed in the previous section.

As shown in Table 7, three symmetric distributions defined by lambdas are considered as case 1, case 2 and case 3. The probability density functions under the alternatives, together with the normal distributions under the null hypothesis which have the same mean and variance as the corresponding lambda distributions are given in Figure 1. Solid lines present the alternative GLDs, dashed lines present the normal distributions.

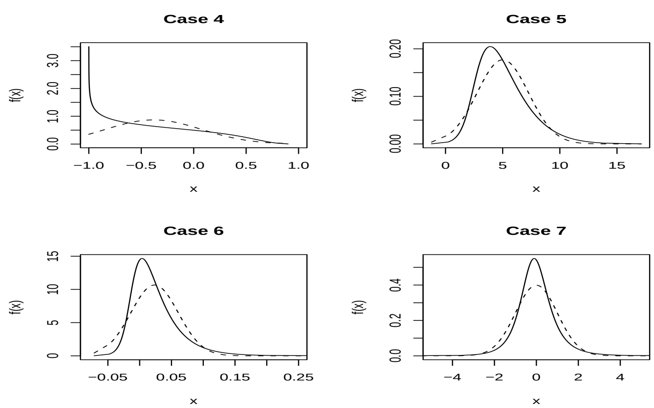

Figure 1: PDFs of Normal Distributions vs. Those of First Three GLDs

The GLD as alternative in case 1 is actually the uniform distribution $U(-0.5, 0.5)$. The null hypothesis we considered in this case is the normal distribution with mean 0 and variance 0.0833 which are the same as case 1 distribution. As we can see in Table 8, at sample size n = 60, the estimated powers of the statistics are less than 76% except for Song's statistic I which is 96.1%. It is much higher than the others and attains 1 at sample size n =100. The most used KS statistic appears the least powerful against this alternative in case 1.

The alternative distributions in case 2 and case 3 have similar shapes. From Figure 1, we can see clearly that both of them have a peak near the mean and fat tails. This can also been seen from the associated kurtosis values in Table 7, which are 11.61 and 126.9 respectively. The estimated powers of the six statistics are reported in Table 8. It shows that the LR statistic is more powerful against the alternatives in case 2 and case 3. The EDF statistics perform good detections as well. In these two cases, Fan's statistic T has little power for small sample size n =60, 100.

Four asymmetric GLD are considered as cases 4-7. The solid lines in Figure 2 present the plots of these distributions. The dashed lines present normal distributions under the corresponding null hypothesis. Case 6 and case 7 distribution are heavy tailed with the kurtosis value 13.66 and 11.43 respectively.

Table 8: Percentage Rejections: Three Alternatives for Cases 1-3 (at the 5% Level)

n	EDF			Directed Divergence Measure		Non-directional
	KS	CVM	AD	Spacing Method	MinxEnt	Divergence
	D	W^2	A^2	I	LR	Measure T
	Case 1: mean 0, variance 0.08, skewness 0, and kurtosis 1.8					
60	35.5	59.3	75.3	96.1	32.2	68.7
100	60.5	87.0	95.9	100	89.8	98.7
200	96.1	99.8	100	100	100	100
600	100	100	100	100	100	100
	Case 2: mean 0, variance 1.0001, skewness 0, and kurtosis 11.6					
60	35.0	51.6	50.8	40.2	64.7	0.0
100	55.9	69.0	74.0	56.2	80.2	0.6
200	83.7	93.7	94.2	78.2	97.3	6.6
600	99.9	100	100	90.5	100	89.5
	Case 3: mean 0, variance 0.53, skewness 0, and kurtosis 126.9					
60	50.5	71.1	68.1	57.7	79.1	0.2
100	76.3	87.0	89.6	76.9	92.9	2.6
200	95.6	99.3	99.0	93.5	99.4	27.1
600	100	100	100	96.0	100	99.9

Table 9: Percentage Rejections: Four Alternatives for Cases 4-7 (at the 5% Level)

n	EDF			Directed Divergence Measure		Non-directional
	KS	CVM	AD	Spacing Method	MinxEnt	Divergence
	D	W^2	A^2	I	LR	Measure T
	Case 4: mean -0.38, variance 0.21, skewness 0.51, and kurtosis 2.2					
60	67.6	82.1	93.6	99.1	55.8	57.2
100	92.1	98.1	99.7	100	95.3	97.3
200	99.9	100	100	100	100	100
600	100	100	100	100	100	100
	Case 5: mean 5.01, variance 5.09, skewness 0.89, and kurtosis 4.28					
60	47.1	51.2	62.8	43.1	60.5	0.6
100	68.0	77.7	84.0	64.7	85.8	10.9
200	95.0	98.6	98.0	90.3	98.7	66.7
600	100	100	100	99.7	100	100
	Case 6: mean 0.02 variance 0.0014, skewness 1.5, and kurtosis 13.66					
60	70.5	77.8	86.7	69.6	85.9	0.6
100	90.5	95.9	96.3	90.5	97.9	22.8
200	99.8	99.9	100	99.6	100	90.1
600	100	100	100	100	100	100
	Case 7: mean 0, variance 1.0001, skewness 0.76, and kurtosis 11.43					
60	39.0	46.5	52.2	37.4	60.6	0.0
100	58.0	68.3	74.0	54.8	80.0	0.7
200	84.8	92.6	94.3	76.7	96.4	6.7
600	100	100	100	90.3	100	92.4

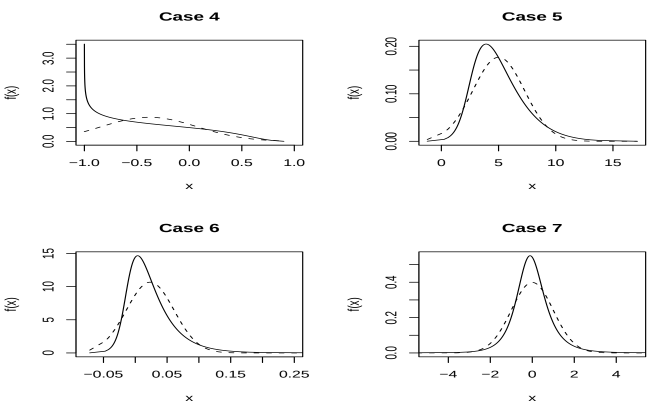

Figure 2: PDFs of Normal Distributions vs. Those of Last Four GLDs

As shown in Table 9, most of the estimated powers attain 1 at sample size 200 for case 4. At small sample size n = 60, 100, Song's test statistics I is more powerful against the alternative in this case. For the fifth and the sixth cases, the AD test is more powerful than the test based on the LR statistic when sample size is 60, but the test based on LR dominates when the sample size n =100, 200, 600. Statistic LR shows the highest estimated power against the alternative in case 7 as well. Fan's statistic T appears to have no power in the last three cases at sample size 60.

4 Conclusion and Discussion

The goodness of fit problem is very important in experimental analysis. Among a large number of tests for judging normality, no one test is optimal for all possible deviations from normality. In this paper, we compared six statistics for testing the goodness of fit of a parametric density function. The first three statistics, based on EDF, are simple and widely used. The remaining three are based on the divergence measure and applied using the MinxEnt principle, the m-spacing method and the kernel method. We investigated the sensitivity of these tests by applying the tests to a generalized lambda distribution under the alternatives.

In evaluating their performance under the null hypothesis, the critical values obtained from the Monte Carlo simulations are shown to give a closer rejection rates to the significant levels for tests I and LR. However, comparing Table 5 and Table 6, it is hard to say by which method the critical values work well for the statistic T.

To examine the power performances, we compared the powers against seven alternatives in Tables 8 and 9. Several conclusions can be drawn from this table. First, for a fixed sample size n, the power depends on how far the alternative distribution is away from the normal distribution. Second, for a given alternative, the power increases with increasing the sample size. Third, for a large sample size, the powers of most statistics in this study increased to one. The statistic T has no power against the last six alternatives at sample size $n = 60$. The Kullback-Leibler divergence based statistics I works well for the first and the fourth distributions and LR statistic works well for the other distributions in our study. According to the associated kurtosis, it seems that LR test based on Kullback-Leibler divergence will better detect a distribution with a peak near the mean and with heavy tails.

This paper investigated several statistics for testing for normality. It is of interest to consider the null hypothesis for other than the family of normal distributions such as the null of uniform distribution which can be equivalent to a test for the form of a parametric density function. If F is the true distribution of x, then $F(x) \sim U(0,1)$. The exponential family is also of interest to be considered as the null distribution because it is of fundamental importance in the modeling and analysis of transition data. It is also important to examine how these tests work for testing multivariate distributions.

References

[1] Ait-Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrics, 64, 385-426.*

[2] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr., 19, 716-723.*

[3] Akaike, H. (1985). Prediction and Entropy. In A Celebration of Statistics, The ISI Centenary Volume, A.C. Atkinson and S.E. Fienberg, Eds. New York: Springer-Verlag, pp. 1-24.

[4] Akaike, H. (1992). Information theory and an extension of the maximum likelihood principle. in Breakthroughs in Statistics, Foundations and Basic Theory, S. Kotz and N.L. Jognson, Eds. New York: Springer-Verlag, vol. I, pp 610-624.

[5] Akaike, H. (1994). Implications of informational point of view on the development of statistical science. In Proc. 1st US/Japan Conf. Frontiers of Statistical Modeling: An Informational Approach, pp. 23-38.

[6] Cressie, N. and T.R.C., Read (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, 46 (Ser. B), 440-464.*

[7] D'Agostino, R.B., and M.A. Stephens (1986). Goodness-of-Fit Techniques. New York: Marcel Dekker.

[8] Dudewicz, E.J., and E.C. van Der Meulen (1981). Entropy-based tests of uniformity. *Journal of the American Statistical Association, 76, 967-974.*

[9] Fan, Y.Q. (1994). Testing the goodness of fit of a parametric density function by kernel method. *Econometric Theroy, 10, 316-356.*

[10] Fredholm, B., K. Gummarson, G. Jensen, and J. Muntzing (1978). Mammary tumor inhibition and subacute toxicity in rats of prdrimostine and of its molecular components chlorambuic and prednisolone. *Acta Pharmacol. Et Toxical., 42, 159-163.*

[11] Freeman, M.F. and J.W. Tuke (1950). Transformations related to the angular sand the square Root. *Annals of Mathematical Statistics, 21, 607 -611.*

[12] Griffin, M.J. and E.M. Whitham (1978). Individual variability and its effect on subjective and biodynamic response to whole body vibration. *Journal of Sound and Vibration, 58, 239-250.*

[13] Kapur, J.N. and H.K. Kesavan (1992). Entropy Optimization Principles with Applications. Academic Press, San Diego.

[14] Kullback, S. and R.A. Leibler (1951). On information and sufficiency. *Ann. Math. Statist., 22, 79-86.*

[15] Ozturk, O. and T.P. Hettmansperger (1997). Generalised weighted Cramer-von Mises distance estimators. *Biometrika, 84,283-294.*

[16] Pagan A. and A. Ullah (1999). Nonparametric Econometrics. Cambridge University Press, Cambridge, UK.

[17] Parzen, E. (1985). Entropy interpretation of goodness-of-fit tests. Manuscript, Texas A&M University.

[18] Ramberg, J.S. and B.W. Schmeiser (1974). An approximate method for generating asymmetric random variables. *Communications of the Association for Computing Machinery, 17, 78-82.*

[19] Rao, K.C. and D.S. Robson (1974). A chi-square statistic for goodness-of-fit within the exponential family. *Communications Statisitcs, 3, 1139-1153.*

[20] Rosenblatt, M.A. (1975). Quadratic measure of deviation of two-dimensional density estimated and a test of independence. *Annals of Statistics, 3, 1-14.*

[21] Shannon, C.E. (1948). The Mathematical Theory of Communication. Bell System Tech. J. and in Shannon and Weaver, (1949). The Mathematical Theory of Communication. University of Illinois, Urbana, 3-91.

[22] Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis. New York, Chapman and Hall.

[23] Smith, T.J., A.R. Temple, and J.C. Reading (1976). Cadmium, lead and copper blood levels in normal children. *Clinical Toxicology, 9, 75-87.*

[24] Song, K. (2002). Goodness-of-fit test based on Kullback-Leibler discrimination information. *IEEE Transactions on Information Theory, 48, 1103-1117.*

[25] Stephens, M.A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association, 69, 730-737.*

[26] Ullah, A. (1996). Entropy, divergence and distance measures with econometric applications. *Journal of Statistical Planning and Inference, 49, 137-162.*