Läuter, Henning; Sachsenweger, Cornelia

Working Paper
Comparison of nonparametric goodness of fit tests

SFB 373 Discussion Paper, No. 1999,2

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin


This Version is available at:
http://hdl.handle.net/10419/61723

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Comparison of Nonparametric Goodness of Fit Tests

Henning Läuter, Cornelia Sachsenweger

University of Potsdam, Institute of Mathematics
PF 60 15 53, 14415 Potsdam

Abstract

We consider two tests for testing the hypothesis that a density lies in a parametric class of densities and compare them by means of simulation. Both considered tests are based on the integrated squared distance of the kernel density estimator from its hypothetical expectation. However, different kernels are used. The unknown parameter will be replaced by its maximum-likelihood-estimation (m.i.e.).

The power of both tests will be examined under local alternatives. Although both tests are asymptotically equivalent, it will be shown that there is a difference between the power of both tests when a finite number of random variables is used. Furthermore it will be shown that asymptotically equivalent approximations of the power can differ significantly when finite sample sizes are used.

Keywords: Goodness of fit, kernel estimator, local alternatives, simulation

AMS 1991 Subject Classification: Primary 62G10, 62G20.

The research for this paper was carried out within Sonderforschungsbereich 373 and was printed using funds made available by the Deutsche Forschungsgemeinschaft.
1 Introduction

Let $X_1, \ldots, X_n$ be a sample of i.i.d. random variables with density $f_n$. We consider the problem of testing the hypothesis

$$H : f_n \in \mathcal{F} = \{ f_n(\cdot, \theta) | \theta \in \Theta \subseteq \mathbb{R}^k \}$$

against the nonparametric alternative

$$K : f_n \notin \mathcal{F}$$

where $\mathcal{F}$ is a parametric class of density functions. Our test statistic is the $L_2$-norm of the distance of the kernel density estimator

$$\hat{f}_n(t) = (nh_n)^{-1} \sum_{i=1}^n K \left( \frac{t - X_i}{h_n} \right)$$

from its expectation under the hypothesis density $f_n(\cdot, \theta) = f_\theta$

$$E \hat{f}_n(t) = e_{h_n}(t; f_\theta) = \int K(x)f(t - xh_n, \theta) \, dx.$$

Under certain assumptions on the density of the random variables, the kernel function, the bandwidth and the weight function of the kernel density estimator follows with Liero/Läuter/Konakov (1998):

**Theorem 1.1** If for a parameter estimator $\hat{\theta}_n$ under the hypothesis $H$ with $X_i \sim f_\theta$ for each $\theta \in \Theta$

$$\sqrt{n} |\hat{\theta}_n - \theta| = O_P(1)$$

is true and if $h_n(\ln n)^\xi \rightarrow 0$ for a $\xi > \frac{k}{2}$, then follows for all $n \rightarrow \infty$ for all $x$

$$P_{f_\theta} \left( \frac{h_n^{-\frac{1}{2}}(Q_n(f_{\hat{\theta}_n}) - \mu(f_{\hat{\theta}_n}))}{\sigma(f_{\hat{\theta}_n})} \leq x \right) \rightarrow \Phi(x)$$
with

\[ Q_n(f_{\hat{\vartheta}_n}) = n h_n \int (\hat{f}_n(t) - \varepsilon_h(t, f_{\hat{\vartheta}_n}))^2 a(t) \, dt \quad \text{with} \]

\[ \varepsilon_h(t, f_{\hat{\vartheta}_n}) = \int K(x) f_{\hat{\vartheta}_n}(t - x h_n) \, dx, \]

\[ a(t) = \text{weight function} \quad \text{and} \]

\[ \hat{f}_n(t) = (n h_n)^{-1} \sum_{i=1}^{n} K \left( \frac{t - X_{ni}}{h_n} \right), \]

\[ \mu(f_{\hat{\vartheta}_n}) = \kappa \int f_{\hat{\vartheta}_n}(t) a(t) \, dt \quad \text{with} \]

\[ \kappa = \int K^2(x) \, dx, \]

\[ \sigma^2(f_{\hat{\vartheta}_n}) = 2 \kappa^* \int f_{\hat{\vartheta}_n}^2(t) a^2(t) \, dt \quad \text{with} \]

\[ \kappa^* = \int (K*K)^2(x) \, dx, \]

\[ \Phi(u_n) = 1 - \alpha \]

and \( \int := \int_{-\infty}^{\infty} \).

This way we get an asymptotic \( \alpha \)-test, if we reject the hypothesis if

\[ Q_n(f_{\hat{\vartheta}_n}) \geq \mu(f_{\hat{\vartheta}_n}) + u_n \frac{\sqrt{2}}{h_n} \sigma(f_{\hat{\vartheta}_n}). \]

The power of a test is the probability that the test rejects the hypothesis. That is

\[ P \left( Q_n(f_{\hat{\vartheta}_n}) \geq \mu(f_{\hat{\vartheta}_n}) + u_n \frac{\sqrt{2}}{h_n} \sigma(f_{\hat{\vartheta}_n}) \right). \]

We compare the power of the test with the Epanechnikov kernel to that with the Gauß kernel, always using \( a(t) \equiv 1 \). Therefore we use local alternatives of the form

\[ K_n : \quad f_n(\cdot, \vartheta_n) = f(\cdot, \vartheta_n) + N_n w(\cdot - c) b_n^{-1} \]

with \( \vartheta_n = \vartheta + n^{-\beta} \psi \) with an unknown vector \( \psi \), \( \beta > 0 \) and \( \vartheta \) fixed. Let \( N_n \) be a sequence of positive numbers converging to zero and \( c \) in the support of \( f \).

Let \( w \) be a function with limited support, for which \( \int w(x) \, dx = 0 \). For \( b_n \equiv b \) we get the Pitman alternative and for \( b_n \rightarrow 0 \) the sharp peak alternative.

In Figure 1 there are three of such local alternatives in comparison with the normal density \( f(x; 0.2, 1) \).
2 The Maximum Likelihood Estimation

Because the parameter $\theta$ is unknown, it has to be estimated. If the hypothesis holds then

$$\hat{\theta}_n - \theta = n^{-\frac{1}{4}} Z_n$$

is true for the m.l.e. $\hat{\theta}_n$ of the parameter. $Z_n$ is asymptotically normal distributed with expectation zero. It can be shown that $\hat{\theta}_n$ under the local alternatives behave similar as under the hypothesis under following weak assumptions. The open kernel of the parameter space $\Theta$ will be denoted by $\Theta^o$.

A1 For all $x \in \mathbb{R}$ there exist the second derivative $\frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x, \theta)$, $i, j = 1, \ldots, k$. They are continuous for all $\theta \in \Theta^o$ and it is

$$\int \nabla \phi f(x, \theta) \, dx = 0, \quad \int \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x, \theta) \, dx = 0.$$

(2)

A2 For each $\theta \in \Theta^o$ exists such a $\delta = \delta(\theta)$ and such a ball $U_\delta(\theta) \subset \Theta^o$ with the radius $\delta$ that for all $i, j = 1, \ldots, k$

$$\left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(x, \theta') \right| \leq m(x, \theta) \quad \forall \theta' \in U_\delta(\theta)$$

(3)

for a measurable function $m$ with $\int m(x, \theta) f(x, \theta) \, dx < \infty$.

A3 The determinant of the Fisher Information

$$I(\theta) = \int \nabla \phi \ln f(x, \theta)(\nabla \phi \ln f(x, \theta))^t f(x, \theta) \, dx$$

(4)

is positive for each $\theta \in \Theta^o$. 

Figure 1: Local alternatives $f_n$ with $N_n = 0.38, b_n = 1$, with $N_n = 0.038, b_n = 1$, with $N_n = 0.38, b_n = 0.1$ and hypothesis $f$. 

$\int_{n,0.38,1} \int_{n,0.038,1} \int_{n,0.38,0.1}$
We put
\[ \lambda_n(t) = -n^{-\beta} \psi^t \nabla \varphi f(t, \tilde{\vartheta}) - N_n w((t - c)b_n^{-1}) \]
with a suitable \( \tilde{\vartheta} \) between \( \hat{\vartheta}_n \) and \( \vartheta \).

Then it follows:

**Theorem 2.1** If the assumptions A1, A2 and A3 hold then

\[ \int \nabla \vartheta \ln f(t, \vartheta) \lambda_n(t) \, dt = O(\tau_n) \]

for the sequence \( \tau_n = n^{-\beta} + N_n b_n \) and we get under \( K_n \) for the m.l.e.

\[ \hat{\vartheta}_n - \vartheta = n^{-\frac{1}{2}} Z_n + \tau_n S + o_P(\tau_n). \]  

(5)

\( Z_n \) is asymptotically normal distributed with expectation zero and \( S \) a non-random vector.

The proof can be found in Liero/Läuter/Konakov (1998). In the family \( \mathcal{F} \) of normal distribution densities with the \( k(=2) \)-dimensional unknown parameter \( (\vartheta_{(1)}, \vartheta_{(2)}) \) with expectation \( \vartheta_{(1)} = \mu \) and variance \( \vartheta_{(2)} = \sigma^2 \), which we will consider in the following, the assumptions A1, A2 and A3 hold.

Following Theorem 2.1 for the sequence \( \tau_n = n^{-\beta} + N_n b_n \)

\[ \int \nabla \vartheta \ln f(t, \vartheta) \lambda_n(t) \, dt = \int \left( \frac{(x - g(\vartheta_{(1)}))g'(\vartheta_{(1)})}{(x - g(\vartheta_{(1)})^2 - \sigma^2) \sigma^2} \right) \lambda_n(t) \, dt = O(\tau_n). \]

Consequently, with Theorem 2.1 follows that the m.l.e. behave under the local alternatives similar as under the hypothesis.

It is obvious that the m.l.e. for \( g(\vartheta_{n(1)}) \) equals the arithmetic mean of the observations. For getting the values of the m.l.e. \( \hat{\vartheta}_{n(1)} \) out of \( g(\hat{\vartheta}_{n(1)}) \) we demand unique invertibility of the function \( g \). In the following examples the function \( g \) has this property.

The m.l.e. of \( \vartheta_2 = \sigma^2 \) is always

\[ \hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - g(\hat{\vartheta}_{n(1)}))^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - g(\hat{\vartheta}_{n(1)})^2. \]

Therefore we will only give the values of the m.l.e. for \( \vartheta_{(1)} \) in the following.

Now we want to determine the nonrandom vector \( S \) in Theorem 2.1 which is independent of \( n \). From the proof of the theorem we get

\[ nT_n^{-\frac{1}{2}} (X_n, \vartheta) \nabla \vartheta \ln f(X_n, \vartheta) = n^{\frac{1}{2}} \tau_n (S + o_P(1)). \]
THE MAXIMUM LIKELIHOOD ESTIMATES

Now for the vector \( \mathbf{z} \) and with the law of large numbers we see
\[
n^{-1}T_n(X_n, \theta) \xrightarrow{n \to \infty} -I(\theta)
\]
with the Fisher Information \( I(\theta) \). It follows
\[
-n^{-1} : I(\theta)^{-1} E_n \nabla \ln f(X_n, \theta) =: \hat{S}_n \xrightarrow{n \to \infty} S.
\]
We get
\[
I(\theta)^{-1} = \frac{2\sigma^6}{g'(\theta(1))^2} \begin{pmatrix}
\frac{1}{2\sigma^4} & 0 \\
0 & \frac{\sigma^2}{\sigma^2} & 0 \\
\end{pmatrix} \begin{pmatrix}
\frac{\sigma^2}{\sigma^2} & 0 \\
0 & 2\sigma^4 \\
\end{pmatrix}
\]
and with \( \psi' = (\psi_1, \psi_2) \)
\[
E_n \nabla \ln f(X_n, \theta)
\]
\[
\begin{pmatrix}
\int \frac{(t-g(\theta(1)))g'(\theta(1))}{\sigma^2} \lambda_n(t) \, dt \\
\int \frac{(t-g(\theta(1)))^2}{\sigma^2} \lambda_n(t) \, dt
\end{pmatrix}
\]
\[
\begin{pmatrix}
-\frac{n^{-1}g(\theta(1))^2}{\sigma^2} \psi_1 - \frac{N_n g(\theta(1))^1}{\sigma^2} \int t w((t - \psi(1))b_n^{-1}) \, dt \\
-\frac{N_n^2 g(\theta(1))^2}{\sigma^2} \int t w((t - \psi(1))b_n^{-1}) \, dt
\end{pmatrix}
\]
Now for the vector \( \hat{S}_n \) we get
\[
\hat{S}_n = -n^{-1} T_n(X_n, \theta)
\]
\[
\approx (n^{-\beta} + N_n b_n)^{-1} \begin{pmatrix}
n^{-\beta} \psi_1 + \frac{N_n}{\sigma^2} \int t w((t - \psi(1))b_n^{-1}) \, dt \\
\frac{1}{2} n^{-\beta} \psi_2 + N_n \int (t - g(\theta(1)))^2 w((t - \psi(1))b_n^{-1}) \, dt
\end{pmatrix}
\]
With the substitution \( z := \frac{t - \psi(1)}{b_n} \) and \( \int w(z) \, dz = 0 \) it is
\[
\hat{S}_n \approx (n^{-\beta} + N_n b_n)^{-1} \begin{pmatrix}
n^{-\beta} \psi_1 + \frac{N_n}{\sigma^2} \int z w(z) \, dz \\
\frac{1}{2} n^{-\beta} \psi_2 + N_n b_n^3 \int z^2 w(z) \, dz + 2N_n b_n^3 (c - g(\theta(1))) \int z w(z) \, dz
\end{pmatrix}
\]
Taking limits we determine the vector \( S \) which is independent of \( n \). Therefore we consider various cases according to the two local alternatives. We will find out that depending on \( \tau_n \) we get different vectors \( S \). Consequently, the
vector $S$ depends on the local alternatives. Let the vector $\tilde{S}_n$ for $n \to \infty$ converge to $S_{1j}$ in case of Pitman alternatives and to $S_{2j}$ in case of sharp peak alternatives. Let $j = 1$ for $\frac{n^{-\beta}}{N_n} \to \infty$, $j = 2$ for $\frac{n^{-\beta}}{N_n} \to 0$ and $j = 3$ for $\frac{n^{-\beta}}{N_n} \to \text{const.}$ First let $K_n$ be Pitman alternatives, that is $b_n \equiv b$.

1.1 If $\frac{n^{-\beta}}{N_n} \to \infty$ we get

$$\tau_n = n^{-\beta} \left( 1 + \frac{N_n}{n^{-\beta} b} \right) \sim n^{-\beta}$$

and the terms of the vector $\tilde{S}_n$ containing the factor $\frac{N_n}{n^{-\beta}}$ can be eliminated for determining the vector $S$. We get

$$S_{11} = \begin{pmatrix} \psi_1 \\ \frac{2b}{\sigma^2} \end{pmatrix}.$$  

This vector $S$ is consequently independent of the considered model, that is of $g(\theta_{(1)})$.

1.2 If $\frac{n^{-\beta}}{N_n} \to 0$ we get $\tau_n \sim N_n b$ and the terms of the vector $\tilde{S}_n$ containing the factor $\frac{n^{-\beta}}{N_n}$ can be eliminated for determining the vector $S$. In this case we get

$$S_{12} = \begin{pmatrix} \frac{b}{g'(\theta_{(1)})} \int z w(z) \, dz \\ \frac{1}{\sigma^2} \int z^2 w(z) \, dz + \frac{4b^2 (1 - g'(\theta_{(1)}))}{\sigma^2} \int z w(z) \, dz \end{pmatrix}.$$  

1.3 If $\frac{n^{-\beta}}{N_n} \to \text{const.} = L_1$ we get $n^{-\beta} \sim L_1 \cdot N_n$ and $\tau_n \sim (L_1 + b) N_n$. In this case we have

$$S_{13} = \begin{pmatrix} \frac{L_1 \cdot \psi_2}{L_1 + b} + \frac{b^2}{(L_1 + b) g'(\theta_{(1)})} \int z w(z) \, dz \\ \frac{2L_2 \cdot \psi_2}{(L_1 + b) \sigma^2} + \frac{2b^2}{\sigma^2} \int z^2 w(z) \, dz + \frac{4b^2 (1 - g'(\theta_{(1)}))}{\sigma^2 (L_1 + b)} \int z w(z) \, dz \end{pmatrix}.$$  

For sharp peak alternatives, that is $b_n \xrightarrow{n \to \infty} 0$, we get:

2.1 If $\frac{n^{-\beta}}{N_n b_n} \to \infty$ then follows $S_{21} = \begin{pmatrix} \psi_1 \\ \frac{2b}{\sigma^2} \end{pmatrix}$.

2.2 If $\frac{n^{-\beta}}{N_n b_n} \to 0$ then follows $S_{22} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

2.3 If $\frac{n^{-\beta}}{N_n b_n} \to L_2$ then follows $S_{23} = \begin{pmatrix} \frac{L_2 \cdot \psi_2}{L_2 + b} \\ \frac{2L_2 \cdot \psi_2}{(L_2 + b) \sigma^2} \end{pmatrix}$.

Obviously under the sharp peak alternatives always the vector $S$ is independent of $g(\theta_{(1)})$. 
Now the m.l.e. and the vectors $S_{12}$ and $S_{13}$ will be determined for two models of normal distribution under the special local alternatives with the error function

\[
w(t) = \begin{cases} 
\sin(-2\pi t) & \text{for } t \in [-\frac{1}{2}, \frac{1}{2}] \\
0 & \text{else} 
\end{cases}.
\]

- We have independent, identically normal distributed random variables $Y_i = \mu + \varepsilon$. Let $\varepsilon$ be normally distributed with expectation zero and variance $\sigma^2$, that is $Y_i \sim N(\mu, \sigma^2)$. It is $g(\hat{\theta}_{(1)}) = \mu$ and the m.l.e.

\[
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i.
\]

For the vectors $S_{12}$ and $S_{13}$ we get

\[
S_{12} = \left( \begin{array}{c}
-\frac{1}{2\pi} \\
2b(\varepsilon - \mu) \\
\frac{L_1 \psi_2}{L_2 + b} - \frac{L_2^2}{2\pi(L_1 + b)} \\
\frac{2L_1 \psi_2}{\sigma^2(L_2 + b)} - \frac{2L_2^2(\varepsilon - \mu)}{\pi\sigma^2(L_1 + b)} 
\end{array} \right),
\]

\[
S_{13} = \left( \begin{array}{c}
\frac{L_1 \psi_2}{L_2 + b} - \frac{L_2^2}{2\pi(L_1 + b)} \\
\frac{2L_1 \psi_2}{\sigma^2(L_2 + b)} - \frac{2L_2^2(\varepsilon - \mu)}{\pi\sigma^2(L_1 + b)} 
\end{array} \right).
\]

- Now we have independent, identically normal distributed random variables with $EY = e^{-\gamma t}$. The parameter $\hat{\theta}_{(1)} = \gamma$ is to be estimated by

\[
\hat{\gamma}_n = -\frac{\ln \left( \frac{\sum_{i=1}^{n} x_i}{t} \right)}{t}.
\]

For the vectors $S_{12}$ and $S_{13}$ we get

\[
S_{\gamma_{12}} = \left( \begin{array}{c}
-\frac{1}{2\pi} e^{-\gamma t} \\
2b(\varepsilon - e^{-\gamma t}) \\
\frac{L_1 \psi_2}{L_2 + b} - \frac{2L_1 \psi_2}{2\pi(L_1 + b)t} \\
\frac{2L_1 \psi_2}{(L_1 + b)^2\sigma^2} - \frac{2L_2^2(\varepsilon - e^{-\gamma t})}{\pi\sigma^2(L_1 + b)} 
\end{array} \right),
\]

\[
S_{\gamma_{13}} = \left( \begin{array}{c}
\frac{L_1 \psi_2}{L_2 + b} - \frac{L_2^2}{2\pi(L_1 + b)t} \\
\frac{2L_1 \psi_2}{(L_1 + b)^2\sigma^2} - \frac{2L_2^2(\varepsilon - e^{-\gamma t})}{\pi\sigma^2(L_1 + b)} 
\end{array} \right).
\]

**3 Power of the tests**

When we use the test which rejects the hypothesis if

\[
Q_n(f_{\hat{\theta}_n}) \geq \mu(f_{\hat{\theta}_n}) + u_n h_{\alpha}^{1/2} \sigma(f_{\hat{\theta}_n})
\]
then \( \hat{\theta}_n \) will be the m.l.e. in the following.

For the asymptotic power under the local alternatives we can find explicit approximations. In Liero/Läuter/Konakov (1998) the following result was proved:

**Theorem 3.1** Given

\[ n h_n^2 N_n^2 b_n = c_{1n}, \quad n^{1-2\beta} h_n^2 = c_{2n}, \quad n h_n^2 \tau_n^2 = c_{3n}, \]
\[ h_n^{-2} N_n b_n = c_{4n}, \quad h_n^{-2} n^{-\beta} = c_{5n}, \quad h_n^{-2} \tau_n = c_{6n}. \]

Under similar assumptions as in Theorem 1.1 and \( h_n b_n^{-1} \to 0 \) we get for the power \( \Pi(f_n) := P(Q_n(f_{\hat{\theta}_n}) \geq \mu(f_{\hat{\theta}_n}) + u_\alpha h_n^2 \sigma(f_{\hat{\theta}_n})(f_n) \)

\[
\lim_{n \to \infty} (\Pi(f_n) - \Phi(U_{1n,\alpha})) = 0
\]

with

\[ U_{1n,\alpha} = \sigma^{-1}(f_{\hat{\theta}}) \left[ c_{1n} \int w^2(t) a(e - tb_n) \text{d}t + \int \left[ (\sqrt{c_{2n} \psi} - \sqrt{c_{3n} S}) \nabla_{\vartheta} f(t, \vartheta) \right]^2 a(t) \text{d}t \right. 
\]
\[ + 2 \sqrt{c_{1n} b_n} \int (\sqrt{c_{2n} \psi} - \sqrt{c_{3n} S}) \nabla_{\vartheta} f(e - tb_n, \vartheta) w(t) a(e - tb_n) \text{d}t 
\]
\[ + \kappa(c_{4n} \int w(t) a(e - tb_n) \text{d}t + \int (c_{5n} \psi - c_{6n} S) \nabla_{\vartheta} f(t, \vartheta) a(t) \text{d}t) - u_\alpha. \]

Considering the asymptotic behaviour we get \( c_{1n} = n h_n^2 N_n^2 \) and \( c_{4n} = h_n^{-2} N_n \) as leading coefficients in the approximation of the power. Therefore we get the following simpler approximation of power \( U_{2n,\alpha} \) which is asymptotically equivalent to \( U_{1n,\alpha} \).

**Theorem 3.2** *(Pitman alternative)* Given \( b_n \equiv b \). For

\[ U_{2n,\alpha} = \sigma^{-1}(f_{\hat{\theta}}) \left[ c_{1n} \int \left( \frac{t}{b} - S \nabla f(e - tb_n, \vartheta) \right)^2 \text{d}t \right] - u_\alpha 
\]

we have

\[ \lim_{n \to \infty} (\Pi(f_n) - \Phi(U_{2n,\alpha})) = 0. \]

**Theorem 3.3** *(Sharp peak alternative)* Given \( b_n \to 0 \) For

\[ U_{2n,\alpha} = \sigma^{-1}(f_{\hat{\theta}}) \left[ c_{1n} b_n \int w^2(t) a(t) \text{d}t + T_n \right] - u_\alpha 
\]

with

\[ T_n = \kappa c_{4n} \left[ b_n \int w(t) t \text{d}t a'(c) - \int \nabla_{\vartheta} f_{\hat{\theta}}(t) a(t) \text{d}t S \right] \]
we have

$$\lim_{n \to \infty} (\Pi(f_n) - \Phi(U_{2,n,0})) = 0.$$ 

Under assumption A1 and the choice \(a(t) = 1\) we have \(T_n = 0\). Consequently we get for Pitman and sharp peak alternatives:

**Theorem 3.4** If \(\frac{b_n}{h_n} \to 0\) and \(nh_n^2 \to \infty\) then

$$\lim_{n \to \infty} \Pi(f_n) \begin{cases} = \alpha \\ > \alpha, < 1 \quad \text{for} \ b_n c_1 n \to \begin{cases} 0 \\ \xi > 0 \quad \infty \end{cases} \end{cases}.$$ 

4 Exactness of the approximations

In the following we consider how \(\Phi(U_{1,n,0})\) and \(\Phi(U_{2,n,0})\) approximate the power when we have a finite number of observations. Let \(\mathcal{F}\) be the class of normal distributions \(N(\mu, \sigma^2)\) for \(1^t = (1, \ldots, 1)\) and \(\mu = g(\vartheta_{(1)})\). For constructing the local alternatives we choose

$$w(t) = \begin{cases} \sin(-2\pi t) & \text{for} \ t \in [-\frac{1}{2}, \frac{1}{2}] \\ 0 & \text{else} \end{cases},$$

c = 0 and \(\beta = 1000\). With \(S = (S_{(1)}, S_{(2)})\), \(f_1 = \frac{\partial f(j, \vartheta)}{\partial \vartheta_{(1)}}\), \(f_2 = \frac{\partial^2 f(j, \vartheta)}{\partial \vartheta_{(1)}^2}\), \(f_{11} = \frac{\partial f(j, \vartheta)}{\partial \vartheta_{(1)}}\), and \(f_{22} = \frac{\partial^2 f(j, \vartheta)}{\partial \vartheta_{(1)}^2}\) as well as assumption A1 about the hypothesis density we get

$$U_{1,n,0} = \sigma^{-1}(f_j) \left( \frac{c_{1n}}{2} + \int (\sqrt{\varphi_n S_{(1)}} f_1 + (\sqrt{\varphi_n} - \sqrt{\varphi_n S_{(2)}}) f_2)^2 dt \\ + 2 \sqrt{c_{1n} b_n} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\sqrt{\varphi_n S_{(1)}} f_1 + (\sqrt{\varphi_n} - \sqrt{\varphi_n S_{(2)}}) f_2) \sin(-2\pi t) dt \right) - a_n,$$

$$U_{2,n,0} = \sigma^{-1}(f_j) \left[ c_{1n} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left( \sin \left( -2\pi \left( \frac{t}{b} \right) \right) - S_{(1)} f_1 - S_{(2)} f_2 \right)^2 dt \right] - a_n \quad \text{(Pitman alternative)}$$

$$U_{2,n,0} = \frac{\sigma^{-1}(f_j) c_{1n} b_n}{2} - a_n \quad \text{(sharp peak alternative)}.$$ 

As values for \(N_n\) and \(b_n\) we choose 0.38, 0.2, 0.038 and 1.0, 0.5, 0.1, respectively. Let \(n\) be 50, 100 or 150.

Every one of these cases we can embed in the Pitman or sharp peak alternatives. The classification of a special case with fixed \(n\) into either one or the
other of both classes is somewhat subjective. We choose sharp peak alternatives as model for our considerations.

For some $N_n$ and $b$ one can find the approximations $\Phi(U_{1,n,0.12})$ in Table 1 and the simpler approximations $\Phi(U_{2,n,0.12})$ in Table 2.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$N_n$</th>
<th>$b_n$</th>
<th>$b_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>50</td>
<td>0.380</td>
<td>0.9924</td>
<td>0.5921</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>0.4358</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.380</td>
<td>1.0000</td>
<td>0.9957</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>0.200</td>
<td>0.9957</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Approximations of power $\Phi(U_{1,n,0.12})$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$N_n$</th>
<th>$b_n$</th>
<th>$b_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>50</td>
<td>0.380</td>
<td>0.9958</td>
<td>0.9260</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>0.5117</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.380</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>0.9957</td>
<td>0.9106</td>
</tr>
<tr>
<td></td>
<td>0.038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>0.200</td>
<td>0.9957</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Approximations of power $\Phi(U_{2,n,0.12})$

In simulations we considered the power of the test using the Epanechnikov kernel or the Gauß kernel.

- Power when using the Epanechnikov kernel

In Figure 2 one can see the hypothesis, the local alternative for $N_n = 0.38$
and $b_n = 1$, the density estimator with the Epanechnikov kernel

$$K(x) = \begin{cases} \frac{3}{2}(1 - x^2) & \text{for } |x| \leq 1 \\ 0 & \text{else} \end{cases}$$

with a creation of 50 random numbers and the expectation of the density estimator. The results of the simulation can be found in Table 3. In

![Figure 2: Local alternative $f_{50}$ with $N_{50} = 0.38$, $b_{50} = 1$, hypothesis $f$, expectation $\epsilon_{E,0.15}$ and density estimator $\hat{f}_{E,50}$](image)

Table 4 are the differences $\Phi(U_{i,50,0.12}) - \Pi_i(f_{50})$, $i = 1, 2$ and in Table 5 are the relative distances $\frac{\Pi_i(f_{50}) - \Phi(U_{i,50,0.12})}{\Pi_i(f_{50})}$.

<table>
<thead>
<tr>
<th>$N_{50}$</th>
<th>$b_{50}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>0.79</td>
</tr>
<tr>
<td>0.200</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Table 3: Power $\Pi_1(f_{50})$ using the Epanechnikov kernel

- Power when using the Gauß kernel

In Figure 3 one finds the hypothesis, the local alternative for $N_n = 0.38$ and $b = 1$, the density estimator with Gauß kernel with a creation of 50 random numbers and the expectation of the density estimator. In Table 6 are the results of the simulations using the Gauß kernel.

We got the differences shown in Table 7 and the relative distances shown in Table 8.
Table 4: Differences between the power $\Pi_1(f_{50})$ and the power approximation $\Phi(U_{1,50,0.12})$ using the Epanechnikov kernel

<table>
<thead>
<tr>
<th>$N_{50}$</th>
<th>$b_{50}$</th>
<th>$\Phi(U_{1,50,0.12}) - \Pi_1(f_{50})$</th>
<th>$b_{50}$</th>
<th>$\Phi(U_{2,50,0.12}) - \Pi_1(f_{50})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.380</td>
<td>1.0</td>
<td>0.2024</td>
<td>1.0</td>
<td>0.2058</td>
</tr>
<tr>
<td>0.200</td>
<td>0.5</td>
<td>0.2221</td>
<td>0.5</td>
<td>0.5560</td>
</tr>
<tr>
<td>0.200</td>
<td>0.5</td>
<td>0.2317</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Relative distances between the power $\Pi_1(f_{50})$ and the power approximations $\Phi(U_{1,50,0.12})$ using the Epanechnikov kernel

| $N_{50}$ | $b_{50}$ | $|\Phi(U_{1,50,0.12}) - \Pi_1(f_{50})|/\Pi_1(f_{50})$ | $b_{50}$ | $|\Phi(U_{2,50,0.12}) - \Pi_1(f_{50})|/\Pi_1(f_{50})$ |
|-------|--------|----------------------------------|--------|----------------------------------|
| 0.380 | 1.0    | 0.2562                           | 1.0    | 0.2605                           |
| 0.200 | 0.5    | 0.6003                           | 0.5    | 0.8275                           |

For the case $b_n = 0.1$, $N_n = 0.038$ we repeated the simulation as well with $n = 50$ as with $n = 100$ random numbers. Using $n = 50$ random numbers we got the same power $\Pi_2(f_{50}) = 0.15$ as in the first simulation. However, using $n = 100$ random numbers we got the power $\Pi_2(f_{100}) = \frac{7}{50} \approx 0.2333$ which is still higher. From Figure 4 we can guess the reason for this phenomenon. In the figure one finds a density estimation with $n = 50$ random numbers, its expectation, the hypothesis, the alternative for $b_{50} = 0.1$ and $N_{50} = 0.038$ as well as the random numbers $x_{50}$, which fall into the considered interval. It is obviously that the density estimator is not near the alternative. The reason for this is that the number of random variables is too small for a convenient choice of the bandwidth $b_n$ when we have such a tiny $b_n$. Consequently, a simulation with $b_n = 0.1$ and $N_n = 0.038$ demands a much greater number of random variables.

- Interpretation of the results

With the simulations we could confirm the fact, that both $L_2$-tests recognize a distance from the hypothesis density worse with declining interval of distance and declining distance. Very small distances in small intervals will be interpreted as random errors and consequently the hy-
Figure 3: Local alternative $f_{50}$ with $N_{50} = 0.38$, $b_{50} = 1$, hypothesis $f$, expectation $\epsilon_{G,0.15}$ and density estimator $\hat{f}_{G,50}$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$N_n$</th>
<th>$b_n$</th>
<th>1.0</th>
<th>0.5</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.380</td>
<td>0.72</td>
<td>0.27</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>0.19</td>
<td>0.13</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.038</td>
<td>0.06</td>
<td>0.07</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.380</td>
<td>$\frac{20}{30} \approx 0.6667$</td>
<td>$\frac{11}{30} \approx 0.3667$</td>
<td>$\frac{7}{30} \approx 0.2333$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>$\frac{8}{30} \approx 0.2667$</td>
<td>$\frac{1}{30} \approx 0.0333$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.038</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>0.200</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Power $\Pi_2(f_n)$ using the Gauß kernel

hypothesis will not be rejected. Both tests discerned a distance from the hypothesis worse when the intervals of distance were longer ($b_n = 1$) but the distance itself smaller ($N_n = 0.2$) than when the intervals of distance were smaller ($b_n = \frac{1}{2}$) and the distance greater ($N_n = 0.2$). That is due to the squaring of the distance in $L_2$-tests. If the distances are less than one they become still less this way and so a rejection of the hypothesis becomes improbable.

Furthermore, the simulations show significant differences between both tests. The power is always higher when using the Epanechikov kernel. Though the decisions based on these tests are the same when we have an unlimited number of random variables, the differences should
Table 7: Differences between power $\Pi_2(f_n)$ and power approximations $\Phi(U_{i,n,0.12})$ using the Gauß kernel

<table>
<thead>
<tr>
<th>$n$</th>
<th>$N_n$</th>
<th>$b_n$</th>
<th>$\Phi(U_{1,n,0.12}) - \Pi_2(f_n)$</th>
<th>$\Phi(U_{2,n,0.12}) - \Pi_2(f_n)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$b_n$</td>
<td>$b_n$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>50</td>
<td>0.380</td>
<td>0.2800</td>
<td>0.6258</td>
<td>0.1214</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>0.5270</td>
<td>0.1802</td>
<td>0.0833</td>
</tr>
<tr>
<td></td>
<td>0.038</td>
<td>0.0753</td>
<td>0.0572</td>
<td>-0.0272</td>
</tr>
<tr>
<td>100</td>
<td>0.380</td>
<td>0.0333</td>
<td>0.6290</td>
<td>0.0333</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>0.7233</td>
<td>0.5366</td>
<td>0.7920</td>
</tr>
<tr>
<td></td>
<td>0.038</td>
<td>-0.1101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>0.200</td>
<td>0.4957</td>
<td>0.4957</td>
<td>0.4957</td>
</tr>
</tbody>
</table>

Table 8: Relative distances between power $\Pi_2(f_n)$ and power approximation $\Phi(U_{i,n,0.12})$ using the Gauß kernel

<table>
<thead>
<tr>
<th>$n$</th>
<th>$N_n$</th>
<th>$b_n$</th>
<th>$\frac{\Phi(U_{1,50,0.12}) - \Pi_1(f_{50})}{\Pi_1(f_{50})}$</th>
<th>$\frac{\Phi(U_{2,50,0.12}) - \Pi_1(f_{50})}{\Pi_1(f_{50})}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$b_n$</td>
<td>$b_n$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>50</td>
<td>0.380</td>
<td>0.3889</td>
<td>2.3178</td>
<td>2.4280</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>2.7737</td>
<td>1.3862</td>
<td>1.6660</td>
</tr>
<tr>
<td></td>
<td>0.038</td>
<td>1.2550</td>
<td>0.8171</td>
<td>0.1813</td>
</tr>
<tr>
<td>100</td>
<td>0.380</td>
<td>0.0344</td>
<td>1.7153</td>
<td>0.0344</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>2.7112</td>
<td>16.1141</td>
<td>2.7112</td>
</tr>
<tr>
<td></td>
<td>0.038</td>
<td>0.4719</td>
<td>0.4719</td>
<td>0.4719</td>
</tr>
</tbody>
</table>

be considered when the number of random variables is limited.

Moreover, the asymptotically equivalent approximations of power $\Phi(U_{1,n,0})$ and $\Phi(U_{2,n,0})$ differ when the number of random variables is finite. We see, that with a finite number of random variables $\Phi(U_{1,n,0})$ approximates the power nearly always better and never worse than $\Phi(U_{2,n,0})$. Consequently, the simplifications in $\Phi(U_{2,n,0})$ do not seem convenient for a small number of random variables.
References


