15th Triennial World Congress, Barcelona, Spain

NONPARAMETRIC RATIO ESTIMATION OF A MULTIVARIATE DENSITY DERIVATIVES FROM DEPENDENT OBSERVATIONS

Gennady Koshkin and Vyacheslav Vasil'iev

Department of Applied Mathematics and Cybernetics,
Tomsk University, 36 Lenin str., 634050 Tomsk, Russia
e-mail: vas@vmm.tsu.ru

Abstract: The ratio estimation problem of probability density function partial derivatives under the assumption of asymptotic decay of the dependence between observations is solved.
The convergence rate for estimators of probability density function partial derivatives and its ratios in the metrics $L_m,\ m \geq 2$, are established. The main part of asymptotic mean square error of the piecewise smooth approximation of the ratio substitution estimator is found. These results are applied to the ratio estimation of derivatives of the probability density of errors in stochastic regression processes.

Keywords: Non-parametric identification, ratios estimation, mean square convergence, asymptotic properties, autoregressive models.

1. INTRODUCTION[1]

An important problem in applied and theoretical research is to study the properties of nonparametric estimators of multivariate probability density functions (p.d.f.) and their derivatives from dependent samples.

Usually the dependence conditions are formulating in terms of mixing conditions, which may be combined with conditions on a deterministic p.d.f.

Note also the papers of Doukhan and Ghindes (1983), Robinson (1986), Boldin (1982) and Vasil'iev (1986), treating specific forms of dependence in estimation problems of the p.d.f. and the distribution function of the errors in an autoregression process.

From among recent papers on this subject one can point out the papers of Masry (1991) and Bosq and Cheze-Payaud (1999), where for mixing processes an exact asymptotic expression for the mean square error (MSE) of a kernel estimate for the p.d.f. is obtained, and the papers of Tran (1990) and Honda (1998), which show that in some cases the dependence has no effect on the asymptotic variance and mean square error.

Along with estimation of the p.d.f., the estimation of partial derivatives ratios of a multivariate p.d.f. is of interest. These ratios are needed in many statistical problems, for example:

— finding the extremal points and verifying sufficient conditions for maximum and minimum of a multimodal p.d.f.;

— estimation of the Fisher information of p.d.f.;

— optimal Bayes estimation of the parame-

[1]Research supported by the Russian Foundation of Fundamental Research, grants 98–07–03194 and 00–01–00880

ter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)$ of an exponential distribution when the prior distribution is unknown, see (Singh, 1976) and the estimator is of the form

$$\widehat{\boldsymbol{\theta}} = \left(\frac{\partial f(\mathbf{x})}{\partial x_1} \Big/ \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_s} \Big/ \frac{\partial^2 f(\mathbf{x})}{\partial x_s^2} \right);$$

— optimal control of a multivariate autoregression process (Nemirovskii and Tsypkin, 1984) etc.

Considered problems have a special interest for dependent observations. For example, the knowledge of logarithmic derivative of p.d.f. is necessary for the construction of optimal algorithms of nonlinear filtration (Dobrovidov, 1984) and adaptive control (Nemirovskii and Tsypkin, 1984) of stochastic processes. Another applications of similar type nonparametric estimators were considered by Boldin (1994, 1996).

A more detailed bibliography can be found in Koshkin and Vasil'iev (1997, 1998), Vasil'iev (1997), Vasil'iev and Koshkin (1998).

For the ratio estimation of p.d.f. partial derivatives are often used the substitution estimators. The difficulty relates to possible unboundedness of the ratio estimators at some points, see (Cramer, 1946) for details.

In this paper the ratio estimation problem of p.d.f. partial derivatives under the assumption of asymptotic decay of the dependence between observations is considered. Similar to Koshkin (1999), a piecewise smooth approximation of the substitution estimators is used.

The main part of the MSE of ratios with improved rate of convergence is found. The rate of convergence of p.d.f. partial derivatives and its ratios in the metrics $L_m,\ m \geq 2$, are established.

2. PROBLEM SETTING

Let $\{\mathcal{F}_n, n \geq 0\}$ be a filtration in a probability space $(\Omega, \mathcal{F}, P)$, and let a sequence $\boldsymbol{\varepsilon} = \{\boldsymbol{\varepsilon}_n,\ n \geq 1\}$ of independent identically distributed random vectors (r.v.'s) $\boldsymbol{\varepsilon}_n = (\varepsilon_{n1}, \ldots, \varepsilon_{ns})'$ with p.d.f. $f(\mathbf{t})$, $\mathbf{t} \in R^s$, adapted to $\{\mathcal{F}_n\}$ be given (a prime denotes the transposition).

Denote for a vector $\mathbf{a}$ with integer components $\mathbf{a} = (\alpha_1, \ldots, \alpha_s)$, such that $\alpha_1 + \alpha_2 + \cdots + \alpha_s = \alpha$, the partial derivative $f_{\mathbf{a}}^{(\alpha)}(\mathbf{t})$ by the formula

$$f_{\mathbf{a}}^{(\alpha)}(\mathbf{t}) = \frac{\partial^\alpha \mathbf{f}(\mathbf{t})}{\partial \mathbf{t}_1^{\alpha_1} \ldots \partial \mathbf{t}_s^{\alpha_s}}, \quad \mathbf{f}_{\mathbf{0}}^{(\mathbf{0})}(\mathbf{t}) = \mathbf{f}(\mathbf{t}).$$

Consider for the fixed vectors $\mathbf{a}$ and $\mathbf{b}$ the estimation problem of the ratio

$$T(\mathbf{t}) = \frac{f_{\mathbf{a}}^{(\alpha)}(\mathbf{t})}{f_{\mathbf{b}}^{(\beta)}(\mathbf{t})} \qquad (1)$$

of partial derivatives of p.d.f. $f(\mathbf{t})$ from observations of the process $\mathbf{z} = \{\mathbf{z}_n,\ n \geq 1\}$,

$$\mathbf{z}_n = \boldsymbol{\varepsilon}_n + \mathbf{g}_{\lambda, n-1},\ n \geq 1. \qquad (2)$$

Here $\mathbf{g}_\lambda = \{\mathbf{g}_{\lambda,n}, n \geq 0\}$ is a sequence of unobservable, adapted to $\{\mathcal{F}_n\}$ and possibly mutually dependent s-dimensional r.v.'s; $\boldsymbol{\lambda} \in \mathcal{A}$ is an unknown nuisance vector parameter, with $\mathcal{A}$ being the set of its possible values.

As an estimate for the ratio of the partial derivatives $\hat{f}_{\mathbf{a}}^{(\alpha)}(\mathbf{x})$ and $\hat{f}_{\mathbf{b}}^{(\beta)}(\mathbf{x})$ from the observations $\mathbf{z}_n = \boldsymbol{\varepsilon}_n + \mathbf{g}_{\lambda,n},\ n \geq 1$, the ratio of a statistics of the form

$$\hat{f}_{\mathbf{a}}^{(\alpha)}(\mathbf{t}) = \frac{1}{N h_N^{s+\alpha}} \sum_{n=1}^{N} K_{\mathbf{a}}^{(\alpha)} \left(\frac{\mathbf{t} - \mathbf{z}_n}{h_N} \right) \qquad (3)$$

will be used. Here $K(\mathbf{z}) : R^s \to R^1$ is a kernel, which needs not possess the properties of a p.d.f.; $h = \{h_N,\ N \geq 1\}$ is a sequence of positive numbers.

The estimators of the type (3) for the p.d.f. $f(\mathbf{t})$ from the observations (2) have been considered in the papers of Koshkin and Vasil'iev (1998), Vasil'iev and Koshkin (1998). According to these results the substitution estimators

$$T_N(\mathbf{t}) = \frac{\hat{f}_{\mathbf{a}}^{(\alpha)}(\mathbf{t})}{\hat{f}_{\mathbf{b}}^{(\beta)}(\mathbf{t})} \qquad (4)$$

have the properties of asymptotic normality and almost sure convergency. But the investigation of the MSE for $T_N(\mathbf{t})$ has some difficulties through the possible smallness of the denominator $\hat{f}_{\mathbf{b}}^{(\beta)}(\mathbf{t})$. Therefore, similar to Koshkin (1999), the estimator

$$\tilde{T}_N(\mathbf{t}) = \frac{T_N(\mathbf{t})}{(1 + \delta_N |T_N(\mathbf{t})|^q)^\rho} \qquad (5)$$

will be used. Here $\delta = \{\delta_N\}_{N \geq 1}$ is a sequence of positive integers; and for the constants ρ and q the inequalities

$$\rho q \geq 1,\ \rho > 0,\ q > 0$$

are fulfilled.

Let ν be a quantity related to the maximal order of differentiability of $f(\mathbf{x})$.

In this paper the exact asymptotic expression for the MSE of estimators (3) and (5) with improved rate of convergence for $\nu > 2$ is obtained.

Thus for a sufficiently smooth distribution ($\nu >> 1$) the rate of convergence of nonparametric estimators (3) and (5) approachs to the rate $N^{-1/2}$ of parametric estimates for independent observations.

Similar results are also obtained for the ratios estimators of the p.d.f. partial derivatives of the noises in multivariate deterministic and stochastic regression processes, including the autoregression processes.

3. PROPERTIES OF $\hat{f}_{\mathbf{a}}^{(\alpha)}(\mathbf{t})$

In this section the basic asymptotic properties of the estimators (3) are given.

Denote by $\mathcal{H}_1(\alpha)$ the set of monotonically decreasing sequences $h = \{h_N, N \geq 1\}$ of real numbers $h_N > 0$ satisfying the condition

$$\lim_{N\to\infty} \left(h_N + (Nh_N^{s+2\alpha})^{-1}\right) = 0.$$

Let

$$\mathcal{H}_2(\alpha) = \mathcal{H}_1(\alpha) \cap \Big\{h : \sum_{N\geq 1} (Nh_N^{s+2\alpha})^{-2} < \infty\Big\},$$

$$\mathcal{H}_3(\alpha) = \mathcal{H}_1(\alpha) \cap \{h : \lim_{N\to\infty} Nh_N^{s+2(\alpha+\nu)} = 0\}$$

where $\nu \geq 2$ is an integer.

Further, omitting the subscript $\mathbf{b} = (\beta_1, \ldots, \beta_s)'$ of partial derivatives $f_{\mathbf{b}}^{(k)}(\mathbf{x})$ and $f_{n,\mathbf{b}}^{(k)}(\mathbf{x})$ will mean that the set of indices $\beta_1, \ldots, \beta_s$ is not specified.

Definition 1.
(i) The p.d.f. $f(\mathbf{t})$ is said to belong to the set $\mathcal{N}_1(n)$ if the function $f(\mathbf{t})$ is continuous and $\alpha+\nu$ times differentiable on $R^s, \nu \geq 2$, and all the partial derivatives of order $\alpha + \nu$ satisfy the Lipschitz condition of degree $0 < \gamma \leq 1$ with constant $\mathcal{L} > 0$, i.e., for all $\mathbf{y} \in R^s$

$$|\rho(\mathbf{t}, \mathbf{y})| \leq \mathcal{L}\|\mathbf{t} - \mathbf{y}\|^\gamma$$

where

$$\rho(\mathbf{t}, \mathbf{y}) = f^{(n+\nu)}(\mathbf{t}) - f^{(n+\nu)}(\mathbf{y}), \ \|\mathbf{z}\|^2 = \sum_{j=1}^{s} z_j^2;$$

(ii) the p.d.f. $f(\mathbf{t})$ is said to belong to the set $\mathcal{N}_2(n)$ if $f(\mathbf{t}) \in \mathcal{N}_1(n)$ and $\sup_{\mathbf{t}} f(\mathbf{t}) \leq C$.

Definition 2. A finitely supported function $K(\mathbf{u})$ belongs to the class $\mathcal{B}(\alpha)$ if it is continuously differentiable up to the order α (inclusive); $K(\mathbf{u}) \in \mathcal{B}^+(\alpha)$ if $K(\mathbf{u}) \in \mathcal{B}(\alpha)$ and $\int_{R^s} K(\mathbf{u})\, d\mathbf{u} = 1$.

Definition 3. For an even $\nu \geq 2$ we say that $K(\mathbf{u}) \in \sigma_\nu(\alpha)$ if $K(\mathbf{u}) \in \mathcal{B}^+(\alpha)$, $K(\mathbf{u}) = K(-\mathbf{u})$, $\int \| \mathbf{u} \|^\nu | \ K(\mathbf{u}) \ | \ d\mathbf{u} < \infty$, $\int u_i^j K(\mathbf{u}) d\mathbf{u} = 0$, $\int u_i^\nu K(\mathbf{u}) d\mathbf{u} \neq 0$, $i = \overline{1, s}$, $j = \overline{1, \nu - 1}$.

Definition 4. The family $g(\mathcal{A})$ is said to belong to $G(m_1, m_2, \alpha)$ if as $N \to \infty$

$$\sup_{\mathcal{A}} < M_\lambda \| \mathbf{g}_{\lambda,N-1} \|^{2m_1} >= o(1/(Nh_N^{s+2\alpha})^{m_1})$$

and

$$\sup_{\mathcal{A}} < M_\lambda \| \mathbf{g}_{\lambda,N-1} \|^{2m_1 m_2} >= o(1/(Nh_N^{s+2\alpha})^{m_1}).$$

Here $< \mathbf{v}_N >= \dfrac{1}{N} \sum_{n=1}^{N} \mathbf{v}_n$.

Put

$$S_{\mathbf{a},\mathbf{b}}^\alpha(\hat{f}) = M_\lambda(\hat{f}_{\mathbf{a}}^{(\alpha)} - f_{\mathbf{a}}^{(\alpha)})(\hat{f}_{\mathbf{b}}^{(\alpha)} - f_{\mathbf{b}}^{(\alpha)}),$$

$$u^{2m}(\hat{f}_{\mathbf{a}}^{(\alpha)}) = M_\lambda(\hat{f}_{\mathbf{a}}^{(\alpha)} - f_{\mathbf{a}}^{(\alpha)})^{2m},$$

$$\omega_{\mathbf{a},\mathbf{b}}^\alpha(\mathbf{t}) = \omega_{\mathbf{a}}^\alpha(\mathbf{t})\omega_{\mathbf{b}}^\alpha(\mathbf{t}),$$

$$\omega_{\mathbf{a}}^\alpha(\mathbf{t}) = \frac{1}{\nu!} \sum_{i=1}^{s} f_{\mathbf{a}+\mathbf{b}_i(\nu)}^{(\alpha+\nu)}(\mathbf{t}) \int u_i^\nu K(\mathbf{u}) d\mathbf{u},$$

$$\mathbf{b}_i(\nu) = \nu(\delta_{i1}, \ldots, \delta_{is})'$$

with δ_{ij} denoting the Kronecker delta,

$$L_{\mathbf{a},\mathbf{b}}^\alpha = \int K_{\mathbf{a}}^{(\alpha)}(\mathbf{u}) K_{\mathbf{b}}^{(\alpha)}(\mathbf{u}) d\mathbf{u},$$

$$v_N^2(\alpha) = h_N^{2\nu} + (Nh_N^{s+2\alpha})^{-1}, \ N \geq 1.$$

Theorem 1. The estimators (3) of the function $f^{(\alpha)}(\mathbf{t})$ have the following asymptotic properties:

(i) if $f(\mathbf{t}) \in \mathcal{N}_1(\alpha)$, $K(\mathbf{u}) \in \sigma_\nu(\alpha)$, $h \in \mathcal{H}_1(\alpha)$, $g(\mathcal{A}) \in G(1, \nu_\alpha, \alpha)$, $\nu_\alpha = \max(\nu + 1, \ (\alpha + \nu + 1)/2)$, $\nu \geq 2$, then as $N \to \infty$

$$\sup_{\mathcal{A}} | \ S_{\mathbf{a},\mathbf{b}}^\alpha(\hat{f}) - \frac{L_{\mathbf{a},\mathbf{b}}^\alpha f(\mathbf{t})}{Nh_N^{s+2\alpha}} - \omega_{\mathbf{a},\mathbf{b}}^\alpha(\mathbf{t}) h_N^{2\nu} \ | = o(v_N^2(\alpha));$$

for the bias b_N of the estimator $\hat{f}_{\mathbf{a}}^{(\alpha)}(\mathbf{t})$ by $g(\mathcal{A}) \in G(1, \nu + 1, \alpha)$ as $N \to \infty$ we have

$$\sup_{\mathcal{A}} | \ b_N - \omega_{\mathbf{a}}^\alpha(\mathbf{t}) h_N^\nu \ | = o(v_N(\alpha));$$

(ii) if $f(\mathbf{t}) \in \mathcal{N}_2(\alpha)$, $K(\mathbf{u}) \in \sigma_\nu(\alpha)$, $g(\mathcal{A}) \in G(m, \nu + 1, \alpha)$, $m \geq 1$, $h \in \mathcal{H}_1(\alpha)$, then

$$\sup_{\mathcal{A}} u^{2m}(\hat{f}_{\mathbf{a}}^{(\alpha)}) = O(v_N^{2m}(\alpha)), \ N \to \infty.$$

4. ESTIMATION OF $T(\mathbf{t})$

It is well known that the substitution estimators of the type $T_N(\mathbf{t})$ are unstable, and theorems for MSE using the majorizing sequences can not be apply, see (Koshkin, 1999; Dobrovidov and Koshkin, 1997). The problem can be resolved by using a piecewise smooth approximation of the form (5).

Put

$$H(\boldsymbol{\phi}): \; R^p \to R^1, \; p \geq 1;$$

$\hat{\boldsymbol{\phi}} = \{\boldsymbol{\phi}_N\}_{N\geq 1}$ is the sequence of estimators of the argument $\boldsymbol{\phi}$;

$$Q(m) = \{(q,k) : q \geq q(m) = 2k/(m-k-1) > 0,$$
$$m \geq m_0 = [3, \; k=1; \; 2k, \; k \geq 2]\},$$

k, m are some positive integers;

$$\tilde{H}(\boldsymbol{\phi}_N) = H(\boldsymbol{\phi}_N)/(1 + \delta_N |H(\boldsymbol{\phi}_N)|^q)^\rho$$

where $q > 0, \; \rho > 0, \; \rho q \geq 1, \; \delta_N > 0.$

Let us formulate the theorem for the estimator $\tilde{H}(\boldsymbol{\phi}_N)$ (see Corollary 4 in Koshkin (1999)).

Theorem 2. Assume that

(i) for some $m \geq 3$

$$M\|\boldsymbol{\phi}_N - \boldsymbol{\phi}\|^m = O(d_N^{-m/2}), \; d_N \uparrow \infty;$$

(ii) function $H(\mathbf{z}) \in W_p(\boldsymbol{\phi})$, i.e., there is an neighbourhood in which the function $H(\mathbf{z})$ and all its partial derivatives up to the second order are continuous and bounded in some neighbour of the point $\boldsymbol{\phi}$;

(iii) $\delta_N = C d_N^{-1}, \; 0 < C < \infty;$

(iv) $H(\boldsymbol{\phi}) \neq 0$ or q is natural even.

Then for any $(q,k) \in Q(m)$

$$|M[\tilde{H}(\boldsymbol{\phi}_N) - H(\boldsymbol{\phi})]^k - M[\nabla H(\boldsymbol{\phi})(\boldsymbol{\phi}_N - \boldsymbol{\phi})]^k| =$$
$$= O\left(d_N^{-\frac{k+1}{2}}\right)$$

where

$$\nabla H(\boldsymbol{\phi}) = \left(\frac{\partial H}{\partial \phi_1}, \ldots, \frac{\partial H}{\partial \phi_p}\right).$$

Put

$$\chi(x \leq a) = (1, \; x \leq a; \; 0, \; x > a),$$

$$\tilde{L}^{\alpha,\beta}_{\mathbf{a},\mathbf{b}} = f(\mathbf{t})(f^{(\beta)}_{\mathbf{b}}(\mathbf{t}))^{-2}[L^{\alpha}_{\mathbf{a},\mathbf{a}}\chi(\alpha \geq \beta) -$$
$$-2T(\mathbf{t})L^{\alpha}_{\mathbf{a},\mathbf{b}}\chi(\alpha = \beta) + T^2(\mathbf{t})L^{\beta}_{\mathbf{b},\mathbf{b}}\chi(\alpha \leq \beta)],$$

$$\tilde{\omega}^{\alpha,\beta}_{\mathbf{a},\mathbf{b}} =$$
$$= \left(f^{(\beta)}_{\mathbf{b}}(\mathbf{t})\right)^{-2} [\omega^{\alpha}_{\mathbf{a}}\chi(\alpha \geq \beta) - T(\mathbf{t})\omega^{\beta}_{\mathbf{b}}\chi(\alpha \leq \beta)]^2,$$
$$\varpi = \max(\alpha, \beta).$$

The next result follows from Theorems 1,2 for the estimators of $T(\mathbf{t})$.

Theorem 3. Assume that $f(\mathbf{t}) \in \mathcal{N}_2(\varpi), \; \nu \geq 2$, for some $m \geq 3 \;\; g(\mathcal{A}) \in G(m, \nu_\varpi, \varpi), \; h \in \mathcal{H}_1(\varpi)$. Then for the estimators $\tilde{T}_N$, defined by (7), for any $(q,k) \in Q(m), \; \delta_N = O(v_N^2(\varpi))$ as $N \to \infty$ the next equation

$$\sup_{\mathcal{A}} M_\lambda |\tilde{T}_N(\mathbf{t}) - T(\mathbf{t})|^k = O(v_N^k(\varpi))$$

holds; the main part of the MSE of the estimator $\tilde{T}_N$ has the form

$$\sup_{\mathcal{A}} |M_\lambda[\tilde{T}_N(\mathbf{t}) - T(\mathbf{t})]^2 - \frac{f(\mathbf{t})\tilde{L}^{\alpha,\beta}_{\mathbf{a},\mathbf{b}}}{N h_N^{s+2\varpi}} - \tilde{\omega}^{\alpha,\beta}_{\mathbf{a},\mathbf{b}} h_N^{2\nu}| =$$
$$= o(v_N^2(\varpi)).$$

Note that the convergence rate of the MSE for the estimator $\tilde{T}_N(\mathbf{t})$ coincides with the convergence rate of MSE for estimators of greatest derivative of the p.d.f. $f(\mathbf{t})$ in the ratio $T(\mathbf{t})$.

5. DERIVATIVES ESTIMATION OF THE P.D.F. OF THE NOISE IN AUTOREGRESSION

Consider the scalar random process $(x_n)_{n>-p}$ described by the system of equations

$$x_n = \lambda_1 x_{n-1} + \ldots + \lambda_p x_{n-p} + \varepsilon_n, \; n = 1,2,\ldots \quad (6)$$

Here $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)' \in \mathcal{A}$ is the vector of unknown parameters with $\mathcal{A}$ being the set of all vectors $\boldsymbol{\lambda}$, for which the process (6) is stable, i.e. all roots of the polynomial

$$\mathcal{P}(z) = z^p - \lambda_1 z^{p-1} - \ldots - \lambda_p$$

lay inside the unit circle. Note that $\mathcal{A}$ is a bounded set in R^p. Assume that $(\varepsilon_n, \mathcal{F}_n)_{n\geq 1}$ is a sequence of independent identically distributed values with p.d.f. $f(\mathbf{t})$.

The problem is to construct estimators for partial derivatives $f^{(\alpha)}_{\mathbf{a}}(\mathbf{t})$ of p.d.f. $f(\mathbf{t})$ and its ratios with improved rate of convergence. Since ε_n cannot be observed directly, we will use the observable variables

$$z_n = x_n - \boldsymbol{\lambda}'(n-1)\mathbf{x}(n-1) =$$

$$= \varepsilon_n + (\boldsymbol{\lambda} - \boldsymbol{\lambda}(n-1))'\mathbf{x}(n-1),\ n \geq 1, \quad (7)$$
$$\boldsymbol{\lambda}(n) = (\lambda_1(n), \ldots, \lambda_p(n))',$$
$$\mathbf{x}(n) = (x_n, \ldots, x_{n-p+1})'$$

for the estimation of $f_{\mathbf{a}}^{(\alpha)}(\mathbf{t})$, as it was done in Koshkin and Vasil'iev (1997, 1998), Vasil'iev (1986, 1997), Vasil'iev and Koshkin (1998). The observations (7) have the form (2) if put

$$g_{\lambda,n} = \boldsymbol{\Delta}'(n)\mathbf{x}(n), \quad \boldsymbol{\Delta}(n) = \boldsymbol{\lambda} - \boldsymbol{\lambda}(n). \quad (8)$$

Here $\boldsymbol{\lambda}(n)$ are some $\mathcal{F}_n$-measurable estimators for the vector parameter $\boldsymbol{\lambda}$.

Note that the variables $y_n = \boldsymbol{\lambda}'(n)\mathbf{x}(n)$ may be considered as the predictions of x_n, and $z_n = x_n - y_n$ are the errors of these predictions.

The estimators $\boldsymbol{\lambda}(n)$ will be constructed on the basis of the least square estimators $\boldsymbol{\lambda}_k$, which are calculated at some unbounded increasing sequence $(r_k)_{k\geq 0}$ of natural numbers $r_k,\ r_0 > 0$:

for $k \geq 0$

$$\boldsymbol{\lambda}_k = \mathbf{F}^{-1}(r_k)\mathbf{B}(r_k),\ \boldsymbol{\lambda}_0 = 0, \quad (9)$$

$(\mathbf{F}(n))_{ij} = S_{|i-j|}(n),\ i,j = \overline{1,p},\ S_l(n) = \langle x_n x_{n-l}\rangle$, $\mathbf{B}(n) = (S_1(n), \ldots, S_p(n))'$ with $(\mathbf{F})_{ij}$ denoting the ij-th element of matrix $\mathbf{F}$.

Denote by $\tilde{\boldsymbol{\lambda}}_k = (\tilde{\lambda}_{k1}, \ldots, \tilde{\lambda}_{kp})'$ the regularized estimators

$$\tilde{\lambda}_{ki} = \lambda_{ki}(1 + r_k^{-1}|\lambda_{ki}|^{\tau})^{-\rho}, \quad (10)$$
$$\tau > 0,\ \rho > 0,\ \rho\tau \geq 1.$$

Let $S(\mathcal{A})$ be a closed ball in R^p with the radius c_p containing the set $\mathcal{A}$. Put for all $k \geq 1$ and $r_{k-1} \leq n < r_k$

$$\boldsymbol{\lambda}(n) = \mathrm{proj}_{S(\mathcal{A})}\boldsymbol{\lambda}^*(k-1) =$$
$$= [1 + (c_p/||\tilde{\lambda}_k|| - 1)\chi(||\tilde{\lambda}_k|| > c_p)]\tilde{\lambda}_k. \quad (11)$$

The following Lemma contains the properties of estimators $\tilde{\lambda}_k$ and predictions y_n.

Lemma. Assume that the sequences $(\boldsymbol{\lambda}(n))$ and $\mathbf{g}_\lambda,\ \boldsymbol{\lambda} \in \mathcal{A}$, are defined in (6)–(11), and for some $m \geq 1$ the moments $E||\mathbf{x}(0)||^{4m} < \infty$ and $E\varepsilon_1^{4m} < \infty$ exist. Then

i) for all $(\tau, i) \in Q(m)$ as $k \to \infty$ the estimators (10) have the property

$$\sup_{\mathcal{A}} M_\lambda \|\tilde{\lambda}_k - \lambda\|^i = O(r_k^{-i/2});$$

ii) as $N \to \infty$

$$\sup_{\mathcal{A}} \sum_{n=1}^{N} M_\lambda \|\mathbf{g}_{\lambda,n-1}\|^m = \begin{cases} O(\ln N), & m = 2, \\ O(1), & m > 2. \end{cases}$$

The assertions of Theorems 1,3 for the observation model (7) will be true if the assemblage $g(\mathcal{A})$ of sequences $\mathbf{g}_\lambda$ defined by (8) satisfies the conditions G.

The next Proposition follows from the assertion 2 of the Lemma.

Proposition. Let the autoregressive process (6) be stable and for some integer numbers $m_1 \geq 1$, $m_2 \geq 1$ the moments $E||\mathbf{x}(0)||^{8m_1m_2}$ and $E\varepsilon_1^{8m_1m_2}$ exist. Then the family $g(\mathcal{A})$ of functions $(\mathbf{g}_{\lambda,n})$ defined by (8) has the property

$$g(\mathcal{A}) \in G(m_1, m_2, \alpha)$$

if $r_k = k^\rho,\ \rho > 4/3$ and $h \in \mathcal{H}_1(\alpha)$.

Remark 1. Theorem 2 is used both for the parameters estimation of autoregressive process and for the ratio estimation of the p.d.f. derivatives of autoregressive noise. The results of Section 4 may be also applied for some multidimensional linear stochastic regression models considered by Koshkin and Vasil'iev (1998), Vasil'iev and Koshkin (1998). Note that for these models the using of sequential estimators of dynamic system parameters proposed by Borisov and Konev (1977) and Vorobeichikov and Konev (1980) is preferable.

Remark 2. Similar to Vasil'iev (1997) it can be shown that the second assertion of Lemma and Proposition hold under somewhat weaker assumptions on the moments of $\mathbf{x}(0)$ and ε_1.

6. CONCLUSION

The kernel estimators (3) and (5) for the derivatives of multivariate p.d.f. $f(\mathbf{t})$ and its ratio $T(\mathbf{t})$ by dependent observations have the improved rate of convergence $N^{\nu/(s+2(\alpha+\nu))}$ and $N^{\nu/(s+2(\overline{\omega}+\nu))}$ correspondingly.

This improvement of the convergence rate is provided by the requirements σ_ν with $\nu > 2$ on the kernel $K(\cdot)$, see (Müller and Gasser, 1979; Koshkin, 1990).

The results are applied to the dynamic systems, in particular, for the estimation of p.d.f. derivatives of the noises of stochastic regression and autoregression processes with unknown parameters. The proposed method may be applied to the more general models, for example, to the nonlinear ARCH time series considered by Neumann and Kreiss (1998), Masry and Tjøstheim (1995) or for the control systems (Konev and Vasil'iev, 1997).

Theorems 1,2 give the possibilities to estimate more general functions $H(\cdot) \in W_p(\phi)$ of p.d.f. derivatives in the metrics L_m, $m \geq 2$, in comparison with the ratio-function considered in this paper.

7. REFERENCES

Boldin, M.V. (1982). 'Estimation of the distribution of noise in an autoregression scheme'. *Theory Probab. Appl.*, **27**, 866–871.

Boldin, M.V. (1994). 'On medium estimates and tests in autoregressive schemes'. *Math. Methods Statist.*, 3, 114–129.

Boldin, M.V. (1996). 'On nonparametric sign estimators in multiparameter autoregression'. *Math. Methods Statist.*, 5, 154–172.

Borisov, V.Z. and V.V. Konev (1978). 'Sequential estimation of parameters of discrete processes'. *Autom. Remote Control*, **38**, 1475–1480.

Bosq, D. and N. Cheze-Payaud (1999). 'Optimal asymptotic quadratic error of nonparametric regression function estimates for a continuous-time process from sampled-data'. *Statistics*, **32**, 229–247.

Cramer, H. (1946). *Mathematical methods of statistics.* Princeton University Press.

Dobrovidov, A.V. (1984). 'Asymptotically ϵ-optimal nonparametric procedure for nonlinear filtering of stationary sequences with unknown statistical characteristics'. *Autom. Remote Control*, **45**, 1569–1576.

Dobrovidov, A.V. and G.M. Koshkin (1997). *Nonparametric signal estimation.* (Russian. English summary). Nauka, Fizmatlit, Moscow.

Doukhan, P. and M. Ghindes (1983). 'Estimation de la transition de probabilité d'une chain de Markov–Doeblin recurrente. Etude du cas processus autoregressif general d'ordre 1'. *Stochastic Process. Appl.*, **15**, 271–293.

Honda, T. (1998). 'Sequential estimation of the marginal density function for a strongly mixing process'. *Sequen. Anal.*, **17**, 3-4, 239-251.

Konev, V.V. and S.M. Pergamenshchikov (1986). 'On the duration of sequential estimation of parameters of stochastic processes in discrete time'. *Stochastics*, **18**, 133–154.

Konev, V. and V. Vasil'iev (1997). 'On optimal adaptive control of a linear stochastic process'. In: *Proc.15th IMACS World Congress on Scientific Computation, Modelling and Applied Mathematics*, August 24-29, Berlin, Germany, **5**, 87-91.

Koshkin, G.M. (1990). 'Asymptotic properties of functions of statistics and their application to nonparametric estimation'. *Autom. Remote Control*, **51**, 3, 345–357.

Koshkin, G.M. (1999). 'Deviation moments of the substitution estimator and its piecewise smooth approximations'. *Sib.Math.Zh.*, **40**, 3, 515–527.

Koshkin, G.M. and V.A. Vasil'iev (1997). 'An estimation of a multivariate density and its derivatives by weakly dependent observations'. In: *Proc. Steklov Math. Inst. Seminar Statistics and Control of Stochastic Processes.* The Liptser Festschrift (Yu.M. Kabanov, B.L. Rozovskii, A.N. Shiryaev. (Ed)), 229–241. World Scientific, Singapore e.a., New Jersey, London, Hong Kong.

Koshkin, G.M. and V.A. Vasil'iev (1998). 'Nonparametric estimation of derivatives of a multivariate density from dependent observations'. *Math. Methods of Statistics*, **7**, 4, 361–400.

Masry, E. (1991). 'Multivariate probability density deconvolution for stationary random processes'. *IEEE Trans. Inform. Theory*, **37**, 1105–1115.

Masry, E. and D. Tjøstheim (1995). 'Nonparametric estimation and identification of nonlinear ARCH time series'. *Econometric Theory*, **11**, 258–289.

Müller, H.-G. and T. Gasser (1979). 'Optimal convergence properties of kernel estimates of derivatives of a density function'. *Lect. Notes Math.*, 757, 144–154.

Nemirovskii, A.S. and Ya.Z. Tsypkin (1984). 'On optimal algorithms of adaptive control'. *Avtom. Telemekh.*, 12, 64–77 (in Russian).

Neumann, M.H. and J.-P. Kreiss (1998). 'Regression-type inference in nonparametric autoregression'. *Ann. Statist.*, **26**, 4, 1570–1613.

Robinson, P.M. (1986). 'Nonparametric estimation from time series residuals'. *Can. Cent. Etud. Rech. Oper.*, **28**, 197–202.

Singh, R.S. (1976). 'Nonparametric estimation of mixed partial derivatives of a multivariate density'. *Multivariate Analysis*, **6**, 111–122.

Tran, L.T. (1990). 'Kernel density estimation of random fields'. *J. Multiv. Anal.*, **34**, 37–53.

Vasil'iev, V.A. (1986). 'Estimation of the distribution of disturbances in an autoregression process'. In: *Matematicheskaya Statistika i ee Prilozheniya*, Tomsk: Tomsk Univ. Press, **10**, 9–24 (in Russian).

Vasil'iev, V.A. (1997). 'On identification of dynamical systems of autoregressive type'. *Avtom. Telemekh.*, 12, 106–118 (in Russian).

Vasil'iev, V.A. and G.M. Koshkin (1998). 'Nonparametric identification of autoregression'. *Teor. Veroyatn. Primen.*, **43**, 3, 577–588 (in Russian).

Vorobejchikov, S.E. and V.V. Konev (1981). 'On sequential identification of stochastic systems'. *Eng. Cybern.* **18**, 4, 121–127.