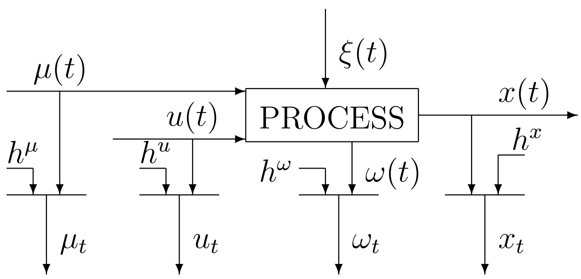
NON-PARAMETRIC MODELS OF STOCHASTIC SYSTEMS WITH DELAY

R.S. Boyko, Ya.I. Demchenko, A.V. Medvedev
Siberian State Aerospace University named after academician M.F. Reshetnev
Krasnoyarsk, RUSSIA
e-mail: saor_medvedev@sibsau.ru

Abstract

The paper considers an identification problem for stochastic static systems with delay. The identification problem is investigated in a broad sense, specially under conditions of non-parametric uncertainty, i. e. in the case when parametric structure of the model is known to within parameters vector. The new classes of non-parametric estimations of regression line of observations is proposed. The paper gives the corresponding theorems of convergence for non-parametric estimation and models.

A general scheme of the discrete-continuous process may be presented in the following way, picture 1.



Picture 1. The scheme of usual process

Lets bring in the following symbols: $x(t)$ – output variable of the process that can be considered scalar without generality violation, $u(t)$ and $\mu(t)$ – corresponding controlled and uncontrolled operated variables, $\omega(t)$ – controlled variable while the process, h^μ, h^u, h^x, h^ω random errors of dimension corresponding to $\mu(t-\tau_\mu)$, $u(t-\tau_u)$, $\omega(t-\tau_\omega)$ and $x(t-\tau)$, $\xi(t-\tau_\xi$ – random distributors, t – time. So, $x(t)$ – can be presented as objectively existed dependence:

$$x(t) = A\left(u(t-\tau_u), \mu(t-\tau_\mu), \omega(t-\tau_\omega), \xi(t-\tau_\xi), t\right) \tag{1}$$

where A - unknown process operator,

$$u(t-\tau_u) \in R^k, \mu(t-\tau_\mu) \in R^m, \omega(t-\tau_\omega) \in R^n, x(t-\tau_x) \in R^1$$

Dimension of variables $\mu(t-\tau_\mu)$, $u(t-\tau_u)$, $\omega(t-\tau_\omega)$ and $x(t-\tau_x)$ is realized with random errors having zero mathematical expectation and restricted variance; probability density is unknown. Lets denote these observations x_t, u_t, μ_t, ω_t, $t=1,2,\ldots,$ here t – discrete time. An investigator while modeling similar processes has a purpose to build mathematical model

$$\tilde{x}(t-\tau_x) = B\left(u(t-\tau_u), \mu(t-\tau_\mu), \omega(t-\tau_\omega), t\right) \tag{2}$$

where B – an operator class, determined on the basis of the existed a priori information, $\tilde{x}(t)$ – output of the model. It is evident that in this case $\tilde{x}(t)$ is tending to proximity of $x(t)$ in the sense of the accepted optimization criteria. A modeling problem of similar processes is aggravated due to the lack of a prior information about an operator A and high dimension of variables $u(t-\tau_u)$ and $\mu(t-\tau_\mu)$.

Identification in a broad sense consumes the step absence of parametric operator class B if, of course, there are not enough a priory data. Often it is much easier to determine an operator class (2) on the basis of qualitative information, for example, process linearity or a type of non-linearity, uniqueness or lack of uniqueness etc. In this case an identification problem is in this operator estimation on the basis of a sample $\left\{x_t, u_t, \mu_t, \omega_t, t=\overline{1,S}\right\}$ in the form

$$\tilde{x}_s(t-\tau_{xs}) = B_s\left(u(t-\tau_u), \mu(t-\tau_\mu), \omega(t-\tau_\omega), t, \vec{x}_s, \vec{u}_s, \vec{\mu}_s, \vec{\omega}_s\right) \tag{3}$$

where $\vec{x}_s=(x_1,\ldots,x_s)$, $\vec{u}_s=(u_1,\ldots,u_s)$, $\vec{\mu}_s=(\mu_1,\ldots,\mu_s)$, $\vec{\omega}_s=(\omega_1,\ldots,\omega_s)$ - time vector. For some operator classes identification theory in a broad sense was developed on the basis of non-parametric statistics methods [6]. In this case formulas like (3) are non parametric models of the process (1), for the scalar $u(t)$,$\mu(t)$ and $\omega(t)$, $x(t)$ have a view:

$$x_s(t) = \frac{\sum_{i=1}^{s} x\Phi\left(c_s^{-1}\left(u-u_{i-\tau_u}\right)\right)\Phi\left(c_s^{-1}\left(\mu-\mu_{i-\tau_\mu}\right)\right)\Phi\left(c_s^{-1}\left(\omega-\omega_{i-\tau_\omega}\right)\right)}{\sum_{i=1}^{s} \Phi\left(c_s^{-1}\left(u-u_{i-\tau_u}\right)\right)\Phi\left(c_s^{-1}\left(\mu-\mu_{i-\tau_\mu}\right)\right)\Phi\left(c_s^{-1}\left(\omega-\omega_{i-\tau_\omega}\right)\right)} \tag{4}$$

where τ_u, τ_μ, τ_ω - delays for different inputs. Further we will not write the delay sign in order to simplify the text.

The given research paper is devoted to the case when a sample (u_1,x_1), (u_2,x_2), ... , (u_s,x_s) in the space of input-output variables is ordered in some specific way, determined by a real situation we often meet with in real situations. It is stipulated by the following features:

1. compressibility of variables measured in some subfields;
2. emptiness lack of variables measured in some subfields;
3. rarefaction few variables of measured parameters distributed in the field irregularly.

Lets use the new class of non-parametric estimations (in order to simplify, we shall use the scalar process input and output). Let (x,u) - a random variable with values in the space $\Omega(x;u)\subset R^2$, $x\in\Omega(x)\subset R^1$ and $p(x,u)$ - distribution density 2- of dimensional random variable (x, u) is unknown. Let (x_1, u_1), (x_2, u_2), ... , (x_s, u_s) - a sample from s statistically independent observations $(n+1)$-of dimensional random variable (x, u). As it is known from [7], the non-parametric estimation of regression line is:

$$\overline{x}_s(u) = \frac{\sum_{i=1}^{s} x_i \Phi_1\left(c_s^{-1}\left(u_j - u_j^i\right)\right) \Phi_2\left(c_s^{-1}\left(u_j - u_j^i\right)\right)}{\sum_{i=1}^{s} \Phi_1\left(c_s^{-1}\left(u_j - u_j^i\right)\right) \Phi_2\left(c_s^{-1}\left(u_j - u_j^i\right)\right)} \tag{5}$$

$$\tilde{x}_s(u) = \frac{\sum_{i=1}^{s} x_i \left(\Phi_1\left(c_s^{-1}\left(u_j - u_j^i\right)\right) + \Phi_2\left(c_s^{-1}\left(u_j - u_j^i\right)\right)\right)}{\sum_{i=1}^{s} \left(\Phi_1\left(c_s^{-1}\left(u_j - u_j^i\right)\right) + \Phi_2\left(c_s^{-1}\left(u_j - u_j^i\right)\right)\right)} \tag{6}$$

The concrete view of the functions 1 and 2 is represented in [4].

In the case when u-in dimensional vector, the estimations (5), (6) will be:

$$\overline{x}_s(u_1, \ldots, u_n) = \sum_{i=1}^{s} y_i \prod_{j=1}^{n} \prod_{k=1}^{N} \Phi_j^k\left(c_s^{-1}\left(u_j - u_j^i\right)\right) \Bigg/ \sum_{i=1}^{s} \prod_{j=1}^{n} \prod_{k=1}^{N} \Phi_j^k\left(c_s^{-1}\left(u_j - u_j^i\right)\right) \tag{7}$$

$$\tilde{x}_s(u_1, \ldots, u_n) = \sum_{i=1}^{s} y_i \prod_{j=1}^{n} \prod_{k=1}^{T} \Phi_j^k\left(c_s^{-1}\left(u_j - u_j^i\right)\right) \Bigg/ \sum_{i=1}^{s} \prod_{j=1}^{n} \prod_{k=1}^{T} \Phi_j^k\left(c_s^{-1}\left(u_j - u_j^i\right)\right) \tag{8}$$

N and T are constants.

Let $\left\{x_i, u_i, i = \overline{1, s}\right\}$ – a sample, optionally dependent but equally distributed n-dimensional random variables, $x_i \in R^k$, $u_i \in R^1$, $i = 1, \ldots, s$ [4, 8].

Theorem 1 *Let $\Phi(z)$ Borel's function, satisfying the following conditions:*

$$\sup_z |\Phi(z)| < \infty, \int_{R^k} |\Phi(z)| dz < \infty, \int_{R^k} \Phi(z) dz = 1, \Phi(z) = \Phi(-z), \lim_{s \to \infty} C_s = 0,$$

$$\lim_{s \to \infty} s C_s^k = \infty$$

and function $x = f(u)$ is measurable by Borels, then

$$\lim_{s \to \infty} M\left\{(x - \overline{x}_s(u))^2\right\} = 0, \qquad \forall u \in \Omega(u) \tag{9}$$

Theorem 2 *In terms of Theorem 1 the statistics $\tilde{x}_s(u_1, \ldots, u_n)$ come all into one in midlesquare $x(u_1, u_2, \ldots, u_n)$ when $s \to \infty$*

During making of non-parametric models in some cases it is more worth wile to use estimations like [5, 1], which are:

$$x_x(u) = \left[\sum_{j=1}^{s} \left(\varphi_j\left(\alpha, x^l, u^l, u\right)\right) \prod_{i=1}^{n} \Phi\left(\frac{u^i - u_j^i}{c_s^i}\right)\right] \cdot \left[\sum_{j=1}^{s} \prod_{i=1}^{n} \Phi\left(\frac{u^i - u_j^i}{c_s^i}\right)\right]^{-1}, \tag{10}$$

here $\left(\varphi_j\left(\alpha, x^l, u^l, u\right)\right)$ *is a hyper plane, through l points, which are situated in area of j-point* $(x;u)$*,* α *- a vector of parameters, defined by the method of less squares.*

Further we will view two real processes of bacterial reproduction in fast food production and accumulation of carcinogenic substances during the heating process. The main problem of food contaminations is bacterial contamination, i.e. penetration of bacteria and their endotoxines with infected food, which guide to sharp inflammation process in the stomach and intestine. The endotoxines, absorbed in blood summon some disorder of water-electrolyte exchange, disorder of heart and vessels system, disorder of kidneys and above-kidneys.

The illness is summoned by various kinds of bacteria from intestine-family, more often by salmonella, which are more than 1700 types. Some infections may be caused by staphylococcus, streptococcus, and also some conditionally pathogenic bacteria; it may be also the intestine-stic, proteus etc.

The first process is connected with accumulations of harmful toxic substances in product because of bacteria reproduction. So, the quantity of toxic substances is directly proportional to pathogenic bacteria quantity in product. And the quantity of bacteria is proportional to storage time of the product or its ingredients in conditions, which are favorable for bacteria reproduction (temperature higher than +5 degrees celsium). In the work the bacteria reproduction process was modeled at a practical example of reproduction of E. Coli, S. Aureus Salmonella at a nourish substrate in conditions of room-temperature and humidity. Measuring of bacteria quantity in example during the experiment was made in each 3 hours after the nourish substrate infection moment during 24 hours. So, the selection was made, which was increased with a help of bootstrap-process [2].

Modeling of harmful substances accumulation as a result of pathogenic bacteria activity was made using parametric and non-parametric models (3) of class (5), (6). Coefficient estimation in the parametric models was made using method of fewer squares.

In different variants of calculation the relative square predicting error of bacteria quantity was not more than 3%. Parametric model in such case is: $x = 0,02t^4 + 3,3t$. Where x – bacteria quantity, t – time. Non-parametric models of type (4) using bootstrap-process [2] show better results (error of prediction was less than 2,5%).

The second process, identified in the work, is connected with carcinogenic substances accumulation (substances, which assist cancer development) in product during the process of its heating (frying with oil). The carcinogenic substances quantity in the technological environment is proportional to the heating time during high temperature (more than 100 degrees). As a source of statistical data in the work we used an experiment of Research Institute of Oil Industry [3].

Approximation of harmful substances accumulation as a result of carcinogens increasing during long oil heating [3]. Lets consider y (carcinogens quantity) from x (heating time) as several formulas, using piece approximation of graph of fast food harmful substances accumulation in product during heating:

$$y = \begin{cases} const, & x \in [0;4] \\ 0,128x^2 - 0,352, & x \in [4;10] \\ 12,5 + \sqrt[1,3]{x-10}, & x \in [10;\text{futher}] \end{cases}$$

(relative square prediction error was 7%, non-parametric models of type (4) using bootstrap-process show better results error is less than 5,5%). So, the made models may be used in research of the heating processes in oil and the processes of storage, in the cases when the real experiment is difficult to make.

During the researches of harmful substances accumulation in product as a result of pathogenic bacteria activity and carcinogens accumulation during heating process, practical experiments and modeling, it was shown that bacteria reproduction in product during storage, and also carcinogens accumulation during frying influences negatively to the food production safety and quality.

As a conclusion, it is necessary to mark that the above viewed new type of non-parametric estimations of regression line with a help of observations is a foundation for building non-parametrical models of stochastic processes with delay. There are also some private results of usage of non-parametrical models, and models of parametrical class during modeling of processes of harmful substances accumulation in food production during its preparation and storage.

References

[1] Demchenko Ya.I., Medvedev A.V. (2010) Some non-parametric estimation in identification problems for stochastic systems. *Probability theory, random processes, mathematical statistics and applications. Proceedings of the International Scientific Conference. Minsk.*

[2] Efron B. (1988) *Non-traditional methods of multivariate statistic analysis.* Mir, Moscow.

[3] Guravleva L.N. (2009) Study of oxidizing of oils during high-temperature heating while deep frying and development of the way to increase their stability. *Abstract of PhD thesis.* St.-Petersburg.

[4] Gutshmidt V.A. Demchenko Ya.I., Kurachenko M.V., Terenteva E.S., Faustov A.V. (2008) Non-parametric modeling of stochastic systems. *Probability theory, random processes, mathematical statistics and supplements. Proc. of international conference. Minsk* pp. 66-73.

[5] Katkovnik V.Ya. (1985) *Non-parametric identification and data smoothing.* Nauka, Moscow.

[6] Medvedev A.V. (1995) Data analysis in identification task. *Data computer analysis and modeling: Proc. of International conference. Minsk.* - Vol. **2**, pp. 201-206.

[7] Medvedev A.V. (1996) Non-parametric managing systems theory elements *Actual problems of informatics, applied mathematics and mechanics. Novosibirsk-Krasnoyarsk: Publishing house of Siberian Branch of Russian Science Academy* Part 3. Informatics. pp. 87-112.

[8] Vasilev V.A. Dobrovidov A.V., Koshkin G.M. (2004) *Non-parametric estimation of functional of sucessions stationery distributions.* Nauka, Moscow.