SORT 29 (2) July-December 2005, 201-216

On the role played by the fixed bandwidth in the Bickel-Rosenblatt goodness-of-fit test

Carlos Tenreiro

Universidade de Coimbra

Abstract

For the Bickel-Rosenblatt goodness-of-fit test with fixed bandwidth studied by Fan (1998) we derive its Bahadur exact slopes in a neighbourhood of a simple hypothesis $f = f_0$ and we use them to get a better understanding on the role played by the smoothing parameter in the detection of departures from the null hypothesis. When f_0 is a univariate normal distribution and we take for kernel the standard normal density function, we compute these slopes for a set of Edgeworth alternatives which give us a description of the test properties in terms of the bandwidth h. A simulation study is presented which indicates that finite sample properties are in good accordance with the theoretical properties based on Bahadur local efficiency. Comparisons with the quadratic classical EDF tests lead us to recommend a test based on a combination of bandwidths in alternative to Anderson-Darling or Cramér-von Mises tests.

MSC: 62G10, 62G20

Keywords: goodness-of-fit test, kernel density estimator, Bahadur efficiency.

1 Introduction

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of independent and identically distributed d-dimensional random vectors with unknown density function f. As it has been shown by Bickel and Rosenblatt (1973), a test of the simple hypothesis $H_0 : f = f_0$ against the alternative $H_a : f \neq f_0$, where f_0 is a fixed density function on $\mathbb{R}^d$, can be based on the L_2 distance between the kernel density estimator of f introduced by Rosenblatt (1956)

Address for correspondence: Carlos Tenreiro. Departamento de Matemática, Universidade de Coimbra, Apartado 3008, 3001-454 Coimbra, Portugal. Phone: (351) 239 791 155. Fax: (351) 239 832 568. E-mail: tenreiro@mat.uc.pt

Received: July 2004
Accepted: April 2005

and Parzen (1962), and its mathematical expectation under the null hypothesis (see also Fan (1994) and Gouriéroux and Tenreiro (2001)):

$$I_n^2(h_n) = n \int \{f_n(x) - \mathrm{E}_0 f_n(x)\}^2 dx, \tag{1}$$

where, for $x \in \mathbb{R}^d$,

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(x - X_i),$$

$K_{h_n} = K(\cdot/h_n)/h_n^d$ with K a kernel, that is, a bounded and integrable function on $\mathbb{R}^d$, and (h_n) is a sequence of strictly positive real numbers converging to zero, when n goes to infinity (bandwidth). The Bickel-Rosenblatt test is asymptotically consistent and has a normal asymptotic distribution under the null hypothesis.

Following an idea of Anderson, Hall and Titterington (1994) that have used kernel density estimators with fixed bandwidth for testing the equality of two multivariate probability density functions, Fan (1998) uses the statistic (1) with a constant bandwidth for testing the composite hypothesis that f is a member of a general parametric family of density functions. He provides an alternative asymptotic approximation for the finite-sample properties of the Bickel-Rosenblatt test by showing that, for a fixed h, the asymptotic distribution of $I_n^2(h)$ is an infinite sum of weighted χ^2 random variables. Moreover, Fan (1998) proves that $I_n^2(h)$ can be interpreted as a L_2 weighted distance between the empirical characteristic function and the parametric estimate of the characteristic function implied by the null model with weight function $t \to |\phi_K(th)|^2$. In the important case of testing univariate or multivariate normality, and taking for K the standard normal density function, the role played by h in the power performance of the test is assessed in simulation studies by Epps and Pulley (1983), Henze and Zirkler (1990) and Henze and Wagner (1997).

Restricting our attention to the test of a simple hypothesis, the main purpose of this paper is to derive the Bahadur local exact slopes of goodness-of-fit tests based on $I_n^2(h)$, for a fixed $h > 0$, and use them to get a better understanding of the role played by the smoothing parameter in the detection of departures from the null hypothesis. For completeness reasons we give in Section 2 the asymptotic null distribution and the consistency of the test based on kernel density estimators with a fixed bandwidth. Using the integral and quadratic form of $I_n^2(h)$, we derive in Section 3 its Bahadur local exact slopes. They naturally depend on the smoothing parameter, on the kernel, on the null density f_0 and, finally, on the considered departure direction from the null hypothesis. In Section 4, in the particular case of a test for a simple univariate hypothesis of normality and taking for K the standard normal density function, the Bahadur local slopes are numerically evaluated for different values of h for a set of Edgeworth alternatives. These alternatives express departures from the null hypothesis in terms of each one of the first four moments. The tests based on $I_n^2(h)$ for different values of h are compared with the corresponding ones of the quadratic EDF tests of Anderson-Darling (A^2) and Cramér-von Mises (W^2). The results we obtain suggest that a large bandwidth is adequate for detection of location alternatives whereas a small bandwidth is adequate for detection of alternatives for scale, skewness and kurtosis. A simulation study indicating that finite sample properties of tests I^2 are in good accordance with the theoretical properties based on the Bahadur local slopes is also presented. Moreover, if one does not know much about the unknown density function it suggests that a test based on a combination of bandwidths, that establish a compromise between the two opposite effects that the choice of h has in the detection of location and nonlocation alternatives, is a good practical recommendation in alternative to traditional A^2 or W^2 tests.

For convenience of presentation the proofs of some results in this article are given in Section 5. We denote by $\xrightarrow[n\to+\infty]{as}$ the convergence with probability 1 and by $\xrightarrow[n\to+\infty]{d}$ the convergence in distribution.

2 Asymptotic null distribution and consistency

Consider the following assumptions on K which ensure that $d(f,g) = \left(\int \{K_h \star f(x) - K_h \star g(x)\}^2 dx\right)^{1/2}$, where $\star$ denotes the convolution product, is a distance on the set of integrable functions (see Anderson *et al.* (1994)).

Assumptions on K (K)

K is a bounded and integrable function on $\mathbb{R}^d$ with Fourier transform ϕ_K such that $\{t \in \mathbb{R}^d : \phi_K(t) = 0\}$ has Lebesgue measure zero.

In order to derive the asymptotic distribution of $I_n^2(h)$ under H_0 for a fixed $h > 0$, we first note that $I_n^2(h)$ is a V-statistic, that is,

$$I_n^2(h) = \frac{1}{n} \sum_{i,j=1}^{n} Q_h(X_i, X_j), \tag{2}$$

with kernel

$$Q_h(u,v) = \int k(x,u;h)k(x,v;h)dx,$$

where

$$k(x,u;h) = K_h(x-u) - K_h \star f_0(x), \tag{3}$$

for $u,v,x \in \mathbb{R}^d$. From the hypothesis on K, the kernel Q_h is bounded. Therefore the functions $u \to Q_h(u,u)$ and Q_h are P_0 and $P_0 \otimes P_0$ integrable, respectively, where $P_0 = f_0\lambda$ and λ is the Lebesgue measure in $\mathcal{B}(\mathbb{R}^d)$. Moreover, Q_h is symmetric and degenerate, i.e., $\int Q_h(\cdot, v)dP_0(v) = 0,\ a.e.\ (P_0)$. From Gregory (1977), we know that the asymptotic distribution of $I_n^2(h)$ under H_0 can be characterized in terms of the eigenvalues of the symmetric Hilbert-Schmidt operator A_h defined, for $q \in L_2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P_0) =: L_2(P_0)$, by

$$(A_h q)(u) = \int Q_h(u,v)q(v)dP_0(v). \tag{4}$$

In view of the degeneracy property of Q_h, $q_{0,h} = 1$ is an eigenfunction of A_h corresponding to the eigenvalue $\lambda_{0,h} = 0$. Denoting by $\langle\, 1 \,\rangle$ the subspace generated by $q_{0,h}$ and $H(P_0) = \left\{g \in L_2(P_0) : \int g dP_0 = 0\right\}$ the tangent space of P_0, we have $L_2(P_0) = \langle\, 1 \,\rangle \oplus H(P_0)$. The operator A_h is positive definite on $H(P_0)$ as follows from the integral form (3) of Q_h and assumption (K). In fact, if $\langle A_h q, q\rangle = 0$, for some $q \in H(P_0)$, where $\langle\cdot,\cdot\rangle$ denotes the usual inner product in $L_2(P_0)$, we have

$$\begin{aligned} 0 &= \int q(u)k(\cdot, u; h)dP_0(u) \\ &= K \star (qf_0)(\cdot),\ a.e.\ (\lambda), \end{aligned}$$

yielding $\phi_K(t)\phi_{qf_0}(t) = 0$, for all $t \in \mathbb{R}^d$.

From assumption (K) and the continuity of the Fourier transform, we deduce that $\phi_{qf_0}(t) = 0$, $t \in \mathbb{R}^d$, i.e., $q = 0,\ a.e.\ (P_0)$.

Finally, using the the infinite-dimensionality of $H(P_0)$ and the positivity of A_h on $H(P_0)$ we can conclude that A_h has a countable infinite collection $\{\lambda_{k,h}, k \in \mathbb{N}\}$ of strictly positive eigenvalues (see Dunford and Schwartz (1963), Corollary X.4.5).

The following result follows from the limit distribution of degenerate V-statistics (cf. Theorem 4.3.2 of Koroljuk and Borovskich (1989)). Remark that the asymptotic distribution presented by Fan (1998) in Theorem 4.2, is not correct. In general the P_0-integrability of $u \to Q_h(u,u)$ is not a sufficient condition for $\sum \lambda_{k,h} < \infty$.

Theorem 1 *If assumption* (K) *is fulfilled then, under* H_0 *we have*

$$I_n^2(h) \underset{n\to+\infty}{\overset{d}{\longrightarrow}} I_\infty,$$

with

$$I_\infty = \int Q_h(u,u)dP_0(u) + \sum_{k=1}^{\infty} \lambda_{k,h}(Z_k^2 - 1),$$

where the sequence $(\lambda_{k,h})$*, with* $\lambda_{1,h} \geq \lambda_{2,h} \geq \ldots$ *and* $\lambda_{k,h} \to 0, k \to +\infty$*, is described above and* (Z_k) *are i.i.d. standard normal variables. Moreover, the test* $I^2(h) = (I_n^2(h))$

defined by the critical regions $\{I_n^2(h) > c_\alpha\}$*, where* $\mathrm{P}(I_\infty > c_\alpha) = \alpha$*, is asymptotically of level* α *and consistent to test* H_0 *against* H_a*.*

Remark 1 If the density f_0 has a compact support S and Q_h is continuous in $S \times S$, from the Mercer's expansion for Q_h (see Dunford and Schwartz (1963), p. 1088) it follows that $\int Q_h(u,u)dP_0(u) = \sum_{k=1}^\infty \lambda_{k,h}$ and therefore I_∞ takes the form $I_\infty = \sum_{k=1}^\infty \lambda_{k,h} Z_k^2$.

3 Bahadur local efficiency

In order to compare the test $I^2(h)$ with other test procedures, or to compare $I^2(h)$ tests obtained for different values of h, we derive in the following its Bahadur exact slopes $C_{I^2(h)}(f)$, for f in a neighbourhood of f_0. They coincide with the Bahadur approximate slopes (and then with the Bahadur local approximate slopes) derived by Gregory (1980). For the description of Bahadur's concept of efficiency, see Bahadur (1967, 1971) or Nikitin (1995).

Throughout, $|\cdot|_p$ denotes the norm of the Lebesgue space $L_p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda) =: L_p(\lambda)$. The proof of the following result is given in Section 5.

Theorem 2 *We have*

$$C_{I^2(h)}(f) = \frac{b_{I^2(h)}(f)}{\lambda_{1,h}}(1 + o(1)), \; as \; |f - f_0|_1 \to 0,$$

where

$$b_{I^2(h)}(f) = \int \{K_h \star f(x) - K_h \star f_0(x)\}^2 dx,$$

and $\lambda_{1,h}$ *is the largest eigenvalue of the operator* A_h *defined by (4).*

If f_0 belongs to a family of probability density functions of the form $\{f(\cdot;\theta) : \theta \in \Theta\}$, where Θ is a nontrivial closed real interval and $f_0 = f(\cdot;\theta_0)$, for some $\theta_0 \in \Theta$, it is natural to compare a set of competitor tests through its Bahadur local exact slopes when $\theta \to \theta_0$.

Consider the following assumptions on the previous parametric family:

Assumptions on $\{f(\cdot;\theta) : \theta \in \Theta\}$ (P)

For all $x \in \mathbb{R}^d$ the function $\theta \to f(x;\theta)$ is continuously differentiable on Θ, and there exists a neighbourhood $V \subset \Theta$ of θ_0 such that the function $x \to \sup_{\theta \in V} \left|\frac{\partial f}{\partial \theta}(x;\theta)\right|$ is integrable on $\mathbb{R}^d$.

The following result comes easily from Theorem 2, assumption (P) and the dominated convergence theorem.

Corollary 1 *Under assumption* (P)*, we have*

$$\| f(\cdot;\theta) - f(\cdot;\theta_0)\|_1 \to 0, \ \textit{when}\ \theta \to \theta_0,$$

and

$$C_{I^2(h)}(f(\cdot;\theta)) = \frac{b^o_{I^2(h)}(f(\cdot;\theta))}{\lambda_{1,h}}\,(\theta-\theta_0)^2\,(1+o(1)),\ \textit{when}\ \theta\to\theta_0$$

where

$$b^o_{I^2(h)}(f(\cdot;\theta)) = \int \Big(K_h \star \frac{\partial f}{\partial \theta}(\cdot;\theta_0)(x)\Big)^2 dx.$$

Let us denote by $\{q_{k,h}, k \in \mathbb{N}_0\}$ the orthonormal basis for $L_2(P_0)$ corresponding to the infinite collection of eigenvalues of A_h, i.e., for all k and j, $\int Q_h(\cdot,v)q_{k,h}(v)dP_0(v) = \lambda_{k,h}q_{k,h},\ \ a.e.\ (P_0)$ and $\langle q_{k,h}, q_{j,h}\rangle = \delta_{kj}$, where δ_{kj} is the Kronecker symbol. In the following result, we establish a representation for the local slope $C_{I^2(h)}(f(\cdot;\theta))$ when $\theta\to\theta_0$, in terms of the weights $(\lambda_{k,h})$ and the principal components $(q_{k,h})$. It is proven in Section 5.

Corollary 2 *Under assumption* (P)*, if* $\dfrac{\partial \ln f}{\partial\theta}(\cdot;\theta_0) \in L_2(P_0)$*, then*

$$C_{I^2(h)}(f(\cdot;\theta)) = \sum_{k=1}^{\infty} \frac{\lambda_{k,h}}{\lambda_{1,h}}\, a^2_{k,h}(\theta-\theta_0)^2(1+o(1)),\ \textit{when}\ \theta\to\theta_0,$$

where, for $k = 1, 2, \ldots,$

$$a_{k,h} = \Big\langle q_{k,h}, \frac{\partial \ln f}{\partial\theta}(\cdot;\theta_0)\Big\rangle.$$

From the previous representation, in particular from the fact that the weights $(\lambda_{k,h})$ converge to zero, it is clear that only a finite directions of alternatives effectively contribute to $C_{I^2(h)}(f(\cdot;\theta))$. The natural question, that we discuss in the next section for the test of a simple hypothesis of normality, is how rapidly the principal directions loose influence.

4 Testing a simple hypothesis of normality

In this section we consider the test of the simple hypothesis of normality. Without loss of generality we restrict our attention to the test of the hypothesis $H_0 : f = f_{N(0,1)}$ against the alternative hypothesis $H_a : f \neq f_{N(0,1)}$.

4.1 Local alternatives

In order to get a better understanding of the role played by the smoothing parameter in the detection of departures from the null hypothesis, we consider a set of alternatives that satisfy (P) with $f = f_j$ and $\theta_0 = 0$, such that

$$(\mathcal{A}.\,j) \qquad \frac{\partial \ln f_j}{\partial \theta}(\cdot; 0) = H_j(\cdot)/j!, \tag{5}$$

for $j = 1, \ldots, 4$, where H_j is the jth Hermite polynomial defined by:

$$\begin{aligned} H_1(x) &= x; \\ H_2(x) &= x^2 - 1; \\ H_3(x) &= x^3 - 3x; \\ H_4(x) &= x^4 - 6x^2 + 3. \end{aligned}$$

These alternatives are based on the Edgeworth series for the density and the corresponding value of θ indicate departures from the null hypothesis in the jth moment (about Edgeworth expansion see Hall (1997) and the references therein). Remark that the location alternative $f(\cdot;\theta) = f_{N(\theta,1)}(\cdot)$ and the scale alternative $f(\cdot;\theta) = f_{N(0,1+\theta)}(\cdot)$, when $\theta \to 0$, satisfy $(\mathcal{A}.1)$ and $(\mathcal{A}.2)$, respectively. The alternative $f(\cdot;\theta) = 2 f_{N(0,1)}(\cdot) F_{N(0,1)}(\theta \cdot)$, when $\theta \downarrow 0$, considered by Durio and Nikitin (2003), satisfies $(\mathcal{A}.1)$ up to the multiplication by a constant. Finally, the skew and kurtosis alternatives considered by Durbin *et al.* (1975) satisfy $(\mathcal{A}.3)$ and $(\mathcal{A}.4)$, respectively.

4.2 The test statistic

From now on we take for K the standard normal density $K = f_{N(0,1)}$. This choice for the kernel was mainly motivated by the fact that the function $b^o_{I^2(h)}(f(\cdot;\theta))$ given in Corollary 1 can be explicitly evaluated for the set of alternatives described above. Also remark that in this case the calculation of $I_n^2(h)$ does not involve any integration. In fact, the kernel Q_h given by (3) takes the form

$$Q_h(u,v) = f_{N(0,2h^2)}(u-v) - f_{N(0,2h^2+1)}(u) - f_{N(0,2h^2+1)}(v) + f_{N(0,2h^2+2)}(0), \tag{6}$$

for $u, v \in \mathbb{R}$ (see Bowman (1992), Bowman and Foster (1993) and Henze and Wagner (1997)).

4.3 Most significant weights

As described in Section 3, the Bahadur local slope of $I_n^2(h)$ depends on the weights $(\lambda_{k,h})$ and on the principal components $(q_{k,h})$. Numerical evaluations of the most significant weights are shown in Table 1 for four values of h. These approximations have been obtained through the projection method. We have considered the restriction, $A_{h|L}$, of the operator A_h defined by (4) with kernel given by (3) to the finite dimension subspace L of $H(P_0)$ given by $L = \{g \in H(P_0) : g = \sum_{i=1}^n g(\bar{x}_i)\mathbb{I}_{]x_i,x_{i+1}]}\}$, where $n = 1400$, $x_i = -7 + 0.01(i-1)$ and $\bar{x}_i = (x_i + x_{i+1})/2$, for $i = 1, \ldots, n$. The numerical calculation of the eigenvalues of $A_{h|L}$ have been performed using Lapack routines (cf. Anderson *et al.* (1999)).

From these values and the representation for the Bahadur local slopes given in Corollary 2, we expect that test $I_n^2(h)$ for small values of h could use information contained in others components different from the first ones. However, for moderate or large values of h, it appears that $I_n^2(h)$ might exclusively use information contained in the first components.

Table 1: *Weights for $I^2(h)$ with $K = f_{N(0,1)}$ and $f_0 = f_{N(0,1)}$*

	$h = 0.05$	$h = 0.2$	$h = 1.0$	$h = 2.0$
$\lambda_{1,h}$	3.59×10^{-1}	2.61×10^{-1}	5.53×10^{-2}	1.28×10^{-2}
$\lambda_{2,h}$	3.54×10^{-1}	2.36×10^{-1}	2.16×10^{-2}	1.93×10^{-3}
$\lambda_{3,h}$	3.11×10^{-1}	1.49×10^{-1}	3.97×10^{-3}	1.31×10^{-4}
$\lambda_{4,h}$	3.06×10^{-1}	1.29×10^{-1}	1.32×10^{-3}	1.65×10^{-5}
$\lambda_{5,h}$	2.70×10^{-1}	8.46×10^{-2}	2.85×10^{-4}	1.33×10^{-6}
$\lambda_{6,h}$	2.64×10^{-1}	7.15×10^{-2}	8.90×10^{-5}	1.57×10^{-7}
$\lambda_{7,h}$	2.35×10^{-1}	4.82×10^{-2}	2.05×10^{-5}	1.36×10^{-8}
$\lambda_{8,h}$	2.29×10^{-1}	4.00×10^{-2}	6.17×10^{-6}	1.55×10^{-9}
$\lambda_{9,h}$	2.04×10^{-1}	2.74×10^{-2}	1.47×10^{-6}	1.39×10^{-10}
$\lambda_{10,h}$	1.98×10^{-1}	2.24×10^{-2}	4.33×10^{-7}	1.54×10^{-11}
$\lambda_{11,h}$	1.77×10^{-1}	1.56×10^{-2}	1.06×10^{-7}	1.42×10^{-12}
$\lambda_{12,h}$	1.71×10^{-1}	1.26×10^{-2}	3.06×10^{-8}	1.54×10^{-13}

4.4 Bahadur local exact slopes

Similarly to the quadratic EDF tests of Anderson-Darling (A^2) and Cramér-von Mises (W^2) (see Nikitin (1995), p. 73–81), for each one of the alternatives (5) the Bahadur local exact slopes of the tests based on $I^2(h)$ take the form $\theta^2(1+o(1))$, up to the multiplication by a constant, when $\theta \to 0$. Therefore, for the comparison of such tests it is sufficient to compare the coefficients of θ^2. They are usually called local indices and are plotted in Figure 1 for $h \in [0.01, 3]$ and Q_h given by (6). We also plot the local indices for A^2 and W^2 tests.

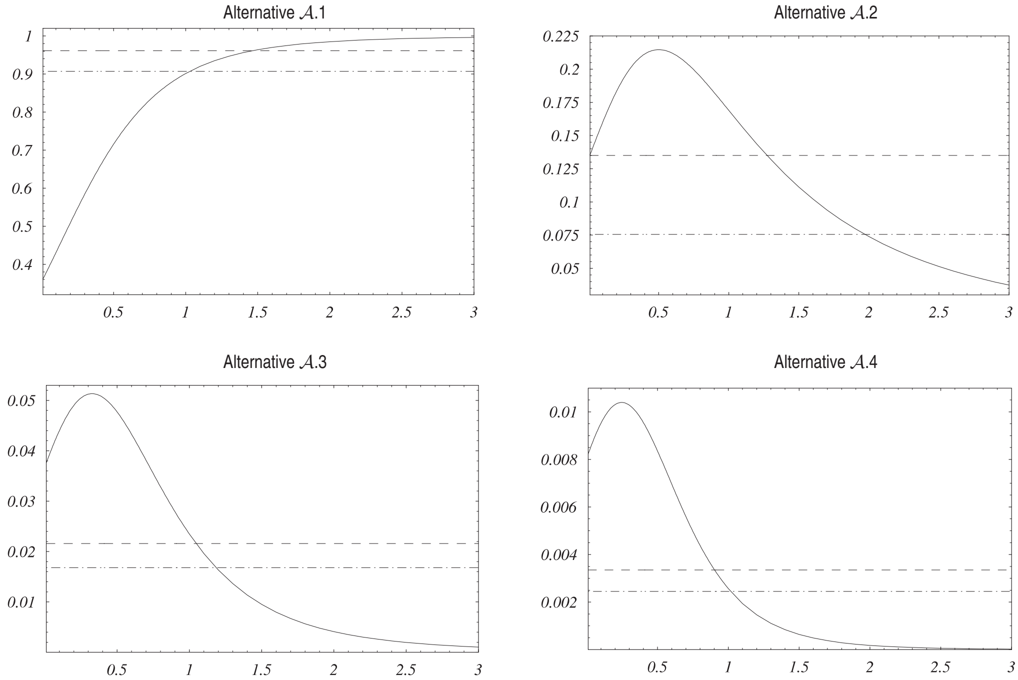

Figure 1: *Local indices for:* $I^2(h)$ *– solid line;* A^2*– broken line;* W^2*– broken and dotted line*

It is clear from Figure 1 that a large bandwidth leads to a strong predominance of the first principal component whereas a small bandwidth leads to a test that uses the information contained in the other components. For the location alternative, we note that the local indices obtained numerically for large values of h are close to one which is, from Bahadur-Raghavachari inequality (see Nikitin (1995), Theorem 1.2.3), the optimal Bahadur local efficiency for this alternative. However, the gain of efficiency in the location alternative by taking a large value of h implies a severe loss of efficiency in the other moment alternatives.

4.5 Combining bandwidths effects

A compromise between the two opposite effects that the choice of h has in the detection of location and nonlocation alternatives can be achieved by considering a test based on a combination of bandwidths, i.e., a test based on the statistic (2) with $K_h = (1-\alpha)K_{h_1} + \alpha K_{h_2}$, where h_1 (small bandwidth) and h_2 (large bandwidth) are two fixed bandwidths, and $\alpha \in [0, 1]$.

Denoting by $I^2(\alpha; h_1, h_2)$ such test and assuming that $\{f(\cdot;\theta) : \theta \in \Theta\}$ satisfies (P), we have

$$
\begin{aligned}
&b^o_{I^2(\alpha;h_1,h_2)}(f(\cdot;\theta))\\
&\quad = \quad (1-\alpha)^2\, b^o_{I^2(h_1)}(f(\cdot;\theta)) + \alpha^2\, b^o_{I^2(h_2)}(f(\cdot;\theta))\\
&\qquad\qquad + 2\alpha(1-\alpha)\int K_{h_1} \star \frac{\partial f}{\partial \theta}(\cdot;\theta_0)(x) K_{h_2} \star \frac{\partial f}{\partial \theta}(\cdot;\theta_0)(x)\, dx.
\end{aligned}
$$

For alternatives (5) we plot in Figure 2 the local indices for the combined test $I^2(\alpha; 0.3, 2.0)$ for $\alpha \in [0.7, 1]$. Notice that $h_1 = 0.3$ and $h_2 = 2.0$ are appropriated bandwidths for the detection of nonlocation and location alternatives, respectively (see Figure 1). It follows that the test $I^2(0.8; 0.3, 2.0)$ is superior to W^2 for all the considered alternatives ($\mathcal{A}$.1-4), and is superior to A^2 for alternatives ($\mathcal{A}$.2-4). Remark that this behaviour cannot be achieved by a test $I^2(h)$ for a fixed h (see Figure 1). The test $I^2(0.9; 0.3, 2.0)$ is superior to A^2 for alternative ($\mathcal{A}$.1) but is inferior to A^2 for alternatives ($\mathcal{A}$.2-4). However, the loss of efficiency for these last alternatives is not as significant as if we take a test $I^2(h)$ with a relative local Bahadur efficiency close to one with respect to $I^2(0.9; 0.3, 2.0)$ for alternative ($\mathcal{A}$.1).

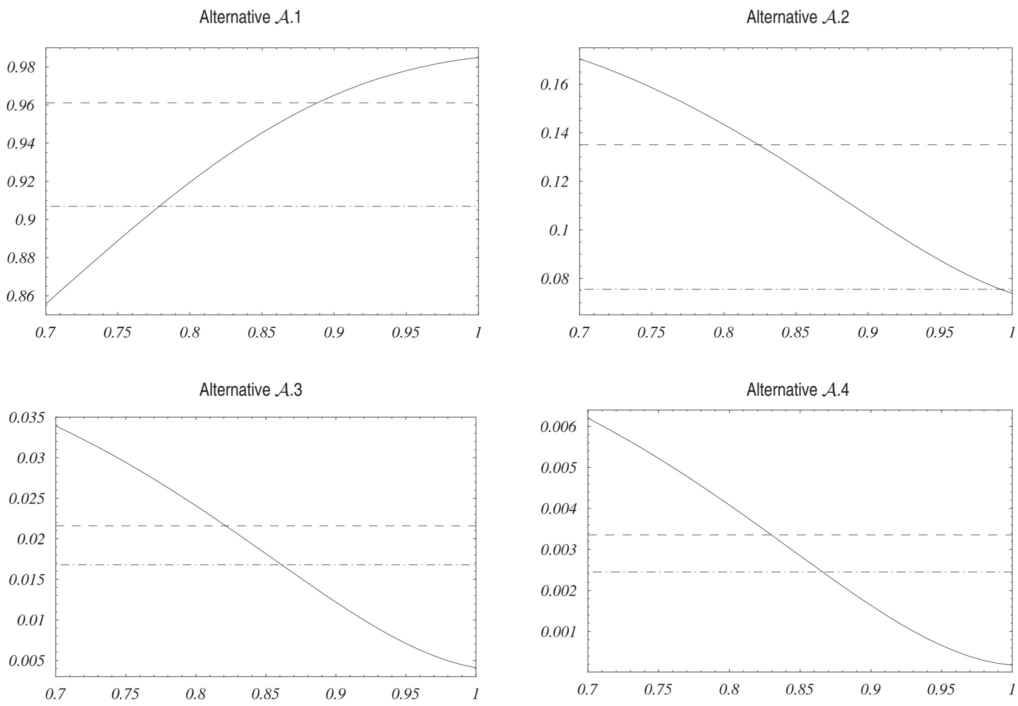

Figure 2: *Local indices for:* $I^2(\alpha; 0.3, 2.0)$ *– solid line;* A^2*– broken line;* W^2*– broken and dotted line*

4.6 Some simulation results

The main purpose of this section is to know if the finite sample properties of the I^2 tests for fixed alternatives are in accordance with the theoretical properties based on Bahadur local efficiency. For that reason we present a simulation study including the tests $I^2(0.3)$ (small bandwidth), $I^2(0.8)$ (medium bandwidth) and $I^2(2.0)$ (large bandwidth) based on fixed bandwidths, and the test $I^2(c) := I^2(0.8; 0.3, 2.0)$ based on a combination of bandwidths. Moreover, as before, the EDF tests A^2 and W^2 will be use for comparison.

To examine the performance of these tests when the null hypothesis is false, we consider three normal alternatives and four nonnormal alternative distribution shapes shown in Figure 3. The nonnormal distributions are members of the generalized lambda family discussed in Ramberg and Schmeiser (1974). The distributions of this family are easily generated because they are defined in terms of the inverses of the cumulative distribution functions: $F^{-1}(u) = \lambda_1 + (u^{\lambda_3} - (1-u)^{\lambda_4})/\lambda_2$, for $0 < u < 1$. The parameters defining the distributions used in the study and the associated mean (μ), variance (σ^2), skewness (α_3) and kurtosis (α_4) values, are given in Table 2. Some of these distributions are used in Fan (1994) to examine the performance of the Bickel-Rosenblatt test with a bandwidth converging to zero as n tends to infinity.

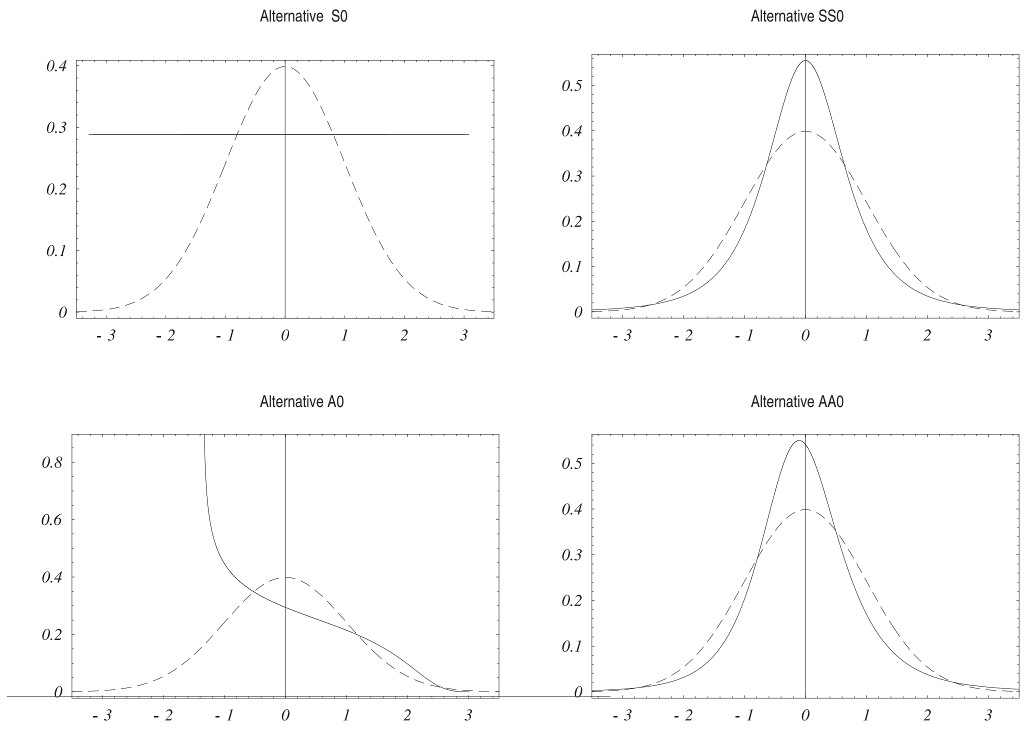

Figure 3: *Distribution shapes considered in the simulation study: Alternative density – solid line; Standard Normal density – broken line*

Table 2: *Distributions used in the simulation study*

Case	μ	σ^2	α_3	α_4	λ_1	λ_2	λ_3	λ_4
Normal distributions								
N1	0.4	1	”	”	—	—	—	—
N2	0	0.36	”	”	—	—	—	—
N3	0.4	0.36	”	”	—	—	—	—
Nonnormal distributions								
Symmetric distributions								
S0	0	1	0	1.8000	0	0.577350	1	1
S1	0.4	1	”	”	0.4	”	”	”
S2	0	0.36	”	”	0	0.962250	”	”
S3	0.4	0.36	”	”	0.4	”	”	”
SS0	0	1	0	11.6136	0	−0.397012	−0.16	−0.16
SS1	0.4	1	”	”	0.4	”	”	”
SS2	0	0.36	”	”	0	−0.663187	”	”
SS3	0.4	0.36	”	”	0.4	”	”	”
Asymmetric distributions								
A0	0	1	0.5129	2.2212	0.835034	0.459063	1.4	0.25
A1	0.4	1	”	”	1.235034	”	”	”
A2	0	0.36	”	”	0.501020	0.765105	”	”
A3	0.4	0.36	”	”	0.901020	”	”	”
AA0	0	1	0.7588	11.4308	−0.116734	−0.351663	−0.13	−0.16
AA1	0.4	1	”	”	−0.283266	”	”	”
AA2	0	0.36	”	”	−0.070040	−0.586106	”	”
AA3	0.4	0.36	”	”	0.329960	”	”	”

In Table 3 we present the Monte-Carlo empirical power results for the previous tests drawn from the considered alternatives. These results are based on 10^4 Monte-Carlo samples of different sizes for a significance level of 0.05. For the evaluation of the critical values of the I^2 tests we have used 10^4 replications. In applying the tests A^2 and W^2 we have followed Stephens (1986).

From Table 3, and Figures 1 and 2, we conclude that the theoretical results based on Bahadur local efficiency are in good accordance with empirical ones. The theoretical properties of I^2 tests are well transferred to finite sample situations.

In practice, the choice among the considered tests depends on the available information about the alternative to the null hypothesis. For alternatives f whose mean and variance satisfy $\mu_f \neq 0$ and $\sigma_f^2 = 1$ (Type I alternatives), A^2 is in general the best test, and each one of the tests $I^2(0.8)$, $I^2(2.0)$ or $I^2(c)$ is better than $I^2(0.3)$ test. For alternatives f satisfying $\mu_f = 0$ or $\sigma_f^2 \neq 1$ (Type II alternatives), $I^2(0.3)$ is globally the best test. Moreover, for these alternatives each one of the tests $I^2(0.3)$, $I^2(0.8)$ or $I^2(c)$ is better or significantly better than A^2 or W^2 tests.

Table 3: *Empirical power at level* 0.05 *for different values of n*

Case	n	$I^2(0.3)$	$I^2(0.8)$	$I^2(2.0)$	$I^2(c)$	A^2	W^2
N1	20	.210	.347	.423	.376	.404	.384
	50	.499	.722	.807	.759	.785	.761
N2	20	.432	.238	.005	.171	.080	.080
	50	.928	.930	.026	.811	.719	.536
N3	10	.362	.357	.183	.319	.245	.335
	20	.745	.805	.551	.750	.702	.743
S0	50	.452	.152	.060	.208	.148	.119
	100	.780	.277	.068	.450	.292	.211
	200	.982	.539	.067	.826	.666	.457
S1	20	.331	.359	.421	.413	.434	.363
	50	.712	.712	.784	.804	.816	.720
S2	20	.272	.138	.007	.087	.048	.052
	50	.986	.917	.030	.760	.636	.244
S3	10	.259	.299	.183	.260	.221	.276
	20	.671	.742	.531	.669	.633	.651
SS0	50	.316	.130	.032	.128	.094	.071
	100	.621	.326	.035	.335	.234	.170
	200	.915	.704	.041	.715	.613	.476
SS1	20	.391	.458	.455	.484	.494	.500
	50	.820	.885	.856	.896	.894	.898
SS2	20	.789	.463	.003	.461	.252	.246
	50	.998	.987	.027	.984	.955	.925
SS3	10	.596	.480	.167	.477	.339	.460
	20	.934	.903	.581	.899	.844	.881
A0	50	.581	.225	.069	.289	.207	.176
	100	.895	.442	.069	.613	.445	.323
	200	.998	.780	.082	.957	.905	.686
A1	20	.217	.223	.368	.315	.358	.251
	50	.588	.516	.740	.763	.790	.592
A2	20	.427	.187	.006	.156	.085	.086
	50	.995	.927	.030	.904	.790	.445
A3	10	.266	.241	.162	.210	.171	.229
	20	.870	.824	.519	.861	.739	.772
AA0	50	.320	.149	.038	.140	.107	.087
	100	.620	.350	.044	.346	.251	.184
	200	.909	.711	.054	.705	.619	.477
AA1	20	.315	.365	.394	.401	.425	.419
	50	.739	.828	.822	.857	.864	.857
AA2	20	.781	.455	.003	.449	.248	.241
	50	.998	.986	.028	.983	.954	.924
AA3	10	.575	.433	.150	.440	.301	.422
	20	.930	.903	.565	.899	.847	.881

If one does not know much about the unknown density function, the undertaken simulation study suggests that the test $I^2(c)$ is a good alternative to both A^2 and W^2 tests. In fact, for Type I alternatives the $I^2(c)$ performance is close to that one of A^2 or W^2, and for Type II alternatives $I^2(c)$ is better or significantly better than A^2 or W^2 tests.

The practical performance shown by the Bickel-Rosenblatt test with fixed bandwidth impels the generalization of the results presented in this paper to the test of a composite null hypothesis. In case of location-scale null families of density functions, this demands the use of a kernel density estimator with data-dependent fixed bandwidth matrix which is out of the scope of this paper. In a future paper we intend to address this subject.

5 Proofs of Theorem 2 and Corollary 2

Proof of Theorem 2: In order to use Theorem 1.2.2 of Nikitin (1995) due to Bahadur (1967, 1971), we first note that from the strong law of large number for U-statistics (cf. Theorem 3.1.1 of Koroljuk and Borovskich (1989)) we have

$$n^{-1} I_n^2(h) \underset{n\to+\infty}{\overset{as}{\longrightarrow}} b_{I^2(h)}(f),$$

for all f, where $b_{I^2(h)}(\cdot)$ is given in Theorem 2. Secondly, it is necessary to solve the problem of determining large deviation asymptotics of the sequence $(I_n^2(h))$ under the null hypothesis. This problem can be solved by using the integral and quadratic form of $I_n^2(h)$ and a generalization of Chernoff large-deviation result due to Sethuraman (1964) (see also Nikitin (1995), p. 23). In fact, we have

$$I_n^2(h) = \left(n^{-1}\|Z_{1,h} + \cdots + Z_{n,h}\|_2\right)^2,$$

where $(Z_{i,h})$ are i.i.d. random variables taking values on $L_2(\lambda)$ given by $Z_{i,h}(x) = K_h(x - X_i) - K_h \star f_0(x)$, for $x \in \mathbb{R}^d$. Moreover, for all $g \in L_2(\lambda)$ we have $\int\int g(x)Z_{1,h}(x)dxdP = \int g(x)\int Z_{1,h}(x)dPdx = 0$, and, for all $z \in \mathbb{R}$, $\int \exp(z\|Z_{1,h}\|_2)dP \le \exp(z\int \|Z_{1,h}\|_2 dP) < +\infty$, since $\|Z_{1,h}\|_2$ is a bounded random variable. The conditions of Sethuraman's theorem are thus fulfilled. Then, for all $a > 0$,

$$\lim_{n\to+\infty} n^{-1} \ln P(n^{-1} I_n^2(h) \ge a) = G(a),$$

where G is a continuous function in a neighbourhood V_0 of zero such that

$$G(a) = -\frac{a}{2\sigma_h^2}(1 + o(1)), \quad \text{as } a \to 0,$$

and

$$\begin{aligned} \sigma_h^2 &= \sup\left\{\int\left(\int g(x)Z_{1,h}(x)dx\right)^2 dP : \|g\|_2 = 1\right\} \\ &= \sup\left\{\int\int g(x)g(y)\int Z_{1,h}(x)Z_{1,h}(y)dPdxdy : \|g\|_2 = 1\right\} \\ &= \sup\left\{\int\int g(x)g(y)\bar{Q}_h(x,y)dxdy : \|g\|_2 = 1\right\}, \end{aligned}$$

with $\bar{Q}_h(x,y) = \int k(x,u;h)k(y,u;h)dP_0(u)$ and k is given by (3).

By the Rayleigh equation (see Dunford and Schwartz (1963)), σ_h^2 is the largest eigenvalue of the integral operator $\bar{A}_h$, with kernel $\bar{Q}_h$, defined on $L_2(\lambda)$. Since the set of eigenvalues of $\bar{A}_h$ coincide with the corresponding one of the operator A_h defined by (4), we get

$$G(a) = -\frac{a}{2\lambda_{1,h}}(1 + o(1)), \quad \text{as } a \to 0,$$

where $\lambda_{1,h}$ is the largest eigenvalue of the operator A_h.

Finally, from the continuity in f_0 of the function $b_{I^2(h)}(\cdot)$ from $L_1(\lambda)$ to $[0, +\infty[$, there exists a neigbourhood V_{f_0} of f_0 such that $\{b_{I^2(h)}(f) : f \in V_{f_0}\} \subset V_0$, and therefore, from Theorem 1.2.2 of Nikitin (1995), we conclude that

$$C_{I^2(h)}(f) = -2\, G(b_{I^2(h)}(f)),$$

for all $f \in V_{f_0}$. □

Proof of Corollary 2: For $b^o_{I^2(h)}(f(\cdot;\theta))$ given in Corollary 1, we have

$$\begin{aligned} b^o_{I^2(h)}(f(\cdot;\theta)) &= \int\!\!\int Q_h(u,v)\frac{\partial \ln f}{\partial \theta}(u;\theta_0)\frac{\partial \ln f}{\partial \theta}(v;\theta_0)\, dP_0(u)dP_0(v) \\ &= \Big\langle A_h \frac{\partial \ln f}{\partial \theta}(\cdot;\theta_0)\, ,\, \frac{\partial \ln f}{\partial \theta}(\cdot;\theta_0)\Big\rangle. \end{aligned}$$

The result follows now from Corollary 1 and the representation $A_h q = \sum_{k=1}^{\infty} \lambda_{k,h} \langle\, q, q_{k,h}\, \rangle\, q_{k,h}$, for all $q \in L_2(P_0)$. □

Acknowledgements

This work was partially supported by CMUC/FCT. The author would like to thank Paulo Oliveira for his helpful comments.

References

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A. and Sorensen, D. (1999). *LAPACK users' guide*, Third edition, SIAM, Philadelphia.

Anderson, N.H., Hall, P., Titterington, D.M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates, *J. Multivariate Anal.*, 50, 41-54.

Bahadur, R.R. (1967). Rates of convergence of estimates and test statistics, *Ann. Math. Statist.*, 38, 303-324.

Bahadur, R.R. (1971). *Some Limit Theorems in Statistics*, SIAM, Philadelphia.

Bickel, P.J., Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *Ann. Statist.*, 1, 1071-1095.

Bowman, A.W. (1992). Density based tests for goodness-of-fit normality, *J. Statist. Comput. Simul.*, 40, 1-13.

Bowman, A.W., Foster, P.J. (1993). Adaptive smoothing and density-based tests of multivariate normality, *J. Amer. Statist. Assoc.*, 88, 529-537.

Dunford, N., Schwartz, J.T. (1963). *Linear Operators, Part II*, Interscience Publishers, New York.

Durbin, J., Knott, M., Taylor, C.C. (1975). Components of Cramér-von Mises Statistics II, *J. Roy. Statist. Soc. Ser. B*, 37, 216-237.

Durio, A., Nikitin, Ya.Yu. (2003). Local Bahadur efficiency of some goodness-of-fit tests under skew alternatives, *J. Statist. Plann. Inference*, 115, 171-179.

Epps, T.W., Pulley, L.B. (1983). A test for normality based on the empirical characteristic function, *Biometrika*, 70, 723-726.

Eubank, R.L., LaRiccia, V.N. (1992). Asymptotic comparison of Crámer-von Mises and nonparametric function estimation techniques for testing goodness-of-fit, *Ann. Statist.*, 20, 2071-2086.

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

Fan, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters, *Econometric Theory*, 14, 604-621.

Gouriéroux, C., Tenreiro, C. (2001). Local power properties of kernel based goodness of fit tests, *J. Multivariate Anal.*, 78, 161-190.

Gregory, G.G. (1977). Large sample theory for U-statistics and tests of fit, *Ann. Statist.*, 5, 110-123.

Gregory, G.G. (1980). On efficiency and optimality of quadratic tests, *Ann. Statist.*, 8, 116-131.

Hall, P. (1997). *The Bootstrap and Edgeworth Expansion*, Springer, New York.

Henze, N., Zirkler, B. (1990). A class of invariante consistent tests for multivariate normality, *Commun. Stat. Theory Methods*, 19, 3595-3617.

Henze, N., Wagner, T. (1997). A new approach to the BHEP tests for multivariate normality, *J. Multivariate Anal.*, 62, 1-23.

Koroljuk, V.S., Borovskich, Yu.V. (1989). *Theory of U-Statistics*, Kluwer, Dordrecht.

Nikitin, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests*, Cambridge University Press.

Parzen, E. (1962). On estimation of a probability density function and mode, *Ann. Math. Statist.*, 33, 1065-1076.

Ramberg, J.S., Schmeiser, B.W. (1974). An approximate method for generating asymmetric random variables, *Commun. ACM*, 17, 78-82.

Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function, *Ann. Math. Statist.*, 27, 832-837.

Sethuraman, J. (1964). On the probability of large deviations of families of sample means, *Ann. Math. Statist.*, 35, 1304-1316.

Stephens, M.A. (1986). Tests based on EDF statistics, In: *Goodness-of-Fit Techniques*, ed. D'Agostino, R.B. and Stephens, M.A., Marcel Dekker, New York, 97-193.