Reports of Enlarged Session of the
Seminar of I. Vekua Institute
of Applied Mathematics
Volume 22, 2008

ON THE NONPARAMETRIC ESTIMATION OF A LOGARITHMIC DERIVATIVE OF A PROBABILITY MEASURE

Nadaraya E.,* Sokhadze G.**

* Iv. Javakhishvili Tbilisi State University
**A. Tsereteli State University

Abstract. The statistical estimating problem for logarithmical derivative of distributions of random elements in Hilbert space is considered.

Keywords and phrases: Estimation, Logarithmic derivative Kernel.

AMS subject classification: 62G08; 62G20.

The statistical estimation of various characteristics of a probability measure by observation data is a topical problem as there is no general solution of the problem connected with estimation of the measure itself (see [1]). Progress made in tackling this question in the theory of nonparametric estimation (the Watson–Nadaraya method) in a finite-dimensional space enables us to extend some results to an infinite-dimensional case as well.

Let $\{\Omega, \mathcal{J}, P\}$ be a fixed probability space, H a separable real Hilbert space, B a σ-algebra of Borel subsets, ξ a random element with values in H and μ its distribution. For the element ξ we only know that its distribution μ is smooth. Here the smoothness implies that there exists a logarithmic derivative of this measure along the constant direction $a \in H$.

Recall that the measure μ has a logarithmic derivative along the constant direction $a \in H$ if it is differentiable along a and $\mu'_a \ll \mu$. In that case $l(x,a) = \frac{d\mu'_a}{d\mu}(x)$. An also the measure μ is said to have a derivative along a if there exists a countably additive measure μ'_a such that the following formula of integration by parts

$$\int\limits_H \big(f'(x), a\big)_H \mu(dx) = -\int\limits_H f(x)\mu'_a(dx)$$

holds true for any bounded and boundedly differentiable function $f(x): H \to R$.

If the logarithmic derivative exists along a, then we have the formula

$$\int\limits_H \big(f'(x), a\big)_H \mu(dx) = -\int\limits_H f(x) l(x,a)\mu(dx),$$

which is sometimes used to define a logarithmic derivative along a.

Let further $X_1, X_2, \ldots, X_n$ be a sampling of independent, equally distributed random elements with values in H. Let the distribution (i.e the corresponding measure) be denoted by μ. Assume that for μ we only know that it has a logarithmic derivative along the constant direction $a \in H$. We are to find an estimate of the unknown logarithmic derivative $l(x,a)$ using the sampling $X_1, X_2, \ldots, X_n$.

Let $\{L_m\}$ be an increasing sequence of finite-dimensional subspaces of the space H such that $\bigcup\limits_{m=1}^{\infty} L_m$ is dense in H. Denote by P_m a finite-dimensional projector H on L_m. Let $\mu_m = \mu \circ P_m^{-1}$, $a_m = P_m a$, $x_m = P_m x$, $l_m^{a_m}(x_m) = P_m l(P_m x, P_m a)$. Then conclude that $l_m^{a_m}(x_m)$ is a logarithmic derivative of the measure μ_m in the space L_m. If $p_m(x_m)$ is an everywhere positive and differentiable density of the measure μ_m, then we have

$$l_m^{a_m}(x_m) = \frac{(\operatorname{grad} p_m(x_m), a_m)_{L_m}}{p_m(x_m)}\,.$$

This logarithmic derivative can be estimated by finite-dimensional techniques using the samplings

$$P_m X_1 = X_1^m, P_m X_2 = X_2^m, \dots, P_m X_n = X_n^m.$$

For each m, using the sampling $X_1^m, X_2^m, \dots, X_n^m$ we construct effective estimates $\widehat{l}_n^m = \widehat{l}_m^{a_m}(x_m)_n$ for $l_m^{a_m}(x_m)$ and show their convergence to $l(x,a)$ as $n, m \to \infty$. In that case, $\widehat{l}_n^m$ can be regarded as an estimate of $l(x,a)$.

Below we obtain a result for an m-dimensional case and then establish a possibility of passage to the limit.

Assume that we have the sampling $X_j = (X_j^1, X_j^2, \dots, X_j^m)$, $j = 1, 2, \dots, n$, of independent and equally distributed random vectors. To estimate the unknown density $p(x)$ we use the statistic (see [2])

$$\widehat{p}_n(x) = \frac{\lambda_n^m}{n} \sum_{i=1}^{n} K(\lambda_n(x - X_i)),$$

where

$$K(x) = \prod_{j=1}^{m} K_j(x_j), \quad x = (x_1, x_2, \dots, x_m), \quad K_j, \quad j = 1, 2, \dots, m,$$

is an arbitrary density function in one-dimensional space. As is well-known, the logarithmic derivative is a vector with components $\frac{1}{p(x)} \frac{\partial p(x)}{\partial x_i}$. Therefore we must estimate

$$l(x) = \frac{1}{p(x)} \operatorname{grad} p(x).$$

As a statistic we take

$$\widehat{l}_n(x) = \frac{\sum\limits_{i=1}^{n} \operatorname{grad} \prod\limits_{j=1}^{m} K_j(\lambda_n(x_i - X_i^j))}{\sum\limits_{i=1}^{n} \prod\limits_{j=1}^{m} K_j(\lambda_n(x_j - X_i^j))}\,. \tag{1}$$

Theorem 1. *Let $K_j(X)$, $j = 1, 2, \dots,$ be a density function,*

$$\lambda_n \to \infty, \quad \frac{\lambda_n^2 \ln n}{n} \to 0 \quad \text{as} \quad n \to \infty.$$

Then (1) converges in $C(R^n)$ to $l(x)$ with probability 1.

In the considered case we take one function $K(x)$. In the conditions of Theorem 1

$$\widehat{l}_n^m(x_m) = \frac{\lambda_n \sum\limits_{i=1}^{n} \sum\limits_{s=1}^{m} a_n^s K'(\lambda_n(x_m^s - X_{is}^m)) \prod\limits_{\substack{j=1 \\ j\neq s}}^{m} K(\lambda_n(x_m^j - X_{ij}^m))}{\sum\limits_{i=1}^{n} \prod\limits_{j=1}^{m} K(\lambda_n(x_m^j - X_{ij}^m))} \tag{2}$$

converges uniformly as $n \to \infty$ to $l_m^{a_m}(x_m)$ with probability 1. As is known (see [3]), $l_m^{a_m}(x_m)$ is a martingale with respect to the system $\{L_m, \mathcal{B}_m, \mu_m\}$ and converges to $l(x,a)$ if and only if it is uniformly integrable with respect to the measure μ, which in our case a priori takes place.

Theorem 2. *Let $K(x)$ be an even, uniformly continuous function,* $0 < K(x) \leq 1$ *and*

$$\int\limits_R K(x)\,dx = 1; \quad \lambda_n \to \infty, \quad \frac{\lambda_n^2 \ln n}{n} \to 0.$$

Then (2) *converges in $C(H)$ to $l(x,a)$ with probability* 1.

REFERENCES

1. Sazonov V.V. Teor. Veroyatnost. i Primenen. **8** (1963), 299–303 (in Russian).

2. Nadaraya E. Nonparametric estimation of probability densities and regression curves, Kluwer Academic Publishers Group, Dordrecht, 1989.

3. Skorokhod A.V. Integration in Hilbert space, M.: 1975 (in ussian).

Received 24.09.2008; revised: 15.12.2008; accepted 24.12.2008.