The Annals of Statistics
1991, Vol. 19, No. 2, 999–1009

THE POWER AND OPTIMAL KERNEL OF THE BICKEL–ROSENBLATT TEST FOR GOODNESS OF FIT

BY B. K. GHOSH AND WEI-MIN HUANG

Lehigh University

Bickel and Rosenblatt proposed a procedure for testing the goodness of fit of a specified density to observed data. The test statistic is based on the distance between the kernel density estimate and the hypothesized density, and it depends on a kernel K, a bandwidth b_n and an arbitrary weight function a. We study the behavior of the asymptotic power of the test and show that a uniform kernel maximizes the power when $a > 0$.

1. Introduction. Let $X_1, \dots, X_n$ be independent and identically distributed random variables with a continuous probability density function f. Rosenblatt's (1956) pioneering work and Parzen's (1962) extension of it introduced the so-called kernel density estimate $f_n(x)$ for estimating $f(x)$ at a fixed point $x \in \mathbb{R}$ using the data $(X_1, \dots, X_n)$,

$$f_n(x) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left[\frac{x - X_i}{b_n}\right] = \frac{1}{b_n} \int K\left[\frac{x-t}{b_n}\right] dF_n(t). \tag{1.1}$$

In (1.1), F_n is the sample distribution function, K is a suitably selected kernel function on $\mathbb{R}$ such that $\int K(x)\,dx = 1$ and $b_n > 0$ is a predetermined bandwidth such that $b_n \to 0$ and $nb_n \to \infty$ as $n \to \infty$. All integrals in this paper, shown without limits, are understood to have limits $-\infty$ to $+\infty$.

In this paper, we consider the problem of testing the hypothesis

$$H\colon f = f_0 \tag{1.2}$$

at a specified significance level α, where f_0 is completely specified. Bickel and Rosenblatt (1973) proposed the test statistic

$$T_n = nb_n \int [f_n(x) - E_0 f_n(x)]^2 a(x)\,dx, \tag{1.3}$$

in which a is a suitably chosen function on $\mathbb{R}$ and $E_0 f_n(x)$ denotes the expectation of $f_n(x)$ under f_0. They show (see their Theorem 4.1) that, under H, T_n is asymptotically normal with mean $\mu(K, a)$ and variance $b_n \sigma^2(K, a)$,

Received September 1989; revised February 1990.
AMS 1980 *subject classifications*. Primary 62G10, 62G20.
Key words and phrases. Tests for goodness of fit, density estimates in tests, optimal kernel for tests, smoothed chi-square tests, asymptotic power.

where

$$\mu(K,a) = I(K)\int f_0(x)a(x)\,dx,$$
(1.4)
$$\sigma^2(K,a) = 2J(K)\int f_0^2(x)a^2(x)\,dx,$$

$$(1.5)\quad I(K) = \int K^2(x)\,dx, \qquad J(K) = \int\left[\int K(x+y)K(x)\,dx\right]^2 dy.$$

Consequently, an asymptotically α-level test of H is provided by

$$(1.6)\qquad \text{Reject } H \text{ if } T_n \geq \mu(K,a) + z_\alpha b_n^{1/2}\sigma(K,a),$$

where z_α is defined by

$$(1.7)\qquad \Phi(z_\alpha) = 1 - a, \qquad \Phi(z) = (2\pi)^{-1/2}\int_{-\infty}^{z}\exp(-x^2/2)\,dx.$$

We shall refer to (1.6) as the *BR-test*. They proposed a second statistic (their $\tilde{T}_n$) which requires additional assumptions. The main conclusion of this paper applies to the power of the second test as well.

The purpose of the present paper is to study the asymptotic power of the BR-test with a view toward an optimum selection of K. We will assume throughout that $b_n = n^{-\delta}$ for some $\delta \in (0,1)$. In Section 2 (Theorems 2.1 and 2.2), we study the asymptotic power of the BR-test against fixed and local alternatives. It is shown that, for any fixed $(a > 0, \delta)$, the power of the BR-test is maximized with respect to K by minimizing $J(K)$ of (1.5). In Section 3, we prove the most interesting result of this paper, which can be stated as the following theorem.

THEOREM 1.1. *Let $\mathscr{K}$ be the set of all $L^1(-\infty,\infty)$ functions K satisfying*

$$K \geq 0, \qquad \int K(x)\,dx = 1, \qquad \int xK(x)\,dx = \xi,$$
(1.8)
$$\int (x-\xi)^2 K(x)\,dx = \tau^2,$$

where $\xi \in (-\infty,\infty)$ and $\tau > 0$ are given numbers. Then the functional $J(K)$ in (1.5) is minimized on $\mathscr{K}$ uniquely by

$$(1.9)\qquad K^*(x) = \begin{cases} (2\sqrt{3}\,\tau)^{-1}, & \text{for } |x-\xi| \leq \tau\sqrt{3}, \\ 0, & \text{for } |x-\xi| > \tau\sqrt{3}, \end{cases}$$

and $\min_K J(K) = J(K^*) = (3\sqrt{3}\,\tau)^{-1}$.

It can be verified that a minimum of J(K) under the first two constraints in (1.8) does not exist. The last two constraints in (1.8) with $\xi = 0$ and $\tau = 1$ are standard practice in the theory of density estimation [see Silverman (1986), Chapter 3].

The conclusion of Theorem 1.1 is in sharp contrast to the Epanechnikov (1969) kernel (see the second kernel in Table 1) which minimizes $\int E_0[f_n(x) - f_0(x)]^2\,dx$ asymptotically and is widely used in the estimation theory of $f_0(x)$. In Section 4, we demonstrate that, from the viewpoint of hypothesis testing, the Epanechnikov kernel can be worse than many other kernels. Apart from its optimality in the present context, K^* in (1.9) renders T_n more easily computable. As a final remark, Bickel and Rosenblatt compared the Pitman efficiency of the BR-test against Pearson's chi-square test when f_0 is a uniform density on $[0,1]$ [any f_0 in (1.2) can, of course, be effectively reduced to this case by the probability integral transformation of the data]. They show that the BR-test based on K^* with $\xi = 0$ and $\tau = (2\sqrt{3})^{-1}$ is strictly better than the chi-square test. Theorem 1.1 gives a better understanding of their choice of the kernel. They chose K^* not because of any optimality consideration in $\mathscr{K}$ (see also their Remark 1 on page 1076) but seemingly because the structure of the chi-square test prompted this. We show in Section 4 that the BR-test based on a nonuniform K can, indeed, be strictly worse than the chi-square test in the sense of Pitman.

Theorem 1.1 has several other applications. First, Beran (1977) proposed a goodness-of-fit test based on Hellinger distances. The power of his test is also a decreasing function of our $J(K)$ and, consequently, K^* should be a good choice for his test as well. Second, if one writes $J(K) = \int G(x)\,dx$, where $G(x)$ is the symmetrization of $K(x)$, one can interpret $J(K)$ as the concentration measure of G [see Grenander (1963)] or the φ-deviation of G from the Lebesgue measure [see Hengartner and Theodorescu (1973)]. Theorem 1.1 then gives a lower bound to these quantities. Finally, Theorem 1.1 supplements known extremal properties of uniform and triangular distributions [see, for instance, Mori and Szekely (1985)].

2. The asymptotic power of the BR-test. The primary objective of this section is to show that the asymptotic power of the BR-test against reasonable alternatives is a decreasing function of $J(K)$. This feature need not, of course, be true in small samples. A second goal is to summarize certain other aspects of the power which follow from the results of Bickel and Rosenblatt (1973) and Rosenblatt (1975).

Bickel and Rosenblatt give regularity conditions under which T_n is asymptotically normal when H holds and later Rosenblatt (1975) relaxed some of them. We synthesize their conditions below because they will be referred to in our subsequent discussions:

(a) *f_0 is bounded, either positive on $\mathbb{R}$ or positive only on some $[c_0, d_0]$, and continuous with a bounded continuous derivative in the interior of its domain of positivity*;
(b) *a is piecewise continuous, bounded and integrable on $\mathbb{R}$*;
(c) $\int K(x)\,dx = 1$, $\int x^2 K(x)\,dx < \infty$, $\int K^2(x)\,dx < \infty$;

and either

(d) $b_n = n^{-\delta}$ *for some* $\delta \in (0, 1/4)$ *and* K *is continuous on* $\mathbb{R}$ *satisfying*

$$\int |K'(x)|^2\,dx < \infty, \qquad \int_{|x|\geq 3} |x|^{3/2}|K'(x)|(\log\log|x|)^{1/2}\,dx < \infty,$$

or

(d′) $b_n = n^{-\delta}$ *for some* $\delta \in (0,1)$ *and* K *is bounded on some* $[C_K, D_K]$ *and* 0 *outside*.

Theorem 4.1 in Bickel and Rosenblatt (1973) then proves that $Z_n(f_0)$ is asymptotically $N(0,1)$ under assumptions (a)–(d), while Corollary 1 in Rosenblatt (1975) does the same under assumptions (a)–(c) and (d′), where

$$Z_n(f_0) = n^{\delta/2}\sigma_0^{-1}\left\{n^{1-\delta}\int [f_n(x) - E_0 f_n(x)]^2 a(x)\,dx - \mu_0\right\}$$

and μ_0 and σ_0 denote $\mu(K,a)$ and $\sigma(K,a)$, respectively.

Let g be an arbitrary alternative to H satisfying assumption (a). Denote the expressions in (1.4) computed under g by μ_g and σ_g^2. It follows from (1.3) and (1.6) that the power of the BR-test against g is

$$\Pi_n(g) = P_g\left(T_n \geq \mu_0 + z_\alpha b_n^{1/2}\sigma_0\right). \tag{2.1}$$

It is not difficult to show that (2.1) reduces to

$$\begin{aligned}\Pi_n(g) = P_g\Bigg(Z_n(g) \geq &- n^{1-\delta/2}\sigma_g^{-1}\\ &\times\left\{\int [g(x) - f_0(x)]^2 a(x)\,dx + o_p(1)\right\}\Bigg).\end{aligned} \tag{2.2}$$

Since $Z_n(g)$ is asymptotically $N(0,1)$ under g, it follows that the power of the BR-test against an arbitrary g tends to 1. One would, of course, expect this from any reasonable test of H. What is to be noted from (2.2), however, is that the asymptotic power is a decreasing function of $J(K)$ for any fixed choice of $(f_0, g, a > 0, \delta)$.

The case of arbitrary g does not shed enough light on the performance of the test since the asymptotic power degenerates to 1. Bickel and Rosenblatt (1973), Theorem 4.2, considered the Pitman alternatives

$$g_{1n}(x) = f_0(x) + n^{-\beta}w(x) + o(n^{-\beta}), \qquad \beta > 0, \tag{2.3}$$

and obtained the following result in a different form.

Theorem 2.1. *Under assumptions* (a)–(d), *the power of the BR-test against* (2.3) *satisfies, for each* $\delta \in (0, 1/4)$,

$$\lim_{n\to\infty} \Pi_n(g_{1n}) = \begin{cases}\alpha, & \text{if } \beta > (2-\delta)/4,\\ \Phi(l), & \text{if } \beta = (2-\delta)/4,\\ 1, & \text{if } 0 < \beta < (2-\delta)/4,\end{cases} \tag{2.4}$$

where

$$l = \sigma_0^{-1} \int w^2(x) a(x)\, dx - z_\alpha.$$

Moreover, under assumptions (a)–(c) *and* (d′), (2.4) *holds for each* $\delta \in (0, 2/3)$.

Three observations can be made from Theorem 2.1 when $a > 0$. First, the asymptotic power of the BR-test against Pitman alternatives is a decreasing function of $J(K)$. Second, as δ gets smaller (from 1/4 or 2/3 to 0), the power improves. This feature, incidentally, conflicts with the fact that the asymptotic normality of T_n under H improves as δ gets larger (from 0 to 1/4 or 1). Finally, the BR-test is less powerful than standard nonparametric tests of H based on F_n (e.g., Kolmogorov and Smirnov) or parametric tests based on the actual form of f_0 (e.g., Wilk and Shapiro). The reason is that the latter can detect (with power greater than α) alternatives at a distance of order $n^{-1/2}$ from H.

In order to emphasize that the last drawback of the BR-test is peculiar to Pitman alternatives, Rosenblatt (1975), Section 3, examined the power against a special case of the following local alternatives:

$$g_{2n}(x) = f_0(x) + n^{-\varepsilon} \sum_{j=1}^{k} w_j(\{x - c_j\} n^\gamma) + o(n^{-\varepsilon-\gamma}), \tag{2.5}$$

$$\varepsilon > 0,\ 0 < \gamma < \delta,$$

which implies that the alternative density has $k \geq 1$ sharp peaks at points c_j. One can generalize his proof for the special case to the following theorem.

THEOREM 2.2. *Under assumptions* (a)–(d), *the power of the BR-test against* (2.5) *satisfies, for each* $\delta \in (0, 1/4)$,

$$\lim_{n\to\infty} \Pi_n(g_{2n}) = \begin{cases} \alpha, & \text{if } 1 - 2\varepsilon - \gamma\delta < \delta/2 < \varepsilon + \gamma, \\ \Phi(l'), & \text{if } \delta/2 = 1 - 2\varepsilon - \gamma < (1+\gamma)/3, \\ 1, & \text{if } \delta/2 < \min[(1-\varepsilon)/2, 1 - 2\varepsilon - \gamma], \end{cases} \tag{2.6}$$

where

$$l' = \sigma_0^{-1} \sum_{j=1}^{k} \left\{ a(c_j) \int w_j^2(x)\, dx \right\} - z_\alpha.$$

Moreover, under assumptions (a)–(c) *and* (d′), (2.6) *holds for each* $\delta \in (0, 1)$.

Here, again, note that the asymptotic power of the BR-test with any $a > 0$ is decreasing in $J(K)$ and δ. Moreover, with a suitable choice of δ, the BR-test can be made more powerful than the standard tests for alternatives (2.6). To see this, consider the triangle in the (δ, γ)-plane with vertices at (0, 0), (2/3, 2/3) and (1, 1/2) when $2\varepsilon = 1 - \gamma - \delta/2$. Expression (2.6) shows that $\Pi_n(g_{2n}) \to \Phi(l') > \alpha$ for every point inside the triangle and, since $\varepsilon + \gamma > 1/2$

in this triangle, the standard tests will have power approaching α. Consequently, the BR-test will be more powerful inside the triangle if one chooses δ suitably. For instance, Rosenblatt's (1975) example of $\delta = 1/2$, $\varepsilon = 1/6$ and $\gamma = 5/12$ under assumption (d') belongs to this category, while $\delta = 17/72$, $\varepsilon = 95/288$ and $\gamma = 2/9$ is an example under assumption (d). What is, perhaps, more striking is that the BR-test can have power 1 for (2.6) when standard tests have no power at all (e.g., $\delta = 17/72$, $\varepsilon = 1/4$, $\gamma = 1/8$). On the other hand, given that a standard test achieves power 1 (i.e., $\beta < 1/2$) against Pitman alternatives (2.3), one can always construct a BR-test with $0 < \delta < 2 - 4\beta$ whose power is greater than α, as Theorem 2.1 shows.

3. Proof of Theorem 1.1. Observe first the $J(K)$ can also be expressed as

$$J(K) = \int \left[\int K(x) K(y - x)\, dx \right]^2 dy. \tag{3.1}$$

This follows by writing $J(K)$ in (1.5) as the triple integral of $K(x)K(y)$ $K(x + z)K(y + z)$ and making the transformations $x = w - u$, $y = v$ and $z = u - v$. We need the following lemma.

LEMMA 3.1. *let $\mathscr{H}_m$ be the set of all $L^1(-\infty, \infty)$-functions h such that the m-fold convolution $h_{(m)} = h * \cdots * h$ is an $L^2(-\infty, \infty)$-function. Then, for each $m \geq 1$, the functional*

$$J_m(h) = \int h^2_{(m)}(y)\, dy \tag{3.2}$$

is strictly convex on $\mathscr{H}_m$.

PROOF. Observe first that $h \in L^1(-\infty, \infty)$ implies that $h_{(m)}$ exists a.e. (in the Lebesgue sense) and $h_{(m)} \in L^1(-\infty, \infty)$ [see Kawata (1972), page 75]. Let

$$\varphi(t) = (2\pi)^{-1/2} \int e^{-ixt} h(x)\, dx, \qquad t \in \mathbb{R}, \tag{3.3}$$

be the Fourier transform of h. Then the Fourier transform of $h_{(m)}$ is $(2\pi)^{(m-1)/2} \varphi^m(t)$. Since $h_{(m)} \in L^2(-\infty, \infty)$, it follows from (3.2) and Parseval's relation [see Kawata (1972), page 205] that

$$\begin{aligned} J_m(h) &= \int \left| (2\pi)^{(m-1)/2} \varphi^m(t) \right|^2 dt \\ &= (2\pi)^{m-1} \int |\varphi(t)|^{2m}\, dt. \end{aligned} \tag{3.4}$$

Let $\varphi_1(t)$ and $\varphi_2(t)$ denote the Fourier transforms of arbitrary h_1 and h_2 in

$\mathscr{H}_m$. Then, for each fixed $t \in \mathbb{R}$ and $\beta \in (0,1)$, we have

$$\begin{aligned} &\beta|\varphi_1(t)|^{2m} + (1-\beta)|\varphi_2(t)|^{2m} \\ (3.5)\qquad &\geq \{\beta|\varphi_1(t)| + (1-\beta)|\varphi_2(t)|\}^{2m} \quad \text{by Jensen's inequality} \\ &\geq |\beta\varphi_1(t) + (1-\beta)\varphi_2(t)|^{2m} \quad \text{by the triangular inequality.} \end{aligned}$$

Moreover, Jensen's equality holds iff $|\varphi_1(t)| = |\varphi_2(t)|$ and the triangular equality holds iff $\varphi_1(t) = k\varphi_2(t)$ for some $k > 0$. These two conditions jointly imply that the left-hand side of (3.5) equals the right-hand side iff $\varphi_1(t) = \varphi_2(t)$ a.e. Since $h_1 \neq h_2 \Rightarrow \varphi_1(t) \neq \varphi_2(t)$ on a set of positive measure, we finally conclude from (3.4) and (3.5) that

$$\begin{aligned} &\beta J_m(h_1) + (1-\beta)J_m(h_2) \\ &\quad = (2\pi)^{m-1}\int\left\{\beta|\varphi_1(t)|^{2m} + (1-\beta)|\varphi_2(t)|^{2m}\right\}dt \\ &\quad > (2\pi)^{m-1}\int|\beta\varphi_1(t) + (1-\beta)\varphi_2(t)|^{2m}\,dt = J_m(\beta h_1 + (1-\beta)h_2). \end{aligned}$$

Thus, $J_m(h)$ is strictly convex on $\mathscr{H}_m$ for every $m \geq 1$. □

Observe from (3.1) and (3.2) that $J(K)$ is the same as $J_2(K)$.

Proof of Theorem 1.1. Simple computations show that K^* of (1.9) indeed satisfies (1.8) and that (3.1) yields $J(K^*) = (3\sqrt{3}\tau)^{-1}$. If any K_1 in $\mathscr{K}$ is such that $J(K_1) = \infty$, then obviously $J(K^*) < J(K_1)$. Such an example arises if $\xi = 1/4$, $\tau = 1/2$ and $K_1(x) = x^{-3/4}e^{-x}/\Gamma(1/4)$ for $x \geq 0$, $K_1(x) = 0$ for $x < 0$. It suffices, therefore, to restrict attention to the subset $\mathscr{K}' \subset \mathscr{K}$ where $J(K) < \infty$. It follows from Lemma 3.1 that J is strictly convex on $\mathscr{K}'$. Now, using (3.3) and (3.4) for $m = 2$, we can express (3.1) as

$$(3.6)\qquad J(K) = \frac{1}{2\pi}\int\left[\left\{\int\cos(tx)K(x)\,dx\right\}^2 + \left\{\int\sin(tx)K(x)\,dx\right\}^2\right]^2 dt.$$

Let $R(x) = K_1(x) - K^*(x)$ for an arbitrary $K_1 \in \mathscr{K}'$, which implies $K^* + \varepsilon R \in \mathscr{K}'$ for each $\varepsilon \in (0,1)$. We will first show by variational calculus that the Gateaux differential $\partial J(K^* + \varepsilon R)/\partial\varepsilon$ at $\varepsilon = 0$ is nonnegative for every K_1 and the assertion of the theorem will then follow from the strict convexity of J [see Luenberger (1969), Chapters 7 and 8]. Let λ_1, λ_2 and λ_3 be real variables (Lagrange multipliers) and define

$$\begin{aligned} &M(K,\lambda_1,\lambda_2,\lambda_3) \\ (3.7)\qquad &\quad = J(K) + \lambda_1\left\{\int K(x)\,dx - 1\right\} + \lambda_2\left\{\int xK(x)\,dx - \xi\right\} \\ &\qquad + \lambda_3\left\{\int(x-\xi)^2K(x)\,dx - \tau^2\right\}. \end{aligned}$$

Then any local (i.e., on a subset of $\mathscr{K}'$) stationary point K^* for M must satisfy the four equations

$$\partial M/\partial\lambda_1 = 0, \qquad \partial M/\partial\lambda_2 = 0, \qquad \partial M/\partial\lambda_3 = 0, \tag{3.8}$$

$$\partial M(K^* + \varepsilon R, \lambda_1, \lambda_2, \lambda_3)/\partial\varepsilon|_{\varepsilon=0} = 0, \tag{3.9}$$

where M in (3.8) refers to (3.7) at $K = K^*$. It is clear from (3.7) that the equations in (3.8) yield the last three constraints in (1.8) and, as mentioned earlier, K^* of (1.9) indeed satisfies these constraints. It follows that the left-hand side of (3.9) under K^* of (1.9) is precisely $\partial J(K^* + \varepsilon R)/\partial\varepsilon$ at $\varepsilon = 0$ for any choice of $(\lambda_1, \lambda_2, \lambda_3)$. Substituting (1.9) and (3.6) in the left-hand side of (3.9), one gets, after some simplifications [see Gradshteyn and Ryzhik (1965), page 452, for the relevant integrals].

$$\begin{aligned} \partial J(K^* + \varepsilon R)/\partial\varepsilon|_{\varepsilon=0} &= \partial M(K^* + \varepsilon R, \lambda_1, \lambda_2, \lambda_3)/\partial\varepsilon|_{\varepsilon=0} \\ &= \int \psi(x, \lambda_1, \lambda_2, \lambda_3) R(x)\, dx, \end{aligned} \tag{3.10}$$

where

$$\psi(x, \lambda_1, \lambda_2, \lambda_3) =$$

$$\begin{cases} \left(\lambda_1 - \dfrac{\xi^2 - 9\tau^2}{6\sqrt{3}\,\tau^3}\right) + \left(\lambda_2 + \dfrac{\xi}{3\sqrt{3}\,\tau^3}\right)x + \left(\lambda_3 - \dfrac{1}{6\sqrt{3}\,\tau^3}\right)x^2, & \text{for } x \in A, \\ \dfrac{\left[3\sqrt{3}\,\tau - |x - \xi|\right]^2}{12\sqrt{3}\,\tau^3} + \lambda_1 + \lambda_2 x + \lambda_3 x^2, & \text{for } x \in B, \\ \lambda_1 + \lambda_2 x + \lambda_3 x^2, & \text{for } x \in C = \mathbb{R} - A - B, \end{cases} \tag{3.11}$$

and A and B refer to the sets $|x - \xi| \le \tau\sqrt{3}$ and $\tau\sqrt{3} < |x - \xi| < 3\sqrt{3}\,\tau$, respectively. Hence, if we choose $\lambda_1 = \lambda_1^*$, $\lambda_2 = \lambda_2^*$ and $\lambda_3 = \lambda_3^*$, where

$$\lambda_1^* = (\xi^2 - 9\tau^2)/(6\sqrt{3}\,\tau^3), \qquad \lambda_2^* = -\xi/(3\sqrt{3}\,\tau^3), \qquad \lambda_3^* = (6\sqrt{3}\,\tau^3)^{-1},$$

then (3.11) reduces to

$$\psi(x, \lambda_1^*, \lambda_2^*, \lambda_3^*) = \begin{cases} 0, & \text{for } x \in A, \\ \left(|x - \xi| - \tau\sqrt{3}\right)^2/(4\sqrt{3}\,\tau^3), & \text{for } x \in B, \\ \left\{(x - \xi)^2 - 9\tau^2\right\}/(6\sqrt{3}\,\tau^3), & \text{for } x \in C, \end{cases} \tag{3.12}$$

and consequently (3.10) yields

$$\begin{aligned} \partial J(K^* + \varepsilon R)/\partial\varepsilon|_{\varepsilon=0} &= (4\sqrt{3}\,\tau^3)^{-1} \int_B \left(|x - \xi| - \tau\sqrt{3}\right)^2 K_1(x)\, dx \\ &\quad + (6\sqrt{3}\,\tau^3)^{-1} \int_C \left\{(x - \xi)^2 - 9\tau^2\right\} K_1(x)\, dx \\ &\ge 0 \quad \text{for all } K_1 \in \mathscr{K}'. \end{aligned} \tag{3.13}$$

The last inequality and the strict convexity of J establish the theorem. Note that the first equation in (3.12) and the convexity of J show that K^* of (1.9) minimizes J locally in the subset $\mathscr{K}'' \subset \mathscr{K}'$, where K_1 has support A or interior to A, while (3.13) globalizes the conclusion to $\mathscr{K}'$. □

4. The relative efficiency of the BR-test. In order to assess the relative gain of K^* over a competing kernel K, denote the statistic (1.3) for the BR-test by $T_n(K)$. We are, of course, assuming that K satisfies (1.8) and both $T_n(K)$ and $T_n(K^*)$ use the same α and δ in (1.3). Then the asymptotic powers of the two tests against Pitman alternatives (2.3) are given by (2.4), in which $\sigma_0(K)$ and $\sigma_0(K^*)$ are defined by (1.4). It is easily shown [Noether (1955)] from the middle part of (2.4) that the Pitman efficiency of $T_n(K)$ relative to $T_n(K^*)$ is

$$R_\delta(K, K^*) = [J(K^*)/J(K)]^{1/(2-\delta)}. \tag{4.1}$$

For each $\delta \in (0,1)$, $R_\delta(K, K^*)$ is positive, ξ-invariant, τ-invariant and by Theorem 1.1 less than 1. Moreover, it is not difficult to show from (2.6) that (4.1) is also the asymptotic relative efficiency of $T_n(K)$ against $T_n(K^*)$ for alternatives (2.5) when $\varepsilon > 0$ and $0 < \gamma/2 < \delta/2 = 1 - 2\varepsilon - \gamma < (1+\gamma)/3$.

Table 1 shows values of $R_\delta(K, K^*)$ for some standard kernels [see Parzen (1962) and Epanechnikov (1969)] and special choices of δ. Values of $I(K)$ and $J(K)$ are also shown merely because they are needed to carry out the BR-test. The chosen values of δ reflect the extreme cases stated in Theorems 2.1 and 2.2. The blank spaces in Table 1 indicate that the corresponding kernels do not satisfy assumption (d′) and therefore the BR-test may not be valid.

The first K in Table 1 is the well-known Epanechnikov (1969) kernel which is asymptotically optimal from the standpoint of estimation of $f_0(x)$. As our table shows, from the standpoint of hypothesis testing, its efficiency is slightly less than K^*. It seems more important to realize that the Epanechnikov kernel can be less efficient than many others. Consider, for instance, the general family of kernels

$$K_c(x) = \begin{cases} (3/8)(c/5)^{1/2}[(3-c) + (3c-5)cx^2/5], \\ \qquad\qquad\qquad\qquad \text{for } |x| \le (5/c)^{1/2}, \\ 0, \qquad\qquad\qquad\quad\; \text{for } |x| > (5/c)^{1/2}, \end{cases} \tag{4.2}$$

where $c \in [1,3]$ is a constant. It is easily verified that for each c, K_c satisfies assumptions (c), (d) and (d′) as well as (1.8) with $\xi = 0$ and $\tau = 1$. Moreover, K_1 is the Epanechnikov kernel, $K_{5/3} = K^*$, and K_c is U-shaped for $c \in (5/3, 3]$. Some algebra with (3.1) and (4.2) yields

$$J(K_c) = \sqrt{c}\,(2715 - 2468c + 1548c^2 - 540c^3 + 81c^4)/3080\sqrt{5}, \tag{4.3}$$

which decreases from $J(K_1)$ to $J(K_{5/3})$, and then increases to $J(K_3)$. Consequently, (4.1) shows that K_1 is less efficient than K_c for every $c \in (1, c')$, where $c' > 5/3$ satisfies $J(K_{c'}) = J(K_1)$. In fact, $c' = 1.9655$ if one solves (4.3) equal to $167(387\sqrt{5})^{-1}$. It is interesting to note here that U-shaped kernels (e.g., K_c for any $5/3 < c < 1.9655$) are usually avoided in the estimation theory of $f_0(x)$.

TABLE 1
Asymptotic efficiency of some kernels relative to K^* for alternatives g_{1n} and g_{2n} ($\xi = 0$, $\tau = 1$)

Kernel	$I(K)$	$J(K)$	$R_\delta(K, K^*)$ $\delta = 0$	$\delta = 1/4$	$\delta = 2/3$	$\delta = 1$
$K^*(x) = (2\sqrt{3})^{-1}$ for $\lvert x\rvert \le \sqrt{3}^{\dagger}$	$(2\sqrt{3})^{-1}$	$(3\sqrt{3})^{-1}$	1	1	1	1
$K(x) = 3(4\sqrt{5})^{-1}(1 - x^2/5)$ for $\lvert x\rvert \le \sqrt{5}^{\dagger}$	$3(5\sqrt{5})^{-1}$	$167(385\sqrt{5})^{-1}$	0.9960	0.9955	0.9941	0.9921
$K(x) = 4^{-1}(\pi^2 - 8)^{1/2}\cos(x(\pi^2 - 8)^{1/2}/2)$ for $\lvert x\rvert \le \pi(\pi^2 - 8)^{-1/2\dagger}$	$(\pi/16)(\pi^2 - 8)^{1/2}$	$\pi(\pi^2 - 8)^{1/2}(2\pi^2 + 15)/768$	0.9952	0.9945	0.9928	0.9905
$K(x) = (\sqrt{6} - \lvert x\rvert)/6$ *for* $\lvert x\rvert \le \sqrt{6}^{\dagger}$	$\sqrt{2}(3\sqrt{3})^{-1}$	$151(315\sqrt{6})^{-1}$	0.9917	0.9905	0.9875	0.9834
$K(x) = (\sqrt{2\pi})^{-1}\exp(-x^2/2)$	$(2\sqrt{\pi})^{-1}$	$(2\sqrt{2\pi})^{-1}$	0.9822	0.9797	—	—
$K(x) = (\sqrt{2})^{-1}\exp(-\sqrt{2}\lvert x\rvert)$	$(2\sqrt{2})^{-1}$	$5(16\sqrt{2})^{-1}$	0.9332	0.9241	—	—
$K(x) = \exp(-x - 1)$ for $x \ge -1^{\dagger}$	2^{-1}	4^{-1}	0.8774	0.8611	—	—

$^{\dagger}K(x) = 0$ otherwise.

The family (4.2) also illustrates a final point. Bickel and Rosenblatt (1973) proved (their page 1083 has several printing errors near the end) that the Pitman efficiency of the chi-square test relative to the BR-test when the latter uses K^* is $(2/3)^{1/(2-\delta)}$ and, therefore, correctly concluded that the BR-test is more powerful for all $\delta \in (0, 2/3)$. On the other hand, it follows from this, (4.1) and $J(K^*) = (3\sqrt{3})^{-1}$ that a BR-test based on some other K will be *less* powerful than the chi-square test for all $\delta \in (0, 2/3)$ if that K satisfies $J(K) > (2\sqrt{3})^{-1}$. Although none of the kernels in Table 1 is of this type, every K_c in (4.2) with $c \in [2.9316, 3]$ satisfies $J(K_c) > (2\sqrt{3})^{-1}$. The point of the example is to emphasize the importance of Theorem 1.1 in assessing the merit of the BR-test.

Acknowledgment. The authors wish to thank a referee for several helpful comments.

REFERENCES

BERAN, R. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095. [Correction: **3** (1975) 1370.]

EPANECHNIKOV, V. A. (1969). Nonparametric estimates of a multivariate probability density. *Theory Probab. Appl.* **14** 153–158.

GRADSHTEYN, I. S. and RYZHIK, I. M. (1965). *Table of Integrals, Series, and Products*. Academic, New York.

GRENANDER, U. (1963). *Probabilities on Algebraic Structures*. Wiley, New York.

HENGARTNER, W. and THEODORESCU, R. (1973). *Concentration Functions*. Academic, New York.

KAWATA, T. (1972). *Fourier Analysis in Probability Theory*. Academic, New York.

LUENBERGER, D. G. (1969). *Optimization by Vector Space Methods*. Wiley, New York.

MORI, T. and SZEKELY, G. J. (1985). An extremal property of rectangular distributions. *Statist. Probab. Lett.* **3** 107–108.

NOETHER, G. E. (1955). On a theorem of Pitman. *Ann. Math. Statist.* **26** 64–68.

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

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

ROSENBLATT, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3** 1–14. [Correction: **10** (1982) 646.]

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.

DEPARTMENT OF MATHEMATICS
LEHIGH UNIVERSITY
BETHLEHEM, PENNSYLVANIA 18015