Georgian Scientists
ქართველი მეცნიერები
Vol. 5 Issue 1, 2023
https://doi.org/10.52340/gs.2023.05.01.27

On the accuracy by the metric L_1 of the density estimation constructed by dependent observations

Beqnu Parjiani[1], Levan Labadze[2] and Tsiala Kvatadze[3]

[1]Associate Professor of Department of Mathematics of Georgian Technical University; [2]Invited Lecturer of Economics, Ilia State University; [3]Invited Lecturer in Statistics, International Black Sea University

Abstract

A narrowly stationary two-component sequence $\{\xi_i, X_i\}_{i\ge1}$ is considered on the probability space (Ω, F, P). The control sequence $\{\xi_i\}_{i\ge1}$ $(\xi_i : \Omega \to \Xi,\ i = 1,2,\ldots)$ is discrete $\Xi = \{b_1, b_2, \ldots, b_r\}$, $P(\xi_i = b_m) = p_m$, $m = \overline{1,r}$, $i = 1,2,\ldots$, $\sum_{m=1}^{r} p_m = 1$. $\{X_i\}_{i\ge1}$ $(X_i : \Omega \to R,\ i = 1,2,\ldots)$ is a conditionally independent sequence, those members represent observations of some random variable X. The conditional distributions $\mathcal{P}_{X|\xi_i=b_m}$, $m = \overline{1,r}$ have unknown densities $f_m(x)$, $m = \overline{1,r}$, respectively. A core Rosenblatt-Parzen-type estimate of the density is $\overline{f}(x) = \sum_{m=1}^{r} p_m f_m(x)$ constructed from the dependent observations. The accuracy of this estimate is determined by the metric L_1. A special case obtained by using the Bartlett core and taking the smoothing coefficient as a specific sequence is considered.

2010 Mathematics Subject Classification. 62G05. 62G07

Introduction

During statistical studies of practical tasks, parametric and non-parametric estimates are obtained. One important issue is the construction of an estimate of the density of the distribution. Until recently, estimates were made by independent observations. Many problems require consideration of dependent observations.

Long-term financial independence studies of investment and insurance companies are constantly conducted in the financial market. It is necessary to assess the risks of banking investments. For this purpose, an analysis of the flow of reinvestments is carried out, and the indicators of the financial stability of the companies are evaluated.

The listed and other tasks require statistical analysis not only with independent, but also with dependent data. Research in this direction has actually just started. And there is a rich historical experience of constructing non-parametric estimates of density through independent observations.

Let's the quantities X_i, ($X_i \in R$), $i=$ 1,2... represent independent observations of a random variable X. Let's say a quantity has an unknown density $g(x)$. In M. Rosenblatt and E. In Parzen's works (see [1], [2]), the class of core estimations is considered as a density $g(x)$ estimation.

$$\hat{g}_n(x,a_n)=\frac{a_n}{n}\sum_{i=1}^{n}k\left(a_n\left(x-X_i\right)\right), \qquad (1)$$

where $\{a_n\}_{n\geq 1}$ is a sequence of positive numbers such that

$$\lim_{n\to\infty}a_n=\infty, \quad a_n=o(n), \qquad (2)$$

and the core $K(x)$ according to Lebesgue is a integrated certain Borel function.

The accuracy of this type of density L_2 approximation constructed by independent observations ([1], [3], [4]) and metrics ([5]) has been determined under different conditions.

It is known by L. Devroye (see [5]) the accuracy of the metric L_1 of the estimate constructed by independent observations of the density. Let's state this result as a lemma. We will use it during the proof of our theorem.

Definition 1 (see [5]). Let us denote by F, the set of such functions $f(x)$ (see [5]) that satisfy the conditions: $f(x)$ is absolutely continuous and has a derivative almost everywhere, f', f' are absolutely continuous and has a derivative almost everywhere, f'', f'' are bounded and continuous.

Definition 2 (see [5]). Let's denote by Φ the set of such functions $\varphi(x)$ (see [5]) that satisfy the conditions: $\varphi(x)$ is a density with a compact carrier having derivatives up to the fourth order (including) $\varphi \in F$, $\varphi'' \in F$ and $\varphi_a(x)=(1/a)\varphi(x/a)$

Definition 3 (see [5]). Let's denote by $K^{\bullet}$ the class of densities bounded on R with a compact carrier on (see [5]), for that $K(-x)=K(x)$

Lemma (see [5]) Let's the quantities X_i, $x_i \in R$, $i=1,2,\ldots$ represent independent observations of some random quantity X having an unknown density $g(x)$ with a compact carrier. Let's say $\hat{g}_n(x,a_n)$ is determined by the equality (1), $K(x)\in K^{\bullet}$ and a_n is a sequence (2), then for the quantity $J^{\bullet}(a_n)=\int_{-\infty}^{\infty}\left|\hat{g}_n(x,a_n)-g(x)\right|dx$ is fair the estimattion

$$EJ^{\bullet}(a_n)\leq\sqrt{\frac{a_n}{n}}\sqrt{\frac{2}{\pi}}\alpha\int_{-\infty}^{\infty}\sqrt{g(x)}dx+\frac{1}{a_n^2}\frac{\beta}{2}\sup_{a>0}\int_{-\infty}^{\infty}\left|\left(g*\varphi_a\right)''(x)\right|dx+o\left(\sqrt{\frac{a_n}{n}}\right), \quad (3)$$

where

$$\alpha=\sqrt{\int_{-\infty}^{\infty}K^2(x)dx}, \quad \beta=\int_{-\infty}^{\infty}x^2K(x)dx,$$

$\varphi\in\Phi$, and the symbol $*$ - is a composition of functions.

If at the same time $g(x) \in F$, then

$$EJ^{\bullet}(a_n) \le \sqrt{\frac{a_n}{n}}\sqrt{\frac{2}{\pi}}\alpha\int_{-\infty}^{\infty}\sqrt{g(x)}dx + \frac{1}{a_n^2}\frac{\beta}{2}\int_{-\infty}^{\infty}\left|g''(x)\right|dx + o\left(\sqrt{\frac{a_n}{n}}\right). \quad (4)$$

Recently, statistical analysis of samples consisting of different types of dependent observations has start.

We will construct a core Rosenblatt-Parzen-type estimation of the density with dependent observations. We consider conditionally independent observations. We will determine the accuracy of the built estimate with a metric L_1.

On the probabilistic space (Ω, F, P) let us consider the two-component stationary (in the narrow sense) sequence of random variables

$$\{\xi_i, X_i\}_{i\ge 1} \quad (5)$$

Where, $\xi : \Omega \to \Xi$, $X_i : \Omega \to R^m$ and Ξ is some space.

Definition 4. The sequence $\{X_i\}_{i\ge1}$ in (5) is called a conditionally independent sequence (see. [6],[7],[8]) controlled by the sequence $\{\xi_i\}_{i\ge1}$ if for any natural n and the fixed trajectory $\bar{\xi}_{1n} = (\xi_1, \xi_2, ..., \xi_n)$ the values $X_1, X_2, ..., X_n$ become independent and for all natural numbers i, l, n, $j_1, j_2, \dots, j_l$, $(2 \le l \le n;\ i \le n;\ 1 \le j_1 < j_2 < \dots < j_l \le n)$ the equalities

$$\mathcal{P}_{(X_{j_1}, X_{j_2}, ..., X_{j_l})|\bar{\xi}_{1n}} = \mathcal{P}_{X_{j_1}|\xi_{j_1}} * \mathcal{P}_{X_{j_2}|\xi_{j_2}} * \dots * \mathcal{P}_{X_{j_l}|\xi_{j_l}},$$

$$\mathcal{P}_{X_i|\bar{\xi}_{1n}} = \mathcal{P}_{X_i|\xi_i},$$

are fulfilled, where $\mathcal{P}_{X|Y}$ is the conditional distribution of the value X under the condition Y. The conditionally independent sequence $\{X_i\}_{i\ge1}$ in (3) is called a sequence with chain dependence (see. [9],[10],[11]) if $\{\xi_i\}_{i\ge1}$ is a finite Markov chain with discrete time.

We are considering the case, when the members of sequence $\{\xi_i\}_{i\ge1}$ are independent and identically distributed discrete random variables

$$\Xi = \{b_1, b_2, ..., b_r\}\ ;\ \ P(\xi_1 = b_i) = p_i\ ,\ i = \overline{1, r}\ ,\ p_1 + p_2 + \dots + p_r = 1.$$

Remark: If the members of a sequence $\{X_i\}_{i\ge1}$ represent elements of a statistical sample or observation, they are called conditionally independent (or, accordingly, chain dependent) observations.

To determine the accuracy of estimates constructed with dependent observations, it is necessary to study the asymptotics of the sums of dependent random variables. In this process, the research methods of distribution of sums of independent random variables are is extended to dependent random variables. Markov dependence is considered in many practical and theoretical problems. It is one of the forms of weak dependence. Other forms of weak dependence are also known. In the works [6] and [9] are considered limit distributions of sums of conditionally independent random variables and limit theorems for functions defined on the Markov chain. Many authors consider sums of random variables whose joint distribution is determined by some "control" sequence of random elements. I. Bokuchava, Z. Kvatadze and T. Shervashidze established limit distributions of normed sums for conditionally independent random variables (see. [7], [8]) and for random variables with chain dependence ([10], [11]).

The study of the asymptotic behavior of the sums of dependent random variables made it possible to consider dependent observations in the theory of statistical estimates. From the second half of the twentieth century, the construction of statistical estimates with dependent observations began. In this regard, the question of constructing a non-parametric estimate of the density on the dependent observations is particularly relevant. In the work of Yakowitz Sidney [12], estimates of density and regression coefficients are constructed from observations bound in a Markov chain. The accuracy of the density estimation constructed by dependent observations by the metric L_2 is known (see [13]). In the series of works [14-16], the Rosenblatt-Parzen-type kernel estimates of the density are constructed with chain dependence observations and conditionally independent observations. Their approximation accuracy with metrics L_2 and L_1 is considered. Z. Kvatadze and B. Pharjiani's case $r=2$, the accuracy of the estimations constructed with both types of dependent observations was determined by metric L_2 (see [14]). In general, the estimate accuracy constructed for observations with chain dependent is established L_1 (see [15]) and L_2 metrics (see [16]).

Methodology

During the proof of the theorem is applied the method I. Bokuchava, T. Shervashidze and Z. Kvatadze presented by in [10, 11]. Using this method, they determined the limit distributions of the normed sums of conditionally independent sequences (see [8]) and sequenge with chain dependence. The asymptotics of the conditional and unconditional distributions of the geometric mean of conditionally independent random variables and random variables with chain dependence were investigated (see [17], [18]). This method became possible to use in the theory of statistical estimations (see [14]). The method uses the decomposition of the sum to be estimated into sums corresponding to the values of the control sequence. On the fixed trajectory $\bar{\xi}_{1n}=(\xi_1,\xi_2,...,\xi_n)$ of the control sequence (5), the second components of the sequence (observations $\{X_i\}_{i\geq 1}$) become independent. The sums obtained from them are uncorrelated and consist of independent and uniformly distributed random variables.

Main Results

Let's consider the sequence (5). ξ_i, $i=1,2,...,$ are independent and identically distributed discrete random variables. Let's assume that

$$\Xi=\{b_1,b_2,...,b_r\}\ ;\ \ P(\xi_1=b_i)=p_i\ ,\ i=\overline{1,r}\ ,\ p_1+p_2+...+p_r=1 .$$

Let us fix the trajectory $\bar{\xi}_{1n}=(\xi_1,\xi_2,...,\xi_n)$ of the sequence $\{\xi_i\}_{i\geq 1}$. In this case, we denote by the values $\nu_n(1)$, $\nu_n(2),\dots,\nu_n(r)$, the frequencies of accepting the values $b_1,b_2,...,b_r$ (respectively) by the members ξ_1, ξ_2,..., ξ_n of the sequence $\{\xi_i\}_{i\geq 1}$.

$$\nu_n(i)=\sum_{k=1}^{n}\mathrm{I}_{(\xi_k=b_i)}\ , \qquad i=\overline{1,r}\ ,$$

where $\mathrm{I}_{(\cdot)}$ is the indicator function. Obviously the equality is fair

$$\nu_n(1)+\nu_n(2)+\ldots+\nu_n(r)=n.$$

Theorem. Let's consider the sequence (5). Let's say that members of the control sequence $\{\xi_i\}_{i\ge1}$ $(\xi_i:\Omega\to\{b_1,b_2,\ldots,b_r\})$ are independent, identically distributed, discrete random variables $(\Xi=\{b_1,b_2,\ldots,b_r\})$, $P(\xi_1=b_k)=p_k$, $k=\overline{1,r}$, $p_1+p_2+\ldots+p_r=1$. Let us say that the sequence of positive numbers a_n satisfies the conditions (2). Let's say that for each function $\Psi:\Xi\to R^1$, for that $E\Psi(\xi_i)<\infty$ when $n\to\infty$ occurs the convergence

$$\frac{1}{n}\sum_{j=1}^{n}\Psi(\xi_i)\to E\Psi(\xi_1) \qquad \text{a. e.} \qquad (6)$$

Let's say that the members of the sequence $\{X_i\}_{i\ge1}$ are conditionally independent observations on some random variable X and the conditional distributions $\mathcal{P}_{X_1|\xi_1=b_i}$, $i=\overline{1,r}$ have unknown densities $f_i(x)$, $i=\overline{1,r}$ with a compact supports. If inequalities

$$D\left(\frac{\nu_n(i)}{n}\right)\le\frac{c_i}{\sqrt{n}}, \qquad i=\overline{1,r}, \qquad (7)$$

are fulfilled for frequencies $\nu_n(i)$, $i=\overline{1,r}$, then for each natural number n, the density estimate $\overline{f}(x)=\sum_{i=1}^{r}p_i f_i(x)$ is the sum $\hat{f}_n(x,a_n)=\frac{a_n}{n}\sum_{i=1}^{n}K(a_n(x-X_i))$, where $K(x)\in K^{\bullet}$, and for the value $J(a_n)=\int_{-\infty}^{\infty}\left|\hat{f}_n(x,a_n)-\overline{f}(x)\right|dx$ the estimate

$$EJ(a_n)\le\sqrt{\frac{a_n}{n}}\alpha\sqrt{\frac{2}{\pi}}\sum_{i=1}^{r}\int_{-\infty}^{\infty}\sqrt{p_i f_i(x)}dx+$$

$$+\frac{\beta}{2a_n^2}\sum_{i=1}^{r}p_i\sup_{a>0}\int_{-\infty}^{\infty}\left|(f_i*\varphi_a)''(x)\right|dx+\frac{1}{\sqrt[4]{n}}\sum_{i=1}^{r}\sqrt{c_i}+o\left(\sqrt{\frac{a_n}{n}}\right), \quad (8)$$

is satisfied, where

$$\alpha=\sqrt{\int_{-\infty}^{\infty}K^2(x)dx}, \qquad \beta=\int_{-\infty}^{\infty}x^2K(x)dx, \quad \varphi_a\in\Phi.$$

If also $f_i(x)\in F$, $i=\overline{1,r}$, then

$$EJ(a_n)\le\sqrt{\frac{a_n}{n}}\alpha\sqrt{\frac{2}{\pi}}\sum_{i=1}^{r}\int_{-\infty}^{\infty}\sqrt{p_i f_i(x)}dx+\frac{\beta}{2a_n^2}\sum_{i=1}^{r}p_i\int_{-\infty}^{\infty}\left|f_i''(x)\right|dx+\frac{1}{\sqrt[4]{n}}\sum_{i=1}^{r}\sqrt{c_i}+o\left(\sqrt{\frac{a_n}{n}}\right) \quad (9)$$

Proof. Of Theorem. Let's apply the method used in [10] [11]. Let's decompose the sum $\hat{f}_n(x,a_n)$ into r sums. Let's fix the trajectory $\overline{\xi}_{1n}=(\xi_1,\xi_2,\ldots,\xi_n)$. For each number i $(1\le i\le r)$, we separately group the terms of those observations $X_1,X_2,\ldots,X_n$ out of the sum of $\hat{f}_n(x,a_n)$, the corresponding $\xi_1,\xi_2,\ldots,\xi_n$ control random variables of which took the value 9. Let's renumber the members of each sum.

$$\tau_0(i)=0,\ \ \tau_m(i)=\min\{j\,|\,\tau_{m-1}<j\le n;\ \xi_i=b_i\};\ i=\overline{1,r},\ m=\overline{1,\nu_n(i)}.$$

It turns out r sequence of indices

$$\tau_1(i),\ \tau_2(i),\ \dots,\ \tau_{\nu_n(i)}(i),\ \ i=\overline{1,r}.$$

The equalities are fulfilled for them

$$\xi_{\tau_m(i)}=b_i,\quad i=\overline{1,r},\quad m=\overline{1,\nu_n(i)}.$$

Now let's present the sum $\hat{f}_n(x,a_n)$ as following

$$\hat{f}_n(x,a_n)=\sum_{i=1}^{r}\frac{\nu_n(i)}{n}\hat{f}_{in}(x,a_n),$$

where

$$\hat{f}_{in}(x,a_n)=\frac{a_n}{\nu_n(i)}\sum_{m=1}^{\nu_n(i)}K(a_n(x-X_{\tau_m(i)})),\ \ i=\overline{1,r}.$$

It is obvious that if $\nu_n(i)=0$, then the corresponding sum $\hat{f}_{in}(x,a_n)$, $i=\overline{1,r}$ does not exist.

Let us show that are finite $E\hat{f}_n(x,a_n)$ and $D\hat{f}_n(x,a_n)$ values.

Let's represent the value $E\hat{f}_n(x,a_n)$ as a conditional mathematical expectation on the fixed trajectory $\bar{\xi}_{1n}$

$$E\hat{f}_n(x,a_n)=E\{E(\hat{f}_n(x,a_n)\,|\,\bar{\xi}_{1n})\}=E\{E(\sum_{i=1}^{r}\frac{\nu_n(i)}{n}\hat{f}_{in}(x,a_n)\,|\,\bar{\xi}_{1n})\}.$$

Let us take into account that, random variables $\nu_n(i)$, ($i=\overline{1,r}$) are commensurate with respect to the σ-algebra induced by the partitioning of the space Ω generated when fixing the trajectory $\bar{\xi}_{1n}$ (see [19]). Therefore, we can take them out of the determined by condition $\bar{\xi}_{1n}$ conditional mathematical expectation sign. Random variables $X_{\tau_m(i)}$, $m=\overline{1,\nu_n(i)}$ are independent when the trajectory $\bar{\xi}_{1n}$ is fixed. They have the same conditional distribution $\mathcal{P}_{X_1|\xi_1=b_i}$ with density $f_i(x)$.

$$E\hat{f}_n(x,a_n)=\sum_{i=1}^{r}E\{\frac{\nu_n(i)}{n}(E\frac{a_n}{\nu_n(i)}\sum_{m=1}^{\nu_n(i)}K(a_n(x-X_{\tau_m(i)}))\,|\,\bar{\xi}_{1n})\}=$$

$$=\sum_{i=1}^{r}E\{\frac{\nu_n(i)}{n}(E\frac{a_n}{\nu_n(i)}\nu_n(i)K(a_n(x-X_{\tau_1(i)}))\,|\,\bar{\xi}_{1n})\}=\sum_{i=1}^{r}a_n\int_{-\infty}^{\infty}K(a_n(x-u))f_i(u)du\,E(\frac{\nu_n(i)}{n})$$

Based on the condition (6), the equality $E\dfrac{\nu_n(i)}{n}=p_i$ is fulfilled. Let's take into account that $K(x)$ is an even function and transform the variable $a_n(u-x)=t$ under the integral sign. Will be obtained the equality

$$E\hat{f}_n(x,a_n)=\sum_{i=1}^{r}p_i\int_{-\infty}^{\infty}K(t)f_i(\frac{t}{a_n}+x)dt.$$

$K(x)$ is a density and $f_i(x)$ is a density bounded by a finite constant. Therefore, $E\hat{f}_n(x,a_n)$ is a finite quantity.

Let us show that the quantity $D\hat{f}_n(x,a_n)$ is finite.

$$D\hat{f}_n(x,a_n)=E\{E([\hat{f}_n(x,a_n)-E\hat{f}_n(x,a_n)]^2\,|\,\xi_{1n})\}=$$

$$= E\{E([\sum_{i=1}^{r}\frac{\nu_n(i)}{n}\hat{f}_{in}(x,a_n) - E\sum_{i=1}^{r}\frac{\nu_n(i)}{n}\hat{f}_{in}(x,a_n)]^2 \mid \xi_{1n})\} =$$

$$= E\{E([\sum_{i=1}^{r}\frac{\nu_n(i)}{n}(\hat{f}_{in}(x,a_n) - E\hat{f}_{in}(x,a_n)]^2 \mid \xi_{1n})\}$$

On the fixed trajectory $\bar{\xi}_{1n}$, the quantities $K\left(a_n\left(x - X_{\tau_m(i)}\right)\right)$, $m = \overline{1,\nu_n(i)}$, $i = \overline{1,r}$ and accordingly, the sums $\hat{f}_{in}(x,a_n)$, $i = \overline{1,r}$ are independent.

$$D\hat{f}_n(x,a_n) = E\{E(\sum_{i=1}^{r}(\frac{\nu_n(i)}{n})^2[\hat{f}_{in}(x,a_n) - E\hat{f}_{in}(x,a_n)]^2 \mid \xi_{1n}) =$$

$$= \sum_{i=1}^{r}E\{(\frac{\nu_n(i)}{n})^2 E([\hat{f}_{in}(x,a_n) - E\hat{f}_{in}(x,a_n)]^2 \mid \xi_{1n})\} =$$

$$= \sum_{i=1}^{r}E\{(\frac{\nu_n(i)}{n})^2 E([\frac{a_n}{\nu_n(i)}\sum_{m=1}^{\nu_n(i)}(K\left(a_n\left(x - X_{\tau_m(i)}\right)\right) - EK\left(a_n\left(x - X_{\tau_m(i)}\right)\right))]^2 \mid \xi_{1n})\} =$$

$$= \sum_{m=1}^{r}E\{(\frac{\nu_n(i)}{n})^2(\frac{a_n}{\nu_n(i)})^2 E(\sum_{j=1}^{\nu_n(i)}[K\left(a_n\left(x - X_{\tau_m(i)}\right)\right) - EK\left(a_n\left(x - X_{\tau_m(i)}\right)\right)]^2 \mid \xi_{1n})\} =$$

$$= \sum_{i=1}^{r}E\{\frac{a_n^2}{n^2}E(\nu_n(i)E[K\left(a_n\left(x - X_{\tau_1(i)}\right)\right) - EK\left(a_n\left(x - X_{\tau_1(i)}\right)\right)]^2 \mid \xi_{1n})\} =$$

$$= \frac{a_n^2}{n}\sum_{i=1}^{r}\int_{-\infty}^{\infty}[K\left(a_n(x-u)\right) - \int_{-\infty}^{\infty}K\left(a_n(x-y)\right)f_i(y)dy]^2 f_i(u)du E(\frac{\nu_n(i)}{n})$$

Apply Equation $E\frac{\nu_n(i)}{n} = p_i$ again. Let's apply the same variable transformation inside the integral sign as when considering the expression $E\hat{f}_n(x,a_n)$. Will be obtained the equality

$$D\hat{f}_n(x,a_n) = \frac{a_n}{n}\sum_{i=1}^{r}p_i\int_{-\infty}^{\infty}[K(t) - \int_{-\infty}^{\infty}K\left(a_n(x-y)\right)f_i(y)dy]^2 f_i(\frac{t}{a_n} + x)dt.$$

Considering the properties of $K(x)$ and equalities (2), it is clear that $D\hat{f}_n(x,a_n)$ is finite.

Let's estimate $EJ(a_n)$.

$$EJ(a_n) = E[E(\int_{-\infty}^{\infty}\left|\hat{f}_n(x,a_n) - \bar{f}(x)\right|dx \mid \bar{\xi}_{1n})] \le$$

$$\le \sum_{i=1}^{r}E[E(\int_{-\infty}^{\infty}\left|\frac{\nu_n(i)}{n}\hat{f}_{in}(x,a_n) - p_i f_i(x)\right|dx\left|\bar{\xi}_{1n}\right.)] \equiv \sum_{i=1}^{r}A_i \qquad (10)$$

Each summand A_i is estimated in the same way.

$$A_i = E[E(\int_{-\infty}^{\infty}\left|\frac{\nu_n(i)}{n}\hat{f}_{in}(x,a_n) - p_i f_i(x)\right|dx\left|\bar{\xi}_{1n}\right.)] \le$$

$$\le E[E(\int_{-\infty}^{\infty}\left|\frac{\nu_n(i)}{n}\hat{f}_{in}(x,a_n) - \frac{\nu_n(i)}{n}f_i(x)\right|dx\left|\bar{\xi}_{1n}\right.)] + E[E(\int_{-\infty}^{\infty}\left|\frac{\nu_n(i)}{n}f_i(x) - p_i f_i(x)\right|dx\left|\bar{\xi}_{1n}\right.)] \le$$

$$\leq E\{E[\int_{-\infty}^{\infty}\left|\frac{\nu_n(i)}{n}\right|\left|\hat{f}_{in}(x,a_n)-f_i(x)\right|dx\left|\bar{\xi}_{1n}\right]\}+E\{E[\int_{-\infty}^{\infty}\left|\frac{\nu_n(i)}{n}-p_i\right|\left|f_i(x)\right|dx\left|\bar{\xi}_{1n}\right]\}=A_{i1}+A_{i2}$$

$\hat{f}_{in}(x,a_n)$ is a density estimate $f_i(x)$ constructed from independent and identically distributed observations on a fixed trajectory $\bar{\xi}_{1n}$. To estimate the quantity $[E\int_{-\infty}^{\infty}\left|\hat{f}_{in}(x,a_n)-f_i(x)\right|dx\left|\bar{\xi}_{1n}\right]\}$, we apply inequality (3).

$$A_{i1}=E\{\left|\frac{\nu_n(i)}{n}\right|[E\int_{-\infty}^{\infty}\left|\hat{f}_{in}(x,a_n)-f_i(x)\right|dx\left|\bar{\xi}_{1n}\right]\}=$$

$$=E\{\left|\frac{\nu_n(i)}{n}\right|E[\int_{-\infty}^{\infty}\left|\frac{a_n}{\nu_n(i)}\sum_{m=1}^{\nu_n(i)}k(a_n(x-X_{\tau_m(i)}))-f_i(x)\right|dx\mid\xi_{1n}]\}\leq$$

$$\leq E\{\left|\frac{\nu_n(i)}{n}\right|[\sqrt{\frac{2}{\pi}}\sqrt{\frac{a_n}{\nu_n(i)}}\alpha\int_{-\infty}^{\infty}\sqrt{f_i(x)}dx+\frac{1}{a_n^2}\frac{\beta}{2}\sup_{a>0}\int_{-\infty}^{\infty}\left|\left(f*\varphi_a\right)''(x)\right|dx+o(\sqrt{\frac{a_n}{\nu_n(i)}})]\}$$

According to the condition of theorem (6), the equations are fulfilled

$$E\frac{\nu_n(i)}{n}=p_i\ ,\ \lim_{n\to\infty}\frac{\nu_n(i)}{n}=p_i\,.$$

Therefore $\nu_n(i)\sim np_i$ and accordingly we have equality.

$$o\left(\frac{a_n}{\nu_n(i)}\right)=o\left(\frac{a_n}{n}\right).$$

The following assessment of the summand A_{i1} is valid

$$A_{i1}\leq\sqrt{\frac{2}{\pi}}\alpha\int_{-\infty}^{\infty}\sqrt{f_i(x)}dxE\{\frac{\nu_n(i)}{n}\sqrt{\frac{a_n}{\nu_n(i)}}\}+\frac{1}{a_n^2}\frac{\beta}{2}\sup_{a>0}\int_{-\infty}^{\infty}\left(f*\varphi_a\right)''(x)dxE(\frac{\nu_n(i)}{n})+$$

$$+o(\sqrt{\frac{a_n}{n}})E\frac{\nu_n(i)}{n}=\sqrt{\frac{a_n}{n}}\frac{\alpha\sqrt{2}}{\sqrt{\pi}}\int_{-\infty}^{\infty}\sqrt{f_i(x)}dxE\sqrt{\frac{\nu_n(i)}{n}}+\frac{\beta p_i}{2a_n^2}\sup_{a>0}\int_{-\infty}^{\infty}\left(f*\varphi_a\right)''(x)dx+p_io(\sqrt{\frac{a_n}{n}})\,.$$

Let's apply the inequality $E\sqrt{\frac{\nu_n(i)}{n}}\leq\sqrt{E\frac{\nu_n(i)}{n}}=\sqrt{p_i}$.

$$A_{i1}\leq\sqrt{\frac{a_n}{n}}\alpha\sqrt{\frac{2p_i}{\pi}}\int_{-\infty}^{\infty}\sqrt{f_i(x)}dx+\frac{\beta p_i}{2a_n^2}\sup_{a>0}\int_{-\infty}^{\infty}\left(f*\varphi_a\right)''(x)dx+p_io(\sqrt{\frac{a_n}{n}})\,.$$

We take the value $(\frac{\nu_n(i)}{n}-p_i)$ out of the sign of the conditional expectation (see [19]). Let's apply condition (6) and inequality (7). The following inequality will be obtained

$$A_{i2}=E[\left|\frac{\nu_n(i)}{n}-p_i\right|E(\int_{-\infty}^{\infty}\left|f_i(x)\right|dx\mid\xi_{1n})]=E\left|\frac{\nu_n(i)}{n}-p_i\right|\leq\sqrt{E(\frac{\nu_n(i)}{n}-p_i)^2}\leq\sqrt{\frac{C_i}{\sqrt{n}}}\,.$$

Finally, the following estimation of the summand A_i will be obtained

$$A_i\leq\sqrt{\frac{a_n}{n}}\alpha\sqrt{\frac{2p_i}{\pi}}\int_{-\infty}^{\infty}\sqrt{f_i(x)}dxE+\frac{\beta p_i}{2a_n^2}\sup_{a>0}\int_{-\infty}^{\infty}\left(f*\varphi_a\right)''(x)dx+p_io(\sqrt{\frac{a_n}{n}})+\frac{\sqrt{C_i}}{\sqrt[4]{n}}\,.$$

We apply this estimation in inequality (10) and will be obtained the estimation of theorem (8).

The estimation of (9) is obtained directly from (8). Let's apply the following fact shown in [5] while proving the inequality of Lemma (4). For $g(x) \in F$ class functions, the expression $\sup\limits_{a>0} \int_{-\infty}^{\infty} \left| (g * \varphi_a)''(x) \right| dx$ does not depend on the selection of the function $\varphi(x) \in \Phi$, and the equality is shown

$$\sup_{a>0} \int_{-\infty}^{\infty} \left| (g * \varphi_a)''(x) \right| dx = \int_{-\infty}^{\infty} \left| g''(x) \right| dx$$

Corollary 1. If under the conditions $K(x)$ of theorem is Bartlett's core

$$K(x) = \bar{K}(x) = \frac{3}{4}(1 - x^2)\mathrm{I}_{[|x| \le 1]}$$

Then for each natural number n is the estimation of the density $\bar{f}(x) = \sum_{i=1}^{r} p_i f_i(x)$ is presented by sum

$$\bar{f}_n(x, a_n) = \frac{3a_n}{4n} \sum_{i=1}^{n} (1 - [a_n(x - X_i)]^2)\mathrm{I}_{[|x - X_i| \le \frac{1}{a_n}]}$$

and for quantity $\bar{J}(a_n) = \int_{-\infty}^{\infty} \left| \bar{f}_n(x, a_n) - \bar{f}(x) \right| dx$, we have the estimate

$$E\bar{J}(a_n) \le \sqrt{\frac{3a_n}{5n}} \sqrt{\frac{2}{\pi}} \sum_{i=1}^{r} \int_{-\infty}^{\infty} \sqrt{p_i f_i(x)} dx +$$

$$+ \frac{0{,}1}{a_n^2} \sum_{i=1}^{r} p_i \sup_{a>0} \int_{-\infty}^{\infty} \left| (f_i * \varphi_a)''(x) \right| dx + \frac{1}{\sqrt[4]{n}} \sum_{i=1}^{r} \sqrt{c_i} + o\left(\sqrt{\frac{a_n}{n}}\right) . \qquad (11)$$

If also $f_i(x) \in F$, $i = \overline{1, r}$, then the inequality is fulfilled

$$E\bar{J}(a_n) \le \sqrt{\frac{3a_n}{5n}} \sqrt{\frac{2}{\pi}} \sum_{i=1}^{r} \int_{-\infty}^{\infty} \sqrt{p_i f_i(x)} dx + \frac{0{,}1}{a_n^2} \sum_{i=1}^{r} p_i \int_{-\infty}^{\infty} \left| f_i''(x) \right| dx + \frac{1}{\sqrt[4]{n}} \sum_{i=1}^{r} \sqrt{c_i} + o\left(\sqrt{\frac{a_n}{n}}\right) . \qquad (12)$$

Proof. It is clear that $\bar{K}(-x) = \bar{K}(x)$, $\bar{K}(x) \le \frac{3}{4}$ and $\bar{K}(x)$ have a compact support. Therefore, it satisfies the conditions of the theorem. The inequalities (11) and (12) are obtained from (8) and (9) if we calculate α and β the quantities

$$\alpha = \sqrt{\int_{-\infty}^{\infty} \bar{K}^2(x) dx} = \frac{3}{4} \sqrt{\int_{-1}^{1} (1 - x^2)^2 dx} = \sqrt{\frac{3}{5}}$$

$$\beta = \int_{-\infty}^{\infty} x^2 \bar{K}(x) dx = \frac{3}{4} \int_{-1}^{1} (x^2 - x^4) dx = \frac{1}{5} .$$

Corollary 2. If we apply the sequence $a_n = \sqrt{n}$ under the conditions of Corollary 1, we obtain the estimation of density $\bar{f}(x) = \sum_{i=1}^{r} p_i f_i(x)$

$$\bar{f}_n^{\bullet}(x,a_n)=\frac{3}{4\sqrt{n}}\sum_{i=1}^{n}(1-[\sqrt{n}(x-X_i)]^2)\mathrm{I}_{[|x-X_i|\le\frac{1}{\sqrt{n}}]}\ .$$

The estimations are obtained from inequalities (11) and (12)

$$E\bar{J}^{\bullet}(a_n)\le\frac{1}{\sqrt[4]{n}}\sqrt{\frac{6}{5\pi}}\sum_{i=1}^{r}\int_{-\infty}^{\infty}\sqrt{p_i f_i(x)}dx+$$

$$+\frac{0,1}{n}\sum_{i=1}^{r}p_i\sup_{a>0}\int_{-\infty}^{\infty}\left|\left(f_i*\varphi_a\right)''(x)\right|dx+\frac{1}{\sqrt[4]{n}}\sum_{i=1}^{r}\sqrt{c_i}+o(\frac{1}{\sqrt[4]{n}})$$

$$E\bar{J}^{\bullet}(a_n)\le\frac{1}{\sqrt[4]{n}}\sqrt{\frac{6}{5\pi}}\sum_{i=1}^{r}\int_{-\infty}^{\infty}\sqrt{p_i f_i(x)}dx+\frac{0,1}{n}\sum_{i=1}^{r}p_i\int_{-\infty}^{\infty}\left|f_i''(x)\right|dx+\frac{1}{\sqrt[4]{n}}\sum_{i=1}^{r}\sqrt{c_i}+o(\frac{1}{\sqrt[4]{n}})$$

Discussion

Let's note that when proving the theorem, the trajectory of the control sequence $\bar{\xi}_{1n}=\left(\xi_1,\xi_2,..,\xi_n\right)$ is fixed and the quantity $EJ(a_n)$ is presented as $EJ(a_n)=E\left\{E(J(a_n)\left|\bar{\xi}_{1n})\right.\right\}$. This method makes it possible to use the independence of observations $X_1,X_2,...,X_n$ on a fixed trajectory. At this time, it becomes possible to expand the estimated sum by the method presented in [11] and [15]. Grouping of identically distributed values into one sum according to the values of the control sequence is used. Each sum is then represented as two sums of centered quantities. Estimations for one of them are written down using the methods used in [10] and [11]. The sums of the second type are evaluated by the classical results obtained in {5}. The measurability of the quantities $\nu_n(i)$ and their compositions with continuous functions (see [19]) with respect to the sigma algebra generated by the division of the probability space is used when fixing the trajectory $\bar{\xi}_{1n}$.

Conclusion.

With the method used in the paper, it is possible to determine the exact upper boundaries of the obtained estimates. The method gives the possibility to be used in determining the accuracy of other (including non-parametric) estimates.

We would like to be thankful to our colleague Prof. Tengiz Shervashidze and will honor his bright memory.

References

1. M. Rosenblatt (1956), Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27,** Chicago, 832-837.
2. E. Parzen (1962), On estimation of a probability density function and mode. *Ann. Math. Statist.*33, Stanford, USA, 1065-1076.
3. G. S. Watson, M.R. Leadbetter (1963) On the estimation of the probability density. *Ann. Math. Statist.* 34, Toronto, : 480-491.
4. E. A. Nadaraya (1983), Nonparametric estimation of the probability density and regression curve. (in Russian) Tbilisi State Univ. Press, 1983.

5. Devroye L., Györfi L. Nonparametric density estimation: the L_1 view. Wiley series in probability and mathematical statistics, Canada: John Wiley & Sons; 1985. p. 367.
6. R. Z. Khasminski, On limiting distributions for sums of conditionally independent random variables. (Russian) Teor. Verojatnosti. i primenen. 6. (1961). №1,.119-125.
7. G. L. O'Brien, Limit theorems for sums of chain-dependent processes. J. Appl. Probability 11 (1974), 582-587.
8. Bokuchava I. V. (1984) Limit theorems for conditionally independent sequences. . (in Russian) Teor. Verojatnost. i Primenen.-MathNet.Ru XXIX. (1984). №1,. p. 192-193.1984. Theory of probability and its applications, 1985, 29:1, 190-196
9. Kvatadze Z., Shervashidze T. (2008) Some limit theorems for I.I.D. and conditionally independent random variables. The Second International Conference, "Problems of Cybernetics and Informatics". September 10-12, 2008. Baku, Azerbaijan. Section #4. "Applied Stochastic Analysis" www.pci2008.science.az/4/12.pdf. Azerbaijan National Academy of Sciences. INSTITUTE OF INFORMATION TECHNOLOGY. Printing House of "Information Technology" Baku. 2008. Vol. II. pp.217-219.
10. I. V. Bokuchava, Z. Kvatadze, T. Shervashidze, On limit theorems for random vectors controlled by a Markov chain. Probability theory and mathematical statistics, vol. I (Vilnius, 1985), 231-250, VNU Sci. Press, Utrecht, 1987.
11. Kvatadze Z., Shervashidze T. (1987) On limit theorems for conditionally independent random variable controlled by a finite Markov chain. Probability theory and mathematical statistics. (Proc. 5th Japan-USSR Symposium on Probability Theory. Kyoto. 1986). Lecture Notes in Mathematics, 1299: 250-259. Springer-Verlag. Berlin (Germany).
12. Yakowitz Sidney (1989) Nonparametric density and regression estimation for Markov sequences without mixing assumptions. 85721–Journal of Multivariate Analysis, 30: 124-136. Arisona, USA.
13. J. Meloche. Asymptotic behavior of the mean integrated squared error of kernel density estimators for dependent observations. Canadian Journal of Statistics. 2009., 18 (3): p. 205-211.
14. Z, Kvatadze, B. Phardjiani. On the Exsactness of Distribution Density Estimates Constructed by Some Class of Dependent Observations. Mathematics and Statistics. 2019 Vol. 7(4), pp. 135-145. SAN JOSE.
15. Kvatadze Z., Pharjiani B., Construction of a kernel density estimator of Rosenblatt-Parzen type by conditionally independent observations. Proceedings of A. Razmadze Mathematical institute. Vol. 173. 2019. issue 3, 93-102.
16. Pharjiani B., Kvatadze TS., Kvatadze Z., On an Application of Density Estimation Constructed by Means of Chain Dependent Samples. Report of XXXIV enlarged session of the seminar of I. Vekua Institute of Applied Mathematics (VIAM), of Ivane Javakhisvili Tbilisi State University (TSU). Sept. 16-19. 2020; Volume, 34. pp. 69-72.
17. Kvatadze Z., Shervashidze T., Some limit theorems for IID and Conditionally independent random variables. The second international Conference, "Problems of Cybernetics and

Informatics". September 10-12, 2008. Baku. Azerbaijan section N 4. "Applied Stochastic Analysis". Science 14/12.pdf. Azerbaijan National Academy of Sciences. INSTITUTE OF INFORMATION TECHNOLOGY. Printing House of "Information Technology" Baku. 2008. Vol. II, 217-219.

18. Kvatadze Z., Shervashidze T. On some Limit Theorems for sums and Products of Random Variables. Proceedings of A. Razmadze Mathematical institute. Vol. 150. 2009. 99-104.
19. Kvatadze Z., Kvatadze TS. Limiting Distribution of a Sequence of Functions Defined on a Markov Chain. Proceedings of A. Razmadze Mathematical institute. Vol. 174. 2020. issue 2, 199-205.

სიმკვრივის დამოკიდებული დაკვირვებებით აგებული შეფასების L_1 მეტრიკით სიზუსტის შესახებ

ბექნუ ფარჯიანი, ლევან ლაზაძე, ციალა ქვათაძე

რეზიუმე

თანამედროვე სტატისტიკურ კვლევებში მრავალი ამოცანის გადაჭრა, (როგორებიცაა მაგალითად საინვესტიციო და სადაზღვევო კომპანიების ფინანსური დამოუკიდებლობის კვლევები, საბანკო ინვესტიციების რისკების შეფასება, კომპანიების ფინანსური სტაბილურობის ინდიკატორების შეფასება და ა. შ.) მოითხოვს დამოკიდებული დაკვირვებების განხილვას. დამოკიდებული მონაცემებით შეფასებების აგება ემყარება დამოუკიდებელი მონაცემებით შეფასებათა აგების მდიდარ ისტორიულ გამოცდილებას. აგებული შეფასებების სიზუსტის დადგენისთვის საჭიროა დამოკიდებული შემთხვევითი სიდიდეების ჯამების ასიმპტოტიკის შესწავლა. ამ პროცესში ხდება დამოუკიდებელ შემთხვევით სიდიდეთა ჯამების განაწილების კვლევის მეთოდების დამოკიდებულ შემთხვევით სიდიდეებზე გავრცობა. განიხილება მარკოვული დამოკიდებულება. ის სუსტად დამოკიდებულების ერთ-ერთი სახეა. მრავალი ავტორი იხილავს ისეთ შემთხვევით სიდიდეთა ჯამებს, რომელთა ერთობლივი განაწილება განისაზღვრება შემთხვევით ელემენტთა რაიმე „მმართველი" მიმდევრობით. ი. ბოკუჩავამ, თ. შერვაშიძემ და ზ. ქვათაძემ დაადგინეს პირობითად დამოუკიდებელი (Bokuchava I. V. (1984) Limit theorems for conditionally independent sequences. (in Russian) Teor. Verojatnost. i Primenen.-MathNet.Ru XXIX. (1984). №1, pp. 192-193.1984. Theory of probability and its applications, 1985, 29:1, 190-196; Kvatadze Z., Shervashidze T. (2008) Some limit theorems for I.I.D. and conditionally independent random variables. The Second International Conference, "Problems of Cybernetics and Informatics". September 10-12, 2008. Baku, Azerbaijan. "Applied Stochastic Analysis" www.pci2008.science.az/4/12.pdf. Azerbaijan National Academy of Sciences. INSTITUTE OF INFORMATION TECHNOLOGY. Printing House of "Information Technology" Baku. 2008. Vol. II. pp. 217-219) და ჯაჭვურად დამოკიდებული

(I. V. Bokuchava, Z. Kvatadze, T. Shervashidze, On limit theorems for random vectors controlled by a Markov chain. Probability theory and mathematical statistics, vol. I (Vilnius, 1985), 231-250, VNU Sci. Press, Utrecht, 1987; Kvatadze Z., Shervashidze T. (1987) On limit theorems for conditionally independent random variable controlled by a finite Markov chain. Probability theory and mathematical statistics. (Proc. 5th Japan-USSR Symposium on Probability Theory. Kyoto. 1986). Lecture Notes in Mathematics, 1299: 250-259. Springer-Verlag. Berlin (Germany)) შემთხვევითი სიდიდეების ნორმირებული ჯამების ზღვარითი განაწილებები.

არაპარამეტრულ შეფასებათა თეორიაში მნიშვნელოვანი ადგილი ეთმობა განაწილების უცნობი სიმკვრივის შეფასებას. მ. როზენბლატის და ე. პარზენის ნაშრომებში (M. Rosenblatt (1956), Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist*. **27,** Chicago, 832-837, E. Parzen (1962), On estimation of a probability density function and mode. *Ann. Math. Statist.*33, Stanford, USA, 1065-1076) განხილულია დამოუკიდებელი დაკვირვებებით აგებული სიმკვრივის გულოვანი შეფასებების კლასი. ცნობილია ამ შეფასებების სიზუსტე L_2 (E. A. Nadaraya (1983), Nonparametric estimation of the probability density and regression curve. (in Russian) Tbilisi State Univ. Press, 1983) და L_1 (Devroye L., Györfi L. Nonparametric density estimation: the L_1 view. Wiley series in probability and mathematical statistics, Canada: John Wiley & Sons; 1985. p. 367) მეტრიკებით.

დამოკიდებულ შემთხვევით სიდიდეთა ჯამების ასიმპტოტური ყოფაქცევის შესწავლამ შესაძლებელი გახადა სტატისტიკურ შეფასებათა თეორიაში დამოკიდებული დაკვირვებების განხილვა. ცნობილია სიმკვრივის არაპარამეტრული შეფასება და რეგრესიის კოეფიციენტების შეფასებები მარკოვის ჯაჭვად შეკრული დაკვირვებებით (Yakowitz Sidney (1989) Nonparametric density and regression estimation for Markov sequences without mixing assumptions. 85721–Journal of Multivariate Analysis, 30: 124-136. Arisona, USA). ასევე დადგენილია დამოკიდებული დაკვირვებებით აგებული სიმკვრივის შეფასების სიზუსტე L_2 მეტრიკით (J. Meloche. Asymptotic behavior of the mean integrated squared error of kernel density estimators for dependent observations. Canadian Journal of Statistics. 2009., 18 (3): p. 205-211). ზ. ქვათაძის და ბ. ფარჯიანის მიერ აგებულია სიმკვრივის გულოვანი შეფასება პირობითად დამოუკიდებელი და ჯაჭვურად დამოკიდებული დაკვირვებებით და დადგენილია მათი სიზუსტე L_2 მეტრიკით იმ კერძო შემთხვევაში, როდესაც მმართველი მიმდევრობა იღებს ორ მნიშვნელობას ($r = 2$) (Z, Kvatadze, B. Phardjiani. On the Exsactness of Distribution Density Estimates Constructed by Some Class of Dependent Observations. Mathematics and Statistics. 2019 Vol. 7(4), pp. 135-145. SAN JOSE). ასევე მიღებულია ჯაჭვურად დამოკიდებული შეფასებებით აგებული გულოვანი შეფასების სიზუსტე L_2 მეტრიკით $r > 2$ შემთხვევაშიც.

წინამდებარე ნაშრომში პირობითად დამოუკიდებელი დაკვირვებებით აგებულია სიმკვრივის როზენბლატ–პარზენის ტიპის გულოვანი შეფასება. გამოყენებულია ი. ბოკუჩავას, თ. შერვაშიძის და ზ. ქვათაძის მიღებული შედეგები პირობითად დამოუკიდებელ შემთხვევით სიდიდეთა ჯამების ზღვარითი

განაწილების შესახებ და დადგენილია აგებული შეფასების სიზუსტე L_1 მეტრიკით. თეორემის დამტკიცების დროს გამოიყენება თ. შერვაშიძის და ზ. ქვათაძის მიერ განხილული მეთოდი. დაფიქსირებულ $\bar{\xi}_{1n}=\left(\xi_1,\xi_2,\ldots,\xi_n\right)$ ტრაექტორიაზე ხდება შესაფასებელი ჯამის დაშლა მმართველი მიმდევრობის მდგომარეობების შესაბამის ჯამებად. ეს ჯამები არაკორელირებულებია. დაფიქსირებულ ტრაექტორიაზე თითოეული ჯამის შემადგენელი შესაკრებები დამოუკიდებლებია. გამოყენებულია ჯაჭვის მდგომარეობებზე განსაზღვრული $\nu_n(i)$ $(i=\overline{1,\ s})$ ფუნქციების ზომადობა (Kvatadze Z., Kvatadze TS. Limiting Distribution of a Sequence of Functions Defined on a Markov Chain. Proceedings of A. Razmadze Mathematical institute. Vol. 174. 2020. issue 2, 199-205) $\bar{\xi}_{1n}$ ტრაექტორიის დაფიქსირებით მიღებული ალბათური სივრცის დაყოფით ინდუცირებული σ ალგებრის მიმართ.

გამოყენებული მეთოდი შესაძლებლობას იძლევა მიღებულ იქნას დამოკიდებული დაკვირვებებით აგებული სხვა ტიპის შეფასების სიზუსტე.