한국통계학회논문집 제2권 1호
1995년 4월 pp. 122-136

Goodness-of-Fit Test Based on Smoothing Parameter Selection Criteria[1]

Jong Tae Kim[2]

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 a new test statistic $\widehat{\lambda_a}$ is investigated.

The test statistic $\widehat{\lambda_a}$ is itself a smoothing parameter which is selected to minimize an estimated MISE for a truncated series estimator 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 $\widehat{\lambda_a}$ is obtained under the null hypothesis. It is also shown that this test is consistent against fixed alternatives.

1. Introduction

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. Set D be the *df* of $Y=G(X)$. Then, $D(u)=F(G^{-1}(u))$ with corresponding density function

$$d(u)=f(G^{-1}(u))/g(G^{-1}(u)), \quad 0<u<1,$$

where g is the density of G and $G^{-1}(u)=\inf\{x: G(x)\geq u\}$. The function d is called the *comparison density function* (Parzen, 1979). In the viewpont of GOF testing H_0 is equivalent testing $d(u)=1$. Thus, one method for H_0 is to find a consistent estimator $\hat{d}$

1) This paper was supported in part by the reserch grant from Institude of Natural Science, Taegu University.
2) Department of Statistics, Taegu University, Taegu, 713-714, KOREA.

of d and then cares $\hat{d}$ and 1. In fact, $\hat{d}(G(x))g(x)$ leads immediately to a estimate the true density for the data.

A generalized cosine Fourier series expansion of d can be written as

$$d(u) = 1 + \sum_{j=1}^{\infty} a_j \cos(j\pi u).$$

where the a_j's are generalized Fourier coefficients, $a_j = \int_0^1 d(u)\cos(j\pi u)du$. Eubank, LaRiccia and Rosenstein (1987) used Pearson's phi-squared distance measure (Lancaster, 1969), $\phi^2 = \int_0^1 (d(u)-1)^2 du$, and they (1992) studied this problem with the Neyman smooth type tests based on nonparametric density function estimation. Therefore, testing H_0 is equivalent to testing $\phi^2 = 0$ or to testing $a_j = 0$ for all j.

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

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

which is called the components of Cramér-von Mises statistic (Durbin and Knott, 1972). If $0 \le \lambda \le n$ is some integer, then a Fourier cosine series estimator $d_{\lambda n}$ of the comparison density d is

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

The integer λ in (1) 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.

In Section 2 we discuss the problem of choosing an optimal smoothing parameter for the optimal comparison density estimator. We also propose test based on nonparametric density estimation and obtain its asymptotic null distribution. This test statistic is, in fact, a data-driven smoothing parameter. Section 3 provides a numerical example using our tests. In Section 4 we study the finite sample power properties from a small scale simulation.

2. The Proposed Tests

2.1 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.

The method for choosing λ that will be focused on here is the data driven procedure proposed by Hart (1985). For this purpose, observe that by Parseval's identity $L(\lambda)=\sum_{j=1}^{\lambda}(\tilde{a}_{jn}^2-2a_j\tilde{a}_{jn})+\sum_{j=1}^{\infty}a_j^2$ and the MISE for the comparison density is given by

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

where $M(\lambda)=\sum_{j=1}^{\lambda}(a_j^2-\sigma_j^2)$. and $\sigma_j^2=\mathrm{var}(\tilde{a}_{jn})=n^{-1}(1+a_{2j}/\sqrt{2}-a_j^2)$. The last term, $\sum_{j=1}^{\infty}a_j^2$, in (2) 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 unbiased estimator $\widehat{M}(\lambda)$ of $M(\lambda)$ can be easily obtained by $\widehat{M}(\lambda)=0$, if $\lambda=0$, and

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

where $\hat{\sigma}_{jn}^2=(1+\tilde{a}_{2jn}/\sqrt{2}-\tilde{a}_{jn}^2)/(n-1)$is an unbiased estimator of $\sigma_j^2=\mathrm{var}(\tilde{a}_{jn})$.

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 $\widehat{M}(\lambda)$ over $0 \le \lambda \le n$. We denote this value by $\hat{\lambda}$ in all that follows.

The estimator of the optimal smoothing parameter obtained by maximizing $\widehat{M}(\lambda)$ 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.

2.2 The Proposed Test

From our previous discussions, we know that $a_j = 0$ for all $j \ge 1$ under the null hypothesis. Thus, $M(\lambda) = -\lambda / n$ 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 $\widehat{\lambda_\alpha}$ to be the maximizer of

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

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

$$\text{Reject} \quad H_0 \quad \text{if} \quad \widehat{\lambda_\alpha} \ge 1. \tag{3}$$

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

$$1 - \alpha = \exp\left\{ - \sum_{j=1}^{\infty} \frac{P(\chi_j^2 > jc)}{j} \right\}, \tag{4}$$

where χ_m^2 is a random variable having a chi-squared distribution with m degrees of freedom. Under this choice for C_α, $P(\widehat{\lambda_\alpha} = 0) \to 1 - \alpha$ as $n \to \infty$. Equation (4) is a consequence of

Theorem 2.3.1 and (4.7) of Spitzer(1956). When $C_\alpha = 2$, the test (3) is using the maximizer of $M(\lambda)$ in (4) and the asymptotic level of the test is about .29. As noted above, this is the reason why we use the test (3).

Table 2.1. Values of C_α which make test (3) asymptotically valid at level α

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

An attractive feature of the test (3) is that it leads immediately to a point estimate of the comparison density in the event that H_0 is rejected. The data analyst generally desires such an estimate if there is reason to believe the null model is inadequate. In our setting, a natural estimator of $d(u)$ is

$$d_{\lambda n}(u) = 1 + \sum_{j=1}^{\widehat{\lambda}_\alpha} \tilde{a}_{jn} \sqrt{2} \cos(j\pi u). \tag{5}$$

In fact, $\hat{d}_{\lambda n}(G(x))g(x)$ leads immediately to a estimate of the true density for the data.

Theorem 2.1. The asymptotic distribution of $\widehat{\lambda}_\alpha$ under H_0 is given by

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

where $q_0 = 1$ 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 $N_1 + 2N_2 + \cdots + rN_r = r$.

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

Theorem 2.2. If $|a_{j_0}|>0$ for some $0<j_0<\infty$, the power of the test based on $\widehat{\lambda_\alpha}$ tends to 1 as $n\to\infty$.

3. A Numerical Examples.

The salinity data is obtained from a large scale environmental inpact study by North Carolina State University concerning the Cape Fear Estuary. The response is water salinity measured in parts per thousand (ppt) at the time of collection of a larval specimen and consists of six samples (cf. D'Agostino and Stephens 1986; Rupport and Carroll 1980).

A question of interest is the following: Is the salinity data uniformly distributed over the time of collection of a larval specimen with six samples?

Table 3.1. Test results for the Salinity Data.

	$\alpha=.01$	$\alpha=.05$	$\alpha=.10$
LM	0 (0)	0 (0)	2 (0)
CVM	0.193 (.743)	0.193 (.461)	0.193 (.347)
AD	1.326 (3.85)	1.326 (2.49)	1.326 (1.93)

The values in parentheses in this table are the percentage points for each significance level α.

Table 3.1 gives the tests result for the salinity data. Only the test based on $\widehat{\lambda_\alpha}$ (LM) indicates that the hypothesis of uniformity should be rejected with significance level of .10. The Cramér-von Mises (CVM) and Anderson-Darling (AD) tests also can not reject H_0.

The values of $\tilde{a}_{jn}=\frac{1}{n}\sum_{r=1}^{n}\sqrt{2}\cos(j\pi Y_r)$ are $-.06864$, -0.33687, $-.02874$, $.00329$, $-.08494$, $-.09312$, $.10837$, $-.084954$, and $-.07350$ for $j=1,2,\ldots,9$, respectively. When $j=2$, we can see that the magnitude of $\tilde{a}_{2n}$ is much bigger than that of the other sample Fourier coefficients and, not surprizingly, the values of $\hat{\lambda}$ and $\hat{\lambda}_{.10}$ both turn out to be 2.

As discussed in Section 2, our tests have the property that they also provide an estimator of the true comparison density function. The value of $\hat{\lambda}_{.10}$ can be used as a smoothing parameter in the estimator $d_{\lambda n}$ of (5). For example, with the salinity data, an estimator of the comparison density is given by

$$d_{2n}(u) = 1 + \sum_{j=1}^{2} \tilde{a}_{jn} \sqrt{2} \cos(j\pi u)$$
$$= 1 - \sqrt{2}(.06864 \cos(\pi u) + .33687 \cos(2\pi u)).$$

4. Simulation

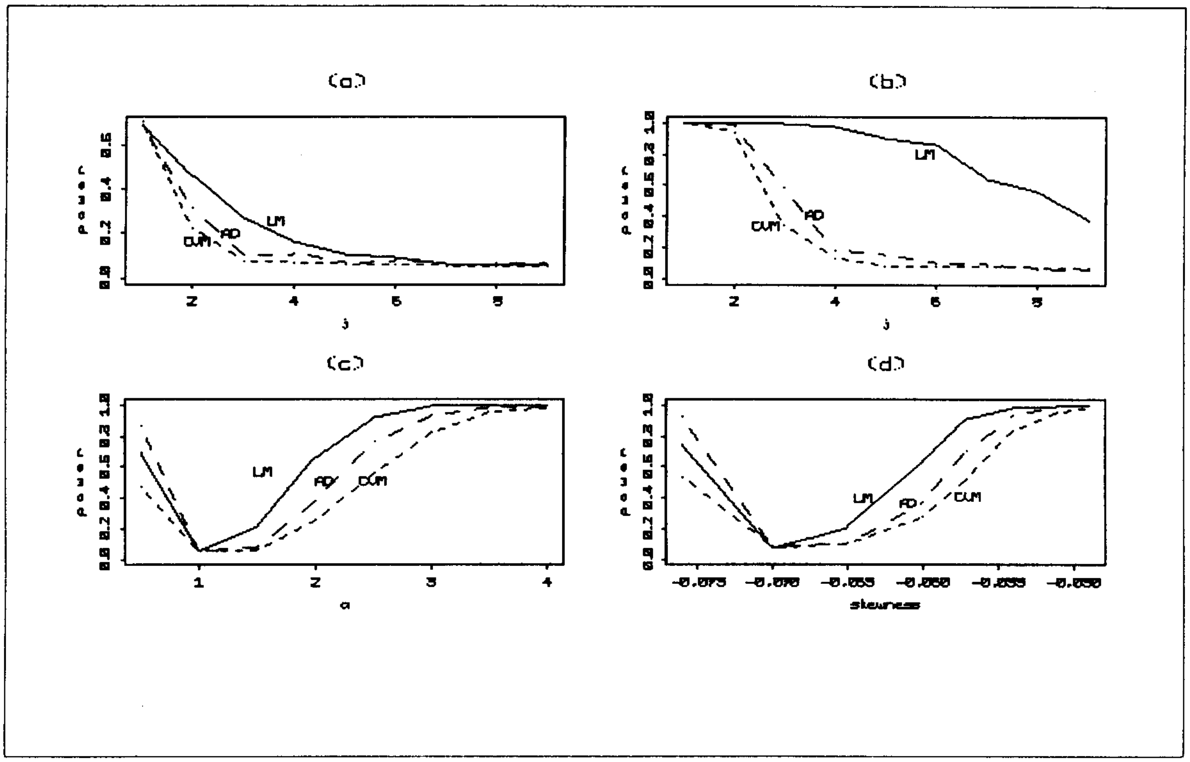

Figure 1. Empirical Power Functions with $\alpha = .05$; (a) Cosine alternatives with $\gamma = 0.5$; (b) cosine alternatives with $\gamma = 1$; (c) symmetric beta alternatives; (d) skewed beta alternatives with mean $= .48$.

We will focus on two types of alternatives; the cosine alternatives

$$d(u) = 1 + \gamma \cos(j\pi u) \quad \text{for } j = 1, 2, \ldots, \quad \text{and } 0 \le \gamma \le 1,$$

and a beta density. For all three cases we generate 1000 replicate samples of size 50 and use these to assess the empirical power of our tests.

The cosine alternatives are included to assess the behavior of the tests for high frequency alternatives. The choice of γ determines the distance of $d(\cdot)$ from the null density, while j can be manipulated to obtain higher or lower frequency departures from uniformity. The beta density alternatives are considered primarily to observe the behavior of the four tests for symmetric and skewed alternatives. The skewed beta density alternatives with mean .48 look like the symmetric beta alternatives. In empirical power studies,we consider only the case of very slightly skewed beta density alternatives, since the powers of all four tests are very sensitive to departures from a symmetric beta density and give powers of one for even mildly skewed alternatives.

The critical values used for CVM and AD are taken from Shorack and Wellner (1986).

Empirical powers are graphed as function of the frequency of the cosine alternatives for $\alpha=.05$ in Figures 1 (a) and (b) with $\gamma=0.5$ and 1.0, respectively. The tests based on $\widehat{\lambda_\alpha}$, CVM and AD have good power at $j=1$, but, as j increases, the power of LM, $\widehat{\lambda_\alpha}$, drops off gradually until $j=5$, while the powers of CVM and AD drop off dramatically for $j>2$.

Figure 1 (c) and (d) are for empirical power against symmetric and skewed beta alternatives, respectively. When the beta alternatives are U - shaped, the AD test, and $\widehat{\lambda_\alpha}$ are significantly better than CVM. When these alternatives are unimodal, $\widehat{\lambda_\alpha}$ have significantly higher power than the other two.

5 Appendix: Proofs

In this section we prove Theorem 2.1 and Theorem 2.2 We begin by establishing several lemmas that are required for the proofs.

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

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

can be represented as

$$p_r=\sum_r^*\left\{\prod_{k=1}^r\frac{1}{N_k!}\left(\frac{a_k}{k}\right)^{N_k}\right\},$$

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

and $\alpha_k = P(S_k > 0)$.

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

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

b) $P(|\frac{1}{m}\sum_{j=1}^{m} V_j^2 - 1| > \varepsilon) \le \frac{2}{m\varepsilon^2} - \frac{3}{2nm\varepsilon^2}$, and, as $n \to \infty$

c) $P(\sup_{1\le m\le n} |\frac{1}{m}\sum_{j=1}^{m} \tilde{a}_{2jn}| > \varepsilon) \to 0$,

d) $P(\sup_{1\le m\le n} |\frac{1}{m}\sum_{j=1}^{m} \tilde{a}_{jn}^2| > \varepsilon) \to 0$.

Proof. For part (b), by using the multinomial thorem $\text{var}(V_j^2) = 2 - 3/2n$ and when $j \ne l$, $\text{cov}(V_j^2, V_l^2) = 0$. Thus, by Markov's inequality, $P(|\frac{1}{m}\sum_{j=1}^{m} V_j^2 - 1| > \varepsilon) \le \frac{1}{m^2\varepsilon^2}\,\text{var}(\sum_{j=1}^{m} V_j^2)$ as was to be shown.

For part (c), Set $\tilde{a}_{2jn} = V_{2j}/\sqrt{n}$ with $V_{2j} = \sqrt{2}\sum_{r=1}^{m}\cos(2j\pi U_r)/\sqrt{n}$ so that $EV_{2j} = 0$ and $\text{cov}(V_{2j}, V_{2l}) = \delta_{jl}$. Then $|\frac{1}{m}\sum_{j=1}^{m}\tilde{a}_{2jn}| \le \frac{1}{\sqrt{n}}\frac{1}{\sqrt{m}}\sqrt{\sum_{j=1}^{m} V_{2j}^2}$ for each $m = 1, 2, \cdots, n$. Let $t_n \to \infty$ as $n \to \infty$ with $t_n = o(n)$ (e.g. $t_n = \log n$). Then

$$\sup_{1\le m\le n}|\frac{1}{m}\sum_{j=1}^{m}\tilde{a}_{2jn}| \le \sup_{1\le m\le t_n}|\frac{1}{m}\sum_{j=1}^{m}\tilde{a}_{2jn}| + \sup_{t_n < m\le n}|\frac{1}{m}\sum_{j=1}^{m}\tilde{a}_{2jn}|.$$

But, for $m \le t_n$, $\frac{1}{\sqrt{n}}\frac{1}{\sqrt{m}}\sqrt{\sum_{j=1}^{m} V_{2j}^2} \le \frac{1}{\sqrt{n}}\sqrt{\sum_{j=1}^{t_n} V_{2j}^2}$ and, from part (b), for any $\eta > 0$

$$P(\frac{1}{n}\sum_{j=1}^{t_n} V_{2j}^2 > \eta) \le P(|\frac{1}{t_n}\sum_{j=1}^{t_n} V_{2j}^2 - 1| > \frac{n}{t_n}\eta - 1)$$
$$\le \frac{2t_n}{(n\eta - t_n)^2} - \frac{3t_n}{2n(n\eta - t_n)^2} \to 0,$$

as $n\to\infty$ if $t_n = o(n)$. For $m > t_n$, $\frac{1}{\sqrt{n}}\frac{1}{\sqrt{m}}\sqrt{\sum_{j=1}^{m} V_{2j}^2} \le \frac{1}{\sqrt{t_n}}\sqrt{\frac{1}{n}\sum_{j=1}^{n} V_{2j}^2}$ and, from part (b), as $n\to\infty$

$$P(\frac{1}{nt_n}\sum_{j=1}^{n} V_{2j}^2\ \eta) \le P(|\frac{1}{n}\sum_{j=1}^{n} V_{2j}^2 - 1| > t_n\eta - 1)$$
$$\le \frac{2}{n(t_n\eta - 1)^2} - \frac{3}{2n^2(t_n\eta - 1)} \to 0.$$

(d) follows by observing that $\frac{1}{m}\sum_{j=1}^{m} \tilde{a}_{jn}^2 = \frac{1}{mn}\sum_{j=1}^{m} V_{jn}^2$.

Proof of Theorem 2.1

Let $\hat{\lambda}$ be the maximizer of $\hat{M}(\lambda)$ over $0 \le \hat{\lambda} \le n$ with $\hat{M}(0) = 0$ and $\hat{M}(\lambda) = \sum_{j=1}^{\lambda}(\tilde{a}_{jn}^2 - c\,\hat{\sigma}_{jn}^2)$ for some constant c. By definition, we choose $\hat{\lambda} = \lambda$ if $\hat{M}(\lambda) - \hat{M}(m) \ge 0$, for all $m = 1, 2, \ldots, n$. To simplify notation, only consider the case $\hat{\lambda} = 0$, as the general case follows similarly.

Now $\hat{\lambda} = 0$ if and only if $\hat{M}(m) \le 0$, for $m = 1, \ldots, n$. Therefore, for any $\varepsilon > 0$,

$$\begin{aligned} P(\hat{\lambda}=0) = \ & P(\frac{1}{m}\sum_{j=1}^{m}(n\,\tilde{a}_{jn}^2 - c) \le \frac{nc}{\sqrt{2m(n+c-1)}}\sum_{j=1}^{m}\tilde{a}_{2jn} - \frac{c(c-1)}{n+c-1}, \\ & m = 1,2,\ldots,n;\ \sup_{1\le m\le n}|\frac{nc}{\sqrt{2m(n+c-1)}}\sum_{j=1}^{m}\tilde{a}_{2jn}| \le \varepsilon) \\ & + P(\frac{1}{m}\sum_{j=1}^{m}(n\,\tilde{a}_{jn}^2 - c) \le \frac{nc}{\sqrt{2m(n+c-1)}}\sum_{j=1}^{m}\tilde{a}_{2jn} - \frac{c(c-1)}{n+c-1}, \\ & m = 1,2,\ldots,n;\ \sup_{1\le m\le n}|\frac{nc}{\sqrt{2m(n+c-1)}}\sum_{j=1}^{m}\tilde{a}_{2jn}| > \varepsilon). \end{aligned} \quad (10)$$

But, by the part (d) of Lemma 2, the second term in (10) is to be 0. Thus it now suffices to work with, as $n\to\infty$,

$$P\Big(\frac{1}{m}\sum_{j=1}^{m}(n\,\tilde{a}^2_{jn}-C^*)\le 0,\quad m=1,2,\ldots,n\Big),$$

for some fixed C^*. Then, we can take $C^*=c-\varepsilon$ and $c+\varepsilon$, and let $\varepsilon\to 0$ to get the desired result. Let M_n be a function of n such that $M_n\to\infty$ as $n\to\infty$. We can always choose M_n to grow sufficiently slowly, for example, $M_n=\log n$. Then we have

$$P\Big(\frac{1}{m}\sum_{j=1}^{m(n}\tilde{a}^2_{jn}-C^*)\le 0,\quad m=1,2,\ldots n\Big)\quad=\quad P(A_n\cap B_n),$$

where $A_n=\sum_{j=1}^{m}(n\,\tilde{a}^2_{jn}-C^*)\le 0\ /\ m,\ m=1,\ldots,M_n$ and $B_n=\sum_{j=1}^{m}(n\,\tilde{a}^2_{jn}-C^*)/M\le 0,\ m=M_{n+1},\ldots,n.$

To prove $P(B_n)\to 1$, define the subsequence of integers $m_k=k^2, k=1,2,\ldots$. Let k_2 be the largest integer k such that $k^2 < M_n$ and k_3 be the largest integer k such that $k^2\le n$. Define $m_{k_3+1}=n$, and for each n and $k=k_2,\ldots,k_3$, note that if m_k is such that $k_2^2\le m_k < n$, and if $m_k < m\le m_{k+1}$, then

$$\frac{\left|\sum_{i=1}^{m}(n\,\tilde{a}^2_{jn}-1)\right|}{m}\le\frac{\left|\sum_{j=1}^{m_k}(n\,\tilde{a}^2_{jn}-1)\right|}{m_k}+\frac{\Psi_{kn}}{m_k},$$

where $\Psi_{kn}=\max_{1\le p\le m_{k+1}-m_k}\left|\sum_{m=m_{k+1}}^{m_{k+p}}(na^2_{kn}-1)\right|$. It follows that

$$B_n\supset\bigcap_{m=M_n}^{n}\Big\{\frac{\left|\sum_{i=1}^{m}(n\,\tilde{a}^2_{jn}-1)\right|}{m}\le(C^*-1)\Big\}$$
$$\supset\bigcap_{k=k_2}^{k_4}\Big\{\Big(\frac{\left|\sum_{i=1}^{m_k}(n\,\tilde{a}^2_{jn}-1)\right|}{m_k}\le\frac{(C^*-1)}{2}\Big)\cap\Big\{\frac{\Psi_{kn}}{m_k}\le\frac{(C^*-1)}{2}\Big\}\Big\}$$

with $k_4 = k_3$ if $k_3^2 < n$ and $k_4 = k_3 - 1$ otherwise. By Markov's inequality and part (a) of Lemma 2,

$$P\Big(\bigcup_{k=k_2}^{k_4} \Big|\sum_{i=1}^{m_k}(n\,\tilde{a}_{jn}^2 - 1)\Big| > \frac{m_k(C^* - 1)}{2}\Big) \le \frac{4}{(C^* - 1)^2} \sum_{k=k_2}^{k_4} k^{-2}, \tag{11}$$

Hence the bound (11) is of the order $\sum_{k=k_2}^{k_4} k^{-2}$, which tends to 0 as $M_n \to \infty$.

For any set of jointly distributed random variables $Z_1, \ldots, Z_n$ with joint df F, let g be the functional $g(F) = \sum_{i=1}^{n} E(Z_i + D)$ with $D = 2$. Define $F_{r,k}$ to be the joint distribution of $a_{(r+1)n}, \ldots, a_{(r+k)n}$ for all r and k. It is clear that $g(F_{r,k}) = 2k$, $g(F_{r,k}) + g(F_{r+k,m}) = g(F_{r,k+m})$, and, by part (a) of Lemma 2, $E(\sum_{j=r+1}^{r+k}(n\,\tilde{a}_{jn}^2 - 1))^2 \le 2k = g(F_{r,k})$. Thus, we may apply Theorem A of Serfling (1970) to obtain

$$E\Psi_{kn}^2 \le \frac{(\log(4k+2))^2(2k+1)}{(\log 2)^2}.$$

Hence, as $M_n \to \infty$,

$$P\Big(\bigcap_{k=k_2}^{k_4} \frac{\Psi_{kn}}{m_k} \le \frac{(c-1)}{2}\Big) \ge 1 - \sum_{k=k_2}^{k_4} \frac{4D(\log(4k+2))^2(2k+1)}{(c-1)^2(\log 2)^2 k^4}$$

which tends to 1. Combining the preceeding results yields $\lim_{n\to\infty} P(B_n) = 1$.

Let $Z_j,\ j = 1, 2, \ldots, M_n$ be iid $N(0,1)$ random variables and set

$$A_n^* = \frac{1}{m}\sum_{j=1}^{m}(Z_j^2 - C^*) \le 0, \quad m = 1, \ldots, M_n.$$

We need to prove $P(A_n)-P(A_n^*) \rightarrow 0$.

Set $C_{in}=(\cos(\pi U_i),\cos(2\pi U_i),\dots,\cos(M_n\pi U_i))'$, for $i=1,2,\dots,n$. We note that $EC_{in}=0$, $\frac{1}{n}\sum_{i=1}^{n}\mathrm{var}(C_{in})=I_{M_n}$, and $\frac{1}{n}\sum_{i=1}^{n}E\|C_{in}\|^4=\frac{1}{4}M_n^2+\frac{1}{8}M_n\le M_n^2$, where $\|\cdot\|$ is the Euclidean norm and I_{M_n} is the M_n dimensional identity matrix. Therefore, by Theorem 13.3 of Bhattacharaya and Ranga Rao (1976),

$$|P(A_n)-P(A_n^*)|\le a(M_n)M_n^2/\sqrt{n},$$

where $a(M_n)$ is a positive constant that depends only on M_n. Since one can always choose M_n to grow sufficiently slowly that $a(M_n)M_n^2/\sqrt{n}\to 0$, it follows that, as $n\to\infty$, $|P(A_n)-P(A_n^*)|\to 0$.

Now, we consider the form of $\lim_{n\to\infty}P(A_n^*)$ using Lemma 1 and results in Spitzer(1956). First, by Lemma 1,

$$P(A_n^*)=\{\sum_{M_n}^{*}\prod_{m=1}^{M_n}\frac{1}{N_m!}(\frac{1-a_m}{m})^{N_m}\},$$

where $\sum_{M_n}^{*}$ extends over all M_n-tuples of integers $(N_1,\dots,N_{M_n})$ such that $N_1+2N_2+\dots+M_nN_{M_n}=M_n$, and, for $m=1,2,\dots,M_n$, $a_m=P(\sum_{j=1}^{m}(Z_j^2-C^*)/m>0)=\Pr(\chi_m^2>C^*m)$. Thus, for $M_n\to\infty$ as $n\to\infty$, by (4.7) of Spitzer (1956),

$$\lim_{n\to\infty}P(A_n^*)=\exp\{-\sum_{m=1}^{\infty}\frac{a_m}{m}\}.$$

To finish the proof of this theorem we get the lower and upper bound for $\lim_{n\to\infty}P(\hat{\lambda}=0)$ by plugging in $C^*=c-\varepsilon$ and $c+\varepsilon$ into our limiting expression for $P(A_n^*)$ and letting $\varepsilon\to 0$. Finally, by taking $c=C_\alpha$, we obtain the conclusion of the theorem.

Proof of Theorem 2.2

To prove Theorem 2.5.2, first suppose there exists some j_0 such that $|a_{j_0}|>0$. Then,

$$P(\hat{\lambda}_a\geq 1)=1-P(\hat{\lambda}_a=0)$$
$$\geq 1-P(\sum_{j=1}^{j_0}(n\tilde{a}_{jn}^2-\frac{n}{n-1}C_a)\leq\frac{C_a n}{(n-1)}\sum_{j=1}^{j_0}(\frac{1}{\sqrt{2}}\tilde{a}_{2jn}-\tilde{a}_{jn}^2)).$$

Now, we see that $\frac{C_a n}{(n-1)}\sum_{j=1}^{j_0}(\frac{1}{\sqrt{2}}\tilde{a}_{2jn}-\tilde{a}_{jn}^2)$ is a $\sqrt{n}$-consistent estimator for $-C_a j_0+C_a\sum_{j=1}^{j_0}2\int_0^1\cos^2(j\pi u)d(u)du-a_j^2$. Thus, for sufficiently large n and any $\varepsilon>0$, we have

$$|\frac{C_a n}{n-1}\sum_{j=1}^{j_0}(\frac{1}{\sqrt{2}}\tilde{a}_{2jn}-\tilde{a}_{jn}^2))-(-C_a j_0+C_a\sum_{j=1}^{j_0}2\int_0^1\cos^2(j\pi u)d(u)du-a_j^2)|<\varepsilon$$

with probability tending to one. By Jensen's inequality, $2\int_0^1\cos^2(j\pi u)d(u)du>a_j^2$ and hence, $C_a\sum_{j=1}^{j_0}2\int_0^1\cos^2(j\pi u)d(u)du-a_j^2+\varepsilon=\varepsilon'>0$. Consequently,

$$P(\hat{\lambda}_a\geq 1)\geq 1-P(\sum_{j=1}^{j_0}n\tilde{a}_{jn}^2\leq\varepsilon')+o(1)$$
$$\geq 1-P(n\tilde{a}_{j_0 n}^2\leq\varepsilon')+o(1).$$

An application of the Berry-Esséen Theorem then gives

$$P(n\tilde{a}_{j_0 n}^2\leq\varepsilon')=\Phi(\frac{\sqrt{\varepsilon'}-\sqrt{n}a_{j_0}}{\sigma_{j_0}})-\Phi(\frac{-\sqrt{\varepsilon'}-\sqrt{n}a_{j_0}}{\sigma_{j_0}})+O(\frac{1}{\sqrt{n}})$$

with Φ the standard normal *df*. The theorem follows by taking limits since $a_{j_0} \neq 0$.

References

[1] Bhattacharya, R. N. and Ranga Rao, R. (1976). *Normal Approximation and Asymptotic Expansions*, Wiley, New York.

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

[3] Diggle, 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.

[4] Durbin, J. and Knott, M.(1972). Components of Cramér-von Mises Statistics, *Journal of the Royal Statistical Society, Series B*, 34, 290-307

[5] 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.

[6] Eubank, R. and LaRiccia, V. N. (1992). Asymptotic Comparison of Cramér-von Mises and Nonparametric Function Estimation Technicques for Testing Goodness-of-Fit, *The Annals of Statistics*, 20, 2071-2086.

[7] 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.

[8] Lancaster, H. O.(1969). *The Chi-Squared Distribution*, Wiley, New York.

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

[10] Ruppport, D. and Carroll, R. J.(1980). Trimmed Least Squares Estimation in the Linear Model, *Journal of the American Statistical Association*, 75, 828-838.

[11] Shorack, G. and Wellner, J.(1986). *Empirical Processes with Applications to Statistics*, Wiley: New York.

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