Mathematics

On Some Goodness-of-Fit Tests Based on Kernel Type Wolverton-Wagner Estimates

Elizbar Nadaraya*, Petre Babilua **, Grigol Sokhadze***

* Academy Member, I. Javakhishvili Tbilisi State University
** I. Javakhishvili Tbilisi State University
*** A. Tsereteli State University, Kutaisi

ABSTRACT. A goodness-of-fit test is constructed by using a Wolverton-Wagner distribution density estimate. The question as to its consistency is studied. The power asymptotics of the constructed goodness-of-fit test is also studied for certain types of close alternatives. © 2009 Bull. Georg. Natl. Acad. Sci.

Key words: distribution density, goodness-of-fit test, power, consistency, limit distribution.

1. Let $X_1, X_2, \ldots, X_n$ be a sequence of independent, equally distributed random values having a distribution density $f(x)$. Using the sampling $X_1, X_2, \ldots, X_n$, it is required to test the hypothesis $H_0: f(x) = f_0(x)$.

Here we consider the test of the hypothesis $H_0$, based on the statistics

$U_n = n a_n^{-1} \int \left( f_n(x) - f_0(x) \right)^2 r(x) dx,$

where $f_n(x)$ is the recurrent Wolverton-Wagner kernel estimate of the probability density defined by

$f_n(x) = n^{-1} \sum_{i=1}^{n} a_i K(a_i(x - X_i)),$

where $\{a_i\}$ is an increasing sequence of positive integers tending to infinity, $r(x) \in \mathbb{R}$ ($\mathbb{R}$ is the set of non-negative, bounded and piecewise-continuous functions on $(-\infty, \infty)$),

$K(x) \in H = \left\{ h : h(x) \geq 0, \sup_{x \in (-\infty, \infty)} h(x) < \infty, \int h(x) dx = 1, x^2 h(x) \in L_4(-\infty, \infty), h_0(ux) \geq h_0(x) \text{ for all } u \in [0,1] \text{ and } \text{ for all } x \in (-\infty, \infty), h_0 = h \ast h, \ast - \text{ convolution operator} \right\}$

Notation.

$L_n = n \int \left( f_n(x) - Ef_n(x) \right)^2 r(x) dx.$


The following statement is true.

**Theorem 1 ([3], [4]).**

a) Let $K \in H$ and $f_0(x)$ be bounded. If $n^{-1}a_n \to 0$ and $a_n^{-s} \gamma_s(n) \to \gamma_s$, $s = 1, 2$ 
$(0 < \gamma_2 < \gamma_1 \leq 1)$ as $n \to \infty$, then

$$
\gamma_1^2 \sigma_n^2 \lesssim \lim_{n \to \infty} \sigma_n^2(f_0) \lesssim \lim_{n \to \infty} \sigma_n^2(f_0) \lesssim \gamma_1 \sigma_0^2
$$

(1)

and

$$
da_n^{-1}(L_n - EL_n) \xrightarrow{D} N(0, 1);
$$

b) Let $K \in H$ and $f_0(x) \in F$ ($F$ is the set of densities having bounded derivatives up to second order), the second order derivative of the density $f_0(x)$ being square integrable. If $n^{-1}a_n \to 0$, $a_n^{-s} \gamma_s(n) \to \gamma_s$, $s = 1, 2$, 
$n^{-1}a_n^{-1/2} \left( \sum_{i=1}^{n} a_i^{-2} \right) \to 0$ as $n \to \infty$, then

$$
da_n^{-1/2} \sigma_n^{-1}(f_0(U_n - \theta_n)) \xrightarrow{D} N(0, 1);
$$

c) Let $K \in H$ and $f_0(x) \in F$, the second derivative of the density $f_0(x)$ being integrable. If to the condition b) for $a_n$ and $\gamma_n$ we add the condition $a_n^{-1} \gamma_1(n) = \gamma_1 + o(a_n^{-1/2})$, then

$$
da_n^{-1/2} \sigma_n^{-1}(f_0(U_n - \Delta f_0)) \xrightarrow{D} N(0, 1),
$$

where $D$ denotes convergence in distribution and $N(0, 1)$ is a random value having a standard normal distribution $\Phi(x)$.

**Remark 1.** If we assume that $r(x) \in R \cap L_1$, then the conditions b) and c) on $f_0(x)$ can be replaced by the only requirement that $f_0(x) \in F$.

It turns out that if $a_n = n^\delta$ and some additional condition is imposed on $K(x)$, then we may succeed in defining

$$
\lim_{n \to \infty} \sigma_n^2(f_0).
$$

Therefore the following statement is true.

**Lemma.** Let $a_n = n^\delta$, $0 < \delta < 1$, $K_0(x)$ have a bounded derivative and $x^2 K_0^{(1)}(x) \in L_1(-\infty, \infty)$. Then we have
\[
\sigma_n^2(f_0) \to \sigma^2(f_0) = 2 \int f_0^2(x)\gamma^2(x)dx \left( \int_0^1 u^\delta K_0(u^{\delta} z)du \right)^2 dz.
\] (2)

as \( n \to \infty \).

**Proof.** It is not difficult to verify that

\[
\sigma_n^2(f_0) = 2 \int f_0^2(x)\gamma(x)dx \left( \frac{1}{n} \sum_{k=1}^{n} \left( \frac{k}{n} \right)^\delta \right) K_0 \left( \frac{k}{n} z \right) \left( x - \frac{z}{a_n} \right) dz.
\]

Using Euler’s summation formula [5]:

\[
\sum_{v=n_1}^{n_2} g(a+hv) = \int_{n_1}^{n_2} g(a+hx)dx + \frac{1}{2} g(a+n_1h) + \frac{1}{2} g(a+n_2h) - h \int_{n_1}^{n_2} P_1(x)g'(x)dx, \quad |P_1(x)| < \frac{1}{2}, \quad x \in (-\infty, \infty),
\]

we insert \( g(x) = x^\delta K_0(x^{\delta} z), \quad a = 0, \quad h = 1/n, \quad n_1 = 1, \quad n_2 = n \), and we obtain

\[
\frac{1}{n} \sum_{k=1}^{n} \left( \frac{k}{n} \right)^\delta K_0 \left( \frac{k}{n} z \right) = \int x^\delta K_0(x^{\delta} z)dx + \frac{1}{2n} \int \left( \frac{1}{n} \right)^\delta K_0 \left( \frac{1}{n} z \right) + \frac{1}{2n} K_0(z) -
\]

\[
- h \int_{n_1}^{n_2} P_1(\nu x)g'(x)dx, \quad |P_1(x)| < \frac{1}{2}, \quad x \in (-\infty, \infty).
\]

Hence, after a few simple estimations, we have

\[
\left( \int_{1/n}^{n_1} \left( \frac{k}{n} \right)^\delta K_0 \left( \frac{k}{n} z \right) r \left( x - \frac{z}{a_n} \right) dz \right)^2 = \int \left( \int_{1/n}^{1} u^\delta K_0(u^{\delta} z)du \right)^2 r \left( x - \frac{z}{a_n} \right) dz + O \left( \frac{\ln n}{n} \right).
\]

Further, by the Lebesgue theorem on passage to the limit under the integral sign, we can write that

\[
\int \left( \int_{1/n}^{1} u^\delta K_0(u^{\delta} z)du \right)^2 r \left( x - \frac{z}{a_n} \right) dz \to \int \left( \int_0^1 u^\delta K_0(u^{\delta} z)du \right)^2 r(x)dz
\]

at \( n \to \infty \). Therefore \( \sigma_n^2(f_0) \to \sigma^2(f_0) \).

**Remark 2.** It is obvious that in the conditions b) and c) of Theorem 1, \( \sigma_n^2(f_0) \) can be replaced by the value \( \sigma^2(f_0) \).

The condition c) of Theorem 1 enables us to construct a test of asymptotic level \( \alpha \), \( 0 < \alpha < 1 \), for checking the hypothesis \( H_0 \) according to which \( f(x) = f_0(x) \). To this end, we should calculate \( U_n \) and repudiate the hypothesis \( H_0 \) when

\[
U_n \geq \lambda_n(\alpha) = \Delta(f_0) + \varepsilon_{\alpha} a_n^{-1/2} \sigma(f_0),
\] (3)

where \( \varepsilon_{\alpha} \) is defined by the equality \( \Phi(\varepsilon_{\alpha}) = 1 - \alpha \).

2. Let us now investigate the asymptotic property of test (3) (i.e. the behavior of the power at *). In the first place, we have to consider the question whether the test is consistent.
Theorem 2. Let \( K(x) \in H \), \( K_n(x) \) satisfy the conditions of the lemma and \( f_0(x) \) the conditions c) of Theorem 1 or \( r(x) \in R \cap L_1 \). If \( 2/9 < \delta \leq 1/2 \), then

\[
\Pi_n(f_1) = P_{H_n}(U_n \geq \lambda_n(\alpha)) \to 1
\]
as \( n \to \infty \), i.e. the test defined by (3) is consistent against any alternative \( H_1 : f(x) = f_1(x) \).

Proof. It is not difficult to check that

\[
\Pi_n(f_1) = P_{H_1} \left\{ a_n^{-1/2} \sigma^{-1}(f_1) \left( U_n^{(l)} - \Delta(f_1) \right) \geq \epsilon_n \frac{\sigma(f_0)}{\sigma(f_1)} + \sigma^{-1}(f_1) \left( \Delta(f_0) - \Delta(f_1) \right) a_n^{1/2} + 2n \right\} + \frac{1}{n} \sigma^{-1}(f_1) \left[ \int (f_n(x) - f_1(x)) \phi(x) dx - n \sigma^{-1}(f_0) \int (f_1(x) - f_0(x)) r(x) dx \right] =
\]

\[
= P_{H_1} \left\{ a_n^{-1/2} \sigma^{-1}(f_1) \left( U_n^{(l)} - \Delta(f_1) \right) \geq - \sqrt{\frac{n}{\sqrt{2}}} \left( \sigma^{-1}(f_1) \Delta + O \left( \sqrt{\frac{\lambda_n}{n}} \right) + 2 \sigma^{-1}(f_1) \int (f_n(x) - f_1(x)) \phi(x) dx \right) \right\},
\]
where

\[
U_n^{(l)} = n a_n^{-1} \int (f_n(x) - f_1(x))^2 r(x) dx, \quad \phi(x) = (f_1(x) - f_0(x)) r(x).
\]

Now let us show that

\[
\int (f_n(x) - f_1(x)) \phi(x) dx \to 0.
\]

Indeed,

\[
\int (f_n(x) - f_1(x)) \phi(x) dx = B_{n1} + B_{n2},
\]

\[
B_{n1} = \sum_{j=1}^{n} \xi_j, \quad B_{n2} = \int \left( \frac{1}{n} \sum_{j=1}^{n} a_j EK(a_j (x - X_j)) - f_1(x) \right) \phi(x) dx,
\]

\[
\xi_j = n^{-1} a_j \int [K(a_j (x - X_j)) - EK(a_j (x - X_j))] \phi(x) dx,
\]

where \( E(\cdot) \) is the matrix expectation under \( H_1 \).

By virtue of the assumption as regards \( f_1(x) \) and \( K(x) \) we obtain

\[
B_{n2} = O \left( \frac{1}{n} \sum_{j=1}^{n} a_j^{-2} \right).
\]

Therefore

\[
B_{n2} = O(n^{-2\delta}), \quad \delta < \frac{1}{2}, \quad B_{n2} = O \left( \frac{\ln n}{n} \right), \quad \delta = \frac{1}{2}.
\]

Further it can be easily checked that

\[
\text{Var}B_{n1} = \sum_{j=1}^{n} \text{Var} \xi_j \sim n^{-1} \left[ \int \phi^2(x) f_1(x) dx - \left( \int \phi(x) f_1(x) dx \right)^2 \right]
\]
and
\[ \sum_{k=1}^{n} E \left[ \frac{x_k - M_n x_k}{n \text{Var}_n x_k} \right] x_k^{2+\alpha} \left( \sum_{k=1}^{n} \text{Var}_n x_k \right)^{-1-\alpha/2} \to 0, \ \alpha > 0. \] (5)

Hence
\[ (\text{Var}B_n)^{-1/2} B_n \xrightarrow{D} N(0,1). \] (6)

Also,
\[ \int (f_n(x) - f_1(x)) \varphi(x) dx = \left[ \frac{B_n}{\sqrt{\text{Var}B_n}} \right] \text{Var}B_n + O \left( \frac{1}{n} \sum_{k=1}^{n} a_k^2 \right). \]

This and (4), (5), (6) imply that
\[ \Pi_n(f_1) = P_{H_1} \left[ a_n^{1/2} \sigma^{-1} \left( f_1 \right) (U_n^{(i)} - \Delta(f_1)) \geq - \frac{n}{\sqrt{a_n}} \left( \sigma^{-1}(f_1) \Delta + o_p(1) \right) \right]. \]

Since
\[ a_n^{1/2} \sigma^{-1}(f_1)(U_n^{(i)} - \Delta(f_1)) \xrightarrow{D} N(0,1) \]
for the hypothesis \( H_1 \) and \( n/\sqrt{a_n} \to \infty \), we have \( \Pi_n(f_1) \to 1 \) as \( n \to \infty \).

Thus for any fixed alternative the power of the test based on \( U_n \) tends to 1 as \( n \to \infty \). If however the alternative changes with a change of \( n \) and approaches the basic hypothesis \( H_0 \), then the power of the test will not necessarily converge to 1. Consider, for example, the sequence of Pitman type alternatives close to the hypothesis \( H_0 \)
\[ H_1 : f_1(x) = f_0(x) + \alpha_n \varphi(x) + o(\alpha_n), \ \alpha_n \downarrow 0, \ \int \varphi(x) dx = 0. \] (7)

**Theorem 3.** Let \( K(x) \) and \( K_0(x) \) satisfy the conditions of Theorem 2 and \( f_0(x) \) and \( \varphi(x) \) the conditions c) of Theorem 1. If \( a_n = n^{-\delta} \) and for (7) it can be assumed that \( \alpha_n = n^{-1/2 + \delta/4}, \ 2/9 < \delta \leq 1/2 \), then
\[ P_{H_1} \left[ U_n \geq \lambda_n(\alpha) \right] \to \gamma(U) = 1 - \Phi \left( \epsilon_n - \frac{1}{\sigma(f_0)} \int \phi^2(u) r(u) du \right) \]
as \( n \to \infty \).

**Proof.** We have
\[ P_{H_1} \left[ U_n \geq \lambda_n(\alpha) \right] = \]
\[ = P_{H_1} \left[ d_n^{-1} (L_n - EL_n) \geq \sqrt{a_n} \sigma_f(f_0) \left( \alpha_n - d_n^{-1} EL_n + \sqrt{a_n} \sigma_n^{-1}(f_0) A_n - \sqrt{a_n} \sigma_n^{-1}(f_0) A_{n2} \right) \right], \]
where
\[ A_n = 2n a_n^{-1} \int (f_n(x) - Ef_n(x)) [Ef_n(x) - f_0(x)] r(x) dx, \]
\[ A_{n2} = n a_n^{-1} \int [Ef_n(x) - f_0(x)]^2 r(x) dx. \]

By virtue of the Minkowski integral inequality and (1) it is not difficult to establish that
\[ a_n^{1/2} EL_n = n^{-1} d_n^{-1} \sum_{k=1}^{n} a_k^2 \int EK^2(\alpha_k(x - X_k)) r(x) dx + O(a_n^{-1/2}). \] (9)
Further we have
\[ \frac{\gamma_s(n)}{a_n} = \frac{1}{1 + s\delta} + o(n^{-\delta/2}), \quad s = 1, 2. \]

Therefore (9) readily yields
\[ a_n^{-1}E\lambda_n = n^{\delta/2}\sigma_n^{-1}(f_0)\Delta(f_0) + o(\sigma_n^{-\delta/2}) + O\left(n^{\frac{1}{2} + \frac{3}{4}\delta}\right), \quad \frac{2}{9} < \delta \leq \frac{1}{2}. \]  

(10)

Thus
\[ \sqrt{a_n\sigma_n^{-1}(f_0)}\lambda_n(\alpha) - a_n^{-1}E\lambda_n = \varepsilon_n + o(\sigma_n^{-\delta/2}) + O\left(n^{\frac{1}{2} + \frac{3}{4}\delta}\right) \rightarrow \varepsilon_n \cdot \]  

(11)

Furthermore,
\[ Ef_n(x) - f_0(x) = \alpha_n \phi(x) + \frac{1}{n} \sum_{k=1}^{n} a_k^{-2} \int u^2 K(u) du \int (1-t)f_1^2(x - \frac{tu}{a_k}) \, dt. \]  

(12)

Hence it follows that
\[ \sqrt{a_n\sigma_n^{-1}(f_0)}A_n = \frac{1}{\sigma_n(f_0)} \int \phi^2(x) r(x) \, dx + R_n, \]
\[ R_n = O\left(\frac{\varepsilon_n}{n^{1/4}}\right), \quad \frac{2}{9} < \delta < \frac{1}{2}, \quad R_n = n^{-5/8} \ln n, \quad \text{if} \quad \delta = \frac{1}{2}. \]  

(13)

Now let estimate \( \sqrt{a_n\sigma_n^{-1}(f_0)}E[A_n]. \) Using (12) and the inequality
\[ E[A_n] \leq 2 \frac{n}{a_n} \left\{ n^{-2} \sum_{k=1}^{n} a_k^2 f_1(u) du \left\{ K(a_k(x-u))(Ef_n(x) - f_0(x)) dx \right\}^{1/2} \right\}, \]

which is easy to check, we find
\[ \sqrt{a_n\sigma_n^{-1}(f_0)}E[A_n] \leq R_1(n) + R_2(n), \]
\[ R_1(n) = O\left(\frac{\varepsilon_n}{\sqrt{a_n}}\right) = O(n^{-\delta/2}), \quad R_2(n) = O\left(\frac{1}{\sqrt{a_n}} \sum_{k=1}^{n} a_k^{-2}\right), \]
\[ R_1(n) = O\left(\frac{1-5\delta}{n^{1/2}}\right), \quad \frac{2}{9} < \delta < \frac{1}{2}, \]
\[ R_2(n) = n^{-3/4} \ln n, \quad \text{if} \quad \delta = \frac{1}{2}. \]  

(14)
After combining relations (11), (13), (14) and taking into account $d_n^{-1}(L_n - EL_n) \xrightarrow{D} N(0,1)$ which follows from the case a) of Theorem 1, from (8) we eventually obtain the assertion of Theorem 3.

The Rosenblatt-Bickel test ([6], [7])

$$T_n = \int f_0(x)r(x)dx \left[ K^2(u)du + na^{-\frac{1}{2}}_n e_{\alpha}\sigma_0 \right],$$

$$T_n = na^{-1}_n \int (f_n(x) - f_0(x))^2 r(x)dx,$$

$$f_n(x) = n^{-1} \sum_{i=1}^n K(a_n(x - X_i)),$$

used for checking the hypothesis $H_0$, is well known. The limiting power of this test for the alternative $H_1$ is equal to

$$\gamma(T) = 1 - \Phi \left( \frac{\sigma_0}{\sigma} \right),$$

$$\sigma_0^2 = 2 \int f_0^2(x)r^2(x)dx \int K_0^2(u)du.$$

From inequality (1) and relation (2) it follows that

$$\sigma^2(f_0) \leq \gamma_1 \sigma_0^2 < \sigma_0^2.$$

$$\gamma_1 = \lim_{n \to \infty} \frac{\gamma(n)}{n^\delta} = \frac{1}{1 + \delta}.$$

Therefore, $\gamma(U) > \gamma(T)$, i.e. the tests based on $U_n$ for the alternative $H_1$ are more powerful in the limit than the tests based on $T_n$. Furthermore, it can be easily noticed that (16) can be immediately obtained from

$$\left( \int_0^1 u^\delta K_0(u^\delta z)du \right)^2$$

by using the Cauchy-Schwartz inequality.

Acknowledgement

This work was supported by the GNSF, Grant No. GNSF/PRES08/3-319.
18

Elizbar Nadaraya, Petre Babilua, Grigol Sokhadze

შეჯიბრილი თანახმად კრიტერიუმთა დაფუძნებლად
საშუალებით კვარტალ-კვარტალის მიღწევის სქემაცხება

ე. ნადარაია*, ჰ. ბაბილუა**, გ. სოკხაძე***

* აკადემიის წევრი, ა. ჯაროდისას სახ. ოფიციალი საქართველო
** ა. ჯაროდისას სახ. ოფიციალი საქართველო
*** ჰ. ჭარილის სახ.ღალათი საქართველო, ჭარილი

ამომრჩევით თანახმად კრიტერიუმთა დაფუძნებლად კვარტალ-კვარტალის საშუალებით სქემაცხება.
შეჯიბრილი თანახმად კვარტალ-კვარტალის მიღწევის სქემაცხება.

REFERENCES

5. G. Kramer (1975), Matematicheskie metody statistiki, Moscow (in Russian).

Received May, 2009