

Article

Parameterization of the Stochastic Model for Evaluating Variable Small Data in the Shannon Entropy Basis

Oleh Bisikalo ¹, Vyacheslav Kharchenko ², Viacheslav Kovtun ^{3,*}, Iurii Krak ⁴ and Sergii Pavlov ⁵

- ¹ Department of Automation and Intelligent Information Technologies, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Khmelnske Shose Str. 95, 21000 Vinnytsia, Ukraine
- ² Department of Computer Systems, Networks and Cybersecurity, Faculty of Radio Electronics, Computer Systems and Information Communications, National Aerospace University KhAI, Chkalov Str. 17, 610700 Kharkiv, Ukraine
- ³ Department of Computer Control Systems, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Khmelnske Shose Str. 95, 21000 Vinnytsia, Ukraine
- ⁴ Department of Theoretical Cybernetics, Faculty of Computer Sciences and Cybernetics, Taras Shevchenko National University of Kyiv, Volodymyrska Str. 60, 01033 Kyiv, Ukraine
- ⁵ Laboratory of Biomedical Optics, Department of Biomedical Engineering, Faculty for Infocommunications, Radioelectronics and Nanosystems, Vinnytsia National Technical University, Khmelnske Shose Str. 95, 21000 Vinnytsia, Ukraine
- * Correspondence: kovtun_v_v@vntu.edu.ua

Abstract: The article analytically summarizes the idea of applying Shannon's principle of entropy maximization to sets that represent the results of observations of the "input" and "output" entities of the stochastic model for evaluating variable small data. To formalize this idea, a sequential transition from the likelihood function to the likelihood functional and the Shannon entropy functional is analytically described. Shannon's entropy characterizes the uncertainty caused not only by the probabilistic nature of the parameters of the stochastic data evaluation model but also by interferences that distort the results of the measurements of the values of these parameters. Accordingly, based on the Shannon entropy, it is possible to determine the best estimates of the values of these parameters for maximally uncertain (per entropy unit) distortions that cause measurement variability. This postulate is organically transferred to the statement that the estimates of the density of the probability distribution of the parameters of the stochastic model of small data obtained as a result of Shannon entropy maximization will also take into account the fact of the variability of the process of their measurements. In the article, this principle is developed into the information technology of the parametric and non-parametric evaluation on the basis of Shannon entropy of small data measured under the influence of interferences. The article analytically formalizes three key elements: -instances of the class of parameterized stochastic models for evaluating variable small data; -methods of estimating the probability density function of their parameters, represented by normalized or interval probabilities; -approaches to generating an ensemble of random vectors of initial parameters.

Keywords: Shannon entropy; machine learning; evaluation of small data; measurement errors; stochastic model; parametric optimization; normalized probabilities; interval probabilities



Citation: Bisikalo, O.; Kharchenko, V.; Kovtun, V.; Krak, I.; Pavlov, S. Parameterization of the Stochastic Model for Evaluating Variable Small Data in the Shannon Entropy Basis. *Entropy* **2023**, *25*, 184. <https://doi.org/10.3390/e25020184>

Academic Editor: Yong Deng

Received: 26 December 2022

Revised: 10 January 2023

Accepted: 11 January 2023

Published: 17 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

One of the most relevant problems of modern science is the extraction of useful information from available data. In various fields of science, methodologies aimed at solving this problem are being developed. Each such methodology is based on a certain hypothesis about the properties of the data and the real or hypothetical source of their origin. In the context of the data evaluation problem, two fundamental hypotheses can be distinguished [1–5]. The first hypothesis focuses on directly measurable, deterministic parameters to identify potential functional dependencies between them. All data that

cannot be attributed to one or more defined parameters are considered influences in this hypothesis and are rejected. Naturally, such an approach is adequate and productive only if the information is extracted from data obtained from a known, sufficiently investigated the source of origin. The second hypothesis focuses on the analysis of the data as such and is focused on identifying patterns in them, the presence of which can be assessed using a certain defined metric. This can be, for example, a measure of data sufficiency, a property of a sample from the general population, the normality of probability distribution densities, etc. It is practically impossible to guarantee the characteristics of these properties for specific data. However, the improbable becomes common if we analyze not data, but Big Data. This trend is the basis for the progress of such methodologies as mathematical statistics [2,6–8], machine learning [9–12], econometrics [13–16], financial mathematics [17–19] and control theory [20–23].

In recent decades, the first two of the methodologies just mentioned have been heard. Machine learning is based on the axiomatic perception of probability spaces, as outlined in the paradigm of the theory of statistical learning developed in the 1960s [24–26]. There are several dominant categories of machine learning, but the most common is tutored learning [9,10,27,28]. In this category, researchers work with symmetric finite datasets, summarized in the “input” and “output” entities. The purpose of data analysis is to identify the functional dependence between these entities. The set of admissible types of functions forms the hypothesis space of this category of machine learning. The machine learning algorithm consistently evaluates the expected risks of describing the dependence of the existing “input” and “output” entities by each type of function from the hypothesis space. The evaluation is carried out by calculating a single loss function for the entire research. The expected risk is understood as the product of the sum of the estimates and the probability distribution of the data. If the compatible mapping probability distribution is known, then finding the best hypothesis is a trivial task. In the general case, the distribution is unknown, so the machine learning algorithm chooses the most appropriate hypothesis according to a certain rule and proves this thesis by calculating the empirical risk. In addition to the computational complexity, the disadvantage of machine learning is the tendency of the algorithms of this methodology to minimize the loss function by overfitting the potentially best hypothesis to the available data (so-called overtraining [9,27,29]). A typical way to detect (but not prevent) overtraining is to test the best hypothesis on data that the algorithm has not yet worked on (the control sample). Methods of mathematical statistics are not subject to retraining, because they do not assess empirical risk as such.

A typical example of a problem, in the process of solving which the characteristic features of mathematical statistics and machine learning are manifested, is linear regression [7–11]. In the classic formulation of this problem, we need to find the regression coefficients that minimize the root mean square error between the reference entity “output” and its pattern as generated by the model. Such a problem can be solved in a closed form. The theory of statistical learning states that, if we choose the root mean square error as the loss function and carry out empirical risk optimization, then the obtained result will coincide with the one that we will obtain by applying traditional linear regression analysis. However, the maximum likelihood method [2,6,7,30] characteristic for mathematical statistics will demonstrate a similar result in this situation. By the way, the methods of mathematical statistics do not operate with the concepts of initial and test samples, but use metrics to evaluate the results of the model. In our example, the statistical approach allows us to reach the optimal solution because the solution itself exists in a closed form. The maximum likelihood method does not test alternative hypotheses and does not converge to the optimal solution, unlike a machine learning algorithm. However, if the piecewise linear loss function is used for the machine learning algorithm in the same problem, the final result does not coincide with the maximum likelihood method. The machine learning algorithm allows us to expand the space of relevant hypotheses with an a priori considered loss function. The process of their evaluation is carried out automatically. The maximum likelihood method can estimate the accuracy of the original model but does not allow us

to automatically change its appearance. Therefore, the methods of machine learning and mathematical statistics work in different ways, while producing similar results. If the task of the researcher is to accurately predict the cost of housing, then machine learning tools are exactly what is needed. If a scientist is investigating the relationships between parameters or making scientifically based conclusions about the data, then a statistical model cannot be dispensed with.

Finally, machine learning experts say, “There are no such things as unsolvable problems, either data or computing power is scarce”. Indeed, everyone has heard about Big Data analysis [10–12,31]. Now, however, the issue of analyzing so-called “small data” is becoming increasingly common [32,33]. Classical machine learning approaches are helpless in such a situation. This circumstance prompted the authors to write this article.

Taking into account the strengths and weaknesses of the mentioned methods, we will formulate the necessary attributes of scientific research.

The **object of the research** is the process of the parameterization of the stochastic model for evaluating variable small data for machine learning purposes.

The **research subject** is probability theory and mathematical statistics, evaluation theory, information theory, mathematical programming methods and experiment planning theory.

The **research aims** to formalize the process of finding the best estimates of the probability density functions for the characteristic parameters of instances of the class of stochastic models for evaluating variable small data.

The research objectives are:

- (1) To formalize the process of calculating the variable entropy estimation of the probability density functions of the characteristic parameters of the stochastic variable small data estimation model, represented by normalized probabilities;
- (2) To formalize the process of calculating the variable entropy estimation of the probability density functions of the characteristic parameters of the stochastic variable small data estimation model, represented by interval probabilities;
- (3) To justify the adequacy of the proposed mathematical apparatus and demonstrate its functionality with an example.

The **main contribution** of the research is that the article analytically summarizes the idea of applying the Shannon entropy maximization principle to sets that represent the results of observations of the “input” and “output” entities of the stochastic model for evaluating variable small data. To formalize this idea, a sequential transition from the likelihood function to the likelihood functional and the Shannon entropy functional is analytically described. Shannon’s entropy characterizes the uncertainty caused not only by the probabilistic nature of the parameters of the stochastic data evaluation model but also by influences that distort the results of the measurements of the values of these parameters. Accordingly, based on the Shannon entropy, it is possible to determine the best estimates of the values of these parameters for maximally uncertain (per entropy unit) influences that cause measurement variability. This postulate is organically transferred to the statement that the estimates of the probability distribution density of the parameters of the stochastic model of small data obtained as a result of Shannon entropy maximization will also take into account the fact of the variability of the process of their measurements. In the article, this principle is developed into the information technology of parametric and non-parametric evaluation on the basis of Shannon entropy of small data measured under the influence of interferences.

The highlights of the research are:

- (1) Instances of the class of parameterized stochastic models for evaluating variable small data;
- (2) Methods of estimating the probability density function of their parameters, represented by normalized or interval probabilities;
- (3) Approaches to generating an ensemble of random vectors of initial parameters;

(4) A technique for statistical processing of such an ensemble using the Monte Carlo method to bring it to the desired numerical characteristics.

2. Models and Methods

2.1. Statement of the Research

Evaluation based on data that represent parametric signals or phenomena of physical, medical, economic, biological and other sources of origin is the functional purpose of evaluation theory as a branch of mathematical statistics. To solve the problem of evaluation, parametric and non-parametric approaches are used. In recent decades, the latter has noticeably dominated the former, which has become possible thanks to the “reactive” progress in the field of machine learning and artificial intelligence. At the same time, the focus of researchers’ interest is shifting from the study of the processes represented by Big Data to that of those processes about which the amount of data small, and the data itself contains errors. Such a preamble encourages the perception of the parameters of the small data evaluation model as stochastic quantities. Accordingly, we will call such a model a stochastic model for small data evaluation. The characteristics of such a model are the probability density functions of the stochastic parameters. The primary task in identifying a stochastic estimation model for specific small data is to estimate the parameters of these probability density functions. If this step is passed, then the identified stochastic evaluation model can be taken as a basis for forming moment models of small data, generating an ensemble of random vectors of the initial parameters and carrying out the statistical processing of such an ensemble using the Monte Carlo method [6–8] to bring it to the desired numerical characteristics. The formalization of the way to solve the primary problem formulated above has scientific potential and applied value.

Let there be a stochastic parameterized research object represented by the results of measurements, in which the matrix of values of the input parameters X with the dimension $[o \times n]$ (entity “input”) is matched by a vector of values of the output parameter y with the dimension $[o \times 1]$ (entity “output”), where o is the number of censored observations, and n is the number of input characteristic parameters of the research object.

The process of measuring the values of matrix X and vector y is characterized by errors, which are represented by the symmetrical matrix $N = (v_{ji})$ (variability of the measurement process), $i = \overline{1, n}$, $j = \overline{1, o}$, and vector $v = (v_i)$, where v_{ji} , v_i are independent stochastic values, $\forall i, j$. The value of these stochastic quantities belongs to the intervals $N_{ji} = [v_{ji}^-, v_{ji}^+]$ and $\Upsilon_j = [v_j^-, v_j^+]$, respectively.

The stochastic model of the $\langle X, y \rangle$ data evaluation is represented by an expression

$$s = F(X + N, \alpha) + v, \quad (1)$$

where F is a defined o -dimensional vector function, α is a random n -dimensional vector formed by independent stochastic parameters α_i , $i = \overline{1, n}$, $\forall \alpha_i \in A_i = [\alpha_i^-, \alpha_i^+]$.

Let us assume that the parameters of the stochastic model and the variability of the measurements are continuous stochastic quantities, the values of which belong to the corresponding intervals of the tuple $\langle N_{ji}, \Upsilon_j, A_i \rangle$ (hereinafter—the “genuine” version of the stochastic Model (1) or $GvSM$).

In this case, the probability density functions of the stochastic parameters of $GvSM$ (variability of measurements $P(\alpha)$, input $W(N)$ and output $Q(v)$ parameters) (Independent Stochastic Parameters of the Small Data Estimation Model) are described by the expressions:

$$P(\alpha) = \prod_{i=1}^n p_i(\alpha_i), \quad (2)$$

$$W(N) = \prod_{j=1}^o \prod_{i=1}^n w_{ji}(v_{ji}), \quad (3)$$

$$Q(v) = \prod_{j=1}^o q_j(v_j), \quad (4)$$

where $\alpha_i \in A_i$, $v_{ji} \in N_{ji}$ and $v_j \in Y_j$, respectively. Formulating Expressions (2)–(4), the authors implied a priori that the measurement results were obtained in accordance with the provisions of the experiment planning theory. The corresponding variables are statistically independent.

Functions (2)–(4) will be evaluated based on data $\langle X, y \rangle$ according to Model (1), taking into account the available a priori information summarized by the tuple $\langle P^0(\alpha), W^0(N), Q^0(v) \rangle$.

The stochastic Model (1) generates an ensemble of random vectors s , which can be compared with the vector v obtained as a result of measurements. To carry out such an estimation of the probability density Functions (2)–(4), we will use k moments of the stochastic components of the vector s :

$$m^{(k)} = \left\{ M(s_j^{(k)}) \right\}, j = \overline{1, o},$$

where (Numerical characteristics for estimating these stochastic parameters)

$$M(s_j^{(k)}) = \int_{\substack{\alpha \in A, \\ v \in N, \\ v \in Y}} (F_j(X + N, \alpha) + v_j)^k dP(\alpha) dW(N) dQ(v).$$

Next, we will use moments of the first order ($k = 1$). In accordance:

$$M(s) = \bar{s} = \int_{\substack{\alpha \in A, \\ v \in N, \\ v \in Y}} (F(X + N, \alpha) + v) dP(\alpha) dW(N) dQ(v). \quad (5)$$

Another version of the implementation of the Model (1) will be one in which the parameters of the stochastic model and the variability of the measurements are continuous stochastic values, the belonging of which to the corresponding interval of the tuple $\langle N_{ji}, Y_j, A_i \rangle$ will be characterized by a certain probability (hereinafter—the “quasi” version of the stochastic Model (1) or $QvSM$). In this case:

- (1) the parameters α_i take values in A_i intervals with probabilities $p_i \in [0, 1]$, $i = \overline{1, n}$;
- (2) the parameters v_{ji} take values in intervals N_{ji} with probabilities $w_{ji} \in [0, 1]$, $j = \overline{1, o}$, $i = \overline{1, n}$;
- (3) the parameters v_j take values in intervals Y_j with probabilities $q_j \in [0, 1]$, $j = \overline{1, o}$.

The available a priori information is summarized by the vector (p_i^0, w_{ji}^0, q_j^0) , $j = \overline{1, o}$, $i = \overline{1, n}$.

At the same time, Expressions (2)–(4) retain their legitimacy. We generalize the initial numerical characteristics of $QvSM$ in the form of a vector of quasi-momentums of the first order:

$$\alpha = \alpha^- + PL_\alpha, v = v^- + W \otimes L_v, v = v^- + QL_v, \quad (6)$$

where $L_\alpha = \text{diag}(\alpha_i^+ - \alpha_i^- | i = \overline{1, n})$, $L_v = \text{diag}(v_{ji}^+ - v_{ji}^- | j = \overline{1, o}, i = \overline{1, n})$, $L_\alpha = \text{diag}(v_j^+ - v_j^- | j = \overline{1, o})$ and the $\otimes$ sign represents the element-by-element multiplication operation. Expressions (6) declare the replacement of the elements of the tuple $\langle \alpha, v, v \rangle$ with their quasi-average values.

The analytical expression for the first-order quasi-momentum of the stochastic vector s can be obtained by substituting numerical Characteristics (6) into Expression (1):

$$\tilde{s} = F(v^- + X + W \otimes L_v, \alpha^- + PL_\alpha) + v^- + QL_v. \quad (7)$$

In the context of the proposed statement of the research, we specify its aim and objectivities.

The research aims to formalize the process of finding the best estimates of the probability density functions for the $\langle p, q \rangle$ parameters of $GvSM$ and $QvSM$ represented by Expressions (5) and (7), respectively.

The objectives of the research are:

- (1) To formalize the process of calculating the variable entropy estimation of the probability density functions of characteristic parameters of $GvSM$ represented by normalized probabilities;
- (2) To formalize the process of calculating the variable entropy estimation of the probability density functions of characteristic parameters of $QvSM$ represented by interval probabilities;
- (3) To justify the adequacy of the proposed mathematical apparatus and demonstrate its functionality with an example.

2.2. Parameterization of the Stochastic Model for Evaluating Variable Small Data in the Shannon Entropy Basis

Let us formulate the corresponding probability functionals for the available information about the values of the input and output parameters of the stochastic Model (1).

Taking into account the independence of the parameters of the “input” and “output” entities in the stochastic Model (1) and the variability of their measurement procedure, we determine the compatible probability density function $\Phi(\alpha, v, v)$ and the corresponding logarithmic likelihood ratio $\phi(\alpha, v, v)$ as

$$\Phi(\alpha, v, v) = P(\alpha)W(v)Q(v), \quad (8)$$

$$\phi(\alpha, v, v) = \ln \frac{P(\alpha)}{P^0(\alpha)} + \ln \frac{W(v)}{W^0(v)} + \ln \frac{Q(v)}{Q^0(v)}. \quad (9)$$

Based on Expressions (8) and (9), we formulate the likelihood functional $L(P(\alpha), W(v), Q(v))$:

$$\begin{aligned} L(P(\alpha), W(v), Q(v)) = & \int_{\substack{\alpha \in A, \\ v \in N, \\ v \in Y}} \Phi(\alpha, v, v) \phi(\alpha, v, v) d\alpha dv dv = \int_{\alpha \in A} P(\alpha) \ln \frac{P(\alpha)}{P^0(\alpha)} d\alpha \\ & + \int_{v \in N} W(v) \ln \frac{W(v)}{W^0(v)} dv + \int_{v \in Y} Q(v) \ln \frac{Q(v)}{Q^0(v)} dv. \end{aligned} \quad (10)$$

Expression (10) presented in the $-L(P(\alpha), W(v), Q(v))$ format is the Shannon entropy functional [34,35]. According to its purpose, such a functional is a measure for evaluating the degree of variability of the elements of a tuple $\langle P(\alpha), W(v), Q(v) \rangle$. This fact determines the perspective of using such a functional for evaluating Functions (2)–(4). In the context of this motivation, let us transform Expression (10) into the form

$$\begin{aligned} H(\bar{s}) = & - \sum_{i=1}^n \int_{\alpha_i \in A_i} p_i(\alpha_i) \ln \frac{p_i(\alpha_i)}{p_i^0(\alpha_i)} d\alpha_i - \sum_{j=1}^o \sum_{i=1}^n \int_{v_{ji} \in N_{ji}} w_{ji}(v_{ji}) \ln \frac{w_{ji}(v_{ji})}{w_{ji}^0(v_{ji})} dv_{ji} \\ & - \sum_{j=1}^o \int_{v_j \in Y_j} q_j(v_j) \ln \frac{q_j(v_j)}{q_j^0(v_j)} dv_j. \end{aligned} \quad (11)$$

The Functional (11) is defined for estimating the probability density functions of stochastic parameters of $GvSM$. For $QvSM$, based on Expression (10), we obtain:

$$H(\tilde{s}) = -\sum_{i=1}^n p_i \ln \frac{p_i}{p_i^0} - \sum_{j=1}^o \sum_{i=1}^n w_{ji} \ln \frac{w_{ji}}{w_{ji}^0} - \sum_{j=1}^o q_j \ln \frac{q_j}{q_j^0}.$$

Based on Definition (11), we formulate the problem of finding the optimal estimate of the probability density functions of stochastic parameters of $GvSM$, taking into account the fact of their variability, i.e., $E_{\tilde{s}}$.

We define the objective function of such an optimization problem as:

$$H(\tilde{s}) \rightarrow \max. \quad (12)$$

We define the restrictions of the $E_{\tilde{s}}$ optimization problem as

$$E = P \cup W \cup Q, \quad (13)$$

that is, the probability distribution density of the variability of measurements $P(\alpha) \in P$, input $W(N) \in W$ and output $Q(v) \in Q$ parameters of $GvSM$ must belong to the space E defined by Expression (13), and

$$y = \frac{1}{(M(F^{(k)}(X + v, a) + v^{(k)}))^k}, \quad (14)$$

that is, the elements of the vector with the results of measurements y are equal to the elements of the k th moment of the vector s raised to the k^{-1} th power.

By analogy with the formulation of the optimization Problem (12)–(14), we formulate the problem of finding the optimal estimate of the probability density functions of stochastic parameters of $QvSM$, taking into account the fact of their variability, i.e., $E_{\tilde{s}}$.

We define the objective function of such an optimization problem as:

$$H(\tilde{s}) \rightarrow \max. \quad (15)$$

Recall that the complex parameter $\bar{s}$ generalizes a tuple of interval controlled parameters $\langle P(\alpha), W(v), Q(v) \rangle$ (see Expressions (10) and (5)), and the complex parameter $\tilde{s}$ focuses on the variability of measuring these characteristic parameters (see Expression (7)).

Considerations regarding the formulation of restrictions for finding the extremum of the objective Function (15) are identical to those embodied in Restrictions (13) and (14). At the same time, Restriction (13) fully satisfies the statement of the Problem (15), while Restriction (14) can be written in terms of the definition of $QvSM$:

$$y = F((X + v^- + WL_v), (\alpha^- + PL_\alpha)) + v^- + QL_v. \quad (16)$$

Let us pay attention to the situation when the measurement errors $v(t)$ and the values of the vector of the initial parameters of the stochastic model $s(t)$ are characterized by non-linearity of the r th degree:

$$s(t) = \sum_{r=1}^R \sum_{i=1}^n \alpha_i^h x^{(h_i)}(t) + v(t),$$

where $\alpha = (\alpha_i)$ is a vector of parameters, the independent stochastic elements of which take values from the ranges $A_i = [\alpha_i^-, \alpha_i^+]$ with the probability distribution densities $p_i(\alpha_i)$, $i = \overline{1, n}$.

The measurement of the components of the entities “input” and “output” of the investigated process takes place at moments $t_j, j = \overline{1, o}$. The entity “input” is represented by a set of r -matrices, $r = \overline{1, R}$, of the form

$$X^{(r)} = \begin{pmatrix} x_1^{(r)}(t_1) & \dots & x_n^{(r)}(t_1) \\ \vdots & \ddots & \vdots \\ x_1^{(r)}(t_o) & \dots & x_n^{(r)}(t_o) \end{pmatrix} = \begin{pmatrix} x_{1,1}^{(r)} & \dots & x_{1,n}^{(r)} \\ \vdots & \ddots & \vdots \\ x_{o,1}^{(r)} & \dots & x_{o,n}^{(r)} \end{pmatrix}$$

and the entity “output” is represented by stochastic elements of the vector $s = (s(t_j)), j = \overline{1, o}$.

Denoting $\alpha^{(r)} = (\alpha_i^{(r)}), i = \overline{1, n}; v = (v(t_j)) = (v_j), j = \overline{1, o}$, we present the Expression (17) in the form

$$s = \sum_{r=1}^R X^{(r)} \alpha^{(r)} + v,$$

where the independent elements of the vector of the variability of measurements of the entity “output” v take values in intervals $\Upsilon_j = [v_j^-, v_j^+]$ with the probability density functions $Q(v) = (q_j(v_j)), j = \overline{1, o}$.

Let us identify and investigate the variable entropy estimate of the probability density functions $P(\alpha) = (p_i(\alpha_i)), i = \overline{1, n}$, and $Q(v) = (q_j(v_j)), j = \overline{1, o}$.

We present the objective function of the optimization Problem (12)–(14) in the form

$$H(\bar{s}) = - \sum_{i=1}^n \int_{\alpha_i \in A_i} p_i(\alpha_i) \ln p_i(\alpha_i) d\alpha_i - \sum_{j=1}^o \int_{\alpha_j \in A_j} q_j(v_j) \ln q_j(v_j) dv_j \rightarrow \max.$$

We present the system of Restrictions (13) and (14) in the form

$$\begin{aligned} \bar{P}_i(p_i(\alpha_i)) &= 1 - \int_{\alpha_i \in A_i} p_i(\alpha_i) d\alpha_i = 0, \\ \bar{Q}_j(q_j(v_j)) &= 1 - \int_{v_j \in \Upsilon_j} q_j(v_j) dv_j = 0, \\ \Phi_j(P(\alpha), Q(v)) &= -y_j + \sum_{r=1}^R \sum_{i=1}^n x_{ji}^{(r)} \int_{\alpha_i \in A_i} \alpha_i^r p_i(\alpha_i) d\alpha_i + \int_{v \in \Upsilon_j} v_j q_j(v_j) dv_j, \end{aligned}$$

where $i = \overline{1, n}, j = \overline{1, o}$.

Based on the necessary conditions of stationarity of the Lagrange functional [6–8], we will assert that the entropy estimates $E_s^{(1)}$ of the probability density functions $P(\alpha)$ and $Q(v)$ belong to continuously differentiable functions, respectively:

$$p_i^U \sim a_i \exp\left(-\sum_{r=1}^R b_{ir} \alpha_i^r\right), \quad (17)$$

$$q_j^U(v_j) \sim c_j \exp(-d_j v_j), \quad (18)$$

where a_i, b_i, c_j, d_j are fixed coefficients, $i = \overline{1, n}, j = \overline{1, o}$.

The conclusion generalized by Expressions (17) and (18) can be interpreted as follows:

(1) For a linear stochastic model of estimation of variable small data: entropy estimates $E_s^{(1)}$ are always exponential functions. The results of measuring the entities “input” and “output” of the investigated process determine the form, and not the type, of the $E_s^{(1)}$ -functions of the corresponding linear stochastic model;

(2) For a non-linear stochastic model for evaluating variable small data: the nomenclature of the types of functions of entropy estimates $E_s^{(1)}$ of the “input” and “output” entities

of the investigated process is wider and includes both exponential and power types. The type of $E_{\bar{s}}^{(1)}$ -functions depends on the organization of the measurement process of these “input” and “output” entities.

Therefore, it remains to formalize the variable entropy estimates $E_{\bar{s}}^{(1)}(E_{\bar{s}}^{(1)})$ of the probability density functions p and q of the parameters of $GvSM(QvSM)$, respectively. Let us investigate the linear $GvSM$ without taking into account the variability of the measurement of the “input” entity:

$$\bar{s} = XpL_{\alpha} + qL_v + \Xi(\alpha^-, v^-), \quad (19)$$

where $\Xi(\alpha^-, v^-) = X\alpha^- + v^-$. We define the a priori probabilities by the elements of the tuple $\langle p^0, q^0 \rangle$.

Let us present the objective function of the optimization Problem (15) and (16) in the form

$$H(\bar{s}) = -\sum_{i=1}^n p_i \ln \frac{p_i}{p_i^0} - \sum_{j=1}^o q_j \ln \frac{q_j}{q_j^0} \rightarrow \max, \quad (20)$$

and the system of restrictions we present in the form

$$\sum_{i=1}^n x_{ji} p_i L_{\alpha}^i + q_j L_v^j + \Xi_j, \quad \forall j = \overline{1, o}, \quad (21)$$

at $\sum_{i=1}^n p_i = 1, \sum_{j=1}^o q_j = 1$.

In terms of the Lagrange function, we present the solution of the mathematical programming Problem (20) and (21) as

$$L(\bar{s}) = H(\bar{s}) + \beta \left(1 - \sum_{i=1}^n p_i \right) + \mu \left(1 - \sum_{j=1}^o q_j \right) + \sum_{j=1}^o \psi_j \left(y_j - \sum_{i=1}^n x_{ji} p_i L_{\alpha}^i - q_j L_v^j - \Xi_j \right), \quad (22)$$

where β, μ are fixed coefficients and $\psi = (\psi_1, \dots, \psi_o)$ is a set of Lagrange multipliers.

Entropy estimates $E_{\bar{s}}^{(1)} = \left[(p_i^{\cup}(\psi), i = \overline{1, n}), (q_j^{\cup}(\psi), j = \overline{1, o}) \right]$ are determined based on Expression (22):

$$\begin{aligned} 0 \leq p_i^{\cup}(\psi) &= \frac{p_i^0 \exp \left(- \sum_{j=1}^o x_{ji} \psi_j L_{\alpha}^i \right)}{\sum_{i=1}^n p_i^0 \exp \left(- \sum_{j=1}^o x_{ji} \psi_j L_{\alpha}^i \right)} \leq 1, \\ 0 \leq q_j^{\cup}(\psi) &= \frac{q_j^0 \exp \left(- \psi_j L_v^j \right)}{\sum_{j=1}^o q_j^0 \exp \left(- \psi_j L_v^j \right)} \leq 1, \\ \Phi_j(\psi) &= \frac{1}{y_j - \Xi_j} \sum_{i=1}^n x_{ji} p_i^{\cup}(\psi) L_{\alpha}^i + q_j^{\cup}(\psi) L_v^j = 1. \end{aligned} \quad (23)$$

Now let us investigate how the formulation and solution of the optimization Problem (20) and (21) will change if interval restrictions $0 \leq p_i \leq 1, \forall i \in \overline{1, n}, 0 \leq q_j \leq 1, \forall j \in \overline{1, o}$, are respectively imposed on the values of the elements of the stochastic vectors $\langle p, q \rangle$.

Under such conditions, the variable entropy estimate $E_{\bar{s}}^{(1)}$ of the probability density functions of the parameters p and q of $QvSM$ can be obtained by solving the problem of finding the extreme generalized entropy of the form

$$H(\tilde{s}) = -\sum_{i=1}^n \left(p_i \ln \frac{p_i}{\hat{p}_i^0} + (1-p_i) \ln(1-p_i) \right) - \sum_{j=1}^o \left(q_j \ln \frac{q_j}{\hat{q}_j^0} + (1-q_j) \ln(1-q_j) \right) \rightarrow \max, \quad (24)$$

where $\hat{p}_i^0 = p_i^0 / (1-p_i^0)$, $\hat{q}_j^0 = q_j^0 / (1-q_j^0)$, $i = \overline{1, n}$, $j = \overline{1, o}$.

The objective Function (24) is supplemented by the adapted balance Equation (21):

$$\sum_{i=1}^n x_{ji} p_i L_\alpha^i + q_j L_v^j + \Xi_j = y_j, \quad (25)$$

where $0 \leq p_i \leq 1$, $0 \leq q_j \leq 1$, $i = \overline{1, n}$, $j = \overline{1, o}$.

Applying the method of Lagrange multipliers [7,8,36], the extreme entropy estimates $E_s^{(1)}$ for the optimization Problem (24) and (25) will be obtained as a result of solving the system of equations

$$\begin{aligned} 0 \leq p_i^\cup(\psi) &= p_i^0 / \left(p_i^0 + (1-p_i^0) \exp \sum_{j=1}^o x_{ji} \psi_j L_\alpha^i \right) \leq 1, \\ 0 \leq q_j^\cup(\psi) &= q_j^0 / \left(q_j^0 + (1-q_j^0) \exp(-\psi_j L_v^j) \right) \leq 1, \\ \Phi_j(\psi) &= \frac{1}{y_j - \Xi_j} \sum_{i=1}^n x_{ji} p_i^\cup(\psi) \psi_j + q_j^\cup(\psi) L_v^j = 1, \end{aligned} \quad (26)$$

where $i = \overline{1, n}$, $j = \overline{1, o}$.

The starting point for calculating the variable entropy estimate $E_s^{(1)}$ of the probability density functions of the parameters p and q of $QvSM$, both in the Interpretation (20) and (21), and in the Interpretation (24) and (25), is the calculation of the Lagrange multipliers ψ as a result of solving the systems of equations represented by Expressions (23) and (26), respectively. This process can be arranged, for example, according to the multiplicative algorithm [36]:

$$\varphi_j^{k+1} = \varphi_j^k \Phi_j(\varphi^k),$$

where $\varphi_i = \exp(-\psi_i)$ are exponential Lagrange multipliers, $\varphi_j^0 > 0$, $j = \overline{1, o}$.

3. Experiments

Let us demonstrate the functionality of the mathematical apparatus proposed in Section 2 using the example of calculating the variable entropy estimate of the probability density functions of the characteristic parameters of the linear stochastic small data estimation model with the dimension of the entities “input” $\times$ “output” of $[5] \times [2]$. The matrix of the measurements of the “input” entity looks like this:

$$X = \begin{pmatrix} 1.805 & 2.103 & 3.310 & 2.007 & 1.505 \\ 4.992 & 3.800 & 2.996 & 2.812 & 1.899 \end{pmatrix}.$$

The vector of the measurements of the “output” entity, taking into account variability, looks like this:

$$y = (21.091 \quad 32.814).$$

Quasi-moments of the first order are described by the expressions:

$$\begin{aligned} a_i &= 3.333 p_i, \quad \alpha_i \in A_i, \quad \forall A_i \in A = [0, 10], \quad i = \overline{1, 5}; \\ v_1 &= -1 + 2q_1, \quad v_2 = -2 + 4q_2, \quad v_1 \in Y_1 = [-3, 3], \quad v_2 \in Y_2 = [-6, 6]. \end{aligned}$$

The fixed parameters of the reference model are described by the vector

$$\alpha^0 = (1.011 \quad 2.212 \quad 1.918 \quad 3.986 \quad 0.996).$$

The deviations from the values specified in the vector α^0 caused by the variability of the measurements are characterized by an error $\varepsilon = \|\alpha^0 - \alpha\| / (\|\alpha^0\| + \|\alpha\|)$.

Summarizing the given initial information in the format of Expression (19), we obtain:

$$XL_\alpha p + L_v q = \vec{1},$$

$$\text{where } L_v = \begin{pmatrix} 0.249 & 0 \\ 0 & 0.312 \end{pmatrix}, XL_\alpha = \begin{pmatrix} 0.747 & 0.873 & 1.366 & 0.834 & 0.622 \\ 1.065 & 0.982 & 0.767 & 0.721 & 0.449 \end{pmatrix}, \vec{1} = (1 \ 1).$$

A priori information about the initial values of the vectors $p^0 = (p_i^0)$, $i = \overline{1,5}$, and $q^0 = (q_j^0)$, $j = \overline{1,2}$, is summarized in the corresponding named sets: $p_A^0 = \{1; 1; 1; 1; 1\}$, $p_B^0 = \{0.1; 0.2; 0.3; 0.3; 0.1\}$, $p_C^0 = \{0.3; 0.4; 0.1; 0.05; 0.15\}$, $q_D^0 = \{0.2; 0.8\}$, $q_E^0 = \{1; 1\}$.

The tuple $\langle p_A^0, q_E^0 \rangle$ implies a uniform distribution of the characteristic parameters p and disturbing influences causing measurement variability, q , respectively. Tuples $\langle p_B^0, q_D^0 \rangle$ and $\langle p_C^0, q_E^0 \rangle$ imply uneven distributions of the characteristic parameters and influences, while the latter represents the variant combined according to the a priori probabilities of the corresponding entities.

We obtain optimization problem Statements (20) and (24) for the initial parameters presented above.

The formulation of the optimization Problem (20) and (21) for the above-mentioned initial data has the form:

$$\begin{aligned} H(\tilde{s}) &= - \sum_{i=1}^5 p_i \ln \frac{p_i}{p_i^0} - \sum_{j=1}^2 q_j \ln \frac{q_j}{q_j^0} \rightarrow \max, \\ 0.747p_1 + 0.873p_2 + 1.366p_3 + 0.834p_4 + 0.622p_5 + 0.249q_1 &= 1, \\ 1.065p_1 + 0.982p_2 + 0.767p_3 + 0.721p_4 + 0.449p_5 + 0.312q_1 &= 1; \\ \sum_{i=1}^5 p_i &= 1, p_i > 0; \sum_{j=1}^2 q_j = 1, q_j > 0. \end{aligned} \quad (27)$$

The formulation of the optimization Problem (24) and (25) for the above-mentioned initial data has the form:

$$\begin{aligned} H(\tilde{s}) &= - \sum_{i=1}^5 \left(p_i \ln \frac{p_i}{p_i^0} + (1 - p_i) \ln(1 - p_i) \right) - \sum_{j=1}^2 \left(q_j \ln \frac{q_j}{q_j^0} + (1 - q_j) \ln(1 - q_j) \right) \rightarrow \max, \\ 0.747p_1 + 0.873p_2 + 1.366p_3 + 0.834p_4 + 0.622p_5 + 0.249q_1 &= 1, \\ 1.065p_1 + 0.982p_2 + 0.67p_3 + 0.721p_4 + 0.449p_5 + 0.312q_1 &= 1; \\ \hat{p}_i^0 &= p_i^0 / (1 - p_i^0), \hat{q}_j^0 = q_j^0 / (1 - q_j^0), i = \overline{1,5}, j = \overline{1,2}; \\ 0 \leq p_i \leq 1, 0 \leq q_j \leq 1, i &= \overline{1,5}, j = \overline{1,2}. \end{aligned} \quad (28)$$

Such optimization problems can be solved by methods of non-linear mathematical programming [36]. In particular, for the above optimization problems, the extremum point is analytically identified as $(p_i^* = 0.36p_i^0, q_j^* = 0.36q_j^0)$, $i = \overline{1,5}, j = \overline{1,2}$. So, for our example, the entropy $H(\tilde{s})$ reaches its maximum at the point (p^*, q^*) , where $p^* = f(i, p_i^0)$, $q^* = f(j, q_j^0)$, $i = \overline{1,5}, j = \overline{1,2}$.

Let us examine these dependencies, taking into account that we previously defined schemes for a priori values: $p^0 = \{p_A^0, p_B^0, p_C^0\}$, $q^0 = \{q_D^0, q_E^0\}$. For clarity, we present the dependences $p^* = f(i, p_{\{A,B,C\}}^0)$ and $q^* = f(j, q_{\{D,E\}}^0)$ in the form of diagrams (Figures 1 and 2, respectively).

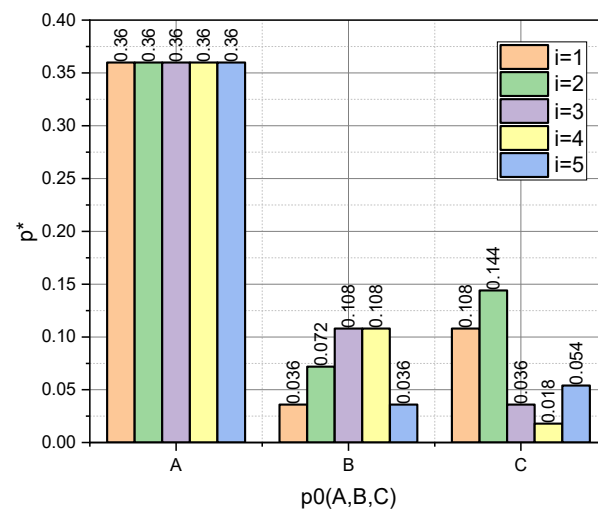


Figure 1. Visualization of dependence $p^* = f(i, p_{\{A,B,C\}}^0)$.

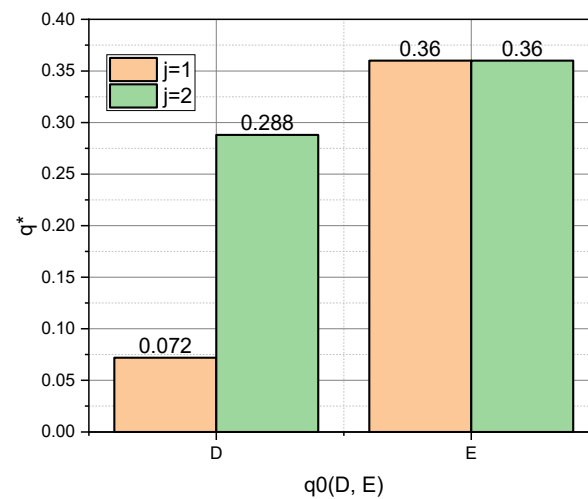


Figure 2. Visualization of dependence $q^* = f(j, q_{\{D,E\}}^0)$.

More detailed information on the values of the characteristic parameters of the investigated linear stochastic model of the small data evaluation presented in Section 3 can be seen in Figures 3 and 4 (for $G\nu SM$ and for $Q\nu SM$, respectively).

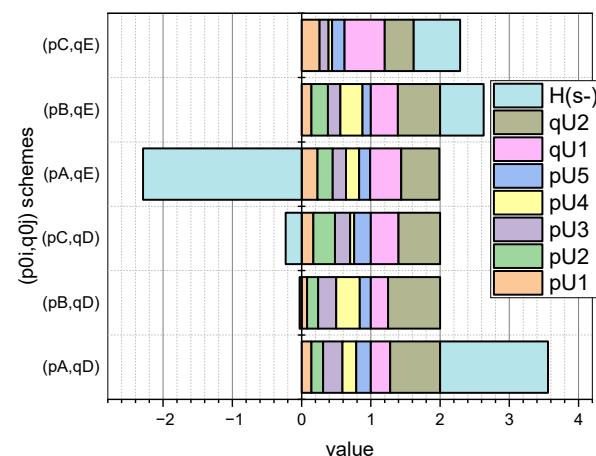


Figure 3. Visualization of dependence $E_s^{(1)}(p_i^U, q_j^U, H(\bar{s})) = f(p_{\{A,B,C\}}^0, q_{\{D,E\}}^0), i = \overline{1,5}, j = \overline{1,2}$.

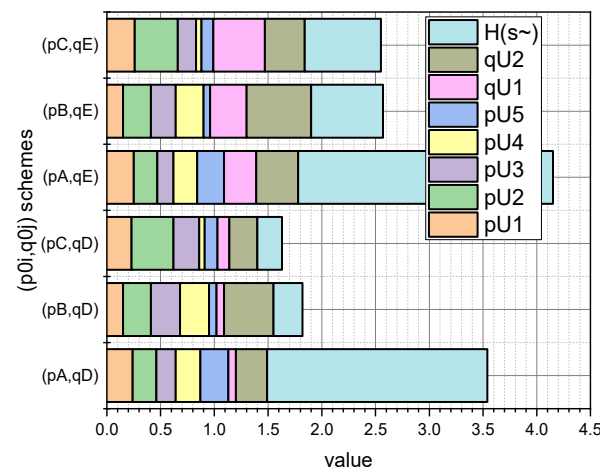


Figure 4. Visualization of dependence $E_s^{(1)}(p_i^U, q_j^U, H(\tilde{s})) = f(p_{\{A,B,C\}}^0, q_{\{D,E\}}^0)$, $i = \overline{1,5}$, $j = \overline{1,2}$.

These figures visualize the values at the extremum point (p^*, q^*) of $E^{(1)}$ -estimates of such characteristic parameters as p_i^U , $i = \overline{1,5}$; q_j^U , $j = \overline{1,2}$, and $H^*(\tilde{s})$ (calculated by Expression (20) adapted to form (27)) and $H^*(\tilde{s})$ (calculated by Expression (24) adapted to form (28)). At the same time, the schemes of the initial values of the vectors $p^0 = (p_i^0)$, $i = \overline{1,5}$, and $q^0 = (q_j^0)$, $j = \overline{1,2}$, are taken into account.

Comparing the symmetrical values visualized in Figures 3 and 4, it can be concluded that the parameter estimates calculated for interval probabilities (i.e., for $QvSM$) are characterized by a larger value of the conditional maximum entropy than that inherited for $GvSM$ (i.e., for the normalized probabilities). The theoretical justification of this empirical fact is presented in Section 4.

Information about the state of the linear stochastic models, summarized by Expressions (27) and (28), is supplemented by such calculated data as:

- (1) the value at the point of extremum (p^*, q^*) of the quasi-moments of the characteristic parameters of $GvSM$ and $QvSM$ (α_i^* , $i = \overline{1,5}$),
- (2) estimates of the variability of the above-mentioned parameters caused by interferences (v_j^* , $j = \overline{1,2}$),
- (3) the errors $\bar{\varepsilon}$ and $\tilde{\varepsilon}$, which characterize the deviation of the measured parameters $\langle \alpha, v \rangle$ from the reference $\langle \alpha^0, v^0 \rangle$ for $GvSM$ and $QvSM$, respectively.

These data are visualized in Figures 5 and 6.

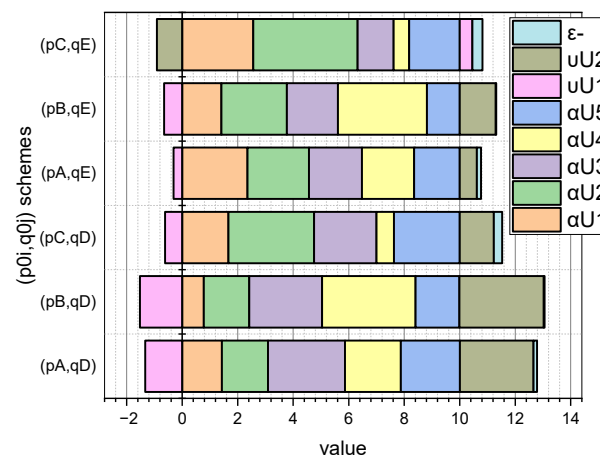


Figure 5. Visualization of dependence $(\alpha_i^*, v_j^*, \bar{\varepsilon}) = f(p_{\{A,B,C\}}^0, q_{\{D,E\}}^0)$, $i = \overline{1,5}$, $j = \overline{1,2}$.

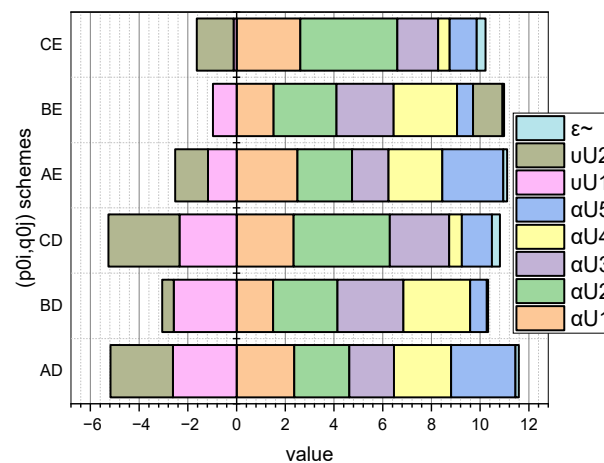


Figure 6. Visualization of dependence $\left(\alpha_i^*, v_j^*, \tilde{\varepsilon}\right) = f\left(p_{\{A,B,C\}}^0, q_{\{D,E\}}^0\right), i = \overline{1,5}, j = \overline{1,2}$.

From the information shown in Figures 5 and 6 (in addition to the information presented in Figures 3 and 4), it can be concluded that the reference parameters and a priori probabilities are correlated. That is, the closer the values in the scheme of a priori probabilities are to the values of the reference parameters, the smaller the value of the error ε . This interpretation, in particular, explains the superiority of the scheme (p_B, q_D) over the scheme (p_C, q_D) , because $\tilde{\varepsilon}_{BD} < \tilde{\varepsilon}_{CD}$.

4. Discussion

Let us begin the analysis of the results presented in Section 3 of the applied use of the mathematical apparatus proposed in Section 2 with the fact that the estimates of the parameters $\langle p, q \rangle$ obtained as a result of solving optimization Problems (27) (derived from Problem (20), (21) and (28)) (derived from Problem (24) and (25)), turn out to be different in terms of the value of the generalized entropy (Expressions (20) and (24), respectively). We will explain this fact on the theoretical basis of the models presented in Section 2.

To simplify the formulations, we will introduce several renovations. Let us redefine entropy $H(s)$ as $H(e) = H(s)$, where $e = \langle p, q \rangle$. Accordingly, e_1^* will be the optimal estimate of the parameters $\langle p^*, q^* \rangle$ represented by normalized probabilities ($H(\bar{s})$ variant) and e_2^* will be the optimal estimate of the parameters $\langle p^*, q^* \rangle$ represented by interval probabilities ($H(\tilde{s})$ variant). Let us denote $\hat{e} = \operatorname{argmax} H(e)$ and define the sets

$$\bar{E} = \{e : \langle p, 1 \rangle, \langle q, 1 \rangle\} \subset \tilde{E} = \{e : 0 \leq e \leq 1\}. \quad (29)$$

Summarizing what has been entered, we formulate the following: if $\hat{e} \in \left(R_+^{(n+o)} \setminus \bar{E}\right)$ then $H(e_1^*) < H(e_2^*)$. The equality $H(e_1^*) = H(e_2^*)$ holds when $e_1^* = \hat{e}$. Let us explain our conclusions. The analysis of the function described by Expression (20) shows that it is a concave function with a single maximum at the point $\hat{e}$. The value of entropy $H(e)$ depends on the distance of a point e from the extreme point $\hat{e}$. In this context, we denote as $\Delta(\hat{e}, e_1^*)$ the distance between the extreme point $\hat{e}$ and the point e_1^* , the coordinates of which we obtain as a result of solving optimization Problem (20) and (21). Accordingly, the parameter $\Delta(\hat{e}, e_2^*)$ characterizes the distance between the extreme point $\hat{e}$ and the point e_2^* , the coordinates of which we obtain as a result of solving optimization Problem (24) and (25). Since Function (20) is strictly concave, based on the Relation (29) we can conclude that $\Delta(\hat{e}, e_1^*) < \Delta(\hat{e}, e_2^*)$. The equality $\Delta(\hat{e}, e_1^*) = \Delta(\hat{e}, e_2^*)$ holds only when $e_1^* = \hat{e}$. The presented theoretical explanations explain the discrepancy between those presented in Figures 3 and 4 empirical values of $E_s^{(1)}(H(\bar{s}))$ and $E_s^{(1)}(H(\tilde{s}))$ for the same schemes $\left(p_{\{A,B,C\}}^0, q_{\{D,E\}}^0\right)$. Comparing the symmetrical values visualized in Figures 3 and 4, it can be concluded that parameter estimates calculated for the interval probabilities (i.e., for

$QvSM$) are characterized by a larger value of the conditional maximum entropy estimate than that characteristic of the normalized probabilities of $GvSM$. Thus, the mathematical apparatus presented in Section 2 was empirically confirmed in Section 3.

In addition, the results of the experiments presented in Section 3 confirmed the conclusion generalized by Expressions (17) and (18) that, for a linear stochastic model of variable small data estimation, entropy estimates $E_{\bar{s}}^{(1)}$ are always exponential functions. The results of measuring the “input” and “output” entities of the investigated process determine the form, and not the type, of the $E_{\bar{s}}^{(1)}$ -functions of the corresponding linear stochastic model of small data estimation.

The results shown in Figures 3 and 4 showed that a priori information about the initial values of the vectors $p^0 = (p_i^0)$, $i = \overline{1,5}$, and $q^0 = (q_j^0)$, $j = \overline{1,2}$, summarized in the corresponding named sets of $p_{\{A,B,C\}}^0$, $q_{\{D,E\}}^0$, has a significant effect on the $E_{\{\bar{s},\bar{s}\}}^{(1)}(p_i^0, q_j^0, H(\{\bar{s}, \bar{s}\}))$ estimates.

In this context, the fact that the author’s mathematical apparatus allows the calculation of the quasi-momentums of the characteristic parameters α_i^* , $i = \overline{1,5}$, of both the $GvSM$ and $QvSM$, as well as the taking into account of their variability v_j^* , $j = \overline{1,2}$, caused by the measurement errors, is very relevant. From those visualized in Figures 5 and 6 of the data, it can be seen that the ε deviations from the values indicated in the vector α^0 caused by the variability of the measurements are most pronounced for the schemes (p_C^0, q_D^0) and (p_C^0, q_E^0) . These schemes are characterized by the fact that the essential parameters of the models are characterized by an uneven distribution (see Figure 1, “C”), and the influence parameters are characterized by both uneven (see Figure 2, “D”) and uniform distributions (see Figure 2, “E”). For both schemes, we obtained: $\bar{\varepsilon}_{CD} = 0.30$, $\bar{\varepsilon}_{CE} = 0.36$; $\tilde{\varepsilon}_{CD} = 0.33$, $\tilde{\varepsilon}_{CE} = 0.36$. Therefore, for the considered example, the unevenness of the distribution of parameters p_i , $i = \overline{1,5}$ provided a significant contribution to the high value of errors ε . Reliable a priori information turned out to be very important in the entropy estimation of variable small data.

5. Conclusions

The article analytically summarizes the idea of applying the Shannon entropy maximization principle to sets that represent the results of observations of the “input” and “output” entities of the stochastic model for evaluating variable small data. To formalize this idea, a sequential transition from the likelihood function to the likelihood functional and the Shannon entropy functional is analytically described. Shannon’s entropy characterizes the uncertainty caused not only by the probabilistic nature of the parameters of the stochastic data evaluation model but also by influences that distort the results of measurements of the values of these parameters. Accordingly, based on the Shannon entropy, it is possible to determine the best estimates of the values of these parameters for maximally uncertain (per entropy unit) influences that cause measurement variability. This postulate is organically transferred to the statement that the estimates of the probability distribution density of the parameters of the stochastic model of small data obtained as a result of Shannon entropy maximization will also take into account the fact of the variability of the process of their measurements. In the article, this principle is developed into the information technology of the parametric and non-parametric evaluation on the basis of Shannon entropy of small data measured under the influence of interferences.

The article also examines the structural properties of stochastic models for variable data evaluation, the parameters of which were represented by normalized or interval probabilities. At the same time, the inherent non-linearity of these models and errors in measuring the values of the “output” entity was taken into account.

The functionality and adequacy of the created mathematical apparatus are proven based on the empirical results obtained during the investigation of the linear stochastic model of evaluating specific variable small data.

The authors acknowledge that the research presented in the article is formulated in academic form. This circumstance complicates the applied use of the obtained results. At the same time, the developed methodological approach can be useful in various important applications. In particular, it concerns the assessment of software reliability, when the sample of data is usually not large due to the difficulties of reliably assessing them during the testing and operation of the system. In this case, the lack of testing data or information about failures during pilot software operation can be compensated for by analyzing the assumptions that are specific to the software and selecting appropriate models using assumption matrices [37]. Thus, studies that combine the analysis of small data and expert methods are interesting.

Another important application is in safety critical systems, which, due to multi-level reserving, have as a rule a low failure rate and small data about them. On the other hand, it is extremely important for such systems to have accurate or at least interval estimates of indicators with an acceptable range. For that, the described method could be combined with the traditional methods of reliability analysis and risk oriented assessing of safety indicators using formal and semi-formal methods [38].

In this regard, **further research** is proposed to formalize the obtained information technology on a UML basis. This will allow the future work to reach the stage of implementing the profile framework. In addition, it would be very interesting and useful from a practical point of view to combine Big and Small Data analysis to create universal or adaptable framework focusing on the assessment of data quality and their selection according to the quality indicator.

Author Contributions: Conceptualization, O.B., V.K. (Vyacheslav Kharchenko), V.K. (Viacheslav Kovtun), I.K. and S.P.; methodology, V.K. (Viacheslav Kovtun); software, V.K. (Viacheslav Kovtun); validation, V.K. (Vyacheslav Kharchenko) and V.K. (Viacheslav Kovtun); formal analysis, V.K. (Viacheslav Kovtun); investigation, V.K. (Viacheslav Kovtun); resources, V.K. (Viacheslav Kovtun); data curation, V.K. (Viacheslav Kovtun); writing—original draft preparation, V.K. (Vyacheslav Kharchenko) and V.K. (Viacheslav Kovtun); writing—review and editing, O.B., V.K. (Vyacheslav Kharchenko), V.K. (Viacheslav Kovtun), I.K. and S.P.; visualization, V.K. (Viacheslav Kovtun); supervision, V.K. (Viacheslav Kovtun); project administration, V.K. (Viacheslav Kovtun); funding acquisition, V.K. (Viacheslav Kovtun). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Most data is contained within the article. All the data available on request due to restrictions, e.g., privacy or ethical.

Acknowledgments: The authors would like to thank the Armed Forces of Ukraine for providing security to perform this work. This work has become possible only because of the resilience and courage of the Ukrainian Army.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ebrahimi, B.; Dellnitz, A.; Kleine, A.; Tavana, M. A novel method for solving data envelopment analysis problems with weak ordinal data using robust measures. *Expert Syst. Appl.* **2021**, *164*, 113835. [\[CrossRef\]](#)
2. Kovtun, V.; Kovtun, O.; Semenov, A. Entropy-Argumentative Concept of Computational Phonetic Analysis of Speech Taking into Account Dialect and Individuality of Phonation. *Entropy* **2022**, *24*, 1006. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Viacheslav, K.; Kovtun, O. System of methods of automated cognitive linguistic analysis of speech signals with noise. *Multimedia Tools Appl.* **2022**, *81*, 43391–43410. [\[CrossRef\]](#)
4. Angeles, K.; Kijewski-Correa, T. Advancing building data models for the automation of high-fidelity regional loss estimations using open data. *Autom. Constr.* **2022**, *140*, 104382. [\[CrossRef\]](#)
5. Hao, R.; Zheng, H.; Yang, X. Data augmentation based estimation for the censored composite quantile regression neural network model. *Appl. Soft Comput.* **2022**, *127*, 109381. [\[CrossRef\]](#)

6. Garza-Ulloa, J. Methods to develop mathematical models: Traditional statistical analysis. In *Applied Biomechanics Using Mathematical Models*; Elsevier: Amsterdam, The Netherlands, 2018; pp. 239–371.
7. Gao, Y.; Shi, Y.; Wang, L.; Kong, S.; Du, J.; Lin, G.; Feng, Y. Advances in mathematical models of the active targeting of tumor cells by functional nanoparticles. *Comput. Methods Programs Biomed.* **2020**, *184*, 105106. [\[CrossRef\]](#)
8. Yang, X.-S.; He, X.-S.; Fan, Q.-W. Mathematical framework for algorithm analysis. In *Nature-Inspired Computation and Swarm Intelligence*; Academic Press: Cambridge, MA, USA, 2020; pp. 89–108. [\[CrossRef\]](#)
9. Wang, X.; Liu, A.; Kara, S. Machine learning for engineering design toward smart customization: A systematic review. *J. Manuf. Syst.* **2022**, *65*, 391–405. [\[CrossRef\]](#)
10. Khan, T.; Tian, W.; Zhou, G.; Ilager, S.; Gong, M.; Buyya, R. Machine learning (ML)-centric resource management in cloud computing: A review and future directions. *J. Netw. Comput. Appl.* **2022**, *204*, 103405. [\[CrossRef\]](#)
11. Gonçalves, L.J.; Farias, K.; Kupssinskü, L.S.; Segalotto, M. An empirical evaluation of machine learning techniques to classify code comprehension based on EEG data. *Expert Syst. Appl.* **2022**, *203*, 117354. [\[CrossRef\]](#)
12. Sholevar, N.; Golroo, A.; Esfahani, S.R. Machine learning techniques for pavement condition evaluation. *Autom. Constr.* **2022**, *136*, 104190. [\[CrossRef\]](#)
13. Alam, J.; Georgalos, K.; Rolls, H. Risk preferences, gender effects and Bayesian econometrics. *J. Econ. Behav. Organ.* **2022**, *202*, 168–183. [\[CrossRef\]](#)
14. Cladera, M. Assessing the attitudes of economics students towards econometrics. *Int. Rev. Econ. Educ.* **2021**, *37*, 100216. [\[CrossRef\]](#)
15. Joubert, J.W. Accounting for population density in econometric accessibility. *Procedia Comput. Sci.* **2022**, *201*, 594–600. [\[CrossRef\]](#)
16. MacKinnon, J.G. Using large samples in econometrics. *J. Econ.* **2022**. [\[CrossRef\]](#)
17. Nazarkevych, M.; Voznyi, Y.; Hrytsyk, V.; Klyujnyk, I.; Havrysh, B.; Lotoshynska, N. Identification of Biometric Images by Machine Learning. In Proceedings of the 2021 IEEE 12th International Conference on Electronics and Information Technologies (ELIT), Lviv, Ukraine, 19–21 May 2021. [\[CrossRef\]](#)
18. Yusuf, A.; Qureshi, S.; Shah, S.F. Mathematical analysis for an autonomous financial dynamical system via classical and modern fractional operators. *Chaos Solitons Fractals* **2020**, *132*, 109552. [\[CrossRef\]](#)
19. Balbás, A.; Balbás, B.; Balbás, R. Omega ratio optimization with actuarial and financial applications. *Eur. J. Oper. Res.* **2021**, *292*, 376–387. [\[CrossRef\]](#)
20. Giua, A.; Silva, M. Petri nets and Automatic Control: A historical perspective. *Annu. Rev. Control* **2018**, *45*, 223–239. [\[CrossRef\]](#)
21. Sleptsov, E.S.; Andrianova, O.G. Control Theory Concepts: Analysis and Design, Control and Command, Control Subject, Model Reduction. *IFAC-PapersOnLine* **2021**, *54*, 204–208. [\[CrossRef\]](#)
22. Knorn, S.; Varagnolo, D. Automatic control: The natural approach for a quantitative-based personalized education. *IFAC-PapersOnLine* **2020**, *53*, 17326–17331. [\[CrossRef\]](#)
23. Rubio-Fernández, P.; Mollica, F.; Ali, M.O.; Gibson, E. How do you know that? Automatic belief inferences in passing conversation. *Cognition* **2019**, *193*, 104011. [\[CrossRef\]](#)
24. Aljohani, M.D.; Qureshi, R. Proposed Risk Management Model to Handle Changing Requirements. *Int. J. Educ. Manag. Eng.* **2019**, *9*, 18–25. [\[CrossRef\]](#)
25. Arefin, M.A.; Islam, N.; Gain, B. Roknujjaman Accuracy Analysis for the Solution of Initial Value Problem of ODEs Using Modified Euler Method. *Int. J. Math. Sci. Comput.* **2021**, *7*, 31–41. [\[CrossRef\]](#)
26. Ramadan, I.S.; Harb, H.M.; Mousa, H.M.; Malhat, M.G. Reliability Assessment for Open-Source Software Using Deterministic and Probabilistic Models. *Int. J. Inf. Technol. Comput. Sci.* **2022**, *14*, 1–15. [\[CrossRef\]](#)
27. Nayim, A.M.; Alam, F.; Rasel; Shahriar, R.; Nandi, D. Comparative Analysis of Data Mining Techniques to Predict Cardiovascular Disease. *Int. J. Inf. Technol. Comput. Sci.* **2022**, *14*, 23–32. [\[CrossRef\]](#)
28. Goncharenko, A.V. Specific Case of Two Dynamical Options in Application to the Security Issues: Theoretical Development. *Int. J. Comput. Netw. Inf. Secur.* **2021**, *14*, 1–12. [\[CrossRef\]](#)
29. Padmavathi, C.; Veenadevi, S.V. An Automated Detection of CAD Using the Method of Signal Decomposition and Non Linear Entropy Using Heart Signals. *Int. J. Image Graph. Signal Process.* **2019**, *11*, 30–39. [\[CrossRef\]](#)
30. Mwambela, A. Comparative Performance Evaluation of Entropic Thresholding Algorithms Based on Shannon, Renyi and Tsallis Entropy Definitions for Electrical Capacitance Tomography Measurement Systems. *Int. J. Intell. Syst. Appl.* **2018**, *10*, 41–49. [\[CrossRef\]](#)
31. Dronyuk, I.; Fedevych, O.; Poplavska, Z. The generalized shift operator and non-harmonic signal analysis. In Proceedings of the 2017 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), Lviv, Ukraine, 21–25 February 2017. [\[CrossRef\]](#)
32. Hu, Z.; Tereykovskiy, I.A.; Tereykovska, L.O.; Pogorelov, V.V. Determination of Structural Parameters of Multilayer Perceptron Designed to Estimate Parameters of Technical Systems. *Int. J. Intell. Syst. Appl.* **2017**, *9*, 57–62. [\[CrossRef\]](#)
33. Izonin, I.; Tkachenko, R.; Shakhovska, N.; Lotoshynska, N. The Additive Input-Doubling Method Based on the SVR with Nonlinear Kernels: Small Data Approach. *Symmetry* **2021**, *13*, 612. [\[CrossRef\]](#)
34. Hu, Z.; Mashtalir, S.V.; Tyshchenko, O.K.; Stolbovyi, M.I. Clustering Matrix Sequences Based on the Iterative Dynamic Time Deformation Procedure. *Int. J. Intell. Syst. Appl.* **2018**, *10*, 66–73. [\[CrossRef\]](#)
35. Hu, Z.; Khokhlovskaya, Y.; Sydorenko, V.; Opirskyy, I. Method for Optimization of Information Security Systems Behavior under Conditions of Influences. *Int. J. Intell. Syst. Appl.* **2017**, *9*, 46–58. [\[CrossRef\]](#)

36. Pineda, S.; Morales, J.M.; Wogrin, S. Mathematical programming for power systems. In *Encyclopedia of Electrical and Electronic Power Engineering*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 722–733. [[CrossRef](#)]
37. Kharchenko, V.S.; Tarasyuk, O.M.; Sklyar, V.V.; Dubnitsky, V.Y. The method of software reliability growth models choice using assumptions matrix. In Proceedings of the 26th Annual International Computer Software and Applications Conference (COMPSAC), Oxford, UK, 26–29 August 2002; pp. 541–546.
38. Babeshko, E.; Kharchenko, V.; Leontiev, K.; Ruchkov, E. Practical Aspects of Operating and Analytical Reliability Assessment of Fpga-Based I&C Systems. *Radioelectron. Comput. Syst.* **2020**, *3*, 75–83. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.