Youngnam Statistical Letters
June 1993, Vol. 4, 137-146

Goodness-of-Fit Test Based on Smoothing Parameter Selection Criteria

Jong Tae Kim[1]

ABSTRACT

The proposed goodness-of-fit test statistic $\hat{\lambda}_\alpha$ derived from the test statistic in Kim (1992) is itself a smoothing parameter which is selected to minimize an estimated MISE for a truncated series estimator, $d_{\hat{\lambda}n}$, of the comparison density function. Therefore, this test statistic leads immediately to a point estimate of the density function in the event that H_0 is rejected. The limiting distribution of $\hat{\lambda}_\alpha$ was obtained under the null hypothesis. It is also shown that this test is consistent against fixed alternatives.

1. Introduction

The goodness-of-fit (GOF) tests have been of interest by numerous statisticians and there are many popular tools, such as the Cramér-von Mises (CVM) statistic, Kolmogorov-Smirnov test statistic, Shapiro-Wilk test statistic and etc.. In this article we describe some new goodness-of-fit tests which are based on nonparametric density estimation with a data-driven smoothing parameter.

Let $X_1, X_2, \cdots, X_n$ be independent, identically distributed (*iid*) random variables with an absolutely continuous distribution function (*df*) F. The classical GOF problem concerns testing $H_0 : F = G$, for some specified absolutely continuous *df* G and the same support as F.

[1] Department of Statistics, Texas A & M University, Texas, U.S.A..

Define the quantile function for G by $G^{-1}(u) = \inf\{x : G(x) \geq u, \quad 0 < u < 1\}$. It is easily obtained that testing $H_0 : F = G$ is equivalent to testing $H_0 : F^{-1} = G^{-1}$. Let D be the *df* of $Y = G(X)$. Then, the GOF hypothesis is equivalent to $H_0 : D(u) = u, \quad 0 \leq u \leq 1$. However, $D(u) = F(G^{-1}(u))$, with corresponding density function $d(u) = f(G^{-1}(u))/g(G^{-1}(u)), \quad 0 < u < 1$. The function d in this case is called the *comparison density function* (Parzen, 1979). Now, by using the comparison density function for comparing F and G, the GOF hypothesis becomes equivalent to $H_0 : d(u) = 1, \quad 0 \leq u \leq 1$.

We discuss the problem of comparison density estimation in Section2 and choosing an optimal smoothing parameter in Section 3. We propose test based on nonparametric density estimation in Section 4 and obtain its asymptotic null distribution. This test statistic is, in fact, a data-driven smoothing parameter. The simulation problem is in Section 5 and its summary and conclusion are in Section 6. The theorems are then proved in Section 7.

2. Comparison Density Estimation

Recall that the hypothesis $F = G$ is equivalent to $D(u) = u$ for all u in [0, 1]. Assume that D is square integrable on [0, 1]. Then, since $D(u) - u$ vanishes at 0 and 1, a natural representation for this difference is its Fourier sine series

$$D(u) - u = \sum_{j=1}^{\infty} \beta_j \sqrt{2} \sin(j\pi u), \quad 0 \leq u \leq 1, \tag{2.1}$$

where the Fourier coefficients β_j are determined by

$$\beta_j = \sqrt{2} \int_0^1 \Big(D(u) - u \Big) \sin(j\pi u) du.$$

On differentiating (2.1), one can formally obtain d, a square integrable density on [0, 1], possessing a Fourier cosine series expansion

$$d(u) - 1 = \sum_{j=1}^{\infty} a_j \sqrt{2} \cos(j\pi u), \tag{2.2}$$

where

$$a_j = \sqrt{2} j\pi \int_0^1 \Big(D(u) - u \Big) \sin(j\pi u) du \tag{2.3}$$
$$= \sqrt{2} \int_0^1 \cos(j\pi u) d(u) du.$$

From this viewpoint, the a_j's emerge as parameters describing the discrepancy between $d(u)$ and 1 as well as between $D(u)$ and u. Thus, the null hypothesis, $F = G$, is again seen as equivalent to $a_j = 0$ for all j and this fact leads to testing subhypotheses about the a_j as advocated by Eubank, LàRiccia and Rosenstein (1987) from a Pearson's ϕ^2 perspective.

The Fourier series expansion in (2.2) actually gives a series expansion for the comparison density of the form

$$d(u) = 1 + \sum_{j=1}^{\infty} a_j \sqrt{2} \cos(j\pi u) \tag{2.4}$$

with the a_j's being unknown parameters. Thus, the comparison density can be estimated by first truncating the series after λ terms and then plugging in estimates for the a_j's. An unbiased, $\sqrt{n}$ -consistent estimator of a_j is provided

$$\tilde{a}_{jn} = \frac{1}{n} \sum_{r=1}^{n} \sqrt{2} \cos(j\pi Y_r), \tag{2.5}$$

which is obtained by replacing $D(u)$ in (2.3) by $D_n(u)$, the empirical distribution function for $Y_1, Y_2, \cdots, Y_n$. If $0 \le \lambda \le n$ is some integer, then a Fourier cosine series estimator $d_{\lambda n}$ of the comparison density d is

$$d_{\lambda n}(u) = 1 + \sum_{j=1}^{\lambda} \tilde{a}_{jn} \sqrt{2} \cos(j\pi u). \tag{2.6}$$

The integer λ in (2.6) is the smoothing parameter for the density estimator. The choice of an optimal smoothing parameter for both density estimation and testing is the central theme of this article.

3. Selecting an Optimal Smoothing Parameter

To assess the performance of a density estimator, we focus on the Mean Integrated Squared Error (MISE), a global measure of the discrepancy between the estimated and true density. Define the MISE to be $R(\lambda) = EL(\lambda)$, where $L(\lambda) = \int_0^1 (d_\lambda(u) - d(u))^2 du$ is often called the Integrated Squared Error (ISE).

The method for choosing λ that will be focused on here is the data driven procedure proposed by Hart (1985). This method chooses a smoothing parameter value that minimizes an estimator of the MISE.

Initially, we derive an alternative expression for $R(\lambda)$ and partially characterize its minimizer. For this purpose, observe that by Parseval's identity

$$L(\lambda) = \sum_{j=1}^{\lambda}(\tilde{a}_{jn}^2 - 2a_j\tilde{a}_{jn}) + \sum_{j=1}^{\infty} a_j^2.$$

This allows us to establish the following result.

Proposition 3.1. The MISE for the comparison density is given by

$$R(\lambda) = -M(\lambda) + \sum_{j=1}^{\infty} a_j^2, \tag{3.1}$$

where

$$\sigma_j^2 = \mathrm{var}(\tilde{a}_{jn}) = \frac{1}{n}\left(1 + \frac{1}{\sqrt{2}}a_{2j} - a_j^2\right) \tag{3.2}$$

and

$$M(\lambda) = \sum_{j=1}^{\lambda}\left(a_j^2 - \sigma_j^2\right).$$

The last term, $\sum_{j=1}^{\infty} a_j^2$, in (3.1) does not depend on λ. Thus, $R(\lambda)$ is minimized by the same value of λ which maximizes $M(\lambda)$. Therefore, an optimality criterion which is equivalent to minimizing MISE is that of maximizing the quantity $M(\lambda)$. To make this feasible, an estimator $\hat{M}(\lambda)$ of $M(\lambda)$ is required.

Proposition 3.2. An unbiased estimator of $M(\lambda)$ is given by

$$\hat{M}(\lambda) = \sum_{j=1}^{\lambda} \tilde{a}_{jn}^2 - 2\sum_{j=1}^{\lambda} \hat{\sigma}_{jn}^2,$$

where

$$\hat{\sigma}_{jn}^2 = \frac{1}{n-1}\left(1 + \frac{1}{\sqrt{2}}\tilde{a}_{2jn} - \tilde{a}_{jn}^2\right)$$

is an unbiased estimator of $\sigma_j^2 = \mathrm{var}(\tilde{a}_{jn})$ in (3.2).

We know that the value of λ which minimizes $R(\lambda)$ is the same as the value which maximizes $M(\lambda)$. Thus, we estimate the maximizer of $M(\lambda)$ by using the maximizer of its unbiased estimator

$$\hat{M}(\lambda) = \begin{cases} 0, & \text{if } \lambda = 0, \\ \sum_{j=1}^{\lambda}\left(\tilde{a}_{jn}^2 - 2\hat{\sigma}_{jn}^2\right), & \text{if } \lambda \geq 1, \end{cases}$$

over $0 \leq \lambda \leq n$. We denote this value by $\hat{\lambda}$ in all that follows.

The estimator of the optimal smoothing parameter obtained by maximizing $\hat{M}(\lambda)$ gives a choice for λ in our density estimator of the comparison density d. It also suggests some possible test statistics for the goodness-of-fit hypothesis.

4. The Proposed Test

From our previous discussions, we know that $a_j = 0$ for all $j \geq 1$ under the null hypothesis. Thus, $M(\lambda) = -\dfrac{\lambda}{n}$ under H_0, giving the trivial maximizer $\lambda = 0$. This suggests that we should reject H_0 if $\hat{\lambda}$ departs far from zero, i.e., if the data indicate we should do significantly less smoothing than should be done when H_0 is true. To test H_0 using $\hat{\lambda}$, we use an alternative strategy.

Define $\hat{\lambda}_\alpha$ to be the maximizer of

$$\hat{M}_\alpha(\lambda) = \begin{cases} 0, & \text{if } \lambda = 0 \\ \sum_{j=1}^{\lambda} \tilde{a}_{jn}^2 - C_\alpha \sum_{j=1}^{\lambda} \hat{\sigma}_{jn}^2 & \text{if } \lambda = 1, 2, \ldots, n, \end{cases} \tag{4.1}$$

where C_α is chosen in such a way that $P(\hat{\lambda}_\alpha = 0)$ is approximately $1-\alpha$ under H_0. Then, the proposed test is formally given by

$$\text{Reject} \quad H_0 \quad \text{if} \quad \hat{\lambda}_\alpha \geq 1. \tag{4.2}$$

Our specific choice for C_α is the value of c for which

$$1-\alpha = \exp\left\{-\sum_{j=1}^{\infty} \frac{P\left(\chi_j^2 > jc\right)}{j}\right\}, \tag{4.3}$$

where χ_m^2 is a random variable having a chi-squared distribution with m degrees of freedom. Under this choice for C_α, $P(\hat{\lambda}_\alpha = 0) \to 1-\alpha$ as $n \to \infty$. Equation (4.3) is a consequence of Theorem 2.3.1 and (4.7) of Spitzer (1956). The approximate values of C_α for various choice of α in Table 3 are from Eubank and Hart (1992). When $C_\alpha = 2$, the test (4.2) is using the maximizer of $M(\lambda)$ in (3.1) and the asymptotic level of the test is about .29. As noted above, this is the reason why we use the test (4.2).

Table 1.

Values of C_α which make test (4.2) asymptotically valid at level α

C_α	6.74	4.18	3.22	2.38	2.00
α	.01	.05	.10	.20	.29

We will begin by establishing several lemmas that are required for the proofs of Theorem 4.1 and Theorem 4.2.

Lemma 4.1. (Spitzer, 1956) Let $Z_1, \cdots, Z_n$ be identically distributed independent random variables and $S_k = Z_1 + \cdots + Z_k, 1 \leq k \leq n$. Then,

$$p_r = P(S_1 > 0, \cdots, S_r > 0)$$

can be represented as $p_r = \sum_r^* \{\prod_{k=1}^r \frac{1}{N_k!}(\frac{\alpha_k}{k})^{N_k}\}$, with $\sum_r^*$ extending over all r-tuples of integers $(N_1, \cdots, N_r)$ such that $N_1 + 2N_2 + \cdots + rN_r = r$ and $\alpha_k = P(S_k > 0)$.

Lemma 4.2. Let $U_r, r = 1, 2, \cdots, n$, be *iid* random variables with a uniform distribution on [0, 1]. Then,

a) $E\left(\sum_{r=1}^{n}\cos(j\pi U_r)\right)^4 = \frac{3}{4}n^2 - \frac{3}{8}n,\ E\left(\sum_{r=1}^{n}\cos(j\pi U_r)\right)^2 = \frac{1}{2}n,$

and for $j \neq l$

b) $E\left((\sum_{r=1}^{n}\cos(j\pi U_r))^2(\sum_{r=1}^{n}\cos(l\pi U_r))^2\right) = \frac{1}{4}n^2.$

Lemma 4.3. Let $\tilde{a}_{jn} = \frac{1}{n}\sum_{r=1}^{n}\sqrt{2}\cos(j\pi U_r)$ with U_r having a uniform distribution on [0, 1]. Then, for any integer m,

$$\text{var}\left\{\sum_{j=1}^{m}(n\tilde{a}_{jn}^2 - 1)\right\} = m\left(2 - \frac{3}{2n}\right).$$

Lemma 4.4. Let $V_j = \frac{\sqrt{2}}{\sqrt{n}}\sum_{r=1}^{n}\cos(j\pi U_r)$ with U_r having a uniform distribution on [0, 1]. Then, for any integer t and any $\epsilon > 0$,

$$P\left(\left|\frac{1}{t}\sum_{j=1}^{t}V_j^2 - 1\right| > \epsilon\right) \leq \frac{2}{t\epsilon^2} - \frac{3}{2nt\epsilon^2}.$$

Lemma 4.5. Let $\tilde{a}_{2jn} = \frac{1}{n}\sum_{r=1}^{n}\sqrt{2}\cos(2j\pi U_r)$ with U_r having a uniform distribution on [0, 1]. Then, for any $\epsilon > 0$, as $n \to \infty$

$$P\left(\sup_{1\leq m\leq n}\left|\frac{1}{m}\sum_{j=1}^{m}\tilde{a}_{2jn}\right| > \epsilon\right) \to 0 \quad \text{and} \quad P\left(\sup_{1\leq m\leq n}\left|\frac{1}{m}\sum_{j=1}^{m}\tilde{a}_{jn}^2\right| > \epsilon\right) \to 0.$$

The main result of this section is the following theorem. It provides an expression for the limiting distribution of our test statistic in (4.2).

Theorem 4.1. The asymptotic distribution of $\hat{\lambda}_\alpha$ under H_0 is given by

$$\lim_{n\to\infty} Pr(\hat{\lambda}_\alpha = \lambda) = q_\lambda(1 - \alpha), \qquad \lambda = 0, 1, 2, \ldots,$$

with

$$q_0 = 1$$

and

$$q_r = \sum_r^* \left\{ \prod_{k=1}^r \frac{1}{N_k!} \left(\frac{P(\chi_k^2 > C_\alpha k)}{k} \right)^{N_k} \right\},$$

with $\sum_r^*$ extending over all r-tuples of integers $(N_1, \cdots, N_r)$ such that $N_1 + 2N_2 + \cdots + rN_r = r$.

Proof. Let $\hat{\lambda}$ be the maximizer of $\hat{M}(\lambda)$ over $0 \le \hat{\lambda} \le n$ with $\hat{M}(0) = 0$ and

$$\hat{M}(\lambda) = \frac{n+1}{n-1} \sum_{j=1}^{\lambda} \tilde{a}_{jn}^2 - \frac{\sqrt{2}}{n-1} \sum_{j=1}^{\lambda} \tilde{a}_{2jn} - \frac{c\lambda}{n-1}, \quad \lambda = 1, 2, \ldots, n,$$

for some constant c. By definition, we choose $\hat{\lambda} = \lambda$ if $\hat{M}(\lambda) - \hat{M}(m) \ge 0$, for all $m = 1, 2, ..., n$. To simplify notation, only consider the case $\hat{\lambda} = 0$, as the general case follows similarly.

Now $\hat{\lambda} = 0$ if and only if

$$\hat{M}(m) \le 0, \qquad \text{for} \quad m = 1, ..., n. \tag{4.4}$$

Therefore, (4.4) is equivalent to working with, for any $\epsilon > 0$,

$$\begin{aligned} P\left(\hat{\lambda} = 0\right) = P\Bigg(& \frac{1}{m} \sum_{j=1}^m \left(n\tilde{a}_{jn}^2 - c \right) \le \frac{nc}{\sqrt{2}m(n+c-1)} \sum_{j=1}^m \tilde{a}_{2jn} - \frac{c(c-1)}{n+c-1}, \\ & m = 1, 2, \ldots, n; \quad \sup_{1 \le m \le n} \left| \frac{nc}{\sqrt{2}m(n+c-1)} \sum_{j=1}^m \tilde{a}_{2jn} \right| \le \epsilon \Bigg) \\ + P\Bigg(& \frac{1}{m} \sum_{j=1}^m \left(n\tilde{a}_{jn}^2 - c \right) \le \frac{nc}{\sqrt{2}m(n+c-1)} \sum_{j=1}^m \tilde{a}_{2jn} - \frac{c(c-1)}{n+c-1}, \\ & m = 1, 2, \ldots, n; \quad \sup_{1 \le m \le n} \left| \frac{nc}{\sqrt{2}m(n+c-1)} \sum_{j=1}^m \tilde{a}_{2jn} \right| > \epsilon \Bigg). \end{aligned}$$

By using Lemma 4.1 - 4.5, and the proofs of Theorem 2.5.1 in Kim (1992), Theorem 4.1 can be easily obtained.

It is also of interest to know how $\hat{\lambda}_\alpha$ fares against alternatives. In this regard we have the following theorem that establishes consistency for the test.

Theorem 4.2. If $|a_{j_0}| > 0$ for some $0 < j_0 < \infty$, the power of the test based on $\hat{\lambda}_\alpha$ tends to 1 as $n \to \infty$.

Proof. Suppose there exists some j_0 such that $|a_{j_0}| > 0$. Then,

$$P\left(\hat{\lambda}_\alpha \geq 1\right) \geq 1 - P\left(\sum_{j=1}^{j_0}\left(n\tilde{a}_{jn}^2 - \frac{n}{n-1}C_\alpha\right) \leq \frac{C_\alpha n}{(n-1)} \sum_{j=1}^{j_0}\left(\frac{1}{\sqrt{2}}\tilde{a}_{2jn} - \tilde{a}_{jn}^2\right)\right).$$

See the proofs of Theorem 2.5.2 in Kim (1992).

5. Summary and Conclusions

Simulation was carried out to study the finite sample power properties of four tests: namely the Cramér-von Mises, W_n^2, and Anderson-Darling tests, A_n^2, the test based on $\hat{\lambda}_\alpha$ and a test obtained from the statistic $T_{\hat{\lambda}n}$ of Kim (1992) with the maximizer of $\hat{M}_\alpha(\lambda)$ for $C_\alpha = 2$. We consider the relative performance of these tests under sine and cosine alternatives.

In the results of simulation studies, for the cases in which the alternative lies in the direction of sine or cosine density alternatives, $T_{\hat{\lambda}n}$ and $\hat{\lambda}_\alpha$ are to be preferred to W_n^2 and A_n^2 for anything but lower frequency alternatives. The power of $T_{\hat{\lambda}n}$ is seen to be both stable and high over all frequencies j. For high frequencies such as $j > 2$, the test using $T_{\hat{\lambda}n}$ is significantly more powerful than the other three tests.

The proposed test statistic $\hat{\lambda}_\alpha$ is itself a smoothing parameter which is selected to minimize an estimated MISE for a truncated series estimator, $d_{\lambda n}$, of the comparison density function. Therefore, this test statistic leads immediately to a point estimate of the density function in the event that H_0 is rejected. The limiting distribution of $\hat{\lambda}_\alpha$ was obtained under the null hypothesis. It was also shown that this test is consistent against fixed alternatives.

$T_{\hat{\lambda}n}$ is essentially a Neyman smooth test that uses an estimated smoothing parameter to choose the number of terms in the statistic. In our simulation study, we found this test to have excellent empirical power properties.

References

1. Eubank, R. and Hart, J. (1992), "Testing Goodness-of-Fit in Regression Via Order Selection Criteria," *Annals of Statistics*, to appear.

2. Eubank, R., LaRiccia, V. N., and Rosenstein, R. (1987), "Test Statistics Derived as Components of Pearson's Phi-Squared Distance Measure," *Journal of the American Statistical Association*, 82, 816-825.

3. Eubank, R. and LaRiccia, V. N. (1992), "Asymptotic Comparison of Cramér-von Mises and Nonparametric Function Estimation Technicques for Testing Goodness-of-Fit," *Annals of Statistics*, to appear.

4. Hart, J. D. (1985), "On the Choice of a Truncation Point in Fourier Series Density Estimation," *Journal of Statistical Computation and Simulation*, 21, 95-116.

5. Kim, J. T. (1992), "Testing Goodness-of-Fit Via Order Selection Criteria," unpublished Ph.D. Dissertation, Texas A&M University, Texas, Dept of Statistics.

6. Parzen, E. (1979), "Nonparametric Statistical Data Modeling," *Journal of the American Statistical Association*, 74, 105-131.

7. Spitzer, F. (1956), "A Combinatorial Lemma and Its Applications to Probability Theory," *Transactions of the American Mathematical Society*, 82, 323-339.