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

ON ESTIMATION OF UNBOUNDED FUNCTIONALS OF PROBABILITY DISTRIBUTION DENSITY

Buadze T., Khechinashvili Z., Sokhadze G.

Abstract. Estimation of a nonlinear, possibly unbounded functional of probability distribution density is studied. The plug-in-estimator is taken for the estimation. The functional can be unbounded, but it cannot exceed polynomial growth. Consistency of the estimator is proved and the convergence order is established. Key Words. Estimation, plug-in-estimator, unbounded functional, consistency.

Keywords and phrases: Estimation, plug-in-estimator, unbounded functional, consistency.

AMS subject classification: 62G07, 62G20, 62E20.

1. Introduction. Let X be the random variable with unknown distribution density $f(x)$. $X_1, X_2, ..., X_n$ is the sample of independent copies of X. Furthermore, $\mathfrak{M}$ is the functional on $L_2(R)$, which is defined on some subspace L and has second order derivative. Assume, that $f \in \mathcal{L}$ and so the $\mathfrak{M}f$ exists. Our aim is to investigate consistency of the estimation for $\mathfrak{M}f$ using Plug-in-Estimator $\mathfrak{M}\hat{f}_n$, where $\hat{f}_n$ is estimation of f.

For estimation of f and its derivatives we use Rosenblatt-Parzen type kernel estimations of probability density (see. [1-3])

$$\hat{f}_n^{(j)}(x) = \frac{1}{nh_n^{j+1}} \sum_{i=1}^{n} K^{(j)}\left(\frac{x - X_i}{h_n}\right), \quad j = 0, 1, ..., m.$$

Here h_n is the sequence of positive numbers which converges to $0, K(x)$ is the function with density properties.

The problem setted includes case of bounded functional (integral and other type) as well as information and entropy functionals.

2. Notations and conditions. In future we need the following notations and conditions. Let X be the random variable with probability distribution density $f(x)$. Suppose, that following conditions are fulfilled

f1) f(x) has continuous derivatives including of order m and some $C_f > 0$, $\sup_{x\in\mathbb{R}} |f^{(i)}(x)| \leq C_f < \infty, i = 1, 2, ..., m$;

f2) there exists a continuous, increasing function $H(x)$, such that

$$\sup_{|y|\leq x} \frac{1}{f(y)} \leq H(x).$$

Consider the real valued function $K(x) \geq 0$ and assume, that following conditions are satisfied

k1) function $K(x)$ has a compact support

k2) $\int_{-\infty}^{\infty} K(x)dx = 1$;
k3) $K(x)$ has continuous derivatives including of order m.

Then for some $C_K > 0, |K^{(i)}(x)| \leq C_K < \infty, \;\; i = 0, 1, ..., m.$

For the sequence h_n we assume, that the following condition is satisfied
h) $h_n, n = 1, 2, ...$ is the sequence of positive numbers, monotonically converging to 0 such, that for some $c > 0$, $h_n \geq \frac{c \log n}{n}$.

It is known (see [4]), that under the conditions **f1), f2), k1)-k3)** and **h)** with the probability 1

$$\sup_{x \in \mathfrak{R}} |\hat{f}_n(x) - E\hat{f}_n(x)| = O\left(\frac{\sqrt{|\log h_n| \vee \log \log n}}{\sqrt{nh_n}}\right).$$

For the functional $\mathfrak{M}$ we assume the following: let $W_m = W_m(R)$ be the Sobolev space of functions from $L_2(R)$ with continuous derivatives including of order m and norm

$$\|g\|_m = \sqrt{\sum_{j=0}^{m} \int_{-\infty}^{\infty} |g^{(j)}(x)|^2 dx}.$$

The space W_n has the scalar product

$$(g_1, g_2)_m = \sum_{j=0}^{m} \int_{-\infty}^{\infty} g_1^{(j)}(x) g_2^{(j)}(x) dx.$$

$\mathfrak{M}\mathfrak{1}$) the functional $\mathfrak{M}$ is defined on the subspace $\mathcal{L} \subset \mathcal{W}_{\Updownarrow}$;
$\mathfrak{M}\mathfrak{2}$) there exists $\mathfrak{M}$;
$\mathfrak{M}\mathfrak{3}$) there exist the functionals $\mathfrak{M}_{\mathfrak{k}}, \mathfrak{k} = \mathfrak{o}, \mathfrak{1}, ...$ such, that
i) Domain of functional $\mathfrak{M}_{\mathfrak{k}}$ is $\mathcal{L}_{\|} = \mathcal{W}_{\Updownarrow}([-\int_{\|}; \int_{\|}])$ for any $k = 1, 2, ...$, where s_k is the sequence diverging to ∞;
ii) $\hat{f}_n \in \mathcal{L}_k$, for each k;
iii) For any $g \in \mathcal{L}, \mathfrak{M}_{\|}\} \to \mathfrak{M}\mathfrak{g}$, when $k \to \infty$;
iv) functionals $\mathfrak{M}$ are smooth in the sense, that there exist derivatives including second order: $\mathfrak{M}'_k$ is a linear functional on $\mathcal{L}_k$, $\mathfrak{M}''_k$ is a belinear functional on $\mathcal{L}_k$ and both satisfy the inequalities

$$\|\mathfrak{M}_k^{(i)} g\|_m \leq C \cdot s_k^{\alpha} \cdot \|g\|^{\beta} \cdot \|g\|_m^2, \quad g \in C([-s_k; s_k]), \;\; \alpha \geq 0, \;\; \beta \leq 0, \;\; i = 1, 2.$$

where $\|g\|$ denotes uniform norm of element g and $\|g\|_m$ is the norm in $\mathcal{L}_k$.

3. Estimation of residual term. Denote $f_n(x) = E\hat{f}_n(x)$. Consider the difference $\mathfrak{M}_n \hat{f}_n - \mathfrak{M}_n f_n$ and using condition $\mathfrak{M}\mathfrak{3}$, **iv)** rewrite it in following form

$$\mathfrak{M}_n \hat{f}_n - \mathfrak{M}_n f_n = S_n(h_n) + R_n, \tag{1}$$

where $S_n(h_n)$ denotes the result of functional derivative (it is linear functional) $\mathfrak{M}_n$ on $\hat{f}_n - f_n$. In (1)

$$R_n = O\left(\|\mathfrak{M}''_n(\hat{f}_n - f_n)\|_m\right).$$

Theorem 1. *If the conditions* **f1),f2),k1)-k4),h)** *and* $\mathfrak{M}1) - \mathfrak{M}3)$ *are fulfilled, then in (1) for the residual component we have*

$$R_n = O\left(\frac{d(s_n)\log n}{nh_n^{2m+1}}\right).$$

4. Consistency. Let $\varepsilon > 0$ be the fixed number. The sequence h_n is chosen such, that

$$\frac{\log n}{nh_n^{2m+1}} \to 0 \quad when\ n \to \infty.$$

Consider s_n as a solution of the equation

$$\frac{\log n}{nh_n^{2m+1}} = \frac{\varepsilon}{d(s_n)}, \tag{2}$$

where

$$d(x) = x^{\alpha} H^{\beta}(x).$$

Theorem 2. *Suppose, that conditions* **f1),f2),k1)-k3),h)** *and* $\mathfrak{M}1) - \mathfrak{M}3)$ *are satisfied. Let* h_n *be the sequence of positive numbers, which monotonically converges to 0 such, that*

$$\frac{\log n}{nh_n^{2m+1}} \to 0, \quad when\ n \to \infty.$$

If for any n, s_n is a solution of equation (2), then with probability 1 we have

$$I(\hat{f}_n, s_n) - I(f) \to 0.$$

R E F E R E N C E S

1. Rosenblatt M. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27** (1956), 832-837.
2. Parzen E. On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33** (1962), 1065-1076.
3. Nadaraya E. Nonparametric Estimation of Probability Densities and Regression Curves. *Kluwer Academic Publishers*, 1988.
4. Gine E., Mason D.M. Uniform in bandwidth estimation of integral functionals of the density function. *Scand. J. Statist.*, **35** (2008), 739-761.

Received 25.05.2014; revised 27.11.2014; accepted 27.12.2014.

Authors' address:

T. Buadze, Z. Khechinashvili, G. Sokhadze
Iv. Javakhishvili Tbilisi State University
2, University St., Tbilisi 0186
Georgia
E-mail: t.buadze@yahoo.com
khechinashvili@gmail.com
giasokhi1@i.ua