საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 6, №2, 2012
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 6, no. 2, 2012

Mathematics

On One Nonparametric Estimate of a Bernoulli Regression Function

Elizbar Nadaraya*, Petre Babilua, Mzevinar Patsatsia†**

* *Academy Member, Department of Exact and Natural Sciences, I. Javakhishvili Tbilisi State University*
***Department of Exact and Natural Sciences, I. Javakhishvili Tbilisi State University*
† *Sokhumi State University, Department of Mathematics and Computer Sciences, Tbilisi*

ABSTRACT. We consider the class of nonparametric estimates of a Bernoulli regression function. We establish the conditions of uniform consistency and give the limit theorems for continuous functionals connected with the functions on $C[a,1-a]$, $0<a<1/2$. *© 2012 Bull. Georg. Natl. Acad. Sci.*

Key words: *Bernoulli regression function, kernel type estimate, asymptotic unbiasedness, consistency, Wiener process.*

Let a random variable Y take two values 1 and 0 with probabilities p ("success") and $1-p$ ("failure"). Assume that the probability of "success" p is a function of an independent variable $x\in[0,1]$, i.e. $p=p(x)=P\{Y=1\mid x\}$ ([1-3]). Let x_i, $i=\overline{1,n}$ be the division points of the interval $[0,1]$ which are chosen from the relation

$$H(x_i)=\int_0^{x_i}h(u)du=\frac{2i-1}{2n},\quad i=\overline{1,n},$$

where $h(x)$ is the known positive distribution density on $[0,1]$. Let further Y_i, $i=\overline{1,n}$, be independent Bernoulli random variables with $P\{Y_i=1\mid x_i\}=p(x_i)$, $P\{Y_i=0\mid x\}=1-p(x_i)$. The problem consists in estimating the function $p(x)$, $x\in[0,1]$, by the sampling $Y_1,Y_2,\ldots,Y_n$. Such problems arise in particular in biology ([1, 3]), in corrosion studies [4] and so on.

As an estimate for $p(x)$ we consider a statistic ([5, 6]) of the form

$$\hat{p}_n(x)=p_{1n}(x)p_{2n}^{-1}(x), \tag{1}$$

$$p_{\nu n}(x)=\frac{1}{nb_n}\sum_{i=1}^{n}h^{-1}(x_i)K\left(\frac{x-x_i}{b_n}\right)Y_i^{2-\nu},\quad \nu=1,2,$$

where $K(x)\ge 0$ is some distribution density (kernel), $K(x)=K(-x)$, $x\in(-\infty,\infty)$, $\{b_n\}$ is a sequence of positive numbers converging to zero.

Lemma 1. *Assume that*

1^0. $K(x)$ *is a function of bounded variation.*

2^0. $p(x)$ *and* $h(x)$ *are also functions of bounded variation on* $[0,1]$, *and* $h(x) \geq \mu > 0$, $x \in [0,1]$.

If $nb_n \to \infty$, *then*

$$\frac{1}{nb_n}\sum_{i=1}^{n} K^{\nu_1}\left(\frac{x-x_i}{b_n}\right) p^{\nu_2}(x_i) h^{-\nu_3}(x_i) = \frac{1}{b_n}\int_0^1 K^{\nu_1}\left(\frac{x-u}{b_n}\right) p^{\nu_2}(u) h^{-\nu_3+1}(u)\,du + O\left(\frac{1}{nb_n}\right), \tag{2}$$

uniformly with respect to $x \in [0,1]$, $\nu_i \in N \cup \{0\}$, $i = \overline{1,3}$.

Proof. Let $H_n(x)$ be an empirical distribution function of the "sampling" $x_1, x_2, \ldots, x_n$, i.e.

$$H_n(x) = n^{-1}\sum_{i=1}^{n} I(x_i < x)$$

It can be easily verified that

$$\sup_{0\leq x\leq 1} |H_n(x) - H(x)| = \sup_{0\leq x\leq 1} \left| \frac{1}{n}\left[nH(x) + \frac{1}{2}\right] - H(x)\right| \leq \frac{1}{2n}. \tag{3}$$

We have

$$\frac{1}{nb_n}\sum_{i=1}^{n} K^{\nu_1}\left(\frac{x-x_i}{b_n}\right) p^{\nu_2}(x_i) h^{-\nu_3}(x_i) - \frac{1}{b_n}\int_0^1 K^{\nu_1}\left(\frac{x-u}{b_n}\right) p^{\nu_2}(u) h^{-\nu_3+1}(u)\ \ du =$$

$$= \frac{1}{b_n}\int_0^1 K^{\nu_1}\left(\frac{x-u}{b_n}\right) p^{\nu_2}(u) h^{-\nu_3}(u)\ \ d\left[H_n(u) - H(u)\right]. \tag{4}$$

Applying the integration by parts formula to the right-hand part of (4) and taking (3) into account, we obtain (2).

Theorem 1. *Let the conditions of Lemma* 1 *be fulfilled. Then the estimate* (1) *is asymptotically unbiased and consistent at all points* $x \in [0,1]$, *where* $p(x)$ *is continuous. Moreover, it has an asymptotically normal distribution, i.e.*

$$\sqrt{nb_n}\left(\hat{p}_n(x) - E\hat{p}_n(x)\right)\sigma^{-1}(x) \xrightarrow{d} N(0,1),$$

$$\sigma^2(x) = \frac{p(x)(1-p(x))}{h(x)}\int K^2(u)\,du,$$

where $p(x)$ *is continuous.*

The proof of the theorem with the aid of Lemma 1 is analogous to the proof of Theorem 1 in [7].

Theorem 2. *Let* $K(x)$ *satisfy condition* 1^0 *of Lemma* 1 *and, besides,* $\varphi(t) = \int e^{itx} K(x)dx$ *be absolutely integrable; then the functions* $p(x)$ *and* $h(x)$ *are continuous and satisfy condition* 2^0 *of Lemma* 1.

(a) *Let* $nb_n^2 \to \infty$, *then*

$$D_n = \sup_{x\in\Omega_n} \left|\hat{p}_n(x) - p(x)\right| \xrightarrow{P} 0\ , \quad \Omega_n = \left[b_n^{\alpha}, 1 - b_n^{\alpha}\right], \quad 0 < \alpha < 1.$$

(b) *If* $\sum_{n=1}^{\infty} n^{-s/2} b_n^{-s} < \infty$, $s > 2$, *then* $D_n \to 0$ *a.s.*

Proof. Using Whittle's inequality [8] for moments of quadratic form, it can be shown [7] that

$$P\left\{\sup_{x\in\Omega_n}\left|\hat{p}_n(x)-E\hat{p}_n(x)\right|\geq\varepsilon\right\}\leq\frac{c_1}{\varepsilon^s\left(\sqrt{n}\ \ b_n\right)^s},\quad s>2,\quad \varepsilon>0. \tag{5}$$

Furthermore, by virtue of Lemma 1,

$$p_{2n}(x)=\frac{1}{b_n}\int_0^1 K\left(\frac{x-u}{b_n}\right)\ du+O\left(\frac{1}{nb_n}\right),$$

and since

$$\sup_{x\in\Omega_n}\left(1-\frac{1}{b_n}\int_0^1 K\left(\frac{x-u}{b_n}\right)du\right)\leq\int_{-\infty}^{-b_n^{\alpha-1}}K(u)\,du+\int_{b_n^{\alpha-1}}^{\infty}K(u)\,du\longrightarrow 0,$$

we can write that $\sup_{x\in\Omega_n}\left|p_{2n}(x)-1\right|\to 0$ as $n\to\infty$, i.e. $p_{2n}(x)\geq 1-\varepsilon_0$, $0<\varepsilon_0<1$, for sufficiently large $n\geq n_0$, uniformly with respect to $x\in\Omega_n$. Therefore from the definition of $\hat{p}_n(x)$ it follows that

$$\sup_{x\in\Omega_n}\left|E\hat{p}_n(x)-p(x)\right|\leq(1-\varepsilon_0)^{-1}\left[\sup_{x\in\Omega_n}\left|Ep_{1n}(x)-p(x)\right|+\sup_{x\in\Omega_n}\left|1-p_{2n}(x)\right|\right], \tag{6}$$

where the second summand in the right-hand part of (6) converges to 0, and the first summand is estimated as follows. It is not difficult to show that

$$\sup_{x\in\Omega_n}\left|Ep_{1n}(x)-p(x)\right|\leq S_n+o(1)+O\left(\frac{1}{nb_n}\right),$$

$$S_n=\sup_{0\leq x\leq 1}\left|\frac{1}{b_n}\int_0^1\left(p(y)-p(x)\right)K\left(\frac{x-y}{b_n}\right)dy\right|.$$

Extend the function $p(x)$ continuously up to the interval $[-1,2]$. Let $p^*(x)$ be the extended function and assume that $p^*(x)\geq 0$ and $\sup_{0\leq x\leq 1}p(x)=\sup_{[-1,2]}p^*(x)$. We easily conclude that

$$S_n\leq\sup_{0\leq x\leq 1}\int_{-1}^{1}\left|p^*(x-y)-p^*(x)\right|\frac{1}{b_n}K\left(\frac{y}{b_n}\right)dy\leq$$

$$\leq\sup_{0\leq x\leq 1}\sup_{|y|\leq\delta}\left|p^*(x-y)-p^*(x)\right|\int K(y)\,dy+2\sup_{-1\leq x\leq 2}p^*(x)\int_{|y|\geq\delta/b_n}K(y)\,dy,\quad\delta>0. \tag{7}$$

The first summand in the right-hand part of (7) can be made arbitrarily small by a choice of $\delta>0$. After choosing $\delta>0$ and assuming that $n\to\infty$, we obtain that the second summand tends to 0. From this and the relations (5) and (6), the proof of the theorem follows.

Theorem 3. *Let the kernel* $K(x)\geq 0$ *be chosen so that it would be a function of finite variation and satisfy the conditions* $K(-u)=K(u)$, $K(u)=0$ *for* $|u|\geq 1$, $\int K(u)du=1$. *Let* $g(x)\geq 0$, $x\in[a,1-a]$,

$0 < a < 1/2$, *be a measurable and bounded function. Let, further,*

$0 < \inf p(x) \le \sup p(x) < 1, \quad x \in [0,1]$.

(a) *If* $p(x)$ *is continuous and* $nb_n^2 \to \infty$, *then*

$$T_n = \sqrt{n} \int_a^{1-a} g_1(x)\left[\hat{p}_n(x) - E\hat{p}_n(x)\right]dx \xrightarrow{d} N\left(0,\sigma^2\right),$$

$$g_1(x) = g(x)\psi(x), \quad \psi(x) = \left(\frac{h(x)}{p(x)(1-p(x))}\right)^{1/2}.$$

(b) *If* $nb_n^2 \to \infty$, $nb_n^4 \to 0$ *and* $p(x)$ *has bounded derivatives up to second order, then for* $n \to \infty$

$$T_n = \sqrt{n} \int_a^{1-a} g_1(x)\ \left[\hat{p}_n(x) - p(x)\right]\ dx \xrightarrow{d} N\left(0,\sigma^2\right),$$

$$\sigma^2 = \int_a^{1-a} g^2(x)\ dx\ .$$

Proof. We have

$$\sigma_n^2 = Var\ \ \overline{T}_n = n^{-1}\sum_{j=1}^{n} \psi^{-2}\left(x_j\right)\left(\frac{1}{b_n}\int_a^{1-a} K\left(\frac{u-x_j}{b_n}\right)g_{2n}(u)\,du\right)^2,$$

$$g_{2n}(x) = g_1(x)\,p_{2n}^{-1}(x).$$

By virtue of Lemma 1, it is not difficult to establish that

$$\sigma_n^2 = \int_a^{1-a} \psi^{-2}(t)dt\left(\frac{1}{b_n}\int_a^{1-a} K\left(\frac{u-t}{b_n}\right)g_1(u)du\right)^2 + o(1) + O\left(\frac{1}{nb_n^2}\right),$$

and also

$$\left|\int_a^{1-a} \psi^{-2}(t)\,dt\ \ \left(\frac{1}{b_n}\int_a^{1-a} K\left(\frac{u-t}{b_n}\right)g_1(u)\,du\right)^2 - \int_a^{1-a} g^2(u)\,du\right| \le$$

$$\le c_2 \int_a^{1-a} dt\left|\frac{1}{b_n}\int_a^{1-a} g_1(u)K\left(\frac{u-t}{b_n}\right)du - g_1(t)\int_a^{1-a}\frac{1}{b_n}K\left(\frac{u-t}{b_n}\right)du\right| + c_3\int_a^{1-a}\left|\int_a^{1-a}\frac{1}{b_n}K\left(\frac{u-t}{b_n}\right)du - 1\right|dt = A_{1n} + A_{2n}.$$

Since $\int_a^{1-a}\frac{1}{b_n}K\left(\frac{u-t}{b_n}\right)du \to 1$ for all $t \in (a,1-a)$, we see that $A_{2n} \to 0$. Further, extend the function $g_1(u)$ in the exterior of $[a,1-a]$ by zero and denote the extended function by $g_1^*(x)$. Then

$$A_{1n} \le c_4 \int_0^1 \left(\int_{-\infty}^{\infty}\left|g_1^*(x+y) - g_1^*(y)\right|dy\right)\frac{1}{b_n}\ \ K\left(\frac{x}{b_n}\right)dx \le$$

$$\le c_5 \int_{-1}^{1}\left(\int_{-\infty}^{\infty}\left|g_1^*(y+ub_n) - g_1^*(y)\right|dy\right)K(u)\,du = c_5\int_{-1}^{1}\omega(ub_n)K(u)\,du \longrightarrow 0 \quad \text{as} \quad n \to \infty, \qquad (8)$$

where

$$\omega(y)=\int_{-\infty}^{\infty}\left|g_1^*(y+x)-g_1^*(x)\right|dx\,.$$

The inequality (8) holds by virtue of the Lebesgue theorem on majorized convergence and the fact that $\omega(ub_n)\le 2\left\|g_1^*\right\|_{L_1(-\infty,\infty)}$ and $\omega(ub_n)\to 0$ as $n\to\infty$. We have thereby proved that $\sigma_n^2\to\sigma^2$.

Let us now verify the fulfillment of the conditions of the Central Limit Theorem for the sums

$$\overline{T}_n=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}a_{jn}\left(Y_j-p\left(x_j\right)\right),\qquad a_{jn}=\int_a^{1-a}\frac{1}{b_n}K\left(\frac{x-x_i}{b_n}\right)g_{2n}(x)dx$$

Since $a_{jn}\le c_6$, $E\left|Y_j-p\left(x_j\right)\right|^{2+\delta}\le 1$, $j=\overline{1,n}$ and $\sigma_n^2\to\sigma^2$, the Lyapunov fraction $L_n=O\left(n^{-\delta/2}\right)$, $\delta>0$.

Finally, the condition (b) of the theorem follows from (a) if we take into account that

$$\sqrt{n}\int_a^{1-a}g_1(x)\left[E\hat{p}_n(x)-p(x)\right]dx=O\left(\sqrt{n}\,b_n^2\right)+O\left(\frac{1}{\sqrt{n}\,b_n}\right)\tag{9}$$

Lemma 2. *Under the conditions* (a) *and* (b) *of Theorem* 3 *we respectively have*

$$E\left|\overline{T}_n\right|^s\le c_7\left(\int_a^{1-a}g(u)du\right)^{s/2},\quad s>2\tag{10}$$

and

$$E\left|T_n\right|^s\le c_8\left(\int_a^{1-a}g(u)du\right)^{s/2},\quad s>2.\tag{11}$$

Proof. $\overline{T}_n$ is the linear form of $\xi_j=Y_j-p\left(x_j\right)$, $E\xi_j=0$, $1\le j\le n$. It is obvious that $E\left|\xi_j\right|^s\le 1$, $j=\overline{1,n}$. Therefore, by virtue of Whittle's inequality [8] and Lemma 1, we obtain

$$E\left|\overline{T}_n\right|^s\le c(s)2^s\left[\int_0^1\left(\frac{1}{b_n}\int_a^{1-a}K\left(\frac{u-t}{b_n}\right)g_{2n}(u)\,du\right)^2dt+O\left(\frac{1}{nb_n^2}\right)\left(\int_a^{1-a}g_{2n}(u)\,du\right)^2\right]^{s/2},$$

$$c(s)=\frac{2^{s/2}}{\sqrt{\pi}}\ \Gamma\left(\frac{s+1}{2}\right).\tag{12}$$

Further, taking into account that

$$g_{2n}(u)\le g(u)\left[\frac{\left(\max h(x)\right)^{1/2}}{\sqrt{p_1(1-p_2)}}\ \left(1+O\left(\frac{1}{nb_n}\right)\right)\right]\le c_9g(u)\ ,$$

$$a\le u\le 1-a\ ,\quad p_1=\inf p(x)\ ,\quad p_2=\sup p(x)\ ,\quad x\in[0,1]\ ,$$

from (12) we obtain (10). Finally, this and (9) imply (11).

Let us introduce the following random processes:

$$\overline{T}_n(t)=\sqrt{n}\int_a^t\left(\hat{p}_n(u)-E\hat{p}_n(u)\right)\psi(u)du\,,\qquad T_n(t)=\sqrt{n}\int_a^t\left(\hat{p}_n(u)-p(u)\right)\psi(u)du\,.$$

Theorem 4. *Let the conditions* (a) *and* (b) *of Theorem* 3 *be fulfilled. Then for all functional* $f(\cdot)$ *continuous on* $C[a,1-a]$, $0<a<1/2$, *the distributions* $f(\overline{T}_n(t))$ *and* $f(T_n(t))$ *converge to the distribution* $f(w(t-a))$, *where* $w(t-a)$, $a\le t\le 1-a$, *is a Wiener process with the correlation function* $r(s,t)=\min(t-a,s-a)$, $w(t-a)=0$, $t=a$.

The proof of the theorem by Theorem 3 and Lemma 2 is analogous to the proof of Theorem 2 in [9].

მათემატიკა

ბერნულის რეგრესიის ფუნქციის ერთი არაპარამეტრული შეფასების შესახებ

ე. ნადარაია*, პ. ბაბილუა, მ. ფაცაცია†**

* *აკადემიის წევრი, ი. ჯავახიშვილის სახ. თბილისის სახელმწიფო უნივერსიტეტი, ზუსტ და საბუნებისმეტყველო მეცნიერებათა ფაკულტეტი*

** *ი. ჯავახიშვილის სახ. თბილისის სახელმწიფო უნივერსიტეტი, ზუსტ და საბუნებისმეტყველო მეცნიერებათა ფაკულტეტი*

† *სოხუმის სახელმწიფო უნივერსიტეტი, მათემატიკისა და კომპიუტერულ მეცნიერებათა ფაკულტეტი, თბილისი*

განხილულია ბერნულის რეგრესიის ფუნქციის არაპარამეტრული გულოვან შეფასებათა კლასი. შესწავლილია თანაბრად ძალდებულობის საკითხი. გარდა ამისა, განხილულია ამ შეფასებასთან დაკავშირებული უწყვეტ ფუნქციათა $C[a,1-a]$, $0<a<1/2$ კლასში განსაზღვრული ზოგიერთი ფუნქციონალის ზღვარითი განაწილების საკითხი.

REFERENCES

1. *S. Efromovich* (1999), Nonparametric Curve Estimation. Methods, Theory, and Applications. Springer Series in Statistics. New York.
2. *J. B. Copas* (1983), Appl. Statist., **32**, 1: 25-31.
3. *H. Okumora, K. Naito* (2004), J. Nonparametr. Stat., **16**, 1-2: 39-62.
4. *K. B. Manjgaladze* (1986), Soobshch. AN GSSR, 124: 261-263.
5. *E. A. Nadaraya* (1964), Teor. Veroyatn. Primen., 9: 157-159.
6. *G. Watson* (1964), Sankhyâ, Ser. A, 26: 359-372.
7. *E. Nadaraya, P. Babilua, G. Sokhadze* (2010), Bull. Georg. Natl. Acad. Sci., **4,** 3: 5-12.
8. *P. Whittle* (1960), Teor. Veroiatnost. i Primenen., 5: 331-335.
9. *E. Nadaraya, P. Babilua, G. Sokhadze* (2011), Bull. Georg. Natl. Acad. Sci., **5**, 2: 11-18.

Received March, 2012