The Korean Communications
in Statistics Vol. 4, No. 3, 1997
pp. 891 - 901

Goodenss of Fit Test on Density Estimation[1)]

Kim, J. T.[2)], Yoon, Y. H.[3)] and Moon, G. A.[4)]

Abstract

The objective of this research is to investigate the problem of goodness of fit testing based on nonparametric density estimation with a data-driven smoothing parameter. The small and large sample properties of the proposed test statistic Z_{mn} are investigated with the minimizer $\hat{m}$ of the estimated mean integrated squared error by the Diggle and Hall (1986) method.

1. Introduction

The smoothing parameters obtained from stopping rules to find the optimal density estimator have been an important role in the area of the goodenss-of-fit tests (cf.: Fellner (1974), Eubank and LaRiccia (1992), Eubank and Hart (1992), and Kim (1992,1994)). In almost these papers the authors try to solve goodenss-of-fit problem with only the Hart (1985) method for finding smoothing parameter to minimize mean integrated squared error (MISE) for truncated estimator by the stopping rule. But, as denoted by Tarter and Lock (1993), the Hart method has one weakness that it is possible that variance of the chosen smoothing parameters may be inflated by injudicious choice of M_n. While the variance of the smoothing parameters selected by Diggle and Hall (1986) method is smaller than that by Hart method. The goal of this paper is to find the goodness of fit techniques on density estimation by the Diggle and Hall method.

Let $X_1, X_2, \cdots X_n$ be a random samples from an absolutely continuous distribution function F having associated density f. The goodness-of-fit problem concerns testing $H_0 : F = G$, for some specified absolutely continuous distribution function G having density g and same support as F. Set $Y_i = G(X_i)$, for $i=1,2,3,\cdots n$, and then $0 \leq Y_i \leq 1$. Therefore, the testing $H_0 : F = G$ is equivalent to testing $H_0 : Y$ has a

1) This paper was supported in part by the research grant of Institute of National Science, Taegu University
2) Assistant Professor, Department of Statistics, Taegu University, Taegu, Korea
3) Professor, Department of Statistics, Taegu University, Taegu, Korea
4) Assistant Professor, Department of Office Automation, Dong Hae Junior College, Dong Hae, Korea

uniform distribution on [0,1].

In the next section, we consider the Diggle and Hall method of finding the smoothing parameter m on the density estimation.

In section 3, we propose the test statistic for the goodness of fit problem by using the minimizer $R(\widehat{m})$ of the estimated mean integrated squared error in $\hat{R}(m)$ obtained by Diggle and Hall method in section 2. In section 4, we study the finite sample power properties by simulation.

2. Diggle and Hall Method

In the previous section, let $Y_i = G(X_i)$, $i = 1, 2, \cdots, n$, and then be a continuous random sample of the interval [0,1]. Let $\{ \phi_k ; k \geq 0 \}$ be a complete orthonormal sequence on an interval [0,1] and suppose that the square integrable density h admits the expansion

$$h(y) = \sum_{k=0}^{\infty} a_k \phi_k(y), \qquad y \in [0,1], \tag{2.1}$$

where, $a_k = \int_0^1 \phi_k(y) h(y) dy$. An orthogonal series estimator of $h(y)$ based on a sample $Y_1, Y_2, \cdots, Y_n$, is given by

$$h_n(y) = \sum_{k=0}^{m} \hat{a}_k \phi_k(y), \qquad y \in [0,1], \tag{2.2}$$

where the sample coefficients are defined as $\hat{a}_k = n^{-1} \sum_{j=1}^{n} \phi_k(Y_j)$. In this case, ϕ can be chosen as cosine series estimators to have particularly attractive mean squared error properties. Now, the sequence of functions

$$\phi_0(y) = 1,$$
$$\phi_k(y) = \sqrt{2} \cos(k\pi y) \quad \text{for } j \geq 1,$$

are orthonormal on the interval [0,1]. It has mean integrated squared error

$$\begin{aligned} R(m) &= \int_0^1 E(h_n(y) - h(y))^2 dy \\ &= \sum_{k=0}^{m} \sigma_k^2 + \sum_{k=m+1}^{\infty} a_k^2 \end{aligned} \tag{2.3}$$

where,

$$\sigma_k^2 = \text{var}(\hat{a}_k) = n^{-1}(1 + a_{2k}/\sqrt{2} - a_k^2) \tag{2.4}$$

To the estimator of the MISE in (2.3), Diggle and Hall consider the first and second terms on the right side of (2.3) separately. In the first term on the right side of (2.3), since $\sum_{k=0}^{m} a_k^2$ remains bounded as $m \to \infty$, and the value of any σ_k is bounded on [0,1], the first term on

the right side of (2.3) is estimated by m/n.

The second term on the right side of (2.3) is more complicated to estimate since it involves the sum of infinite number of the coefficients. The crucial step of the Diggle and Hall approach is the replacement of an infinite sum by whose limits are allowed to vary. Specifically, if λ is a function of m such that $\lambda(m) \to \infty$, then

$$\sum_{k=m+1}^{\lambda(m)m} a_k^2 \sim \sum_{k=m+1}^{\infty} a_k^2, \quad \text{as } m \to \infty.$$

Thus the $R(m)$ in (2.3) can be estimated by plugging in estimates for the a_k^2. An unbiased estimator of a_k^2 is provided $\hat{a}_k^2 - \hat{\sigma}_k^2$, where

$$\hat{\sigma}_k^2 = (n-1)^{-1}(1 + \frac{1}{\sqrt{2}}\hat{a}_{2k} - \hat{a}_k^2), \tag{2.5}$$

is an unbiased estimator of σ_k^2 in (2.4). Therefore the MISE estimator is obtained as

$$\hat{R}(m) = m/n + \sum_{k=m+1}^{\lambda(m)m} (\hat{a}_k^2 - \hat{\sigma}_k^2), \tag{2.6}$$

where $\hat{\sigma}_k^2 = (n-1)^{-1}(1 + \hat{a}_{2k}/\sqrt{2} - \hat{a}_k^2)$. $\hat{R}(m)$ is a consistent estimate of $R(m)$, in sense that the ratio $\hat{R}(m)/R(m)$ converges in probability to one as m and n increase. Alternatively, since $\hat{a}_k^2 - \hat{\sigma}_k^2$ is designed to estimate a nonnegative quantity, it might be replaced by $\max(\hat{a}_k^2 - \hat{\sigma}_k^2, 0)$. This does not affect the asymptotics. Diggle and Hall also offer some practical suggestions regarding the implementation of their stopping rule.

For their own simulation study, they chose $\lambda(m) = cm^{1/2}$, where c is a small constant number, and experimented with different value of c. It was found that $c = 4$ was a good choice and smaller numbers sometimes produced poor results. With this substitution, the estimated MISE in (2.6) can be expressed

$$\hat{R}(m) = \frac{m}{n} + \sum_{k=m+1}^{4m^{3/2}} (\hat{a}_k^2 - \hat{\sigma}_k^2). \tag{2.7}$$

Thus, one estimates the minimizer of $R(m)$ by using the minimizer of its unbiased estimator $\hat{R}(m)$. The estimator $\hat{m}$ obtained by minimizing $\hat{R}(m)$ has a role for the selecting the optimal density estimator of $h(y)$ in (2.1), and for the goodness-of-fit testing.

3. The Goodness of Fit Test

The viewpoint of goodness of fit taken in this article parallels that of Parzen (1979) : Namely, taking H_0 is equivalent to testing $h(y)=1$, where h is the comparison density function

$$h(y) = f(G^{-1}(y)) / g(G^{-1}(y)), \quad y \in [0,1] \tag{3.1}$$

with g the density of G and $G^{-1} = \inf \{x: G(x) \geq y\}$. Thus, one method for testing H_0 is to estimate h using a consistent estimator h_n and then compare h_n to 1 using some suitable metric.

One natural estimator of h in (3.1) is the truncated Fourier cosine series estimator

$$h_n(y) = 1 + \sum_{k=1}^{m} \hat{a}_k \phi_k(y), \quad y \in [0,1] \tag{3.2}$$

using the squared L_2 $[0,1]$ norm as a measure of distance with

$$\phi_k(y) = \sqrt{2} \cos(k\pi y) \text{ and}$$
$$\hat{a}_k = n^{-1} \sum_{i=1}^{n} \sqrt{2} \cos(k\pi y_i),$$

then one can gives the test statistic

$$T_{En} = n \int_0^1 (h_n(y) - 1)^2 dy = n \sum_{k=1}^{m} \hat{a}_k^2 \tag{3.3}$$

for H_0. Kim (1992) and Eubank and LaRiccia (1992) suggested this type of statistic. However, they did not solve the asymptotic distribution of T_{En} in the case of that $\hat{m}$ is the estimator of smoothing parameters.

Now, we consider the case of that $\hat{m}$ is a random variable of minimizing $\hat{R}(m)$, that is, $\hat{m}$ is the value of finding by Diggle and Hall method in section 2.

The proposed test statistic is formally given by

$$Z_{mn} = \sum_{k=1}^{m} (n\hat{a}_k^2 - 1) / \sqrt{2\hat{m}}. \tag{3.4}$$

The following study is to find the asymptotic properties of Z_{mn} under null hypothesis.

Lemma 3.1. Under the null hypothesis, that is, $a_k = 0$, for all k, then $\sqrt{n}\hat{a}_k$ has the asymptotic normal with mean 0 and variance 1.

proof. Under the assumptions of (2.1) and (2.2), $E(\hat{a}_k) = a_k$ and $\text{var}(\hat{a}_k) = n^{-1}(1 + a_{2k}/\sqrt{2} - a_k^2)$. If $a_k = 0$, for all k, the normality of $\sqrt{n}\hat{a}_k$ can be obtained easily.

Corollary 3.2. Under the assumptions of (2.1) and (2.2), if H_0 holds true, then $P(m = 1) \to 1$, as $n \to \infty$.

If H_0 holds true we detect that the values of the minimizer $\hat{m}$ of $\hat{R}$(m) in (2.6) become almost one by the simulation studies and that $P(m = 1) \to 1$, as $n \to \infty$.

Let χ^2_m denote a central chi-squared random variable with m degrees of freedom.

Lemma 3.3. Under the assumption of (2.1) and (2.2), if the $\hat{m}$ is minimizer of $\hat{R}(m)$ in (2.6), then

$$T_{mn} \xrightarrow{d} \sum_{k=1}^{m} Z_k^2 \sim \chi^2_m ,$$

where Z_k is a standard normal random variable.

proof. From the result of Lemma 3.1, this result is easily obtained. See Eubank and LaRiccia (1992).

Theorem 3.4. Assume that $a_k = 0$ for all k and that $\hat{m} = \hat{m}(n) \to \infty$ in a such that a way that $\hat{m} / n^2 \to 0$. Then the test statistic

$$Z_{mn} = \frac{\sum_{k=1}^{m} (n \hat{a}_k - 1)}{\sqrt{2\hat{m}}} \xrightarrow{d} Z_1,$$

where Z_1 is a N(0, 1) random variable.

proof. From the result of Theorem 1 in Eubank and LaRiccia (1992), Z_{mn} converges to $\sum_{k=1}^{m} (Z_k^2 - 1)/\sqrt{2\hat{m}}$ in distribution. If H_0 hold true, by the Lemma 3.1 and Lemma 3.2 Z_{mn} converges to Z_1 in distribution.

4. Power of the Tests.

In this section, we study the finite sample power properties of three test statistics, including the Cramer-von Mises test statistics,

$$W_n^2 = 1/12n + \sum_{r=1}^{m} [Y_{(r)} - (2r-1)/2n]^2$$

with $Y_{(1)} \le Y_{(2)} \le \cdots \le Y_{(n)}$, the ordered Y_r values, the Anderson and Dariling(1952) test statistic

$$A_n^2 = n^{-1} \sum_{r=1}^{n} (2r-1) \log \frac{1}{Y_{(r)}(1 - Y(n-r))} - n ,$$

and the proposed test statistic

$$Z_{mn} = \sum_{k=1}^{m} (na_k^2 - 1) / \sqrt{2\hat{m}}$$

with $\hat{m}$ the minimizer of $\hat{R}$(m) in (2.3). We consider the relative performance of these test statistics under two types of alternatives; the cosine alternatives

$$h(y) = 1 + \gamma\cos(j\pi y) \quad \text{for } j = 1, 2, \cdots, \text{and } 0 \leq \gamma \leq 1,$$

and a beta density alternatives

$$h(y) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} y^{a-1}(1-y)^{b-1}, \quad a > 0, \quad b > 0$$

with Γ, the gamma function $\Gamma(\cdot)$. For all two cases we generated 1000 replicates from sample of size 50 and used these to assess the empirical power of our tests.

First, we consider the cases of symmetric beta density alternatives , if $a = 1$, then the beta alternative is a uniform density function and the power is just the level of the test, and the empirical power increases as a move away from $a = 1$. From the Table 4. 1., when the symmetric beta alternatives are U-shaped, $a = 0.5$, the Anderson and Darling test and Z_{mn} are better than W_n^2. When these alternatives are unimodal, Z_{mn} have significantly higher power than the other two. Table 4. 2. and Table 4. 3. give the empirical powers against the skewed beta alternative with mean .48 and .52 respectively. We see that the test based on Z_{mn} is clearly more powerful than the other two test.

Finally, the case in which the alternative lies in the direction of cosine density functions were considered with $\alpha = 0.05$. In Table 4.4, the cosine alternatives with $\gamma = 0$ are the null hyperthesis density. Therefore, these powers are the empirical significance level. As the value of γ increases the cosine alternatives become further moved from the null density $h(y) = 1$ and the empirical powers increase accordingly. The power of Z_{mn} is still clearly more powerful than the others with the exception of the case of $j = 1$.

Table 4.1. Proportion of rejections in rejection in 1000 samples with $n = 50$ for symmetric beta alternatives with parameter a and $\alpha = 0.05$.

a	Z_{mn}	W_n^2	A_n^2
0.5	0.677	0.473	0.868
1.0	0.056	0.055	0.057
1.5	0.284	0.071	0.085
2.0	0.685	0.255	0.376
2.5	0.898	0.564	0.763
3.0	0.978	0.832	0.937
3.5	0.994	0.956	0.993
4.0	1.000	0.994	0.999

Table 4.2. Proportion of rejections in rejection in 1000 samples with $n = 50$ for skewed beta alternatives with their mean = 0.48 and skewness β and $\alpha = 0.05$.

β	z_{mn}	W_n^2	A_n^2
-0.76	0.675	0.532	0.929
-0.70	0.075	0.076	0.086
-0.65	0.315	0.102	0.111
-0.60	0.752	0.275	0.374
-0.57	0.959	0.520	0.707
-0.54	0.996	0.853	0.940
-0.51	0.999	0.954	0.995
-0.49	1.000	0.987	0.999

Table 4.3. Proportion of rejections in rejection in 1000 samples with $n = 50$ for skewed beta alternatives with their mean = 0.52 and skewness β and $\alpha = 0.05$.

β	z_{mn}	W_n^2	A_n^2
0.76	0.698	0.430	0.824
0.69	0.074	0.063	0.064
0.63	0.198	0.128	0.140
0.59	0.592	0.339	0.492
0.55	0.887	0.683	0.843
0.52	0.980	0.912	0.978
0.50	0.991	0.988	0.999
0.47	1.000	0.999	1.000

Table 4.4. Proportion of rejections in rejection in 1000 samples with $n = 50$ for skewed beta alternatives with their mean = 0.52 and skewness β and $\alpha = 0.05$.

(γ, j)	z_{mn}	W_n^2	A_n^2
0.0, 1	0.055	0.050	0.045
0.3, 1	0.229	0.286	0.275
0.5, 1	0.604	0.705	0.689
0.7, 1	0.855	0.947	0.947
0.9, 1	0.979	1.000	0.999
1.0, 1	0.999	1.000	1.000
0.0, 2	0.055	0.050	0.045
0.3, 2	0.209	0.111	0.150
0.5, 2	0.472	0.226	0.315
0.7, 2	0.778	0.507	0.656
0.9, 2	0.977	0.841	0.894
1.0, 2	0.999	0.946	0.970
0.0, 3	0.055	0.050	0.045
0.3, 3	0.148	0.056	0.066
0.5, 3	0.376	0.072	0.100
0.7, 3	0.700	0.117	0.193
0.9, 3	0.962	0.248	0.425
1.0, 3	0.992	0.359	0.590
0.0, 4	0.055	0.050	0.045
0.3, 4	0.124	0.060	0.077
0.5, 4	0.359	0.065	0.113
0.7, 4	0.699	0.088	0.176
0.9, 4	0.932	0.121	0.285
1.0, 4	0.975	0.143	0.362
0.0, 5	0.055	0.050	0.045
0.3, 5	0.116	0.052	0.058
0.5, 5	0.249	0.061	0.069
0.7, 5	0.423	0.069	0.091
0.9, 5	0.641	0.079	0.128
1.0, 5	0.671	0.095	0.155

Table 4.4. (Continued.)

(γ, j)	z_{mn}	W_n^2	A_n^2
0.0, 6	0.055	0.050	0.045
0.3, 6	0.132	0.055	0.062
0.5, 6	0.251	0.060	0.085
0.7, 6	0.456	0.068	0.104
0.9, 6	0.558	0.076	0.143
1.0, 6	0.611	0.083	0.175
0.0, 7	0.055	0.050	0.045
0.3, 7	0.112	0.052	0.052
0.5, 7	0.205	0.052	0.061
0.7, 7	0.380	0.054	0.068
0.9, 7	0.508	0.061	0.086
1.0, 7	0.601	0.064	0.097
0.0, 8	0.055	0.050	0.045
0.3, 8	0.106	0.051	0.059
0.5, 8	0.238	0.054	0.069
0.7, 8	0.335	0.060	0.088
0.9, 8	0.450	0.063	0.104
1.0, 8	0.514	0.068	0.114
0.0, 9	0.055	0.050	0.045
0.3, 9	0.103	0.052	0.054
0.5, 9	0.188	0.052	0.058
0.7, 9	0.311	0.055	0.062
0.9, 9	0.438	0.058	0.072
1.0, 9	0.486	0.061	0.077

5. Conclusions

The goal of this research was to study the problem of goodness of fit testing based on nonparametric density estimation with a data-driven smoothing parameter. $\hat{m}$ obtained from the Diggle and Hall method, we studied the small and large sample properties of the proposed test statistic Z_{mn}, and we can see that the test based on Z_{mn} is clearly more powerful than the other two tests, the Cramer-von Mises and Anderson-Darling test.

A referee has pointed out that, when the beta alternatives are U-shaped, the Anderson and Darling test is more powerful than the other two tests and, when the cosine alternative is the case of $j = 1$, W_n^2 is more powerful than the others. With the exception of the above two cases, the power of the proposed test statistic Z_{mn} is still clearly more powerful than the others. As many statistians noticed this point that the Cramer-von Mises type test statistics including the Anderson and Darling test statistic have very good powers at $j = 1$, but as j increases, the powers of W_n^2 and A_n^2 drop off dramatically for $j > 2$. (See Shorack and Wellner (1986), Eubank and LaRiccia (1992), Eubank and Hart (1992), Kim (1992, 1994)).

If there is a test statistic which has the most powerful than any other test statistics in every cases, the test statistic will be an ideal goodness of fit statistic. we have been try to find the ideal test statistic, but this is still open problem.

Acknowledgements

We wish to express our sincere application to referee of this paper for his/her kindness to correct misprints and for nice advices.

Reference

[1] Anderson, T. W. and Darling, D. A. (1952), "Asymptotic Theory of Certain Goodness of Fit Criteria Based On Stochastic Processes", *The Annals of Mathematical Statistics*, 23, 193-212.

[2] Diggle2, P. J. and Hall, P. (1986), "The selection of Terms in an Orthogonal Series Density Estimator", *Journal of the American Statistical Association*, 81, 230-233

[3] Eubank, R, and LaRiccia, V. N. (1992), "Asymptotic Comparison of Cramer-von Mises and Nonparametric Function Estimation Techniques for Testing Goodness-or-Fit", *Annals of Statistics*, 20, 2071-2086.

[4] Eubank, R. and Hart, J. D. (1992), "Testing Goodness-of-Fit in Regression Via Order Selection Criteria", *Annals of Statistics*, 20, 1412-1425.

[5] Fellner, W. H. (1974), "Heuristic Estimation of Probability Densities", *Biometrika*, 61, 485-492.

[6] 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.

[7] Kim, J. T. (1992) "Testing Goodness of Fit via Order Section Criteria", unpublished Ph.D. Dissertation, Texas A&M University, Department of Statistics.

[8] Kim, J. T. (1994) "Goodness of Fit Test Based on Smoothing Parameter Section Criteria", *The Korean Communications in Statistics*, 2, 122 - 136.

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

[10] Shorack, G. R. and Wellner, J. A. (1986) Empirical Preocesses with Applications to Staistics, John Wiley & Sons, Inc.

[11] Tarter, M. E. and Lock, M. D. (1993) *Model - Free Curve Estimation*, Chapman & Hall