The Korean Communications
in Statistics Vol. 8, No. 2, 2001
pp. 483-497

A Study on Goodness-of-fit Test for Density with Unknown Parameters[1]

Changkon Hong[2] and Minyoung Lee[3]

Abstract

When one fits a parametric density function to a data set, it is usually advisable to test the goodness of the postulated model. In this paper we study the nonparametric tests for testing the null hypothesis against general alternatives, when the null hypothesis specifies the density function up to unknown parameters. We modify the test statistic which was proposed by the first author and his colleagues. Asymptotic distribution of the modified statistic is derived and its performance is compared with some other tests through simulation.

Keywords : goodness-of-fit test; kernel density estimation; composite null hypothesis; power comparison.

1. Introduction

Consider a random sample $X_1,\cdots,X_n$ from a population with unknown distribution function $F(x)$. It is often necessary to test hypothesis about $F(x)$ of the form

$$H_0 : F(x) \in \Xi_0,$$

where Ξ_0 is a specified family of distribution functions. The well-known goodness-of-fit tests against general alternatives are usually based on empirical distribution function (EDF), which include Kolmogorov-Smirnov test, Cramér-von Mises test and Anderson-Darling test for the continuous ungrouped data. For the discrete or grouped data, likelihood ratio or Pearson χ^2 provides reasonably good tests. In regression problems, lots of goodness-of-fit test procedures based on the nonparametric estimates of the regression function have been studied by many authors. See Eubank and Hart(1992), Eubank and Spiegelman(1990) and Cox *et al.*(1988). In density estimation problem, however, few studies have been done on test procedures based on

1) The present studies were supported by the research fund of Pusan National University (1995-1998).
2) Associate Professor, Department of statistics, Pusan National University, Pusan 609-735, Korea.
E-mail : ckhong@hyowon.pusan.ac.kr
3) Department of statistics, Pusan National University, Pusan 609-735, Korea

the nonparametric density estimates. Kim *et al.*(1997) proposed as a test statistic the squared L^2-distance between nonparametric density estimates and the null density when the null density is completely specified. More specifically, they suggested a test statistic

$$T_n = \int (\hat{f}_h(x) - f_0(x))^2 w(x) dx,$$

where $w(x)$ is a weight function, $f_0(x)$ is a completely specified null density and $\hat{f}_h(x)$ is a kernel density estimate given by

$$\hat{f}_h(x) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i).$$

Here $K_h(\cdot) = K(\cdot/h)/h$, K is called the kernel function, and h is called the smoothing parameter. In many practical situations, however, the functional form of the null density is specified up to unknown parameter, which must be estimated from the data. In this paper, we study the behavior of the test statistic T_n with estimate $\hat{\theta}$ of θ. In Section 2, the asymptotic distribution of T_n is to be studied. Some simulation results are given in Section 3. And Section 4 gives the concluding remarks.

2. Modification of L^2-Distance Statistic and Its Asymptotic Distribution

There are two types of goodness-of-fit tests; omnibus tests and directional tests. 'Omnibus tests' are designed to be effective against wide classes of alternatives to the null. On the other hand, 'directional tests' are effective at detecting certain type of departures from the null. In particular, Kolmogorov-Smirnov test, Cramér-von Mises test and Anderson-Darling test are 'omnibus tests' . The test statistic T_n suggested by Kim *et al.*(1997) is also an omnibus test statistic. Kim *et al.* (1997) compared the powers of the above four omnibus tests when the null distribution is completely known.

The empirical distribution function(EDF) for the sample $X_1, \cdots, X_n$ denoted by $F_n(x)$ is defined as

$$\begin{aligned} F_n(x) &= \frac{\text{Number of } X_i\text{'s} \le x}{n} \\ &= \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x]}(X_i), \end{aligned}$$

where $I_A(x)$ is an indicator function of set A. When the null distribution function $F_0(x)$ is completely specified, the well known 'omnibus tests' based on EDF are :

The Kolmogorov-Smirnov statistic :

$$D_n = \sup_x |F_n(x) - F_0(x)|.$$

The Cramér-von Mises statistic :

$$W_n^2 = n\int_{-\infty}^{\infty}[F_n(x) - F_0(x)]^2 dF_0(x).$$

The Anderson-Darling statistic :

$$A_n^2 = n\int_{-\infty}^{\infty}\frac{[F_n(x) - F_0(x)]^2}{F_0(x)[1-F_0(x)]}dF_0(x).$$

The above three test statistics are all distribution-free in asymptotic sense and the percentage points are generally known (Lawless(1982)). When the null distribution function has an unknown parameter, the null hypothesis can be expressed as

$$H_0 : F(x) \in \{F_\theta(x); \theta \in \Omega\}, \tag{2.1}$$

and the above test statistics can be modified by replacing $F_0(x)$ with $F_{\hat\theta}(x)$. Here $\hat\theta$ is a reasonable estimate of θ. For some special distributions, the exact percentage points of these modified statistics have been studied by some authors. (See Durbin(1975), Margolin and Maurer(1976)). Stephens(1974) gives the percentage points estimated by the Monte Carlo methods. Similarly, when the null hypothesis specifies only the functional form of the 'densit y' , it can be expressed as

$$H_0 ; f(x) \in \{f_\theta(x); \theta \in \Omega\}, \tag{2.2}$$

where $\{f_0(x;\theta); \theta \in \Omega\}$, is a parametric family of densities. To test (2.2), we suggest a natural modification of Kim *et al.*(1997) statistic, which is given by

$$T_n = \int \{\hat f_h(x) - f_0(x;\hat\theta)\}^2 w(x)dx, \tag{2.3}$$

where $\hat\theta$ is MLE of θ. The null hypothesis is to be rejected for large values of T_n. In the remaining of this section, we will derive the asymptotic distribution of T_n under H_0. We need some assumptions on null density $f_0(x;\theta)$, kernel K, and weight function w.

(A1) Ω is an open interval.

(A2) For each $\theta \in \Omega$, the derivatives

$$\frac{\partial}{\partial\theta}\log f_0(x;\theta), \quad \frac{\partial^2}{\partial\theta^2}\log f_0(x;\theta), \quad \text{and} \quad \frac{\partial^3}{\partial\theta^3}\log f_0(x;\theta) \quad \text{exist for all } x.$$

(A3) For each $\theta_0 \in \Omega$, there exist functions $H_1(x)$, $H_2(x)$, $H_3(x)$ (possibly depending on θ_0) such that for θ in a neighborhood $N(\theta_0)$ of θ_0, the relations holds

$$|\frac{\partial}{\partial\theta} f_0(x;\theta)| \le H_1(x),\ |\frac{\partial^2}{\partial\theta^2} f_0(x;\theta)| \le H_2(x),\ |\frac{\partial^3}{\partial\theta^3} \log f_0(x;\theta)| \le H_3(x),$$

with $\int H_1(x)dx < \infty$, $\int H_2(x)\,dx < \infty$, $\int H_1^2(x)\,w(x)dx < \infty$,

$\int H_2^2(x)\,w(x)dx < \infty$ and $E_\theta[H_3(X)] < \infty$, for $\theta \in N(\theta_0)$.

(A4) For each $\theta \in \Omega$,

$$0 < I(\theta) < \infty,$$

where $I(\theta) = E_\theta\left((\frac{\partial}{\partial\theta} \log f_0(X;\theta))^2 \right)$ is the Fisher Information.

(A5) $f_0(x;\theta)$ and $\frac{\partial}{\partial x} f_0(x;\theta)$ are uniformly continuous in x.

(K) The kernel function K is bounded and nonnegative, such that

$$\int K(z)dz = 1,\ \int zK(z)dz = 0,\ \text{and} \int z^2K(z)dz < \infty.$$

(W) The weight function w(x) is bounded

Remark : assumptions (A1) ~ (A4) guarantee the MLE $\hat{\theta}$ of θ (under the null hypothesis) satisfies

(a) $\widehat{\theta_n} \rightarrow_{a.s} \theta \quad n \rightarrow \infty$ (strong consistency),

(b) $\sqrt{n}(\hat{\theta} - \theta) \rightarrow_D N(0, I(\theta)^{-1})$ (asymptotic normality)

(For the proof, see Serfling(1980) pp.144-145).

Let θ_0 be the true value of θ, then T_n can be decomposed as

$$T_n = T_{1n} + T_{2n} + 2T_{3n} + 2T_{4n} + 2T_{5n} + \int \{E\,\hat{f}_{h(x)} - f_0(x;\theta_0))\}^2 w(x)dx,$$

where

$$\begin{aligned}
T_{1n} &= \int (\hat{f}_h(x) - E\,\hat{f}_h)^2 w(x)dx, \\
T_{2n} &= \int (f_0(x;\theta_0) - f_0(x;\hat{\theta}))^2 w(x)dx, \\
T_{3n} &= \int (\hat{f}_h(x) - E\,\hat{f}_h(x))\,(E\,\hat{f}_h - f_0(x;\theta_0))w(x)dx, \\
T_{4n} &= \int (\hat{f}_h(x) - E\,\hat{f}_h(x))\,(f_0(x;\theta_0) - f_0(x;\hat{\theta}))w(x)dx, \\
T_{5n} &= \int (E\,\hat{f}_h - f_0(x;\theta_0))(f_0(x;\theta_0) - f_0(x;\hat{\theta}))w(x)dx.
\end{aligned}$$

Note that T_n is related to the Kim et al.(1997) test statistic, say T_n^0, in such a way that

$$T_n = T_n^0 + T_{2n} + 2T_{4n} + 2T_{5n}$$

By the Taylor's theorem,

$$
\begin{aligned}
T_{2n} &= \int (f_0(x;\hat\theta) - f_0(x;\theta_0))^2 w(x)dx \\
&= (\hat\theta - \theta_0)^2 \int (\frac{\partial}{\partial\theta} f_0(x;\theta_*))^2 w(x)dx, \\
T_{4n} &= \int (\hat f_h(x) - E\hat f_h(x))(f_0(x;\theta) - (f_0(x;\hat\theta))w(x)dx \\
&= (\theta_0 - \hat\theta)\int (\hat f_h(x) - E\hat f_h(x))\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)dx \\
&\quad + \frac{1}{2}(\theta_0 - \hat\theta)^2 \int (\hat f_h(x) - E\hat f_h(x))\frac{\partial^2}{\partial\theta^2} f_0(x;\theta_*)w(x)dx
\end{aligned}
$$

for some random quantities θ_* and θ^*, which are on the line segment between θ_0 and $\hat\theta$, i.e., $\theta_*, \theta^* \in (\theta_0, \hat\theta)$. Using the above expression, we can prove the following theorem. The proof of the theorem is given in the Appendix.

Theorem 1. Assume that $f_0(x;\theta_0)$ is the true density and $h \to 0$, $nh \to \infty$ *as* $n \to \infty$. Then under the assumptions (A1) $\sim$ (A4),

$$
\begin{aligned}
&(i)\ T_{2n} = O_p(n^{-1}). \\
&(ii)\ T_{4n} = O_p(n^{-1}).
\end{aligned}
$$

Before we state the main theorem, we define some quantities which will be used in the theorem. Let

$$
\begin{aligned}
a_n &= \begin{cases} n^{1/2}h^{-2} & \text{if } nh^5 \to \infty \\ nh^{1/2} & \text{if } nh^5 \to 0 \\ n^{9/10} & \text{if } nh^5 \to \lambda \quad (\neq 0), \end{cases} \\
b_n &= (nh)^{-1}\int f_0(x;\theta_0)w(x)dx \int k^2(z)dz + \int (E\hat f_h(x) - f_0(x;\theta_0))^2 w(x)dx, \\
d_K &= \int u^2 K(u)du, \\
v_1 &= \int (f_0''(x;\theta_0))^2 w(x)^2 f_0(x;\theta_0)dx - (\int (f_0''(x;\theta_0)w(x)f_0(x;\theta_0)dx)^2, \\
v_2 &= \int f_0(x;\theta_0)^2 w(x)dx \int (\int K(z)K(z+u)dz)^2 du, \\
v_3 &= (\int (\frac{\partial}{\partial\theta} f_0(x;\theta_0)\ f_0''(x;\theta_0)w(x)dx)^2 I(\theta_0)^{-1}.
\end{aligned}
$$

Now we state and prove the main theorem.

Theorem 2. Assume that $f_0(x;\theta_0)$ is the true density and $h \to 0$, $nh \to \infty$ *as* $n \to \infty$. Then under the assumptions (A1) ~ (A4), (K) and (W),

$$a_n(T_n - b_n) \to_D \begin{cases} \sigma_1 Z & \text{if } nh^5 \to \infty \\ \sigma_2 Z & \text{if } nh^5 \to 0 \\ \sigma_3 Z & \text{if } nh^5 \to \lambda \quad (\neq 0), \end{cases}$$

where $Z \sim N(0,1)$ and

$$\sigma_1^2 = d_K^2 v_1 + 5 d_K^2 v_3,$$

$$\sigma_2^2 = 2v_2,$$

$$\sigma_3^2 = \lambda^{-1/5}(\lambda \sigma_1^2 + \sigma_2^2).$$

(Proof) T_{1n}, T_{3n} can be expressed as

$$T_{1n} = \int((nh)^{-1}\sum_{i=1}^{n} K((x-X_i)/h) - (nh)^{-1}\sum_{i=1}^{n} EK((x-X_i)/h))^2 w(x)dx$$

$$= (nh)^{-2}\sum_{i=1}^{n}\int(A_n(x,X_i))^2 w(x)dx + 2(nh)^{-2}\sum_{1\le i<j\le n} H_n(X_i,X_j),$$

$$T_{3n} = (nh)^{-1}\sum_{i=1}^{n} Z_{ni},$$

where

$$A_n(u,x) = K((u-x)/h) - EK((u-X_1)/h),$$

$$H_n(x,y) = \int A_n(u,x)A_n(u,y)w(u)du,$$

$$Z_{ni} = \int A_n(x,X_i)(E\hat{f}_h(x) - f_0(x;\theta_0))w(x)dx.$$

And there exists $\theta^{**} \in (\theta_0, \hat{\theta})$ such that

$$T_{5n} = \int((E\hat{f}_h(x) - f_0(x;\theta_0))\{(\theta_0 - \hat{\theta})\frac{\partial}{\partial\theta} f_0(x;\theta_0)$$

$$+ \frac{1}{2}(\theta_0 - \hat{\theta})^2 \frac{\partial^2}{\partial\theta^2} f_0(x;\theta^{**})\}w(x)dx.$$

Kim *et al.*(1997) showed that

$$(nh)^{-2}\sum_{i=1}^{n}\int(A_n(x,X_i))^2 w(x)dx = (nh)^{-1}\int f_0(x;\theta_0)w(x)\int K^2(z)dz$$

$$+ O(n^{-1}) + O_p(n^{-3/2}h^{-1}).$$

And it can be easily shown that

$$T_{5n} = \frac{1}{2}h^2 d_K(\theta_0 - \hat{\theta})\int f_0''(x;\theta_0)\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)dx + O_p(n^{-1})(1+o(1)),$$

$$(\theta_0 - \hat{\theta}) = n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_0(X_i;\theta_0)(I(\theta_0)^{-1} + o_p(1)).$$

These and Theorem 1 together show that the following equation holds

$$T_n - b_n = 2(nh)^{-1}\sum Z_{ni} + 2(nh)^{-2}\sum_{1\le i<j\le n} H_n(X_i, X_j)$$
$$+ h^2 d_K \int (f_0''(x;\theta_0)\frac{\partial}{\partial\theta}(f_0(x;\theta_0)\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)dx$$
$$\times n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_0(X_{i;\theta_0})(I(\theta_0)^{-1} + o_p(1))$$
$$+ O_p(n^{-1}) + O_p(n^{-3/2}h^{-1}).$$

The proof can be completed similarly as in Kim *et al.*(1997).

Remark : Let $f_1(x)$ be the true density and θ_0 be the minimizer of the Kullback-Leibler divergence between $f_1(x)$ and $f_0(x;\theta)$, i.e., $\theta_0 = \arg\min_\theta \int \log(\frac{f_1(x)}{f_0(x;\theta)}) f_1(x)\,dx$. Basu et al. (1998) proved that the consistency (to θ_0) and asymptotic normality of the MLE $\widehat{\theta_n}$. Using this, it can be easily proved that $T_n \to \int (f_0(x;\theta_0) - f_1(x))^2 w(x)dx$ (in probability). This shows that our test is consistent.

3. Monte Carlo Power Study

3.1 Critical values

For the modified classical goodness-of-fit tests D_n, W_n^2 and A_n^2, Stephens(1974) gives the approximate quantiles for some families of null distributions. Table 1 gives the quantiles for some functions of the above statistics under the normal distribution $N(\mu, \sigma^2)$.

Table 1. Quantiles for Functions of D_n, W_n^2, and A_n^2

	Quantile				
Function	.85	.90	.95	.975	.99
$(n^{1/2} - .01 + .85n^{-1/2})D_n$	.775	.819	.895	.955	1.035
$(1 + .5n^{-1})W_n^2$	.091	.104	.126	.148	0.178
$(1 + 4n^{-1} - 25n^{-2})A_n^2$	.576	.656	.787	.918	1.092

In section 2, we proved that the proposed test statistic T_n converges in distribution to normal random variable. But The empirical 95th percentile of $n^{9/10}(T_n - b_n)$ were found to be quite different from the 95th percentile of $N(0,1)$, which is because of the very slow rate of the convergence. Therefore, in the power comparison of the next subsection we use the empirical critical values instead of the percentiles of $N(0,1)$.

3.2 Simulation results

In this subsection, we will compare the performance of the proposed test statistic T_n with the Kolmogorov-Smirnov statistic D_n, the Cramér-von Mises statistic W_n^2 and the Anderson-Darling statistic A_n^2. This was done via simulation. The simulation study consists of four cases of null distribution: standard normal distribution $N(0,1)$, the chi-square distribution $\chi^2(6)$, the Weibull distribution $W(2,1)$, and the exponential distribution $Exp(3)$. For each case, 1,000 replications are done for n=50, 100, 200. All random variables are generated by the IMSL subroutine. To find the kernel density estimates, we use the Gaussian kernel when the null density is normal and use the Epanechinikov kernel when the null density is chi-square, Weibull or exponential. We set the weight function $w(x)$ for T_n the indicator function of null distribution support. The bandwidth h is selected by minimizing the asymptotic mean integrated square error(AMISE).

When the null distribution is normal, as the critical value for T_n we use the empirical percentile, while critical values for D_n, W_n^2, and A_n^2 are taken from Table 1. When the null distribution is chi-square, Weibull, or exponential, the critical values for all of the 4 test statistics are empirical percentiles. To compute the empirical powers of the tests, the number of rejections of H_0 are counted for each test. The empirical power is the number of rejections of H_0 divided by 1,000. For all cases, we set the corresponding alternatives as $(1-\varepsilon)F_Z + \varepsilon F_Y$, where F_Z is cdf of the null distribution and F_Y is cdf of contamination distribution. We use ε = 0., 0.1, 0.2, 0.3, 0.4, 0.5

First, for the normal null distribution we set the corresponding alternatives as $(1-\varepsilon)F_Z + \varepsilon F_Y$, where $Z \sim N(0,1)$ and $Y \sim Beta(2,2)$. The results of the power study are shown in Table 2, which shows that the powers of T_n are worse than those of the other statistics on the whole. But as ε gets larger and n increases, the power of T_n improves and becomes almost the same as the other statistics.

Second, for the case of the χ^2 null distribution we set the alternatives as $(1-\varepsilon)F_Z+\varepsilon F_Y$, where F_Z and F_Y denote cdf. of $Z\sim\chi^2(6)$ and $(Y-6)\sim Beta(2,2)$, respectively. Table 3 shows the empirical powers for the four tests. When the sample size n=100, T_n takes .384 as the power for ε=.1. But under the same condition, D_n, W_n^2 and A_n^2 takes .294, .263 and .189, respectively. From the results we can see that T_n has better power than the other statistics.

Third, for the Weibull null distribution case we set the alternatives as $(1-\varepsilon)F_Z+\varepsilon F_Y$, where F_Z and F_Y denote cdf. of $Z\sim Weibull(2,1)$ and $(Y-1)\sim Beta(3,2)$, respectively. Table 4 shows that T_n is more powerful than the other statistics.

Fourth and Last, when the null distribution is exponential, the alternatives are $(1-\varepsilon)F_Z+\varepsilon F_Y$, where F_Z and F_Y denote cdf. of $Z\sim \mathrm{Exp}(3)$ and $(Y-3)\sim B(2,2)$, respectively. The powers of T_n, D_n, W_n^2 and A_n^2 are given in Table 5, which gives similar result as the previous case.

TABLE 2. Empirical Powers of Tests when $H_0: f\sim Normal$

		ε					
n	Statistic	.0	.1	.2	.3	.4	.5
50	T_n	.050	.075	.199	.433	.616	.836
	D_n	.046	.090	.267	.546	.760	.886
	W_n^2	.045	.101	.340	.635	.849	.949
	A_n^2	.053	.102	.310	.608	.834	.941
100	T_n	.050	.101	.348	.738	.950	.998
	D_n	.052	.158	.493	.841	.974	.997
	W_n^2	.059	.199	.543	.896	.986	1.000
	A_n^2	.057	.169	.497	.873	.985	0.999
200	T_n	.050	.194	.731	.982	1.000	1.000
	D_n	.061	.293	.815	.992	1.000	1.000
	W_n^2	.059	.328	.856	.998	1.000	1.000
	A_n^2	.052	.281	.833	.997	1.000	1.000

TABLE 3. Empirical Powers of Tests when $H_0: f \sim \chi^2$

		ε					
n	Statistic	.0	.1	.2	.3	.4	.5
50	T_n	.050	.206	.630	.916	.998	1.000
	D_n	.051	.169	.542	.878	.989	1.000
	W_n^2	.051	.167	.526	.865	.993	1.000
	A_n^2	.050	.126	.408	.779	.976	.997
100	T_n	.051	.384	.939	.998	1.000	1.000
	D_n	.050	.294	.860	.995	1.000	1.000
	W_n^2	.051	.263	.851	.995	1.000	1.000
	A_n^2	.051	.189	.757	.983	1.000	1.000
200	T_n	.051	.769	.998	1.000	1.000	1.000
	D_n	.051	.590	.990	1.000	1.000	1.000
	W_n^2	.051	.572	.989	1.000	1.000	1.000
	A_n^2	.050	.458	.968	1.000	1.000	1.000

TABLE 4. Empirical Powers of Tests when $H_0: f \sim$ *Weibull*

		ε					
n	Statistic	.0	.1	.2	.3	.4	.5
50	T_n	.050	.301	.683	.950	.996	1.000
	D_n	.051	.188	.527	.879	.985	.999
	W_n^2	.050	.212	.560	.886	.984	.999
	A_n^2	.050	.227	.594	.909	.987	.999
100	T_n	.051	.466	.929	.998	1.000	1.000
	D_n	.050	.325	.852	.998	1.000	1.000
	W_n^2	.051	.372	.869	.995	1.000	1.000
	A_n^2	.050	.405	.884	.998	1.000	1.000
200	T_n	.050	.702	.999	1.000	1.000	1.000
	D_n	.051	.542	.994	1.000	1.000	1.000
	W_n^2	.051	.621	.992	1.000	1.000	1.000
	A_n^2	.051	.663	.996	1.000	1.000	1.000

TABLE 5. Empirical Powers of Tests for $H_0: f \sim$ Exp*onential*

		ε					
n	Statistic	.0	.1	.2	.3	.4	.5
50	T_n	.051	.239	.712	.954	.998	1.000
	D_n	.052	.172	.557	.864	.985	.999
	W_n^2	.050	.195	.597	.882	.988	1.000
	A_n^2	.050	.161	.513	.839	.985	1.000
100	T_n	.051	.417	.945	.999	1.000	1.000
	D_n	.052	.285	.841	.994	1.000	1.000
	W_n^2	.051	.309	.865	.992	1.000	1.000
	A_n^2	.051	.260	.821	.991	1.000	1.000
200	T_n	.051	.727	1.000	1.000	1.000	1.000
	D_n	.050	.500	.991	1.000	1.000	1.000
	W_n^2	.050	.522	.990	1.000	1.000	1.000
	A_n^2	.050	.457	.988	1.000	1.000	1.000

4. Concluding Ramarks

In this paper, we have studied the goodness-of-fit test in density estimation problem when the null hypothesis is composite. We modified the L^2-distance statistic which was proposed by Kim *et al.*(1997), and proved the asymptotic normality of the test statistic for the case of one dimensional parameter theta. For the multi-dimensional parameter theta, the asymptotic normality could also be shown in the similar way. The power comparison between the test statistic T_n and the other classical test statistics is done through the Monte Carlo methods. Because of the slow convergence rate we use the empirical critical values for T_n. For testing normality, T_n shows worse power than those of the other statistics for small n but shows almost the same power for large n and ε. For the other three underlying distribution, T_n shows better powers than the other statistics.

APPENDIX

Proof of Theorem 1.

(i). Since $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$ by Remark (b), it is sufficient to show that

$$\int (\frac{\partial}{\partial\theta} f_0(x;\theta_*))^2 w(x)dx = O_p(1).$$

By the assumption (A3),

$$\int (\frac{\partial}{\partial\theta} f_0(x;\theta_*))^2 w(x)dx \le \int H_1^2(x)w(x)dx \ \equiv \ M < \infty$$

if $|\theta_* - \theta_0| \le c$ for some c. Therefore

$$P\Big(\int (\frac{\partial}{\partial\theta} f_0(x;\theta_*))^2 w(x)dx > M\Big) \ \le \ P(|\theta_* - \theta_0| > c)$$
$$\le \ P(|\hat\theta - \theta_0| > c) \to 0, \quad \text{as } n \to \infty.$$

This shows that $\int (\frac{\partial}{\partial\theta} f_0(x;\theta_*))^2 w(x)dx = O_p(1)$, which completes the proof of (i).

(ii). $E\Big[\int (\hat f_h(x) - E\hat f_h(x))\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)dx\Big] = 0,$

and

$$\begin{aligned}
&Var\Big[\int (\hat f_h(x) - E\hat f_h(x))\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)dx\Big] \\
&= n^{-2}\sum_{i=1}^{n} Var\Big[\int K_h(x-X_i)\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)dx\Big] \\
&= n^{-1}\Big[E\{\int K_h(x-X_1)\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)dx\}^2 \\
&\quad - \{E(\int K_h(x-X_1)\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)dx\}^2\Big].
\end{aligned} \tag{A.1}$$

For the notational convenience, let $\dot f_0(x;\theta_0) = \frac{\partial}{\partial x} f_0(x;\theta_0)$ and $\ddot f_0(x;\theta_0) = \frac{\partial^2}{\partial x^2} f_0(x;\theta_0)$. The second term of (A.1) is

$$\begin{aligned}
&E\Big(\int K_h(x-X_i)\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)dx\Big) \\
&= \int\int K_h(x-u)\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)f_0(u;\theta_0)dxdu \\
&= \int \frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)\int K(t)f_0(x-ht;\theta_0)dtdx,
\end{aligned}$$

and by the Taylor series expansion, this is equal to

$$\begin{aligned}
&\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)\int K(t)(f_0(x;\theta_0) - ht\dot f_0(x;\theta_0) \\
&\qquad + \frac{h^2t^2}{2}\ddot f_0(x;\theta_0) + o(h^2))dtdx \\
&= \int \frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)f_0(x;\theta_0)dx + O(h^2).
\end{aligned}$$

Also the first term of (A.1) is

$$
\begin{aligned}
& E\Big(\int K_h(x-X_i)\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)dx\Big)^2 \\
& = E\Big[\int\int K_h(x-X_i)\frac{\partial}{\partial\theta} f_0(x;\theta_0)\,w(x) \\
& \qquad \times K_h(y-X_i)\frac{\partial}{\partial\theta} f_0(y;\theta_0)\,w(y)dxdy\Big] \\
& = \int\int\int K_h(x-u)\frac{\partial}{\partial\theta} f_0(x;\theta_0)\,w(x) \\
& \qquad \times K_h(y-u)\frac{\partial}{\partial\theta} f_0(y;\theta_0)\,w(y)\,f_0(u;\theta_0)\,dudydx \\
& = \int\int\int K(t)\frac{1}{h}K(\frac{u-x}{h}+t)\frac{\partial}{\partial\theta} f_0(x;\theta_0)\,w(x) \\
& \qquad \times \frac{\partial}{\partial\theta} f_0(y;\theta_0)w(y)f_0(x-ht;\theta_0)dtdxdy \\
& = \int\int\int K(t)K(t+s)\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x) \\
& \qquad \times \frac{\partial}{\partial\theta} f_0(x+hs;\theta_0)w(x+hs)f_0(x-ht;\theta_0)dtdsdx \\
& = \int\int\int K(t)K(t+s)\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x) \\
& \qquad \times \{\frac{\partial}{\partial\theta} f_0(x;\theta_0)+hs\frac{\partial}{\partial\theta} f_0'(x;\theta_0)+\frac{h^2s^2}{2}\frac{\partial}{\partial\theta} f_0''(x;\theta_0)+o(h^2)\} \\
& \qquad \times \{w(x)+hsw'(x)+\frac{h^2s^2}{2}w''(x)+o(h^2)\} \\
& \qquad \times \{f_0(x;\theta_0)-htf_0'(x;\theta_0)+\frac{h^2t^2}{2}f_0''(x;\theta_0)+o(h^2)\}dtdsdx \\
& = \int(\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x))^2 f_0(x;\theta_0)dx+O(h^2),
\end{aligned}
$$

by substituting $t=(x-u)/h$ and $s=(y-x)/h$. Therefore,

$$
\begin{aligned}
& Var\Big[\int(\hat{f}_h(x)-E\,\hat{f}_h(x))\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)dx\Big] \\
& = n^{-1}\Big[\int(\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x))^2 f_0(x;\theta_0)dx \\
& \qquad -\Big(\int\frac{\partial}{\partial\theta} f_0(x;\theta_0)w(x)f_0(x;\theta_0)dx)^2\Big]+o(n^{-1}) \\
& = n^{-1}\Big[E(\frac{\partial}{\partial\theta} f_0(X;\theta_0)w(X))^2-(E\frac{\partial}{\partial\theta} f_0(X;\theta_0)w(X))^2\Big]+o(n^{-1}) \\
& = n^{-1}Var[\frac{\partial}{\partial\theta} f_0(X;\theta_0)w(X)]+o(n^{-1}),
\end{aligned}
$$

where $X \sim f_0(x;\theta_0)$. Thus,

$$\int (\hat{f}_h(x) - E\, \hat{f}_h(x)) \frac{\partial}{\partial \theta} f_0(x;\theta_0) w(x) dx = O_p(n^{-1/2}). \tag{A.2}$$

Now consider the second term of (A.1). It was shown in Kim *et al.* (1997) that

$$\int ((\hat{f}_h(x) - E\, \hat{f}_h(x))^2 w(x) dx = O_p((nh)^{-1}). \tag{A.3}$$

Similarly as in the proof of (i), it can be shown that

$$\int (\frac{\partial^2}{\partial \theta^2} f_0(x;\theta^*))^2 w(x) dx = O_p(1). \tag{A.4}$$

(A.3), (A.4), and the Cauchy-Schwartz inequality together show that

$$\int (f_n(x) - Ef_n(x))(\frac{\partial^2}{\partial \theta^2} f_0(x;\theta^*)) w(x) dx = O_p((nh)^{-1/2}). \tag{A.5}$$

The proof of (ii) is completed by (A.2) and (A.5). □

References

[1] Basu, A., Harris, I. L., Hjort, N. L., and Jones, M. C. (1998). "Robust and efficient estimation by minimizing a density power divergence", *Biometrika*, 85, 549-559.

[2] Cox, D. D., Koh, E., Wahba, G. and Yandell, B. (1988). "Testing the (parametric) null hypothesis in (semiparametric) partial and generalized spline models" , *The Annals of Statistics*, 16, 113-119.

[3] Durbin, J. (1975). "Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests no spacings" , *Biometrika*, 62, 5-22.

[4] Eubank, R. L. (1988). *Spline Smoothing and Nonparametric regression*, New York : Marcel Dekker Inc.

[5] Eubank, R. L. and Hart, J. D. (1992). "Testing goodness-of-fit in regression via order selection criterion" , *The Annals of Statistics*, 20, 1412-1425.

[6] Eubank, R. L. and Spiegelman, C. H. (1990). "Testing the goodness-of-fit of a linear model via nonparametric regression techniques" , *Journal of the American Statistical Association*, 85, 387-392.

[7] Härdle, W. (1990), *Applied Nonparametric Regression*, New York: Cambridge University Press.

[8] Härdle, W. and Mammen, E. (1993). "Comparing nonparametric regression versus parametric regression fits" , *The Annals of Statistics*, 21, 1926-1947.

[9] Kim, C., Hong, C., Jeong, M. and Yang, M. (1997). "Goodness-of-fit test for density estimation" , *Communications in Statistics - Theory and methods*, 26(11), 2725-2741.

[10] Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons Inc.

[11] Margolin, B. H. and Maurer, W. (1976). "Tests of the Kolmogorov-Smirnov type for exponential data with unknown scale and related problems" , *Biometrika*, 63, 149-160.

[12] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, New-York : John Wiley & Sons Inc.

[13] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, London : Chapman and Hall.

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

[15] Stephens, M. A. (1976). "Asymptotic results for goodness-of-fit statistics with unknown parameters" , *The Annals of Statistics*, 4, 357-369.