Reports of Enlarged Session of the
Seminar of I. Vekua Institute
of Applied Mathematics
Volume 28, 2014

ON INTEGRAL SQUARE DEVIATION OF TWO KERNEL ESTIMATORS OF BERNOULLI REGRESSION FUNCTIONS

Nadaraya E., Babilua P., Sokhadze G.

Abstract. The limiting distribution of an integral square deviation between two kernel type estimators of Bernoulli regression functions is established in the case of two independent samples. The criterion of testing is constructed for both simple and composite hypotheses of equality of two Bernoulli regression functions. The question of consistency is studied. The asymptotics of behavior of the power of test is investigated for some close alternatives.

Keywords and phrases: Bernoulli regression function, power of test, consistency, composite hypothesis.

AMS subject classification: 60F05, 62G05, 62G10, 62G20.

Let random variables $Y^{(i)}$, $i=1,2$, take two values 1 and 0 with probabilities p_i (success) and $1-p_i$, $i=1,2$ (failure), respectively. Assume that the probability of success p_i is the function of an independent variable $x\in[0,1]$, i.e. $p_i=p_i(x)=\mathbb{P}\{Y^{(i)}=1\mid x\}$ $(i=1,2)$ (see [1–3]). Let t_j, $j=1,\dots,n$, be the devision points of the interval $[0,1]$:

$$t_j=\frac{2j-1}{2n}\,,\ \ j=1,\dots,n.$$

Let further $Y_i^{(1)}$ and $Y_i^{(2)}$, $i=1,\dots,n$, be mutually independent random Bernoulli variables with $\mathbb{P}\{Y_i^{(k)}=1\mid t_i\}=p_k(t_i)$, $\mathbb{P}\{Y_i^{(k)}=0\mid t_i\}=1-p_k(t_i)$, $i=1,\dots,n$, $k=1,2$. Using the samples $Y_1^{(1)},\dots,Y_n^{(1)}$ and $Y_1^{(2)},\dots,Y_n^{(2)}$ we want to test the hypothesis

$$H_0: p_1(x)=p_2(x)=p(x),\ \ x\in[0,1],$$

against the sequence of "close" alternatives of the form

$$H_{1n}: p_k(x)=p(x)+\alpha_n u_k(x)+o(\alpha_n),\ \ k=1,2,$$

where $\alpha_n\to0$ relevantly, $u_1(x)\neq u_2(x)$, $x\in[0,1]$ and $o(\alpha_n)$ uniformly in $x\in[0,1]$.

The problem of comparing two Bernoulli regression functions arises in some applications, for example, in quantal bioassays in pharmacology. There x denotes the dose of a drug and $p(x)$ the probability of response to the dose x.

We consider the criterion of testing the hypothesis H_0 based on the statistic

$$T_n=\frac12\,nb_n\int_{\Omega_n(\tau)}[\widehat p_{1n}(x)-\widehat p_{2n}(x)]^2p_n^2(x)\,dx=\frac12\,nb_n\int_{\Omega_n(\tau)}[p_{1n}(x)-p_{2n}(x)]^2\,dx,$$
$$\Omega_n(\tau)=[\tau b_n,(1-\tau)b_n],\ \ \tau>0,$$

where

$$\widehat p_{in}(x)=p_{in}(x)p_n^{-1}(x),$$

$$p_{in}(x) = \frac{1}{nb_n}\sum_{j=1}^{n} K\Big(\frac{x-t_j}{b_n}\Big)Y_j^{(i)}, \quad i=1,2, \quad p_n(x) = \frac{1}{nb_n}\sum_{i=1}^{n} K\Big(\frac{x-t_i}{b_n}\Big),$$

$K(x)$ is some distribution density and $b_n \to 0$ is a sequence of positive numbers, $\widehat{p}_{in}(x)$ is the kernel estimator of the regression function (see [4, 5]).

We assume that a kernel $K(x) \geq 0$ is chosen so that it is a function of bounded variation and satisfies the conditions:

$$K(x) = K(-x), \;\; K(x) = 0 \;\text{ for }\; |x| \geq r > 0, \quad \int K(x)\,dx = 1.$$

The class of such functions is denoted by $H(\tau)$.

Lemma 1 ([6]). *Let $K(x) \in H(\tau)$ and $p(x)$, $x \in [0,1]$, be a function of bounded variation. If $nb_n \to \infty$, then*

$$\frac{1}{nb_n}\sum_{i=1}^{n} K^{\nu_1}\Big(\frac{x-t_i}{b_n}\Big)K^{\nu_2}\Big(\frac{y-t_i}{b_n}\Big)p^{\nu_3}(t_i) =$$
$$= \frac{1}{b_n}\int_0^1 K^{\nu_1}\Big(\frac{x-u}{b_n}\Big)K^{\nu_2}\Big(\frac{y-u}{b_n}\Big)p^{\nu_3}(u)\,du + O\Big(\frac{1}{nb_n}\Big)$$

uniformly in $x, y \in [0,1]$, where $\nu_i \in N \cup \{0\}$, $i = 1,2,3$.

Lemma 2. *Let $K(x) \in H(\tau)$ and $p(x) \in C^1[0,1]$ and $u_1(x)$, $u_2(x)$ be continuous functions on $[0,1]$. If $nb_n^2 \to \infty$ and $\alpha_n b_n^{-1/2} \to 0$, then for the hypothesis H_{1n}*

$$b_n^{-1}\sigma_n^2 \longrightarrow \sigma^2(p) = 2\int_0^1 p^2(x)(1-p(x))^2\,dx \int_{|x|\leq 2\tau} K_0^2(x)\,dx,$$
$$b_n^{-1/2}(\Delta_n - \Delta(p)) = O(b_n^{1/2}) + O(\alpha_n b_n^{-1/2}) + O\Big(\frac{1}{nb_n^{3/2}}\Big),$$

where

$$\Delta_n = ET_n^{(1)}, \quad \Delta(p) = \int_0^1 p(x)(1-p(x))\,dx \int_{|x|\leq r} K^2(u)\,du,$$
$$K_0 = K * K, \;\; * \text{ is the convolution operator.}$$

We have the following assertion.

Theorem 1. *Let $K(x) \in H(\tau)$ and $p(x), u_1(x), u_2(x) \in C^1[0,1]$. If $nb_n^2 \to \infty$ and $\alpha_n = n^{-1/2}b_n^{-1/4}$, then for the hypothesis H_{1n}*

$$b_n^{-1/2}(T_n - \Delta(p))\sigma^{-1}(p) \xrightarrow{d} N(a,1),$$

where $\Delta(p)$ and $\sigma^2(p)$ are defined in Lemma 2 and $\xrightarrow{d}$ denotes convergence in distribution and $N(a,1)$ is a random variable having the standard normal distribution with parameters $(a,1)$,

$$a = \frac{1}{2\sigma(p)}\int_0^1 (u_1(x) - u_2(x))^2\,dx.$$

Corollary. *Let $K(u) \in H(\tau)$ and $p(x) \in C^1[0,1]$. If $nb_n^2 \to \infty$, then for the hypothesis H_0*

$$b_n^{-1/2}(T_n - \Delta(p))\sigma^{-1}(p) \xrightarrow{d} N(0,1). \tag{1}$$

Application of the statistic T_n for the hypothesis testing. As an important application of the result of the corollary let us construct the criterion of testing the simple hypothesis $H_0 : p_1(x) = p_2(x) = p(x)$ (this is the case with given $p(x)$); the critical domain is defined by the inequality $T_n \geq d_n(\alpha) = \Delta(p) + b_n^{1/2}\sigma(p)\lambda_\alpha$, and from Theorem we establish that the local behavior of the power $\mathbb{P}_{H_{1n}}(T_n \geq d_n(\alpha))$ is as follows

$$\mathbb{P}_{H_{1n}}(T_n \geq d_n(\alpha)) \longrightarrow 1 - \Phi\Big(\lambda_\alpha - \frac{A(u)}{\sigma(p)}\Big),$$

where

$$A(u) = \frac{1}{2}\int_0^1 (u_1(x) - u_2(x))^2\,dx, \quad u = (u_1, u_2),$$

$\Phi(\lambda_\alpha) = 1 - \alpha$, $\Phi(\lambda)$ is a standard normal distribution.

Note that the statistic function T_n is normalized by the values $\Delta(p)$ and $\sigma^2(p)$ which depend on $p(x)$. If $p(x)$ is not defined by hypothesis, then the parameters $\Delta(p)$ and $\sigma^2(p)$ should be replaced respectively by

$$\widetilde{\Delta}_n = \int_{\Omega_n(\tau)} \lambda_n(x)\,dx \int_{|x|\leq\tau} K^2(x)\,dx, \quad \widetilde{\sigma}_n^2 = 2\int_{\Omega_n(\tau)} \lambda_n^2(x)\,dx \int_{|x|\leq 2\tau} K_0^2(x)\,dx,$$

$$\lambda_n(x) = \Big[p_{1n}(x)(p_n(x) - p_{1n}(x)) + p_{2n}(x)(p_n(x) - p_{2n}(x))\Big]\cdot\frac{1}{2}$$

and we show that

$$b_n^{-1/2}(\widetilde{\Delta}_n - \Delta(p)) \xrightarrow{\mathbb{P}} 0, \quad \widetilde{\sigma}_n^2 \xrightarrow{\mathbb{P}} \sigma^2(p). \tag{2}$$

Let us prove (2). Since $p_n(x) = 1 + O(\frac{1}{nb_n})$ uniformly in $x \in \Omega_n(\tau)$ and $|p_{in}(x)| \leq c_1$, $x \in [0,1]$, $i = 1,2$, we obtain

$$\begin{aligned} &b_n^{-1/2}E|\widetilde{\Delta}_n - \Delta(p)| \\ &\leq c_5 b_n^{-1/2}\Big\{\int_{\Omega_n(\tau)}(E(p_{1n}(x) - Ep_{1n}(x))^2)^{1/2}\,dx + \int_{\Omega_n(\tau)}(E(p_{2n}(x) - Ep_{2n}(x))^2)^{1/2}\,dx\Big\} \\ &\quad + b_n^{-1/2}\int_{\Omega_n(\tau)}|Ep_{1n}(x) - p(x)|\,dx + b_n^{-1/2}\int_{\Omega_n(\tau)}|Ep_{2n}(x) - p(x)|\,dx. \end{aligned}$$

Further, using Lemma 1 and also taking into account that $p(x) \in C^1[0,1]$ and $[\frac{x-1}{b_n}, \frac{x}{b_n}] \supset [-\tau, \tau]$ for all $x \in \Omega_n(\tau)$, it is easy to see that

$$b_n^{-1/2}E|\widetilde{\Delta}_n - \Delta(p)| = O\Big(\frac{1}{\sqrt{n}\,b_n}\Big) + O(b_n^{1/2}) + O\Big(\frac{1}{nb^{3/2}}\Big).$$

Hence $b_n^{-1/2}(\widetilde{\Delta}_n - \Delta(p)) \xrightarrow{\mathbb{P}} 0$. Analogously, it can be shown that $\widetilde{\sigma}_n^2 \xrightarrow{\mathbb{P}} \sigma^2(p)$.

Theorem 2. *Let $K(x) \in H(\tau)$ and $p_1(x) = p_2(x) \in C^1[0,1]$. If $nb_n^2 \to \infty$, then $b_n^{-1/2}(T_n - \widetilde{\Delta}_n)\widetilde{\sigma}_n^{-1} \xrightarrow{d} N(0,1)$ for $n \to \infty$.*

Proof. Follows from (1) and (2).

The theorem enables us to construct an asymptotical criterion of testing the composite hypothesis $H_0 : p_1(x) = p_2(x)$, $x \in [0,1]$. The critical domain for testing this hypothesis is defined by the inequality $T_n \geq \widetilde{d}_n(\alpha) = \widetilde{\Delta}_n + b_n^{-1/2}\widetilde{\sigma}_n\lambda_\alpha$, $\Phi(\lambda_\alpha) = 1-\alpha$.

Theorem 3. *Let $K(x) \in H(\tau)$, $p_1(x), p_2(x) \in C^1[0,1]$. If $nb_n^2 \to \infty$, then $\mathbb{P}_{H_1}(T_n \geq \widetilde{d}_n(\alpha)) \longrightarrow 1$ for $n \to \infty$. Here the alternative hypothesis H_1 is any pair $(p_1(x), p_2(x))$, $p_1(x), p_2(x) \in C^1[0,1]$, $0 \leq p_i(x) \leq 1$, $i = 1,2$, such that $p_1(x) \neq p_2(x)$ on the set of positive measure.*

Remark. Let t_i be the division points of the interval $[0,1]$ which are chosen so that $H(t_j) = \frac{2j-1}{2n}$, $j = 1,\dots,n$, where $H(x) = \int_0^x h(u)\,du$, $h(u)$ is some known continuous distribution density on $[0,1]$. In this case, by a similar reasoning to the above one we can generalize the results obtained in this paper.

Acknowledgement. The work is supported by Shota Rustaveli National Scientific Foundation (Project # FR/308/5-104/12).

R E F E R E N C E S

1. Efromovich S. Nonparametric curve estimation. Methods, Theory, and Applications. *Springer Series in Statistics. Springer-Verlag, New York*, 1999.

2. Copas J. B. Plotting p against x. *Appl. Statist.*, **32**, 2 (1983), 25-31.

3. Okumura H., Naito K. Weighted kernel estimators in nonparametric binomial regression. *The International Conference on Recent Trends and Directions in Nonparametric Statistics.*, *J. Nonparametr. Stat.*, **16**, 1-2 (2004), 39-62.

4. Nadaraya E. A. On estimating regression. (Russian) *Teor. Veroyatn. Primen.*, **9**, (1964), 157-159.

5. Watson G. S. Smooth regression analysis. *Sankhyā Ser. A*, **26**, (1964), 359-372.

6. Nadaraya E., Babilua P., Sokhadze G. Estimation of a distribution function by an indirect sample. *Ukr. Mat. Zh.* **62**, 12 (2010), 1642-1658. English Translation: *Ukr. Math. J.*, **62**, 12 (2010), 1906-1924.

7. Nadaraya E., Babilua P., Sokhadze G. On the integral square deviation of one nonparametric estimation of the Bernoulli regession. *Teor. Veroyatn. Primen.*, **57**, 2 (1964), 322-336.

Received 14.05.2014; revised 25.11.2014; accepted 29.12.2014.

Authors' addresses:

E. Nadaraya, P. Babilua
Faculty Exact and Natural Sciences of
Iv. Javakhishvili Tbilisi State University
2, University St., Tbilisi 0186
Georgia
E-mail: elizbar.nadaraya@tsu.ge;
p_babilua@yahoo.com

G. Sokhadze
I. Vekua Institute of Applied Mathematics of
Iv. Javakhishvili Tbilisi State University
2, University St., Tbilisi 0186, Georgia
E-mail: grigol.sokhadze@tsu.ge