NONPARAMETRIC ANALYSIS OF STOCHASTIC SYSTEMS WITH NONLINEAR FUNCTIONAL HETEROGENEITY

V.I. MALUGIN, M.E.VASILKOV
Belarusian State University
Minsk, BELARUS
e-mail: Malugin@bsu.by

Abstract

The problems of the analysis of stochastic systems described by nonlinear statistical models with heterogeneous functional forms are considered. It is supposed that functional heterogeneity is conditioned by the existing of the different classes of system states. Moreover it being known that every state of a system characterized by un-known nonlinear models, that are different for different classes of states. The methods of estimation and forecasting of the systems states based on multivariate nonparametric density estimate with variable kernel are suggested and examined by means of asymptotic expansions of the conditional risk as well as by statistical modeling experiments.

1 The model of the nonlinear functional heterogeneity

Let stochastic system in the i-th experiment is described by the stacked vector of the features

$$y_i = \begin{pmatrix} x_i \\ z_i \end{pmatrix} \in \Re^p,\ p = N + M\ (N, M \geq 1),\ i = 1, \ldots, n, \tag{1}$$

were $x_i = (x_{i1}, \ldots, x_{iN})' \in \Re^N$, $z_i = (z_{i1}, \ldots, z_{iM})' \in \mathbf{Z} \subset \Re^M$, $\mathbf{Z}$ – bounded space.

The components of $y_i \in \Re^{N+M}$ are satisfied to the nonlinear model

$$T(y_i) = x_i - f(z_i) = \xi_i,\ i = 1, \ldots, n, \tag{2}$$

were $T(\cdot), f(\cdot)$ – unknown sufficiently smooth vector functions; $\xi_i = (\xi_{i1}, \ldots, \xi_{iN})' \in \Re^N$ – independent random vectors with zero mean and non-singular covariance matrix Σ; $z_i \in \mathbf{Z}$ and $\xi_i \in \Re^N$ are independent random vectors with the densities $p_z(z)$ and $p_\xi(\xi)$ correspondingly; decomposition (1) of vector $y_i \in \Re^p$ on subvectors $x_i \in \Re^N$, $z_i \in \mathbf{Z} \subset \Re^M$ is unknown in general case.

According to mentioned above assumptions the density $p(y)$ of the random vector $y_i \in \Re^p\ (i = 1, \ldots, n)$ takes the form

$$p(y) = p_\xi(x - f(z))p_z(z),\ x \in \Re^N,\ z \in Z \subset \Re^M,\ y \in \Re^{N+M}. \tag{3}$$

It is supposed that random values $\xi_i \in \Re^N$ have quite small variance, so according to (3), observations $\{y_i\}(i = 1, \ldots, n)$ are concentrated in a space $\Re^p$ near some

N-dimensional ($N < p$) hyperspace, described by identity $T(y) = 0_N \in \Re^N$, corresponding to some stable state of system. In this connection random values $\{\xi_i\}$ may be interpreted as random deviations from a stable system state.

Functional heterogeneity model. For simplicity it is supposed that there is two classes of system states Ω_0 and Ω_1. Discrete variable $\nu_i = \nu(y_i) \in S \equiv S(2) = \{0, 1\}$ correspondences to the number of state and described by the probability distribution

$$P\{\nu_i = \alpha\} = \pi_\alpha > 0 (\alpha \in S),\ \pi_0 + \pi_1 = 1, \tag{4}$$

were parameters $\pi_\alpha (\alpha \in S)$ are called *a prior probabilities of a system state.*

Nonlinear functions $\{f_\alpha(z)\}$ (or$\{T_\alpha(y)\}$) are unknown and satisfied to the *functional heterogeneity condition*:

$$P(f_1(z) = f_2(z)) = 0,\ \forall z \in \mathbf{Z}. \tag{5}$$

Conditional densities of distribution $\{p_\alpha(y)\}$ for different classes of states $\{\Omega_\alpha\}$ are defined by (3) under $f(z) \equiv f_\alpha(z)(T(y) \equiv T_\alpha(y)),\ \alpha \in S$.

Priory information about the considered model of observation may include the following assumptions:

- probabilistic characteristics $\{\pi_\alpha, p_\alpha(y)\}(\alpha \in S)$ of classes Ω_0 and Ω_1 are unknown;
- endogenous-exogenous structure of the features vector $y = \begin{pmatrix} x \\ z \end{pmatrix} \in \Re^{N+M}$ is known (or is unknown);
- there is classified learning sample of observations $Y = (y_i) \in \Re^{pn}$ from classes Ω_0 and Ω_1, admitted the decomposition: $Y = Y_0 \bigcup Y_1$ were $Y_\alpha = (y_{\alpha i}) \in \Re^{pn_\alpha}$ – the sample from the class $\Omega_\alpha (\alpha \in S,\ n = n_0 + n_1)$.

2 Analysis of stochastic systems with functional heterogeneity models

There are the following actual problems of analysis of stochastic systems with functional heterogeneity models.

The problem of forecasting of endogenous variable. It is supposed that endogenous-exogenous structure of features vector $y = \begin{pmatrix} x \\ z \end{pmatrix} \in \Re^{N+M}$ is known. The problem is to forecast the value of endogenous variable $x \in \Re^N$ when the value of exogenous variable $z \in \mathbf{Z}$ and number of state $\alpha \in S$ are given. The forecasting value $\hat{x}$ for the given $\alpha \in S$ and $z \in \mathbf{Z}$ is calculated by means of the optimization algorithm $\hat{x} = \arg\max_x \hat{p}_\alpha(x, z)$, were $\hat{p}_\alpha(y) \equiv \hat{p}_\alpha(x, z)$ – nonparametric kernel density estimate with variable Gaussian kernel.

The problem of system state forecasting is consisted in the estimation of the class of system state $\nu_i = \nu(y_i) \in S$ for $y_i \in \Re^{N+M} (i = n+1, n+2, \ldots)$ by means of statistical classification algorithms.

There are two kinds of interpretation of this problem:

- *the problem of future system state forecasting* if endogenous-exogenous structure of vector $y = \begin{pmatrix} x \\ z \end{pmatrix} \in \Re^{N+M}$ is known but only vector exogenous variable $z_i \in \mathbf{Z}$ is given, while vector x_i is not observed (in this case the problem of forecasting of endogenous variable should be solved on the preliminary step);
- *the problem of current system state estimation* if full vector of features $y_i = \begin{pmatrix} x_i \\ z_i \end{pmatrix} \in \Re^{N+M}$ is observed, moreover endogenous-exogenous structure of vector $y \in \Re^{N+M}$ may be unknown.

The nonparametric plug-in rules are suggested for current system state estimation problem. These rules are obtained by substituted of the nonparametric density estimates with multivariate Gaussian kernel in Bayesian decision rule [1]:

$$\hat{d}^{(l)}(y) = \mathbf{1}(\hat{G}^{(l)}(y)) + 1, \ \hat{G}^{(l)}(y) = c_2\hat{p}_2^{(l)}(y) - c_1\hat{p}_1^{(l)}(y), \tag{6}$$

were $l = 1$ and $l = 2$ for nonparametric density estimates with fixed and variable Gaussian kernel are distinguished by the choosing of the covariance matrixes H_1 and $\{H^{(i,m(\alpha,i))}\}$ correspondingly:

$$\hat{p}_\alpha^{(1)}(y) = \frac{1}{n_\alpha}\sum_{i=1}^{n_\alpha} n_N\left(y|y_{\alpha i}, h^2H_1\right), \quad \hat{p}_\alpha^{(2)}(y) = \frac{1}{n_\alpha}\sum_{i=1}^{n_\alpha} n_N\left(y|y_{\alpha i}, h^2H^{(i,m(\alpha,i))}\right),$$

$$H^{(i,m(\alpha,i))} = \frac{1}{m(\alpha,i)-1}\sum_{y_{\alpha j}\in S(\alpha,i)\setminus y_{\alpha i}}\left(y_{\alpha j} - y^{(\alpha,i)}\right)\left(y_j - y^{(\alpha,i)}\right)',$$

$$y^{(\alpha,i)} = \frac{1}{m(\alpha,i)-1}\sum_{y_{\alpha j}\in S(\alpha,i)\setminus y_{\alpha i}} y_{\alpha j}.$$

were H_1 is a non-singular fixed matrixes (e.g. full sample covariance matrixes) and $H^{(i,m(\alpha,i))}$ is the matrix statistics calculated on the sample of $m(\alpha,i)$ observations belonged to some local neighborhood $S(\alpha,i)$ of point $y_{\alpha,i}$ [2]. The size $m(\alpha,i)$ of the neighborhood is chousing from the condition of minimum value of Anderson V-statistic [3].

As a criterion of the quality of the decision rules in analytical research it is used the functional of conditional ε-risk described by the formulas

$$r_n^{(l)}(\varepsilon,z) = \mathbf{E}_y\{R_n^{(l)}(\varepsilon,z,Y)\}(l=1,2),$$

$$R_n^{(l)}(\varepsilon,z,Y) = \sum_{\alpha\in S}\pi_\alpha\int\limits_{T(\varepsilon,z)} w\left(\alpha,\hat{d}^{(l)}(y)\right)p_\xi\left(x - f_\alpha(z)\right)dx,$$

$T(\varepsilon,z)\subset\mathbf{X}$ – is a bounded domain satisfied to condition

$$|r_n^{(l)}(\varepsilon,z) - r_n^{(l)}(z)| \le \varepsilon\,(0<\varepsilon<1), \tag{7}$$

were parameter $\varepsilon > 0$ defines the quality of approximation of conditional risk $r_n^{(l)}(z)$ under the given $z\in\mathbf{Z}$ by means of conditional ε-risk $r_n^{(l)}(\varepsilon,z)$: if $\varepsilon\to 0$ then the quality of approximation is increasing.

3 The main results

Analytical investigations of decision rules are conducted for multivariate linear regression models with univariate distributed exogenous variables and Gaussian disturbances in the case of "essential dependent" features, i.e. when $\mathrm{tr}(\Sigma) \to 0$. Alternative decision rules are compared by means of asymptotic expansions of the conditional ε-risk. The preference of the decision rule with variable kernel in the asymptotic of strengthening dependents of features and growing learning sample size ($\mathrm{tr}(\Sigma) \to 0, n \to \infty, \varepsilon \to \infty$) is established [4].The suggested algorithm of forecasting of endogenous variables is compared on accuracy with the known nonparametric estimates of nonlinear regression [5]. The results of statistical experiments are in line with theoretical inferences.

References

[1] Kharin Yu.S. (1996). *Robustness in statistical pattern recognition.* Dordrecht, Boston, London : Kluwer Academic Publishers.

[2] Malugin V.I. (1985). On the estimation of the density of random vectors with essential dependent components. *Vestnik of the BSU.* Ser. 1. N2, pp. 41 - 44.

[3] Anderson T. (2003). *An introduction to multivariate statistical analysis.* 2nd edition. Wiley, New Jersey.

[4] Malugin V.I. (2009). Asymptotic analysis of the risk of nonparametric classification in the case of essential dependent features. *Proceedings of the National Academy of Sciences of Belarus.* Series of Phys.-Mat. Sciences. N3, pp. 10 - 23.

[5] Hardl W. (1994). *Applied nonparametric regression.* Humboldt University, Berlin.