Reports of Enlarged Sessions of the
Seminar of I. Vekua Institute
of Applied Mathematics
Volume 34, 2020

ON AN APPLICATION OF DENSITY ESTIMATION CONSTRUCTED BY MEANS OF CHAIN DEPENDENT SAMPLES

Beqnu Pharjiani Tsiala Kvatadze Zurab Kvatadze

Abstract. We consider r (r >1) sequences of random variables. The members of each sequence are independent and identically distributed random variables. By application of Markov finite stationary regular chain a sequence of chain dependent random variables is obtained. We use this sequence as a sample and the Bartlett kernel to construct Rosenblatt-Parzen type estimation for the density. Its accuracy is determined by L_1 and L_2 metrics. The obtained results are refined in the case of smoothness coefficient $a_n = \sqrt{n}$. One example of an application of this estimation is presented.

Keywords and phrases: Markov Chain, kernel estimation, dependent samples.

AMS subject classification (2010): 62G05, 62G07.

1 Introduction. Nonparametric estimation of distribution density represents one of the topical issues in mathematical statistics. Rosenblatt-Parzen-type kernel estimations ([1-4]) and projection-type estimations ([3, 5]]) are known for independent observations of density. According to the constructed estimations, the accuracy of the density approximation is considered by L_2 ([1-3]) and L_1 ([4, 5]) metrics.

From the 1980s of XX century began the construction of estimations with dependent observations ([6, 7]). It was considered the Markov's dependence, which is one of the forms of weak dependence. After determining the limit distribution of the sums of conditionally independent and chain-dependent random values [8], it became possible to construct a density estimation with such type dependent observations [9].

2 Content. Let's consider in narrow sense stationary two-component random sequence defined on a probability space (Ω,F,P)

$$\{\xi_i, X_i\}_{i\geq 1}\,, \tag{1}$$

where $X_i : \Omega \to R^m$, and $\xi_i : \Omega \to \{b_1, b_2, ..., b_r\}$, $(b_i \in R^1,\ i = \overline{1,r})$ is a finite homogeneous regular Markov chain with initial probabilities $\pi = (\pi_1, \pi_2, ...\pi_r)$, $\pi_i = P\left(\xi_1 = b_i\right)$, $i = \overline{1,r}$ and with the matrix of transition probabilities $P = (p_{ij})_{i,j=\overline{1,r}}$ (see [8]).

Definition 1. A sequence $\{X_i\}_{i\geq 1}$ from (1) is called a chain-dependent sequence if for a natural n the trajectory of the chain $\overline{\xi}_{1n} = (\xi_1, \xi_2, .., \xi_n)$ is fixed, then the random variables $X_1, X_2, \ldots, X_n$ become independent and for all indices $j_1, j_2, \ldots, j_k$ $i,$ $k,$ $(2 \leq k \leq n\ ; i \leq n;\ 1 \leq j_1 < j_2 < \cdots < j_k \leq n)$ the following equalities are valid:

$$P_{\left(X_{j_1},X_{j_2},\ldots,X_{j_k}\right)|\xi_{1n}} = P_{X_{j_1}|\xi_{j_1}} \times P_{X_{j_2}|\xi_{j_2}} \times \cdots \times P_{X_{j_k}|\xi_{j_k}}, P_{X_i|\xi_{1n}} = P_{X_i|\xi_i},$$

where $P_{X|Y}$is the conditional distribution of X for a given Y.

Example. (A construction of a chain-dependent sequence). Let's consider the r sequences of independent random vectors

$$\{X_1^\alpha, X_2^\alpha, ..., X_n^\alpha, ...\}, \alpha = \overline{1, r}.$$

Suppose that for each index α the components of the corresponding member of the sequence have the same distribution P_α, $\alpha = \overline{1, r}$. If in i-th step the chain $\{\xi_i\}_{i\geq 1}$ takes the value: $\xi_i = b_\alpha$, $(1 \leq \alpha \leq r)$, then from $(X_i^1, X_i^2, , ..., X_i^r)^T$ take the value X_i^α, $(1 \leq \alpha \leq r)$ and call it Y_i. Thus we obtain chain-dependent sequence $Y_i = \sum\limits_{\alpha=1}^{r} X_i^\alpha I_{(\xi_i = b_\alpha)}$, where I_A is the indicator of a set A.

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

$$\lim_{n\to\infty} a_n = \infty, \quad \lim_{n\to\infty} \frac{a_n}{n} = 0.$$

Let us interpret the terms of our chain-dependent sequence $\{X_i\}_{i\geq 1}$ as the observations of certain random variable X . Assume further that the conditional distributions $P_{X_i|\xi_i = b_i}$ have unknown densities $f_i(x)$, $i = \overline{1, r}$.

Theorem 1. *Let $f_i \in L_2(-\infty, \infty)$, each f_i, $i = \overline{1, r}$ is continuously differentiable up to order s ($s \geq 2$, s is an even number) and ${f_i}^{(s)}(x)$ is a continuous bounded function. Let $k(x)$ be the function with properties: $k(x) \in L_2(-\infty, \infty)$, $\int_{-\infty}^{\infty} k(x)\, dx = 1$, $k(-x) = k(x)$, $\sup|k(x)| \leq A < \infty$, $\int_{-\infty}^{\infty} x^i k(x)\, dx = 0$, $i = 1, 2, \ldots, s-1$; $\int_{-\infty}^{\infty} x^s k(x)\, dx \neq 0$, $\int_{-\infty}^{\infty} x^s |k(x)|\, dx < \infty$. Then for each natural number n, the density estimation for $\overline{f}(x) = \sum_{i=1}^r \pi_i f_i(x)$ is the sum $\hat{f}_n(x, a_n) = \frac{a_n}{n} \sum\limits_{j=1}^{n} k(a_n(x - X_j))$ and for $u(a_n) := E\int_{-\infty}^{\infty} \left[\hat{f}_n(x, a_n) - \bar{f}(x)\right]^2 dx$ we have the inequality*

$$u(a_n) \leq \left(\sum_{i=1}^{r} M_i\right)^2 + \frac{a_n}{n}\int_{-\infty}^{\infty} k^2(x)\, dx + \left(\sum_{i=1}^{r}\left(C_i(\pi, P)n^{-1} + {\pi_i}^2\right)\right) o\left(\frac{a_n}{n}\right), \qquad (2)$$

where

$$\alpha = \int_{-\infty}^{\infty} x^s k(x)\, dx, \quad M_i = T_i^{1/2} + \left(\frac{c_i(\pi, p)}{n}\int_{-\infty}^{\infty} f_i^2(x)\, dx\right)^{\frac{1}{2}},$$

$$T_i = \left({a_n}^{-2s}\frac{\alpha^2}{(s!)^2}\int_{-\infty}^{\infty}\left[f_i^{(s)}(x)\right]^2 dx + o\left(a_n^{-2s}\right)\right)\left(\frac{c_i(\pi, p)}{n} + {\pi_i}^2\right), \quad i = \overline{1, r}.$$

Theorem 2. *Let's the densities $f_i(x)$ $i=\overline{1,r}$ have a compact support; f_i, $i=\overline{1,r}$ is absolutely continuous and has a derivative f_i' almost everywhere. f_i' is absolutely continuous and has derivative f_i'' almost everywhere. f_i'' is bounded and continuous. Moreover, let $k(x)$ be a bounded by finite constant even density with a compact support. Then the estimation of density $\bar{f}(x)$ is $\hat{f}_n(x,a_n)$ and for $J(a_n):=\int_{-\infty}^{\infty}\left|\hat{f}_n-\bar{f}(x)\right|dx$ we have the inequality*

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

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{r}\sqrt{c_i(\pi,P)}+o\left(\sqrt{\frac{a_n}{n}}\right).$$

where $\alpha:=\sqrt{\int_{-\infty}^{\infty}k^2(x)\,dx}$ and $\beta:=\int_{-\infty}^{\infty}x^2k(x)\,dx$.

Corollary. *If in the conditions of Theorem 1 $k(x)=\bar{k}(x)=\frac{3}{4}(1-x^2)\mathrm{I}_{(|x|\le1)}$ is Bartlett's kernel, $a_n=\sqrt{n}$, and $f_i(x)\in W_2\bigcap L_2(-\infty,\infty)$, $i=\overline{1,r}$. Then for each natural n the sum $\bar{f}_n(x,a_n)=\bar{f}_n(x,\sqrt{n})=\frac{3}{4\sqrt{n}}\sum_{i=1}^{n}(1-n(x-X_i)^2)\mathrm{I}_{[|x-X_i|\le\frac{1}{\sqrt{n}}]}$ is the density estimation of $\bar{f}(x)=\sum_{i=1}^{r}\pi_i f_i(x)$ and according to (2) the following inequality is valid:*

$$\bar{u}(\sqrt{n})=E\int_{-\infty}^{\infty}\left[\bar{f}_n\left(x,\sqrt{n}\right)-\bar{f}(x)\right]^2dx\le\left(\sum_{i=1}^{r}\bar{M}_i\right)^2+\frac{3}{5\sqrt{n}}$$

$$+\left(\frac{1}{n}\left(\sum_{i=1}^{r}c_i(\pi,P)\right)+\sum_{i=1}^{r}{\pi_i}^2\right)o\left(\frac{1}{\sqrt{n}}\right),$$

where

$$\overline{M}_i=\bar{T}_i^{1/2}+\left(C_i n^{-1/2}\int_{-\infty}^{\infty}f_i^2(x)\,dx\right)^{\frac{1}{2}},$$

$$\overline{T}_i=\left(\frac{0.01}{n^2}\int_{-\infty}^{\infty}\left[f_i^{(2)}(x)\right]^2dx+o\left(\frac{1}{n^2}\right)\right)\left(c_i(\pi,P)+\pi_i^2\right),$$

$$i=\overline{1,r}.$$

3 Conclusions. It can be observed that during the proofs of the theorems on the fixed trajectory $\bar{\xi}_{1n}=(\xi_1,\xi_2,...,\xi_n)$ the transition from mathematical expectation to conditional mathematical expectation takes place $(E[\widehat{f}_n(x)]=E\{E[\widehat{f}_n(x)\left|\bar{\xi}_{1n}\right.]\})$. The sum under considered is divided into several summands. One of them will be reduced to a fixed trajectory on such form for that we apply the known results. The inequalities of Fubini, Holder and Jensen and the properties of the ruling sequences are applied for estimation of the remaining sums.

The applied method gives the possibility to construct nonparametric density estimations with other types of dependent observations. It can also be used to construct parametric estimations.

REFERENCES

1. Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, **27**, 3 (1956), 832–637.
2. Mania, G.M. Statistical Estimation of the Probability Distribution (Russian). *Tbilisi State Univ. Press*, Tbilisi, 1974.
3. Nadaraya, E.A. Nonparametric Estimation of the Probability Density and Regression Curve (Russian). *Tbilisi State Univ. Press*, 1983.
4. Mnatsakanov, R.M., Khmaladze, E.M. On the L_1-convergence of statistical kernel estimates of distribution densities (Russian). *Dokl. Akad. Nauk SSSR*, **258**, 5 (1981), 1052–1055.
5. Devroye, L., Dyorfi, L. Nonparametric Density Estimation: the L_1- view. Russian Edition, *Mir, Moscow*, 1988.
6. Yakowitz, S. Nonparametric density and regression estimation for Markov sequenges without mixing assumptions. *Journal of Multivariate Analysis*, **30** (1989), 124–136.
7. Meloce, J. Asymptotic behaviour of the mean integrated squared error of kernel density estimators for dependent observations. *Canadian Journal of Statistics*, **18**, 3 (2009), 205–211.
8. Bokuchava, I.V., Kvatadze, Z.A., Shervashidze, T.L. On limit theorems for random vectors controlled by a Markov chain. *Probability theory and mathematical statistics*, **1** (1985), 231–250.
9. Kvatadze. Z, Phardjiani, B. On the exactness of distribution density estimates constructed by some class of dependent observations. *Mathematics and Statistics*, **7**, 4 (2019), 135–145.

Received 29.05.2020; revised 27.09.2020; accepted 23.10.2020.

Author(s) address(es):

Beqnu Pharjiani
Department of Mathematics, Georgian Technical University
Kostava str. 77, 0175 Tbilisi, Georgia
E-mail: beqnufarjiani@yahoo.com

Tsiala Kvatadze, Zurab Kvatadze
Department of Mathematics, Georgian Technical University
Kostava str. 77, 0175 Tbilisi, Georgia
E-mail: ttkvatadze@gmail.com, zurakvatadze@yahoo.com