ON THE POWER OF THE GOODNESS-OF-FIT TEST BASED ON WOLVERTON-WAGNER DENSITY ESTIMATES

Elizbar Nadaraya[1] and Petre Babilua[2]

I. Javakhishvili Tbilisi State University, Tbilisi, Georgia
[1]*e.nadaraya@math.sci.tsu.ge*, [2]*p_babilua@yahoo.com*

1. Let $X_1, X_2, \dots, X_n$ be a sequence of indepenedent, equally distributed random variables with values in a Euclidean *p*-dimensional space R_p, $p \ge 1$, which have the distribution density $f(x)$, $x = (x_1, \dots, x_p)$. Using the sampling $X_1, X_2, \dots, X_n$, it is required to test the hypothesis

$$H_0 : f = f_0 .$$

We consider the verification test of the hypothesis H_0, based on the statistics

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

where $f_n(x)$ is a kernel estimate of the Wolverton-Wagner probability density,

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

$a_1, a_2, \dots, a_n, \dots$ is a sequence of positive numbers monotonically converging to infinity.

Let us formulate the conjectures concerning $K(x)$ and $f(x)$.

1⁰. The kernel $K(x) = \prod_{j=1}^{p} K_j(x_j)$ and every kernels $K_j(u)$ possesses the following properties:

$$0 \le K_j(u) \le c < \infty, \quad K_j(u) = K_j(-u), \quad u^2 K_j(u) \in L_1(-\infty, \infty),$$

$$\int K_j(u) du = 1, \quad K_j^0(ux) \ge K_j^0(x) \quad \text{for all} \quad u \in [0,1] \quad \text{and all} \quad x \in R_1 = (-\infty, \infty),$$

where $K_j^0 = K_j * K_j$; $*$ is the convolution operator.

2⁰. The distribution density $f(x)$ is bounded and has bounded partial derivatives up to second order.

We have proved the following theorem ([1]-[3]).

Theorem 1. *Let* $K(x)$ *and* $f(x)$ *satisfy conjectures* 1^0 *and* 2^0*, respectively, and, besides, let the second order partial derivatives of the function* $f(x)$ *belong to* $L_1(R_p)$*. If*

$$\frac{a_n^p}{n} \to 0, \quad \frac{\gamma_s(n)}{a_n^{sp}} \to \gamma_s, \quad s = 1,2 \quad (0 < \gamma_2 \le \gamma_1 \le 1)$$

and also

$$\frac{\gamma_1(n)}{a_n^p} = \gamma_1 + o\left(a_n^{-p/2}\right), \quad \left(n a_n^{p/2}\right)^{-1} \sum_{k=1}^{n} a_k^{p-2} \to 0$$

and

$$\left(na_n^{p/2}\right)^{-1}\left(\sum_{j=1}^{n} a_j^{-2}\right)^2 \to 0 \quad as \quad n \to \infty ,$$

then

$$a_n^{p/2}\sigma_n^{-1}(f_0)(U_n - \Theta(f_0)) \xrightarrow{d} N(0,1),$$

$$\gamma_s(n) = \frac{1}{n}\sum_{i=1}^{n} a_i^{ps}, \quad s = 1,2, \quad \Theta(f_0) = \gamma_1 \int f(x) r(x) dx \int K^2(u) du,$$

$$\sigma_n^2(f_0) = a_n^{-p} d_n^2(f_0),$$

$$d_n^2(f_0) = \frac{2}{n^2}\iint f_0^2(x)\left(\sum_{i=1}^{n} a_i^p K_0(a_i(x-y))\right)^2 r(x) r(y) dx dy, \quad K_0 = K * K.$$

Also,

$$2\gamma_1^2 \int f_0^2(x) r^2(x) dx \int K_0^2(u) du \le$$
$$\le \varliminf_{n\to\infty} \sigma_n^2(f_0) \le \varlimsup_{n\to\infty} \sigma_n^2(f_0) \le 2\gamma_1 \int f_0^2(x) r^2(x) dx \int K_0^2(u) du . \quad (1)$$

Theorem 1 allows us to construct a goodness-of-fit test of asymptotic level α for checking the hypothesis H_0 by which $f(x) = f_0(x)$. After that we calculate U_n and reject the hypothesis H_0 if

$$U_n \ge d_n(\alpha) = \gamma_1 \int f_0(x) r(x) dx \int K^2(u) du + \lambda_\alpha a_n^{-p/2} \sigma_n(f_0), \quad (2)$$

where λ_α is a quantile of level α of a standard normal distribution.

2. Now we will investigate the asymptotic property of test (2) (i.e. the behavior of the power function as $n \to \infty$).

The following statement is true.

Theorem 2. *Let all the conditions of Theorem* 1 *be fulfilled. Then*

$$\Pi_n(f_1) = P_{H_1}\{U_n \ge d_n(\alpha)\} \to 1$$

as $n \to \infty$ *, i.e. the test defined in* (2) *is consistent against any alternative* $H_1: \ f_1(x) \ne f_0(x)$,

$$\Delta = \int (f_1(x) - f_0(x))^2 r(x) dx > 0 .$$

Thus for any fixed alternative the power of a test based on U_n tends to 1 as $n \to \infty$. However, if with a change of *n* the alternative changes so that it tends to the basic hypothesis H_0, then the power of the goodness-of-fit test will not necessarily tend to 1. This certainly depends on the convergence of the tendency of the alternative to the zero hypothesis.

Let us assume now that the hypothesis H_0 we are checking is not true; actually, we have the hypothesis

$$H_n: \quad f_{1n}(x) = f_0(x) + \gamma_n \varphi(x) + o(\gamma_n), \quad \gamma_n \downarrow 0, \quad x \in R_1 = (-\infty, \infty), \quad \int \varphi(x) dx = 0 .$$

Theorem 3. *Let* $K(x)$ *and* $f_{1n}(x)$ *satisfy the conditions of Theorem* 1. *If* $a_n = n^\delta$ *and* $\gamma_n = n^{-1/2+\delta/4}$, $\frac{2}{9} < \delta < \frac{1}{2}$, *then*

$$P_{H_n}\{U_n \geq d_n(\alpha)\} \sim 1-\Phi\left(\lambda_\alpha - \frac{1}{\sigma_n(f_0)}\int \varphi^2(u)r(u)du\right),$$

$$\Phi(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}\exp\left\{-\frac{x^2}{2}\right\}dx$$

as $n \to \infty$.

Remark 1. If the smoothness parameter $a_k \equiv a_n$, $k=1,2,\ldots$, and $p=1$, then by Theorem 2 we obtain an analogous limit power function for the alternative H_n of the Rosenblatt-Bickel goodness-of-fit test [4]:

$$P_{H_n}\{T_n \geq d_n^{(1)}(\alpha)\} \sim 1-\Phi\left(\lambda_\alpha - \frac{1}{\sigma_0}\int \varphi^2(x)r(x)dx\right),$$

$$T_n = \frac{n}{a_n}\int (f_n(x)-f_0(x))^2 r(x)dx,$$

where $f_n(x)$ is an estimate of the Rosenblatt-Parzen density,

$$d_n^{(1)}(\alpha)=\int f_0(x)r(x)dx\int K^2(u)du + a_n^{-1/2}\lambda_\alpha\sigma_0,$$

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

If $\lim\limits_{n\to\infty}\sigma_n^2(f_0)=\sigma_1^2$, then inequality (1) implies that $\gamma_1^2\sigma_0^2 \leq \sigma_1^2 \leq \gamma_1\sigma_0^2$, $0<\gamma_1\leq 1$. Therefore, for $\gamma_1 \neq 1$, goodness-of-fit tests based on U_n are more powerful than this based on T_n, moreover, they are asymptotically equivalent for $\gamma_1 = 1$.

3. Let us now introduce into the consideration the alternatives we call "singular" ([5]-[7]):

$$H_n : f_{1n}(x) = f_0(x) + \alpha_n\varphi\left(\frac{x-l}{\gamma_n}\right) + o(\alpha_n\cdot\gamma_n),$$

where $\alpha_n \downarrow 0$, $\gamma_n \downarrow 0$, the function $\varphi(x)$ is bounded and has bounded derivatives up to second order, $\varphi^{(2)}(x)\in L_1(-\infty,\infty)$ and $\int\varphi(x)dx=0$, *l* is some continuity point of $r(x)$ such that $r(l)\neq 0$.

Theorem 4. *Let* $K(x)$ *and* $f_{1n}(x)$ *satisfy the conditions of Theorem* 1. *If*

$$\alpha_n\cdot\gamma_n = o(n^{-1/2}),\quad n a_n^{-1/2}\alpha_n^2\gamma_n \to c_0 \neq 0,\quad \frac{1}{\alpha_n}\frac{1}{n}\sum_{k=1}^{n}a_k^{-2}\to 0\ ,$$

and

$$\gamma_n^{-2}\left(\frac{1}{n}\sum_{k=1}^{n}a_k^{-4}\right)^{1/2}\to 0,$$

then

$$P_{H_n}\{U_n \geq d_n(\alpha)\} \sim 1-\Phi\left(\lambda_\alpha - \frac{c_0 r(l)}{\sigma_n(f_0)}\int \varphi^2(u)du\right)$$

as $n \to \infty$.

The conditions of Theorem 4 for a_n, α_n, γ_n are fulfilled if, say, we assume that

$$a_n = n^{\delta},\ \alpha_n = n^{-\alpha},\ \gamma_n = n^{-\beta}$$

for $\frac{\delta}{2} = 1 - 2\alpha - \beta$, $\alpha + \beta > \frac{1}{2}$, $0 < \delta \le \frac{1}{2}$, $\beta < \delta$, $\delta > \frac{\alpha}{2}$,

while the conditions imposed on α, β and δ are fulfilled, for example, for

$$\delta = \frac{1}{2},\quad \beta = \frac{5}{12},\quad \alpha = \frac{1}{6};\qquad \delta = \frac{1}{4},\quad \beta = \frac{1}{5},\quad \alpha = \frac{27}{80};\qquad \delta = \frac{1}{5},\quad \beta = \frac{1}{6},\quad \alpha = \frac{11}{30}$$

and so on.

Remark 2. If the smoothness parameter $a_k \equiv a_n$, $k = 1,2,\ldots$, then by Theorem 4 we obtain the limit power for alternative H_n of the goodness-of-fit test T_n ([5]-[7])

$$P_{H_n}\left\{T_n \ge d_n^{(1)}(\alpha)\right\} \sim 1 - \Phi\left(\lambda_{\alpha} - \frac{c_0 r(l)}{\sigma_0}\int \varphi^2(u)du\right).$$

If $\lim_{n\to\infty}\sigma_n^2(f_0) = \sigma_1^2$, then from inequality (1) we have $\gamma_1^2\sigma_0^2 \le \sigma_1^2 \le \gamma_1\sigma_0^2$, $0 < \gamma_1 \le 1$. Therefore, for $\gamma_1 \ne 1$, goodness-of-fit tests based on U_n are more powerful than those based on T_n. Moreover, integrating $f_{1n}(x)$, we establish that the alternatives differ from the zero hypothesis by a value of order $\alpha_n\gamma_n = o\left(\frac{1}{\sqrt{n}}\right)$. Therefore goodness-of-fit tests based on a difference between empirical distribution functions like for example, tests ω_n^2 and Kholmogorov-Smirnov type tests cannot differentiate between "singular" hypotheses and the basic one. Hence, by virtue of Theorem 4, for "singular" alternatives the tests based on U_n are more powerful than those of the type mentioned above.

Acknowledgement

This work was supported by the GNSF Grant No. GNSF/ST06/3-039.

References

1. E. Nadaraya, P. Babilua, On the Wolverton-Wagner estimate of a distribution density. *Bull. Georgian National Academy Sci.* **175** (2007), No. 1, 40-44.
2. E. Nadaraya, P. Babilua, On Deheuvels' nonparametric estimate of the distribution density. *Bull. Georgian National Academy Sci.* **175** (2007), No. 2, 15-19.
3. E. Nadaraya, P. Babilua, On the limit distribution of a quadratic deviation of a recurrent kernel estimate of distribution density. *Proc. A. Razmadze Math. Inst.* **144** (2007), 67-89.
4. P. J. Bickel, M. Rozenblatt, On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** (1973), 1071-1095.
5. E. Nadaraya, On some tests based on "kernel"-type estimates of a probability density. (Russian) *Soobshch. Akad. Nauk Gruzin. SSR* **99** (1980), No. 1, 37-40.
6. E. Nadaraya, Nonparametric estimation of probability densities and regression curves. Kluwer Academic Publishers, 1989.
7. M. Rosenblatt, A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3** (1975), 1-14.