2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Input and Output Variables Selection within Non-Parametric System Identification

Kirill R. Chernyshov, *Member, IEEE*

Abstract— **This paper deals with the problem of selecting input and output variables for the system being specified. The paper covers the fairly general case where the output variables of the model are represented as conditional mathematical expectations of the output variables of the system model with respect to the (generalized) input variables. Output variables can also be selected in multiple-input/multiple-output models. This paper shows that the appropriate choice of variables (both inputs and outputs) is done in a way that applies measures consistent with Rényi's notions of random value and vector dependence (all definitions are given in the text). We also introduce the heterogeneity of input and output variables and present a generalization of Rényi's axioms to the case of multivariate dependence.**

I. Introduction

As it is well known, the best approximation of a non-linear dependence, say y of z, is the conditional mathematical expectation $\mathbf{E}\{y/z\}$ [1]. The same is with regard to the multiple dependence, say y of Z, where $Z=(z_1, \dots, z_n)^T$, what means $\mathbf{E}\{y/Z\}=\mathbf{E}\{y/z_1, \dots, z_n\}$. Accordingly, numerical publications are related to the problem, in particular, the ones based on the kernel regression estimation [2-6].

At the same time, few attentions are paid to the selection of namely the variables $Z=(z_1, \dots, z_n)^T$, what means a proper choice of the subset $Z_i=\left(z_{i_1}, \dots, z_{i_m}\right)^T$, $m\le n$, in a certain sense. From the point of view of the identification theory, variables $Z=(z_1, \dots, z_n)^T$ are natural to be referred to as input variables, while y, the output variable. As well, the same is valid for suitable selecting output variables $Y=\left(y_1, \dots, y_p\right)^T$, where $\mathbf{E}\{y_i/Z\}=\mathbf{E}\{y_i/z_1, \dots, z_n\}$, $i=1, \dots, p$. This means selecting a subset $Y_j=\left(y_{i_1}, \dots, y_{i_q}\right)^T$, $q\le p$, being suitable in a certain sense.

The paper is organized as follows. In the next section, few comments on measures of dependence are presented, continued with the notion of the consistent measures of dependence in Section III. Section IV is devoted to constructing Rényi-consistent measures of dependence, and Section V presents the input selection problem, continued with the reducing mutual dependence of the input variables in Section VI. In turn, Section VII is devoted to the extending Rényi axioms to the case of multivariate dependence, and Section VIII presents the selection of output variables. Ways of selecting a suitable consistent measure of dependence is considered in Section IX. Concluding remarks are presented in Section X.

K. R. Chernyshov is with V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences, Moscow, Russia (e-mail: myau@ipu.ru).

II. Comments on Measures of Dependence

In general, the solution of probabilistic system identification problems is always based on the use of random variable measures of dependence, among which the convenient correlation is mostly known, but it can vanish even in deterministic cases of the variables interconnection [7, 8].

Moreover, this applies to the case of stochastic dependence, say a part of the Sarmanov class of probability distribution densities $p_{zy;\lambda}(z,y)$ [9, 10]

$$p_{zy;\lambda}(z,y)=p_z(z)p_y(y)(1+\lambda\phi_1(z)\phi_2(y)), \tag{1a}$$

where

$$\int p_z(z)\phi_1(z)dz=0, \tag{1b}$$

$$\int p_y(y)\phi_2(y)dy=0, \tag{1c}$$

$$\lambda\phi_1(z)\phi_2(y)\ge -1. \tag{2}$$

In (1), $p_z(z)$, $p_y(y)$ are the marginal probability distribution densities, $\phi_1(z)$, $\phi_2(y)$ are functions, meeting condition (2) jointly with the scalar λ.

III. Consistent Measures of Dependence

To overcome the drawback concerned with vanishing a measure of dependence under a dependence, more complex, non-linear measures of dependence have been used to identify the system. And an important feature of this approach is the use of consistent dependence measures. According to Kolmogorov's terminology, a measure of dependence between two random variables is called consistent if the measure vanishes only if these random values are stochastically independent. Therefore, such a measure of dependence should be called Kolmogorov consistent.

At one time, paper [8] formulated seven axioms, which became known as the most appropriate for characterizing measures of dependence between two random values.

A) $\mu(z,y)$ *is defined for any pair of random values* z *and* y, *neither of them being constant with probability* 1.

B) $\mu(z,y)=\mu(y,z)$.

C) $0\le\mu(z,y)\le 1$.

D) $\mu(z,y)=0$ *if and only if* z *and* y *are independent.*

E) $\mu(z,y)=1$ *if there is a strict dependence between* z *and* y, *i.e. either* $y=\varphi(z)$ *or* $z=\psi(y)$ *where* φ *and* ψ *are Borel-measurable functions.*

F) *If a Borel-measurable functions* φ *and* ψ *map the real axis in a one-to-one way onto itself,* $\mu(\varphi(z),\psi(y))=\mu(z,y)$.

G) *If the joint distribution of* z *and* y *is Gaussian, then* $\mu(z,y)=|\boldsymbol{corr}(z,y)|$, *where* $\boldsymbol{corr}(z,y))$ *is the ordinary correlation coefficient of* z *and* y.

Correlation coefficients and correlation ratios are often used as measures of dependence based on comparing the characteristic moments of the joint and marginal probability distributions of a pair of underlying random variables of interest. The mentioned correlation ratio is of the form

$$\theta(z,y)=\frac{\mathbf{var}\left(\mathbf{E}\{{}^{y}/_{z}\}\right)}{\mathbf{var}(y)},\quad \mathbf{var}(y)>0, \tag{3}$$

as well, the maximum correlation coefficient is as follows:

$$R(z,y)=\sup_{\{B\},\{C\}}\mathbf{corr}(B(y),C(z)),$$
$$\mathbf{var}(B(y))>0,\quad \mathbf{var}(C(z))>0. \tag{4}$$

In (3), (4), $\mathbf{var}(\cdot)$ stands for the variance. In (4), the upper bound is taken over the sets of Borel measurable functions, $\{B\}$ and $\{C\}$, and also, $B\in\{B\}$, $C\in\{C\}$.

Of these, only the maximum correlation coefficient satisfies all of the axioms for the above measures of dependence [8], not the usual correlation or correlation ratio, and especially the correlation coefficient does not satisfy axioms D, E, and F.

Along with the maximum correlation coefficient, a wide class of dependence measures is constructed by direct comparison of the joint $p_{yz}(y,z)$ and marginal $p_y(y)$, $p_z(z)$ probability distribution densities of random values. Such a class is called divergence measures of probability distributions. The best-known representatives of the class are differential mutual information

$$0\le I(z,y)=\mathbf{E}\ln\frac{p_{zy}(z,y)}{p_z(z)p_y(y)}\le\infty, \tag{5}$$

and contingency coefficient

$$0\le\Delta^2(z,y)=$$
$$=\mathbf{E}\frac{(p_{zy}(z,y)-p_z(z)p_y(y))^2}{p_{zy}(z,y)p_z(z)p_y(y)}\le\infty. \tag{6}$$

Conventionally, in (5), (6) $\mathbf{E}\{\cdot\}$ stands for the mathematical expectation.

These measures of dependence are consistent according to Kolmogorov, but at the same time they do not satisfy Rényi's axioms C, E and G. These measures of dependence satisfy all of Rényi's axioms except perhaps axiom F, namely the invariance to one-to-one transformations of random variables, which we will refer to as the Rényi consistency.

IV. Constructing Rényi-Consistent Measures of Dependence

On the other hand, in the general case, Kolmogorov consistency is weaker than Rényi consistency, as is evident. The latter notion is more stringent. In this paper, we propose a constructive procedure to construct the Rényi consistent measure of dependence using the Kolmogorov consistent measure of dependence. This procedure modifies the Kolmogorov consistent measure of dependence to obtain a measure of dependence satisfying all Renyi's axioms except, possibly, axiom F, and consists of the following steps. Namely,

1) For any consistent in the Kolmogorov sense measure of dependence M_{zy} between random values z and y, calculate this measure for the two-dimensional standard Gaussian density depending on the correlation coefficient $\mathbf{corr}(z,y)$).

2) Represent the expression obtained as a function, say f, in modulo of the correlation coefficient

$$f(|\mathbf{corr}(z,y))|), \tag{7}$$

and invert expression (7).

3) The expression obtained,

$$f^{-1}\left(\mathrm{M}_{zy}\right), \tag{8}$$

(as a function of the initial measure of dependence M_{zy}) defines the measure of dependence between two random values z and y, meeting all Rényi axioms, with the exception, may be, axiom F.

In particular, for the maximum correlation coefficient, the corresponding function is the identity transformation. Therefore, this method assumes the following transformation for differential mutual information,

$$\iota(z,y)=f^{-1}\big(I(z,y)\big)=\sqrt{1-e^{-2I(z,y)}},$$

and for the contingency coefficient

$$\delta^2(z,y)=\Theta^{-1}_{\Delta(z,y)}\big(\Delta(z,y)\big)=$$
$$=\sqrt{\frac{\Delta^2\ (z,y)}{\Delta^2\ (z,y)+1}}.$$

These two equations are known in the literature as equations obtained by various non-universal methods and are presented here as examples to confirm the applicability of the general proposed technology.

V. Selecting Input Variables

Returning directly to system identification, the problem of probabilistic system identification and structural identification can actually be traced back to the possibility of identifying analytical relationships between the output and input variables of a system. Thus, only consistent measures of dependence can reveal such dependencies and describe them quantitatively. A system can then be considered as unidentifiable only if the value of the consistent measure of dependence is equal to zero for all pairs of input and output variables. On the other hand, the fact that the consistent measure of dependence is typically “only” positive may not be sufficient to decide to include a given input variable in the model, because its magnitude is so small that the corresponding effect of that input variable on the output variable may not be sufficient in the sense of the researcher. This intuitively leads to the conclusion that the identification of a system should take into account not only the consistent value of the measure of dependence, but also the number of input variables and the diversity of their effects on the corresponding output variable. This is also an incentive to use consistent measures of dependence in the context of the identification problem described earlier.

Consider the following simple example. If the output variable of a system is expressed as a linear dependence on, say, 100 input variables, then the Rényi consistent measure of dependence between the output and each input variable is 0.1, which intuitively would seem very small. Therefore, a natural conclusion would be to take into account the heterogeneity of the effect of the input variables on the output variables in order to make an informed decision on the choice of input variables.

The following approach can be applied to describe the heterogeneity of the impact of input variables on the output one. Let the input variables of the system $z_1, \dots, z_n$ be ordered in ascending order by the values of the Rényi-consistent measure of dependence of the output and input variables $z_{i_1}, \dots, z_{i_n}$

$$\mu_1 = \mu_R(y, z_{i_1}) \leq \cdots \leq \mu_n = \mu_R(y, z_{i_n}).$$

Then such heterogeneity can be *quantitatively* expressed by the following measure,

$$\eta^{inp} = \frac{\sum_{k=1}^{n} \nu_k \mu_k}{\sum_{k=1}^{n} \mu_k}, \qquad (9)$$

where

$$\nu_k = 2\frac{k-1}{n-1} - 1,$$

and take their values in the unit of interval: the closer the values of the Rényi-consistent measure of dependence are to each other, the closer to zero is the value of the heterogeneity measure characterizing the actual participation of the input variable and the uniformity of its effect on the output variable. The characteristics of the heterogeneity measure are as follows:

Taking values in a unit interval

$$0 \leq \eta^{inp} \leq 1;$$

It turns to zero only if the values of the measure of dependence according to Rényi between the output and input variables are identical,

$$\eta^{inp} = 0 \text{ if and only if } \mu_1 = \cdots = \mu_n;$$

It achieves unity only if the value of the Rényi consistent measure of dependence between the output and input variables is other than zero for an input variable and the value of the dependent measure of the consistent Rényi measure of dependence is zero for all other input variables.

$$\max_{\mu_1, \dots, \mu_n} \eta^{inp} = 1,$$

$$\underset{\mu_{i_1}, \dots, \mu_{i_n}}{\operatorname{argmax}} \eta^{inp} = \{0, \dots, 0, \mu\} \quad \forall \mu > 0.$$

To prove relationship (9), it is sufficient to look in more detail at the equation, where the coefficients are of the form:

$$2\frac{k-1}{n-1} - 1 = \frac{n}{n-1}\frac{2k-1}{n} - 1.$$

After that, (9) can be expressed as follows:

$$\eta^{inp} = \frac{n}{n-1}\frac{A_n}{B_n},$$

where: A_n is the square of a convex polygon formed by a segment $[A, C]$ and a broken line $AA_1A_2 \dots A_{n-1}C$, and B_n is the square of a triangle ABC. This view is the end proof result.

This therefore, the heterogeneity measure describes the diversity of the system structure in terms of input variables and provides a quantitative assessment that helps researchers decide whether to include one or more input variables in the final model. In the example under consideration, the heterogeneity measure is zero, which means that all input variables are equally important, which means that all of them should be included in the model.

Based on this consideration, each input variable in the system can be associated with a point on the unit square (Figure 1).

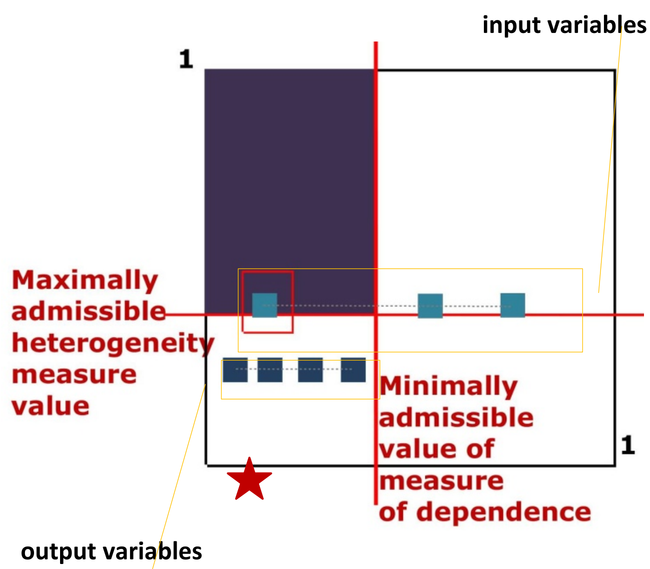

Figure 1. The choice of variables based on their measure of dependence and heterogeneity values.

In this case, the coordinates of such a point are the value of the Rényi-consistent measure of dependence plotted along the abscissa, $\mu_R(y, z_{i_j})$, individual for each z_{i_j}, $j = 1, \dots, n$, and the value of the heterogeneity measure is along the ordinate, η^{inp}, common for all z_{i_j}, $j = 1, \dots, n$. In this case, the researcher sets the minimum allowable value of the measure of dependence, say $\mu_{\min}$, and the maximum allowable value of the heterogeneity measure, say $\eta^{\max}$. These values define a rectangle in the upper left corner of the unit square.

$$\Gamma = [0, \mu_{\min}] \otimes [\eta^{\max}, 1]. \quad (10)$$

Thus, if a point lies within rectangle (10), then the corresponding output variable should not be included in the constructed system model, and if not, then the input variable should be included. Figure 1 shows the corresponding diagram for the three variables "above", where the asterisks indicate the case described in the example above for all input variables to be used in the model.

VI. Reducing The Mutual Dependence of Input Variables

On the other hand, there may be important relationships with the input variables. These relationships are naturally reflected in the value of the Rényi-consistent measure of dependence, which represents the relationship between the output variables and the input variables. Therefore, the following algorithm is proposed to appropriately exclude the input variables from the model based on the above ideas. For the input variable with the largest Rényi-consistent measure of dependence on the output variable

$$\mu_R(y, z_{i_0}) \geq \mu_R\left(y, z_{i_j}\right) \quad \forall j \neq 0$$

Rényi-consistent measures of dependence with each of the other *input* variables are calculated,

$$\mu_R\left(z_{i_0}, z_{i_j}\right) \quad \forall j \neq 0.$$

If the values of these measure of dependence are greater than the value κ specified by the researcher,

$$\mu_R(z_{i_0}, z_{i_k}) \geq \kappa,$$

this is because the contribution to the output variable was made by the input variable which, relatively, has the highest value of the measure of dependence with the output variable. Consequently, these variables z_{i_k} are excluded from the model.

VII. Extending Rényi Axioms to the Case of The Multivariate Dependence

The presented approach for constructing Rényi-consistent measures of dependence allows a detailed extension of Rényi's axioms to the multivariate dependence case and, hence, Rényi-consistent dependence measures for random vectors. In the literature, an extension of Rényi's axioms to the multivariate case was presented in [11]. However, it does not provide the exact agreement with Rényi's axioms for the two-dimensional distribution case that is generally required.

In particular, it concerns axioms C, E and G, which are related to the normalization conditions. On the other hand, the normalization condition is very important, since taking only the semi-positive values does not actually provide a basis for comparing and evaluating the significance of certain quantities. The axioms proposed in this paper, in a suitable comparison with the axioms of [11], are as follows, with indicating the changes in red.

A) $\mu(z_1, \dots, z_n)$ is defined for any random vector $Z = (z_1, \dots, z_n)$, neither of the components of the vector Z being a constant with probability 1.

B) For any permutation $\sigma = (i_1, \dots, i_n)$ of the indexes $\{1, 2, \dots, n\}$ the measure is invariant, i.e. $\mu(Z) = \mu(z_1, \dots, z_n) = \mu(z_{i_1}, \dots, z_{i_n})$.

C) $0 \leq \mu(z_1, \dots, z_n) \leq 1$ versus $0 \leq \mu(z_1, \dots, z_n) = \gamma$ (can be ∞) in [11].

D) $\mu(z_1, \dots, z_n) = 0$ if and only if the random variables $z_1, \dots, z_n$ are independent.

E) $\mu(z_1, \dots, z_n) = 1$ versus $0 \leq \mu(z_1, \dots, z_n) = \gamma$ (can be ∞) in [11] if and only if there exists a deterministic dependence between components of the random vector Z.

F) Invariance to any one-to-one transformation $\Psi = (\psi_1, \dots, \psi_n)$ of $Z = (z_1, \dots, z_n)$ onto R^n, i.e. $\Psi(Z) = \left(\psi_1(z_1), \dots, \psi_n(z_n)\right)$, namely:

$$\mu\left(\psi_1(z_1), \dots, \psi_n(z_n)\right) = \mu(z_1, \dots, z_n).$$

G) If the joint distribution of the vector $Z = (z_1, \dots, z_n)$ is Gaussian, then for the case of $n = 2$, $\mu(z_1, z_2) = |\mathbf{corr}(z_1, z_2)|$ versus $\mu(z_1, z_2)$ is an increasing function of $\mathbf{corr}(z_1, z_2)$ in [11].

VIII. Selecting Output Variables

Similar to the approach described above for characterizing heterogeneity in the significance of input variables, the Rényi-consistent multiple measure of dependence can be used to properly construct the heterogeneity of the output variables of the system under study and reveal the most significant ones. That is, suppose that the output variables of the system are sorted in ascending order based on the value of the Rényi-consistent multiple measure of dependence for each output and all input variables

$$\mu_1 = \mu_R(y_{i_1}, z_1, \dots, z_{p_1}) \leq \dots \leq \leq \mu_m = \mu_R(y_{i_M}, z_1, \dots, z_{p_l}). \quad (11)$$

In (11), the lower scripts $1, \dots, p_1, \dots, 1, \dots, p_l$, are used to denote input variables that have been selected in accordance to the above procedure for the input variables selection.

The heterogeneity of the effects of all input variables (selected) on each output variable of the target system can then be expressed quantitatively by the following measure, which takes values in the units interval

$$\eta^{out} = \frac{\sum_{k=1}^{p} \nu_k \mu_k}{\sum_{k=1}^{p} \mu_k}, \quad (12)$$

where

$$\nu_k = 2\frac{k-1}{p-1} - 1.$$

The characteristics of heterogeneity measure of the output variable (12) are the same as that of the input variable.

It takes values in unit interval,

$$0 \le \eta^{out} \le 1.$$

It becomes zero only if the value of the Rényi multiple measure of dependence between all output and input variables matches

$$\eta^{out} = 0 \text{ if and only if } \mu_1 = \cdots = \mu_p.$$

Unity is achieved only if the measure of multiple dependence between an output and all input variables is not zero, its value is consistent with the Rényi concept for all other output variables, and the measure of multiple dependence is zero,

$$\max_{\mu_1,\dots,\mu_m} \{\eta^{out}\} = 1,$$

$$\underset{\mu_{i_1},\dots,\mu_{i_p}}{\operatorname{argmax}} \eta^{out} = \{0, \dots, 0, \mu\} \quad \forall \mu > 0.$$

Accordingly, the solution of the problem of inclusion or exclusion of one or another output variable is solved similarly to the case of choosing input variables. Namely, the researcher presets the minimally admissible magnitude of the Rényi-consistent measure of the multiple dependence $\mu_{\min}$, and the maximally admissible magnitude of the measure of the heterogeneity $\eta^{\max}$. These magnitudes forms a rectangle

$$\Delta = [0, \mu_{\min}] \otimes [\eta^{\max}, 1]. \quad (13)$$

and if there is an output variable within rectangle (13), it is excluded from the model. Figure 1 shows an example of selecting output variables within the unit square (bottom four dots), which means that all four output variables are included in the model.

IX. Towards the Choice of the Rényi Consistent Measure of Dependence

Strictly speaking, the choice of a Rényi consistent measure of dependence is a separate problem. In this case, Rényi axiomatics [8] plays the most significant role. Thus, the maximum correlation is redundantly satisfying the Rényi axioms, insofar as, in accordance with the Rényi axioms, the measure of dependence must satisfy the axiom of invariance with respect to the one-to-one transformation of each of the pair of random variables. But the maximum correlation satisfies such an axiom for any transformations. On the other hand, the calculation of the maximum correlation is a rather complicated problem associated with the iterative calculation of the first eigenfunctions of the stochastic kernel, which has the form

$$\frac{p_{yz}(y,z)}{\sqrt{p_y(y)p_z(z)}}.$$

Naturally, under the conditions of solving the identification problem itself, when neither the joint nor the marginal probability distribution densities of random variables are known, the construction of the first eigenfunctions of a given stochastic kernel makes the solution of such a problem much more complicated.

As is well known, a wide class of consistent measures of dependence can be built on the basis of divergence measures of probability distributions, of which the Kullback-Leibler divergence is the best known. But at the same time, it is not symmetrical. On its basis, a well-known dependence measure is built – differential mutual information – and, accordingly, the information correlation coefficient

$$\sqrt{1 - e^{-2I_{yz}(y,z)}},$$

where $I_{yz}(y,z)$ is the differential mutual information (5). In turn, the construction of estimates of the differential mutual information based on sample dependent observations is based on L'Hospital rule [12] applied for each of the differential entropies, which, ultimately, is a serious problem.

This problem disappears when Rényi and Tsallis divergence measures of any order are used. On the other hand, the computational complexity of the Rényi and Tsallis divergences increases with increasing the order, but it is generally accepted that order two, i.e. quadratic divergence, is quite sufficient. On the other hand, another advantage of Rényi divergence over Kullback-Leibler divergence is that Rényi divergence contains the logarithm of the integral rather than the integral of the logarithm. But the Tsallis divergence does not contain a logarithm at all. At the same time, a symmetric measure of divergence was constructed in [13] based on Tsallis entropy of order two. It can be shown that its version, normalized to unity,

$$\sqrt{1 - \frac{1}{\left(8\pi I_2^T(y,z) + 1\right)^2}},$$

where $I_2^T(y,z)$ is the symmetric mutual information of Tsallis of order two [13], satisfies all Rényi axioms, except for the axiom about invariance with respect to a one-to-one transformation of each of a pair of random variables. Thus, this measure of dependence is Rényi consistent and is the most appropriate for the purposes outlined in this paper.

X. Conclusions

The paper has been devoted to a most general case of selecting variables in the non-linear system model represented in the form of the conditional mathematical expectation. Meanwhile, the input variable $Z = (z_1, \dots, z_n)^T$ can involve both system input variables and preceding output variables, and the problems of selecting these variables (their quantity and quality) have been considered in the paper.

At the heart of the whole approach is the consistent application of the concept of Rényi measures of dependence and heterogeneity measures that allow for the selection of input and output variables. And of course, the role of the researcher remains in the foreground, appropriately choosing the minimum allowable size of the measure of dependence and the maximum allowable size of the heterogeneity measure.

The approach is distribution free, and as a measure of dependence *any* consistent in the Rényi sense measure of dependence can be applied. Nevertheless, the inference of Section IX about selecting a Rényi consistent measure of dependence should not be excluded.

REFERENCES

[1] G. S. Watson. "Smooth regression analysis", *Sankhya. Ser. A*, vol. 26, 1964, pp. 355-372.

[2] A. V. Dobrovidov, G. M. Koshkin, and V. A. Vasiliev. *Non-parametric State Space Models*, Kendrick Press, 2012, 501 p.

[3] L. Fahrmeir, T. Kneib, S. Lang, and B. D. Marx. *Regression. Models, Methods and Applications*, Springer, 2021, 746 p.

[4] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*, Springer, 2002, 650 p.

[5] O. Nelles. *Nonlinear System Identification. From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes*, Springer, 2020, 1226 p.

[6] G. Pillonetto, Tianshi Chen, A. Chiuso, G. De Nicolao, and L. Ljung. *Regularized System Identification. Learning Dynamic Models from Data*, Springer, 2022, 394 p.

[7] N. S. Rajbman. "Extensions to nonlinear and minimax approaches", *Trends and Progress in System Identification*, ed. P. Eykhoff, Pergamon Press, Oxford, 1981, pp. 185-237.

[8] A. Rényi. "On measures of dependence", *Acta Mathematica Academiae Scientiarum Hungarica*, vol. 10, no. 3–4, 1959, pp. 441-451.

[9] N. Balakrishnan, and Chin-Diew Lai. *Continuous Bivariate Distributions*, Second Edition, Springer, 2009, 712 p.

[10] O. V. Sarmanov. "Remarks on uncorrelated Gaussian dependent random variables", *Theory Probab. Appl.,* vol. 12, no. 1, 1967, pp. 124-126.

[11] A. C. Micheas and K. Zografos. "Measuring stochastic dependence using φ-divergence", *J. Multivariate Analysis*, vol. 97, 2006, pp. 765-784.

[12] A. Mokkadem. "Estimation of the entropy and information of absolutely continuous random variables", *IEEE Transactions on Information Theory*, 1989, vol. IT-35, pp. 193-196.

[13] K. R. Chernyshov. "Constructing a Symmetric Tsallis Divergence as a System Identification Criterion", *IFAC-PapersOnLine*, 2018, vol. 51, no. 32, pp. 550-555.