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.l.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 $X_{n1}, \dots, X_{nn}$ be a sample of i.i.d. random variables with density f_n. We consider the problem of testing the hypothesis

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

against the nonparametric alternative

$$K: \quad f_n \notin \mathfrak{F}$$

where $\mathfrak{F}$ 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 $f_n(\cdot, \vartheta) = f_\vartheta$

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

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 $\hat{\vartheta}_n$ under the hypothesis H with $X_i \sim f_\vartheta$ for each $\vartheta \in \Theta$*

$$\sqrt{n}|\hat{\vartheta}_n - \vartheta| = \mathrm{O}_\mathrm{P}(1)$$

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

$$\mathsf{P}_{f_\vartheta}\left(\frac{h_n^{-\frac{1}{2}}(Q_n(f_{\hat{\vartheta}_n}) - \mu(f_{\hat{\vartheta}_n}))}{\sigma(f_{\hat{\vartheta}_n})} \le x\right) \to \Phi(x)$$

with

$$\begin{aligned}
Q_n(f_{\hat\vartheta_n}) &= nh_n \int (\hat f_n(t) - e_h(t, f_{\hat\vartheta_n}))^2 a(t)\,\mathrm{d}t \quad \textit{with} \\
e_h(t, f_{\hat\vartheta_n}) &= \int K(x) f_{\hat\vartheta_n}(t - xh_n)\,\mathrm{d}x, \\
a(t) &= \textit{weight function} \quad \textit{and} \\
\hat f_n(t) &= (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{t - X_{ni}}{h_n}\right), \\
\mu(f_{\hat\vartheta_n}) &= \kappa \int f_{\hat\vartheta_n}(t) a(t)\,\mathrm{d}t \quad \textit{with} \\
\kappa &= \int K^2(x)\,\mathrm{d}x, \\
\sigma^2(f_{\hat\vartheta_n}) &= 2\kappa^* \int f^2_{\hat\vartheta_n}(t) a^2(t)\,\mathrm{d}t \quad \textit{with} \\
\kappa^* &= \int (K * K)^2(x)\,\mathrm{d}x\,, \\
\Phi(u_\alpha) &= 1 - \alpha
\end{aligned}$$

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

This way we get an asymptotic α-test, if we reject the hypothesis if

$$Q_n(f_{\hat\vartheta_n}) \ge \mu(f_{\hat\vartheta_n}) + u_\alpha h_n^{\frac{1}{2}} \sigma(f_{\hat\vartheta_n})\,.$$

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

$$P\left(Q_n(f_{\hat\vartheta_n}) \ge \mu(f_{\hat\vartheta_n}) + u_\alpha h_n^{\frac{1}{2}} \sigma(f_{\hat\vartheta_n})\right)\,.$$

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

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

with $\vartheta_n = \vartheta + n^{-\beta}\psi$ with an unknown vector ψ, $\beta > 0$ and ϑ 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 $\int w(x)\,\mathrm{d}x = 0$. For $b_n \equiv b$ we get the Pitman alternative and for $b_n \xrightarrow[n\to\infty]{} 0$ the sharp peak alternative. In Figure 1 there are three of such local alternatives in comparison with the normal density $f(x; 0.2, 1)$.

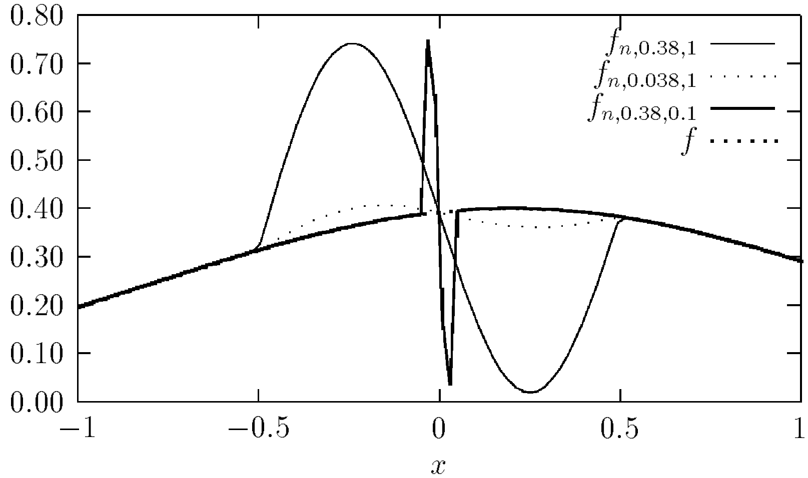

Figure 1: Local alternatives f_n with $N_n = 0.38, b_n = 1$, with $N_n = 0.038, b_n = 1$, with $N_n = 0.38, b_n = 0.1$ and hypothesis f.

2 The Maximum Likelihood Estimation

Because the parameter ϑ is unknown, is has to be estimated. If the hypothesis holds then

$$\hat{\vartheta}_n - \vartheta = n^{-\frac{1}{2}} Z_n$$

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

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

$$\int \nabla_\vartheta f(x,\vartheta)\,\mathrm{d}x = 0, \quad \int \frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j} f(x,\vartheta)\,\mathrm{d}x = 0\,. \tag{2}$$

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

$$\left| \frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j} \ln f(x,\vartheta') \right| \leq m(x,\vartheta) \quad \forall \vartheta' \in U_\delta(\vartheta) \tag{3}$$

for a measurable function m with $\int m(x,\vartheta) f(x,\vartheta)\,\mathrm{d}x < \infty$.

A3 The determinant of the Fisher Information

$$I(\vartheta) = \int \nabla_\vartheta \ln f(x,\vartheta) (\nabla_\vartheta \ln f(x,\vartheta))^t f(x,\vartheta)\,\mathrm{d}x \tag{4}$$

is positive for each $\vartheta \in \Theta^\circ$.

We put

$$\lambda_n(t) = -n^{-\beta}\psi^t\nabla_\vartheta f(t,\tilde{\vartheta}) - N_n w((t-c)b_n^{-1})$$

with a suitable $\tilde{\vartheta}$ between $\hat{\vartheta}_n$ and ϑ.

Then it follows:

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

$$\int \nabla_\vartheta \ln f(t,\vartheta)\lambda_n(t)\,\mathrm{d}t = \mathrm{O}(\tau_n)$$

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

$$\hat{\vartheta}_n - \vartheta = n^{-\frac{1}{2}} Z_n + \tau_n S + \mathrm{o_P}(\tau_n)\,. \tag{5}$$

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 $\mathfrak{F}$ 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 $\tau_n = n^{-\beta} + N_n b_n$

$$\int \nabla_\vartheta \ln f(t,\vartheta)\lambda_n(t)\,\mathrm{d}t = \int \begin{pmatrix} \frac{(x-g(\vartheta_{(1)}))g'(\vartheta_{(1)})}{\sigma^2} \\ \frac{(x-g(\vartheta_{(1)}))^2-\sigma^2}{2\sigma^4} \end{pmatrix} \lambda_n(t)\,\mathrm{d}t = \mathrm{O}(\tau_n)\,.$$

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

$$nT_n^{-1}(X_n,\vartheta)n^{\frac{1}{2}}\mathrm{E}_n\nabla_\vartheta \ln f(X_{nl},\vartheta) = n^{\frac{1}{2}}\tau_n(S + \mathrm{o_P}(1))\,.$$

$T_n = T_{nij}$ is the matrix

$$T_{nij}(x,\vartheta) = \frac{\partial^2}{\partial\vartheta_i \partial\vartheta_j} \ln \prod_{l=1}^{n} f(x_{nl},\vartheta)$$

and with the law of large numbers we see

$$n^{-1} T_n(X_n,\vartheta) \xrightarrow[n\to\infty]{} -I(\vartheta)$$

with the Fisher Information $I(\vartheta)$. It follows

$$-\tau_n^{-1} \cdot I(\vartheta)^{-1} \mathrm{E}_n \nabla_\vartheta \ln f(X_{nl},\vartheta) =: \tilde{S}_n \xrightarrow[n\to\infty]{} S\,.$$

We get

$$I(\vartheta)^{-1} = \frac{2\sigma^6}{g'(\vartheta_{(1)})^2} \begin{pmatrix} \frac{1}{2\sigma^4} & 0 \\ 0 & \frac{g'(\vartheta_{(1)})^2}{\sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{\sigma^2}{g'(\vartheta_{(1)})^2} & 0 \\ 0 & 2\sigma^4 \end{pmatrix}$$

and with $\psi^t = (\psi_1, \psi_2)$

$$\begin{aligned} & \mathrm{E}_n \nabla_\vartheta \ln f(X_{nl},\vartheta) \\ = & \begin{pmatrix} \int \frac{(t-g(\vartheta_{(1)}))g'(\vartheta_{(1)})}{\sigma^2} \lambda_n(t)\,\mathrm{d}t \\ \int \left(\frac{(t-g(\vartheta_{(1)}))^2}{\sigma^6} - \frac{1}{2\sigma^2}\right) \lambda_n(t)\,\mathrm{d}t \end{pmatrix} \\ = & \begin{pmatrix} -\frac{n^{-\beta} g'(\vartheta_{(1)})^2 \psi_1}{\sigma^2} - \frac{N_n g'(\vartheta_{(1)})}{\sigma^2} \int t w((t-c)b_n^{-1})\,\mathrm{d}t \\ -\frac{n^{-\beta}\psi_2}{\sigma^6} - \frac{N_n}{\sigma^6} \int (t-g(\vartheta_{(1)}))^2 w((t-c)b_n^{-1})\,\mathrm{d}t \end{pmatrix}. \end{aligned}$$

Now for the vector $\tilde{S}_n$ we get

$$\begin{aligned} \tilde{S}_n &= -\tau_n^{-1} I(\vartheta)^{-1} \mathrm{E}_n \nabla_\vartheta \ln f(X_{nl},\vartheta) \\ &\approx (n^{-\beta} + N_n b_n)^{-1} \cdot \begin{pmatrix} n^{-\beta}\psi_1 + \frac{N_n}{g'(\vartheta_1)} \int t w((t-c)b_n^{-1})\,\mathrm{d}t \\ \frac{2}{\sigma^2}\left(n^{-\beta}\psi_2 + N_n \int (t-g(\vartheta_1))^2 w((t-c)b_n^{-1})\,\mathrm{d}t\right) \end{pmatrix}. \end{aligned}$$

With the substitution $z := \frac{t-c}{b_n}$ and $\int w(z)\,\mathrm{d}z = 0$ it is

$$\tilde{S}_n \approx (n^{-\beta} + N_n b_n)^{-1} \cdot \begin{pmatrix} n^{-\beta}\psi_1 + \frac{b_n^2 N_n}{g'(\vartheta_{(1)})} \int z w(z)\,\mathrm{d}z \\ \frac{2}{\sigma^2}\left(n^{-\beta}\psi_2 + N_n b_n^3 \int z^2 w(z)\,\mathrm{d}z \right. \\ \left. + 2N_n b_n^2 (c - g(\vartheta_{(1)})) \int z w(z)\,\mathrm{d}z\right) \end{pmatrix}.$$

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 τ_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 $\frac{n^{-\beta}}{N_n} \to$ 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

$$S_{11} = \begin{pmatrix} \psi_1 \\ \frac{2\psi_2}{\sigma^2} \end{pmatrix} .$$

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

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

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

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

$$S_{13} = \begin{pmatrix} \frac{L_1 \cdot \psi_1}{L_1 + b} + \frac{b^2}{(L_1+b) g'(\vartheta_{(1)})} \int z w(z)\,\mathrm{d}z \\ \frac{2 \cdot L_1 \cdot \psi_2}{(L_1+b)\sigma^2} + \frac{2b^3}{\sigma^2 (L_1+b)} \int z^2 w(z)\,\mathrm{d}z + \frac{4b^2(c - g(\vartheta_{(1)}))}{\sigma^2 (L_1+b)} \int z w(z)\,\mathrm{d}z \end{pmatrix} .$$

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 $S_{21} = \begin{pmatrix} \psi_1 \\ \frac{2\psi_2}{\sigma^2} \end{pmatrix}$.

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 $S_{23} = \begin{pmatrix} \frac{L_2 \cdot \psi_1}{L_2 + 1} \\ \frac{2 \cdot L_2 \cdot \psi_2}{(L_2+1)\sigma^2} \end{pmatrix}$.

Obviously under the sharp peak alternatives always the vector S is independent of $g(\vartheta_{(1)})$.

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 ε be normally distributed with expectation zero and variance σ^2, that is $Y_i \sim N(\mu, \sigma^2)$. It is $g(\vartheta_{(1)}) = \mu$ 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

$$S_{12} = \begin{pmatrix} -\frac{b}{2\pi} \\ -\frac{2b(c-\mu)}{\pi\sigma^2} \end{pmatrix} ,$$

$$S_{13} = \begin{pmatrix} \frac{L_1 \cdot \psi_1}{L_1+b} - \frac{b^2}{2\pi(L_1+b)} \\ \frac{2 \cdot L_1 \cdot \psi_2}{\sigma^2(L_1+b)} - \frac{2b^2(c-\mu)}{\pi\sigma^2(L_1+b)} \end{pmatrix} .$$

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

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

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

$$S_{\gamma 12} = \begin{pmatrix} -\frac{b\,\mathrm{e}^{-\gamma t}}{2\pi t} \\ -\frac{2b\left(c - \mathrm{e}^{\gamma t}\right)}{\pi\sigma^2} \end{pmatrix} ,$$

$$S_{\gamma 13} = \begin{pmatrix} \frac{L_1 \cdot \psi_1}{L_1+b} - \frac{b^2\,\mathrm{e}^{-\gamma t}}{2\pi(L_1+b)t} \\ \frac{2 \cdot L_1 \cdot \psi_2}{(L_1+b)\sigma^2} - \frac{2b^2\left(c - \mathrm{e}^{\gamma t}\right)}{\pi\sigma^2(L_1+b)} \end{pmatrix} .$$

3 Power of the tests

When we use the test which rejects the hypothesis if

$$Q_n(f_{\hat{\vartheta}_n}) \geq \mu(f_{\hat{\vartheta}_n}) + u_\alpha h_n^{\frac{1}{2}} \sigma(f_{\hat{\vartheta}_n})$$

then $\hat{\vartheta}_n$ 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*

$$\begin{aligned} nh_n^{\frac{1}{2}} N_n^2 b_n &= c_{1n}, \quad n^{1-2\beta} h_n^{\frac{1}{2}} = c_{2n}, \quad nh_n^{\frac{1}{2}} \tau_n^2 = c_{3n}, \\ h_n^{-\frac{1}{2}} N_n b_n &= c_{4n}, \quad h_n^{-\frac{1}{2}} n^{-\beta} = c_{5n}, \quad h_n^{-\frac{1}{2}} \tau_n = c_{6n} \,. \end{aligned}$$

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

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

with

$$\begin{aligned} U_{1,n,\alpha} &= \sigma^{-1}(f_\vartheta) \Big(c_{1n} \int w^2(t) a(c - tb_n)\, \mathrm{d}t + \int [(\sqrt{c_{2n}}\psi - \sqrt{c_{3n}} S)^t \nabla_\vartheta f(t,\vartheta)]^2 a(t)\, \mathrm{d}t \\ &\quad + 2\sqrt{c_{1n} b_n} \int (\sqrt{c_{2n}}\psi - \sqrt{c_{3n}} S)^t \nabla_\vartheta f(c - tb_n, \vartheta) w(t) a(c - tb_n)\, \mathrm{d}t \\ &\quad + \kappa (c_{4n} \int w(t) a(c - tb_n)\, \mathrm{d}t + \int (c_{5n}\psi - c_{6n} S)^t \nabla_\vartheta f(t,\vartheta) a(t)\, \mathrm{d}t) \Big) - u_\alpha \,. \end{aligned}$$

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

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

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

we have

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

Theorem 3.3 *(Sharp peak alternative) Given* $b_n \xrightarrow[n\to\infty]{} 0$. *For*

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

with

$$T_n = \kappa c_{4n} b_n \left[b_n \int w(t) t\, \mathrm{d}t\, a'(c) - \int \nabla_\vartheta f_\vartheta(t)^t a(t)\, \mathrm{d}t\, S \right]$$

we have

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

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

Theorem 3.4 *If* $\frac{h_n}{b_n} \xrightarrow[n\to\infty]{} 0$ *and* $nh_n^{\frac{3}{2}} \xrightarrow[n\to\infty]{} \infty$ *then*

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

4 Exactness of the approximations

In the following we consider how $\Phi(U_{1,n,\alpha})$ and $\Phi(U_{2,n,\alpha})$ approximate the power when we have a finite number of observations. Let $\mathfrak{F}$ be the class of normal distributions $N(\mu\mathbf{1}, \sigma^2 I)$ for $\mathbf{1}^t = (1,\ldots,1)$ and $\mu = g(\vartheta_{n(1)})$. 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 $S = (S_{(1)}, S_{(2)})^t$, $f_1 = \frac{\partial f(t,\vartheta)}{\partial \vartheta_{(1)}}$, $f_2 = \frac{\partial f(t,\vartheta)}{\partial \sigma^2}$, $f_{11} = \frac{\partial f(c-tb_n,\vartheta)}{\partial \vartheta_{(1)}}$ and $f_{22} = \frac{\partial f(c-tb_n,\vartheta)}{\partial \sigma^2}$ as well as assumption A1 about the hypothesis density we get

$$\begin{aligned} U_{1,n,\alpha} &= \sigma^{-1}(f_\vartheta)\left(\frac{c_{1n}}{2} + \int (-\sqrt{c_{3n}}S_{(1)}f_1 + (\sqrt{c_{2n}} - \sqrt{c_{3n}}S_{(2)})f_2)^2\,\mathrm{d}t\right. \\ &\qquad \left. + 2\sqrt{c_{1n}b_n}\int_{-\frac{1}{2}}^{\frac{1}{2}} (-\sqrt{c_{3n}}S_{(1)}f_{11} + (\sqrt{c_{2n}} - \sqrt{c_{3n}}S_{(2)})f_{22})\sin(-2\pi t)\,\mathrm{d}t\right) \\ &\qquad - u_\alpha\,, \\ U_{2,n,\alpha} &= \sigma^{-1}(f_\vartheta)\left[c_{1n}\int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\sin\left(-2\pi\left(\frac{t}{b}\right)\right) - S_{(1)}f_1 - S_{(2)}f_2\right)^2 \mathrm{d}t\right] - u_\alpha \\ &\qquad \text{(Pitman alternative)} \\ U_{2,n,\alpha} &= \frac{\sigma^{-1}(f_\vartheta)c_{1n}b_n}{2} - u_\alpha \quad \text{(sharp peak alternative)}\,. \end{aligned}$$

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 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_n one can find the approximations $\Phi(U_{1,n,0.12})$ in Table 1 and the simplier approximations $\Phi(U_{2,n,0.12})$ in Table 2.

		Epanechnikov kernel		Gauß kernel		
n	N_n	b_n		b_n		
		1.0	0.5	1.0	0.5	0.1
50	0.380	0.9924	0.5921	1.0000	0.8958	0.1714
	0.200	0.4358		0.7170	0.3102	0.1333
	0.038			0.1353	0.1272	0.1228
100	0.380			1.0000	0.9957	
	0.200			0.9900	0.5699	
	0.038					0.1232
150	0.200			0.9957		

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

		Epanechnikov kernel		Gauß kernel		
n	N_n	b_n		b_n		
		1.0	0.5	1.0	0.5	0.1
50	0.380	0.9958	0.9260	1.0000	0.9958	0.8359
	0.200	0.5117		0.8287	0.5361	0.2843
	0.038			0.1369	0.1317	0.1268
100	0.380			1.0000	1.0000	
	0.200			0.9957	0.9106	
	0.038					0.1313
150	0.200			0.9957		

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

$$K(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{for } |x| \leq 1 \\ 0 & \text{else} \end{cases}$$

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

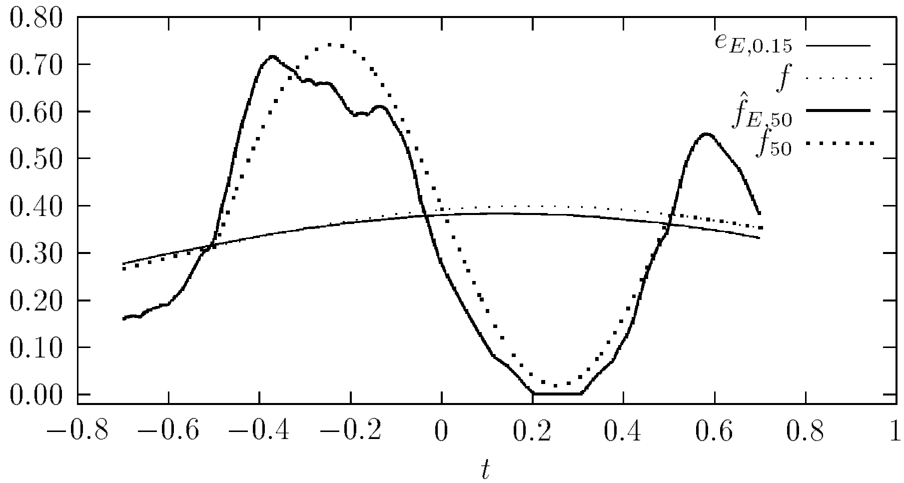

Figure 2: Local alternative f_{50} with $N_{50} = 0.38$, $b_{50} = 1$, hypothesis f, expectation $e_{E,0.15}$ and density estimator $\hat{f}_{E,50}$

Table 4 are the differences $\Phi(U_{i,50,0.12}) - \Pi_1(f_{50})$, $i = 1, 2$ and in Table 5 are the relative distances $\frac{|\Pi_1(f_{50}) - \Phi(U_{i,50,0.12})|}{\Pi_1(f_{50})}$.

N_{50}	b_{50}	
	1.0	0.5
0.380	0.79	0.37
0.200	0.28	

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.

N_{50}	$\Phi(U_{1,50,0.12}) - \Pi_1(f_{50})$		$\Phi(U_{2,50,0.12}) - \Pi_1(f_{50})$	
	b_{50}		b_{50}	
	1.0	0.5	1.0	0.5
0.380	0.2024	0.2221	0.2058	0.5560
0.200	0.1558		0.2317	

Table 4: Differences between the power $\Pi_1(f_{50})$ and the power approximation $\Phi(U_{i,50,0.12})$ using the Epanechnikov kernel

N_{50}	$\frac{\lvert\Phi(U_{1,50,0.12}) - \Pi_1(f_{50})\rvert}{\Pi_1(f_{50})}$		$\frac{\lvert\Phi(U_{2,50,0.12}) - \Pi_1(f_{50})\rvert}{\Pi_1(f_{50})}$	
	b_{50}		b_{50}	
	1.0	0.5	1.0	0.5
0.380	0.2562	0.6003	0.2605	1.5030
0.200	0.5564		0.8275	

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

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 $\Pi_2(f_{100}) = \frac{7}{30} \approx 0.2333$ 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_{50i} 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-

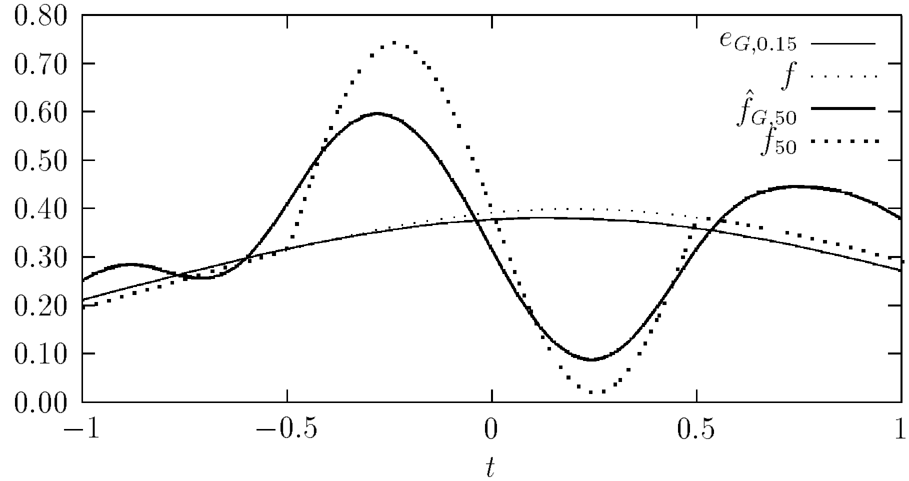

Figure 3: Local alternative f_{50} with $N_{50} = 0.38$, $b_{50} = 1$, hypothesis f, expectation $e_{G,0.15}$ and density estimator $\hat{f}_{G,50}$

n	N_n	b_n		
		1.0	0.5	0.1
50	0.380	0.72	0.27	0.05
	0.200	0.19	0.13	0.05
	0.038	0.06	0.07	0.15
100	0.380	$\frac{29}{30} \approx 0.9667$	$\frac{11}{30} \approx 0.3667$	
	0.200	$\frac{8}{30} \approx 0.2667$	$\frac{1}{30} \approx 0.0333$	
	0.038			$\frac{7}{30} \approx 0.2333$
150	0.200	0.50		

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

pothesis 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 Epanechnikov kernel. Though the decisions based on these tests are the same when we have an unlimited number of random variables, the differences should

		$\Phi(U_{1,n,0.12}) - \Pi_2(f_n)$			$\Phi(U_{2,n,0.12}) - \Pi_2(f_n)$		
		b_n			b_n		
n	N_n	1.0	0.5	0.1	1.0	0.5	0.1
50	0.380	0.2800	0.6258	0.1214	0.2800	0.7258	0.7859
	0.200	0.5270	0.1802	0.0833	0.6387	0.4061	0.2343
	0.038	0.0753	0.0572	−0.0272	0.0769	0.0617	−0.0232
100	0.380	0.0333	0.6290		0.0333	0.6333	
	0.200	0.7233	0.5366		0.7290	0.8773	
	0.038			−0.1101			−0.1020
150	0.200	0.4957			0.4957		

Table 7: Differences between power $\Pi_2(f_n)$ and power approximations $\Phi(U_{i,n,0.12})$ using the Gauß kernel

		$\frac{\lvert\Phi(U_{1,50,0.12}) - \Pi_1(f_{50})\rvert}{\Pi_1(f_{50})}$			$\frac{\lvert\Phi(U_{2,50,0.12}) - \Pi_1(f_{50})\rvert}{\Pi_1(f_{50})}$		
		b_n			b_n		
n	N_n	1.0	0.5	0.1	1.0	0.5	0.1
50	0.380	0.3889	2.3178	2.4280	0.3889	2.6881	15.7180
	0.200	2.7737	1.3862	1.6660	3.3616	3.1238	4.6860
	0.038	1.2550	0.8171	0.1813	1.2817	0.8814	0.1547
100	0.380	0.0344	1.7153		0.0344	1.7270	
	0.200	2.7112	16.1141		2.7334	26.3453	
	0.038			0.4719			0.4372
150	0.200	0.9914			0.9914		

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

be considered when the number of random variables is limited.

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

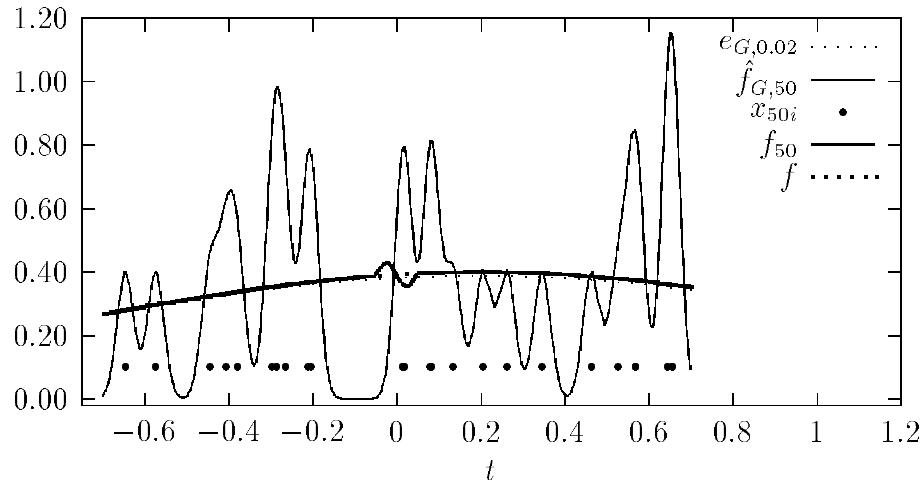

Figure 4: Local alternative f_{50} with $N_{50} = 0.038$, $b_{50} = 0.1$, hypothesis f, expectation $e_{G,0.02}$, density estimator $\hat{f}_{G,50}$ and random numbers x_{50i}

References

[1] Bickel, P. J./Rosenblatt, M. (1973): *On some Global Measures of the Deviations of Density Function Estimates.* Ann. Statist. **1**, 1071–1095.

[2] Bratley, P./Fox, B. L./Schrage, E. L. (1983): *A Guide to Simulation.* New York: Springer.

[3] Ghosh, B. K./Huang, W.-M. (1991): *The Power and Optimal Kernel of the Bickel-Rosenblatt Test for Goodness of Fit.* Ann. Statist. **19**, 999–1009.

[4] Devroye, L./Györfi,L. (1985): *Nonparametric Density Estimation: The L_1 View.* New York: Wiley.

[5] Liero, H./Läuter, H./Konakov, V. (1998): *Nonparametric versus parametric goodness of fit.* Statistics **31**, 115–149.

[6] Rio, E. (1994): *Local invariance principles and their application to density estimation.* Prob. Theory Rel. Fields, 21–45.

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