Entropy **2014**, *16*, 675-698; doi:10.3390/e16020675

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

New Methods of Entropy-Robust Estimation for Randomized Models under Limited Data

Yuri Popkov [1,2,3,*] and Alexey Popkov [1]

[1] Institute for Systems Analysis of Russian Academy of Sciences, 9 prospect 60-let Octyabrya, Moscow 117312, Russia; E-Mail: apopkov@isa.ru

[2] Moscow Institute of Physics and Technology, 9 Institutskiy pereulok, g. Dolgoprudny, Moskovskaya oblast 141700, Russia

[3] Higher School of Economics, 20 Myasnitskaya, Moscow 101000, Russia

* Author to whom correspondence should be addressed; E-Mail: popkov@isa.ru.

Received: 17 October 2013; in revised form: 17 December 2013 / Accepted: 14 January 2014 / Published: 23 January 2014

Abstract: The paper presents a new approach to restoration characteristics randomized models under small amounts of input and output data. This approach proceeds from involving randomized static and dynamic models and estimating the probabilistic characteristics of their parameters. We consider static and dynamic models described by Volterra polynomials. The procedures of robust parametric and non-parametric estimation are constructed by exploiting the entropy concept based on the generalized informational Boltzmann's and Fermi's entropies.

Keywords: randomized data models; robustness; entropy function and entropy functional; entropy functional variation; likelihood function and likelihood functional; Volterra polynomials; multiplicative algorithms; symbolic computing

1. Introduction

The problem of useful information retrieval (we comprehend it as parametric and nonparametric estimation based on real data) is a major one in modern science. Different scientific disciplines suggest numerous methods of solving this problem. Each method stems from certain hypotheses regarding the properties of data accumulated during the normal functioning of their source. Among advanced scientific disciplines in this field, we mention mathematical statistics [1–4], econometrics [5–7], financial mathematics [8,9], control theory [10,11] and others.

Methods developed within their frameworks rest upon two groups of fundamental hypotheses. The first one relates to *models*, whereas the other concerns *data*. Notably, models are supposed to have well-defined parameters (we call them deterministic). Parameter values appear unknown and unmeasurable directly.

The second group of hypotheses applies to data and plays an essential role. In fact, these hypotheses are stated in terms of the statistical properties of data arrays (e.g., a sufficient number of data arrays, a property of a sample from a universal set, normal probability density). In practice, it seems impossible to check such properties and assumptions in concrete problems.

The described situation happens in a class of problems, where the sizes of real data arrays are limited and data incorporate errors [5,6,12].

Consequently, the characteristics (parameters) of a model are estimated by a small amount of incompletely reliable data. We can treat them as random objects. In this case, the estimated characteristics of a model acquire the properties of random variables.

Therefore, one naturally arrives at the idea of considering model parameters as random quantities. This idea transforms the model with deterministic parameters to the model with random parameters. In the sequel, we adopt the term of a randomized model (RM). The characteristics of an RM include the probability density functions (pdfs) of the model parameters. Thus, one should estimate the pdfs of model parameters (not the estimates of their values) based on available data. Having such estimates at one's disposal, one can apply an RM for:

- constructing = moment models (MMs), where appropriate moments of random parameters serve as the model parameters;
- generating an ensemble of random vectors (ERV) of RM "output" with a pdf estimate (by the Monte Carlo method) and performing the statistical processing of the ensemble to form the desired numerical characteristics (including moment ones).

Both directions of RM usage enlarge appreciably the application domains of such models (especially, the ones with a high level of uncertainties). However, a researcher still faces a certain problem. How could the probability density functions of parameters be estimated in randomized models?

This paper proposes involving the informational entropy maximization principle on sets defined by the "input-output" measurements of an RM. The proposal originates from a couple of considerations discussed and formalized below.

The first consideration is connected with the generalized notion of likelihood, *viz.*, with transition from likelihood functions to likelihood functionals and, subsequently, to informational entropy functionals.

The second consideration is based on methodological interpretations of the notion of informational entropy as a measure of uncertainty. Entropy maximization guarantees the best solutions under the maximal uncertainty. This line of reasoning was pioneered in [13]. Informational entropy characterizes uncertainty caused by random parameters of an RM and by measurement noises. The last property of informational entropy ensures the best estimates for maximally uncertain noises (in units of entropy).

Hence, the pdf estimates resulting from informational entropy maximization can be viewed as robust. This interpretation varies from the classic conception of robustness suggested in [14].

2. Randomized Models

2.1. Static Objects

Consider a static parameterized object with a measurable "input" described by a matrix, X, of sizes $(s \times n)$ and a measurable "output" characterized by a vector, **y**, of length s. Here, s indicates the amount of input observations, and n stands for the number of the object's parameters.

The relationship between the "input" and "output" (including measurement errors) is defined by the randomized static model (RSM):

$$\mathbf{v} = \mathbf{F}[X + \eta, \mathbf{a}] + \xi \tag{1}$$

We adopt the following notation: **F**is a given s-dimensional vector-function; **a**designates a random n-dimensional vector of some parameters with independent components a_i, $i = \overline{1, n}$, possessing values in the ranges $\mathcal{A}_i = [a_i^-, a_i^+]$, $i = \overline{1, n}$.

The linear modification of the RSM acquires the form:

$$\mathbf{v} = (X + \eta)\mathbf{a} + \xi \tag{2}$$

Measurements incorporate errors modeled by a matrix, η, of sizes $(s \times n)$ ("input" errors) and by a vector, ξ, of length s ("output" errors). The elements $\eta_{ji}\,(j = \overline{1, s},\ i = \overline{1, n})$ and the components $\xi_j\ (j = \overline{1, s})$ represent independent random variables. Their values lie within the intervals $\mathcal{E}_{ji} = [\eta_{ji}^-, \eta_{ji}^+]$ and $\Xi_j = [\xi_j^-, \xi_j^+]$, $j \in [1, s]$, respectively.

2.2. Dynamic Objects

Consider a discrete dynamic object with a finite "memory", m. The object's input, $x[k]$, is measured precisely, while the output, $y[k]$, is measured with an additive noise, $\xi[k]$. Here, $k \in \mathcal{T} = [m, m + s]$, and s corresponds to the number of measurements. Suppose that the process, $\xi[k]$, is random with independent values.

The connection between the observed input and output (including output measurement errors) is defined by a randomized dynamic model (RDM). This model is described by the discrete functional Volterra polynomial of degree, R [15]:

$$v[k] = \sum_{h=1}^{R} \sum_{(n_1,\ldots,n_h)=0}^{m} \left(w^{(h)}[n_1, \ldots, n_h] \prod_{r=1}^{h} x[k - n_r] \right) + \xi[k] \tag{3}$$

Equality Equation (3) employs the weight functions (actually, impulse responses) $w^{(h)}[n_1, \ldots, n_h]$. These functions are random with independent ordinates belonging to the intervals:

$$\mathcal{W}^{(h)} = [\beta_-^{(h)} \exp\left(-\alpha^{(h)}(n_1 + \cdots + n_h)\right),\ \beta_+^{(h)} \exp\left(-\alpha^{(h)}(n_1 + \cdots + n_h)\right)] \tag{4}$$

where $\beta_\pm^{(h)}, \alpha_\pm^{(h)}$ correspond to given constants.

The discrete dynamic model has the nonlinear expression Equation (3). Nevertheless, it can be linearized by the lexicographic ordering of the variables $\{n_1, \dots, n_h\}$. Renumber the resulting sets from zero to $t_h = (m+1)^h - 1$ according to a lexicographic rule. Introduce a local index, $i^{(h)} \in [0, t_h]$, and a set, such that:

$$\{n_1, \dots, n_h\} \to i^{(h)}, \qquad i^{(h)} \in [0, t_h] \tag{5}$$

According to the accepted numbering, adopt the following indexing of random parameters that correspond to the values of weight functions in Equation (3):

$$a^{(h)}_{i^{(h)}} \to w^{(h)}[n_1, \dots, n_h], \qquad i^{(h)} \in [0, t_h] \tag{6}$$

Construct the random vector:

$$\mathbf{a}^{(h)} = \{a^{(h)}_0, \dots, a^{(h)}_{t_h}\}$$

Components of the vector, $\mathbf{a}^{(h)}$, belong to the ranges:

$$\mathcal{A}^{(h)}(i^{(h)}) = [a^{(h)}_-(i^{(h)}), a^{(h)}_+(i^{(h)})] \tag{7}$$

By virtue of Equation (4), we have:

$$\begin{aligned} a^{(h)}_-(i^{(h)}) &= \beta^{(h)}_- \exp[-i^{(h)}\alpha^{(h)}_-] \\ a^{(h)}_+(i^{(h)}) &= \beta^{(h)}_+ \exp[-i^{(h)}\alpha^{(h)}_+] \end{aligned} \tag{8}$$

Similarly to Equation (5), introduce the lexicographic ordering of the variables $\{k-n_1, \dots, k-n_h\}$, where k indicates a fixed parameter and indexes $n_1, \dots, n_h$ possess values in the interval $[0, m]$. For each fixed k, renumber the resulting sets according to Equation (5):

$$\{k-n_1, \dots, k-n_h\} \to (k, i^{(h)}) \tag{9}$$

Consequently, for fixed k in formula Equation (3), the lexicographically ordered products of the variables, $x[k-n_r]$, form the vector:

$$\mathbf{x}^{(h)}[k] = \{x^{(h)}_{k,0}, \dots, x^{(h)}_{k,t_h}\} \tag{10}$$

where:

$$\mathbf{x}^{(h)}[k] = \mathbf{x}^{(h)}[m+k], \qquad k \in [0, s]$$

Therefore, equality Equation (3) can be rewritten as:

$$v[k] = \sum_{h=1}^{R} \langle \mathbf{a}^{(h)}, \mathbf{x}^{(h)}[k] \rangle + \xi[k], \quad v[k] = v[m+k],\ \xi[k] = \xi[m+k]; \qquad k \in [0, s] \tag{11}$$

Consider the interval $\mathcal{T} = [m, m+s]$, which corresponds to the measurements of the RDM input and output at instants $m+k$, $k \in [0, s]$. The observed output and noise are characterized by the vectors:

$$\mathbf{v} = \{v[0], \dots, v[s]\}, \qquad \xi = \{\xi[0], \dots, \xi[s]\} \tag{12}$$

Define the input measurement matrices, whose rows are formed from vector Equation (10):

$$X^{(h)} = \left[\mathbf{x}^{(h)}[k],\ k \in [0, s]\right], \qquad h \in [1, R]$$

Build the block matrix, X, of sizes $[(s+1)\times u]$, where $u=\sum_{h=1}^{R}(t_h+1)$:

$$X = \left[X^{(1)}, \ldots, X^{(R)}\right]$$

Finally, construct the block random vector of model parameters (of length u):

$$\mathbf{a} = \{\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(R)}\} \tag{13}$$

Thus, dynamic model Equation (3) belongs to the class of randomized models with the random parameters, **a**, and the output noise, ξ. It can be reduced to the linear form:

$$\mathbf{v} = X\,\mathbf{a} + \xi \tag{14}$$

Structurally, this expression is analogous to linear static data model Equation (2). However, the input matrix embraces the nonlinearity and dynamics of model Equation (14).

3. Probabilistic Characteristics of RMs

Numerical characteristics of RMs are understood as the probability density functions (alternatively, probability functions) of model parameters and noise components.

In the sequel, we believe that RMs whose random components are described by probability density functions belong to the class RM-PWQ. By analogy, RMs described by the probabilities of lying within appropriate intervals will form the class RM-pwq.

3.1. RM-PWQ

The parameters of an RDM from this class and the measurement noises represent continuous random variables. They take values in intervals, where there exist probability density functions (pdfs):

-for the random parameters of an RSM:

$$P(\mathbf{a}) = \prod_{i=1}^{n} p_i(a_i), \qquad a_i \in \mathcal{A}_i \tag{15}$$

- for the random parameters of an RDM:

$$P(\mathbf{a}) = \prod_{h=1}^{R}\prod_{i=1}^{t_h} p_i^{(h)}(a_i^{(h)}) \tag{16}$$

- for the "input" measurement noises:

$$W(\eta) = \prod_{j=1}^{s}\prod_{i=1}^{n} w_{ji}(\eta_{ji}), \qquad \eta_{ji} \in \mathcal{E}_{ji} \tag{17}$$

- for the "output" measurement noises:

$$Q(\xi) = \prod_{j=1}^{s} q_j(\xi_j), \qquad \xi_j \in \Xi_j \tag{18}$$

The above-mentioned pdfs of the random parameters and noises should be estimated using measurements of the RM "input" and "output" and *a priori* information on the pdfs. Formalization of such information involves the *a priori* pdfs of the parameters and noises, ($P^0(\mathbf{a})$, $W^0(\eta)$ and $Q^0(\xi)$), as well as the classes of pdfs.

An RM generates an ensemble, $\mathcal{V}$, of the random vectors, $\mathbf{v}$, Equations (2) and (14). Measurements make up the vector, $\mathbf{y}$, with measured components. Therefore, the estimation of pdfs lies in the forming of the vectors of the appropriate numerical characteristics of this ensemble. We employ moments of random components of the vector $\mathbf{v}$:

$$\mathbf{m}^{(k)} = \{\mathcal{M}(v_1^{(k)}), \ldots, \mathcal{M}(v_s^{()})\} \tag{19}$$

where k denotes the order of the moment,

$$\mathcal{M}(v_j^{(k)}) = \int_{\mathbf{a}\in\mathcal{A},\, \eta\in\mathcal{E},\, \xi\in\Xi_j} \left(F_j[X+\eta, \mathbf{a}] + \xi_j\right)^k P(\mathbf{a})W(\eta)Q(\xi)d\mathbf{a}d\eta d\xi, \quad j = \overline{1, s} \tag{20}$$

This paper utilizes the first moments, *viz.*, the average values of components from the vector, $\mathbf{v}$. For static model RSM-PWQ Equation (1), we obtain:

$$\mathcal{M}(\mathbf{v}) = \bar{\mathbf{v}} = \int_{(\mathbf{a}\in\mathcal{A},\eta\in\mathcal{E},\xi\in\Xi)} \left(\mathbf{F}[X+\eta, \mathbf{a}] + \xi\right) P(\mathbf{a})W(\eta)Q(\xi)d\mathbf{a}d\eta d\xi \tag{21}$$

For the dynamic model RM-PQ Equation (14), we similarly have:

$$\mathcal{M}(\mathbf{v}) = \bar{\mathbf{v}} = X \int_{\mathbf{a}\in\mathcal{A}} \mathbf{a}\, P(\mathbf{a})\, d\,\mathbf{a} + \int_{\xi\in\Xi} \xi\, Q(\xi)\, d\,\xi \tag{22}$$

3.2. RM-pwq

The parameters of an RM from this class and the measurement noises turn out continuous random variables. Their *belonging* to an appropriate interval is characterized by some probability.

The parameters $a_1, \ldots, a_n$ possess values within the intervals $\mathcal{A}_1, \ldots, \mathcal{A}_n$ with the probabilities $p_1, \ldots, p_n$, respectively. Actually, the probabilities $p_i \in [0, 1]$, $i = \overline{1, n}$. By analogy, the elements, η_{ji}, of the "input" measurement noise matrix take values from the intervals, $\mathcal{E}_{ji}$, with the probabilities, $w_{ji} \in [0, 1]$, respectively. Finally, the components, ξ_j, of the "output" measurement noises lie inside the intervals, Ξ_j, with the probabilities, $q_j \in [0, 1]$, respectively. Denote by p_i^0, w_{ji}^0, q_j^0 the *a priori* values of the listed probabilities ($j = \overline{1, s}$, $i = \overline{1, n}$).

Just like models from the class RM-PWQ, RMs belonging to the class considered to reproduce an ensemble, $\mathcal{V}$, of the random vectors, $\mathbf{v}$. This is done by generating the random parameters with the probabilities, $\mathbf{p}$, the elements of the "input" measurement noise matrix with the probabilities, W, and the components of the "output" measurement noises with the probabilities, $\mathbf{q}$.

As a numerical characteristic for the ensemble generated by an RM of this class, select the vector of the first *quasi-moments* (see [5]):

$$\mathbf{a} = \mathbf{a}^- + L_{\mathbf{a}}\mathbf{p}, \quad \eta = \eta^- + L_\eta \otimes W, \quad \xi = \xi^- + L_\xi \mathbf{q} \tag{23}$$

In the previous formula, $\otimes$ stands for element-wise multiplication,

$$\begin{aligned} L_{\mathbf{a}} &= \text{diag}\,[(a_i^+ - a_i^-)|i=\overline{1,n}];\;\; L_\xi = \text{diag}\,[(\xi_j^+ - \xi_j^-)|j=\overline{1,s}] \\ L_\eta &= [(\eta_{ji}^+ - \eta_{ji}^-)|i=\overline{1,n},\; j=\overline{1,s}] \end{aligned} \tag{24}$$

The transform Equation (23) can be interpreted as a substitution of the random parameters and noises by their "quasi-average" values.

By applying Equation (23) to Equation (1), we obtain the following expression for the first quasi-moment of the random vector, $\mathbf{v}$, for RSM Equation (1):

$$\tilde{\mathbf{v}} = \mathbf{F}\left[X + (\eta^- + L_\eta \otimes W), \mathbf{a}^- + L_{\mathbf{a}}\mathbf{p}\right] + \xi^- + L_\xi \mathbf{q} \tag{25}$$

In the case of RDMs, the first quasi-moment formula of the random vector, $\mathbf{v}$, Equation (10), acquires the form:

$$\tilde{\mathbf{v}} = X\mathbf{a} + L_{\mathbf{ap}} + \xi^- + L_\xi \mathbf{q} \tag{26}$$

Concluding this section, we emphasize a relevant aspect. For the class RM-PWQ, it is necessary to estimate probability density functions. For the class RM-pwq, one has to estimate vectors characterizing probability distributions.

4. Principles of Entropy-Robust Estimation

We propose to introduce a *likelihood functional* for the estimation of probability density functions for RM-PWQ (RM-PQ). So long as the model parameters and measurement noises are independent, their joint distribution function takes the form:

$$\Phi(\mathbf{a}, \eta, \xi) = P(\mathbf{a})W(\eta)Q(\xi) \tag{27}$$

Suppose that we know the *a priori* density functions, $P^0(\mathbf{a})$, $W^0(\eta)$, $Q^0(\xi)$.

Following [4], specify the log-likelihood ratio (LLR) by:

$$\varphi(\mathbf{a}, \eta, \xi) = \ln \frac{P(\mathbf{a})}{P^0(\mathbf{a})} + \ln \frac{W(\eta)}{W^0(\eta)} + \ln \frac{Q(\xi)}{Q^0(\xi)} \tag{28}$$

Hence, the LLR represents a nonrandom function of random arguments.

The probabilities, $\mathbf{p}, W, \mathbf{q}$, of certain events (e.g., entering appropriate intervals by the parameters, "input" and "output" noises) being estimated, one should define the LLR by:

$$\varphi(\mathbf{p}, W, \mathbf{q}) = \ln \frac{\mathbf{p}}{\mathbf{p}^0} + \ln \frac{W}{W^0} + \ln \frac{\mathbf{q}}{\mathbf{q}^0} \tag{29}$$

where $\mathbf{p}^0$, W^0 and $\mathbf{q}^0$ are *a priori* probabilities.

Now, introduce the likelihood functional (the generalized informational Boltzmann entropy functional), $\mathcal{L}$ (see Equation (29)), in the form:

$$\begin{aligned} &\mathcal{L}[P(\mathbf{a}, W(\eta), Q(\xi)] = -\mathcal{H}[P(\mathbf{a}), W(\eta), Q(\xi)] = \\ &= \int_{\mathbf{a}\in\mathcal{A}} P(\mathbf{a})\, \ln \frac{P(\mathbf{a})}{P^0(\mathbf{a})} d\mathbf{a} + \int_{\eta\in\mathcal{E}} W(\eta)\, \ln \frac{W(\eta)}{W^0(\eta)} d\eta + \\ &+ \int_{\xi\in\Xi} Q(\xi)\, \ln \frac{Q(\xi)}{Q^0(\xi)} d\xi \end{aligned} \tag{30}$$

Obviously (see Equation (28)), the likelihood functional with the minus mark represents the generalized Boltzmann entropy functional. This function has numerous interpretations. In particular, it is treated as the "distance" between pdfs [16], as an uncertainty measure [12,13,17] or as a robustness measure [6] (the degree of invariance of the pdfs $P(\mathbf{a}), W(\eta), Q(\xi)$, with respect to observations). In this paper, we involve the last interpretation of the entropy functional to construct estimates of the pdfs.

Define the likelihood function (the generalized informational Boltzmann entropy function) by (see [18]):

$$H(\mathbf{p}, W, \mathbf{q}) = -\sum_{i=1}^{n} p_i \ln \frac{p_i}{p_i^0} - \sum_{j=1}^{s}\sum_{i=1}^{n} w_{ji} \ln \frac{w_{ji}}{w_{ji}^0} - \\ - \sum_{j=1}^{s} q_j \ln \frac{q_j}{q_j^0} \quad (31)$$

The estimation quality for pdfs or probability vector components is characterized by the maximal value of the generalized informational Boltzmann entropy functional (or function). In the sequel, we distinguish between two types of estimates.

1. The $\mathcal{S}^k_{PWQ}$-robust entropy estimate (the $\mathcal{S}^k_{PQ}$-robust entropy estimate)of the pdfs of the parameters and measurement noises results from solving the problem:

$$\mathcal{H}[P(\mathbf{a}, W(\eta), Q(\xi)] \Rightarrow \max \quad (32)$$

provided that:

- the probability density functions belong to the class:

$$\mathcal{S} = \mathcal{P} \bigcup \mathcal{W} \bigcup \mathcal{Q} \quad (33)$$

  where $\mathcal{P}, \mathcal{W}$ and $\mathcal{Q}$ denote the classes of the pdfs of the parameters, "input" and "output" noises, respectively;
- the balance condition holds true for the degree $(1/k)$ of the k-th moment, $\mathbf{v}$, and the measured vector, $\mathbf{y}$. For RSM-PWQ, this condition becomes:

$$\left(\mathcal{M}\{\mathbf{F}^{(k)}[(X+\eta), \mathbf{a}] + \xi^{(k)}\}\right)^{(1/k)} = \mathbf{y} \quad (34)$$

For RDM-PQ, the balance condition takes the form:

$$(\mathcal{M}\{X\mathbf{a} + \xi\})^{(1/k)} = \mathbf{y} \quad (35)$$

2. The $\mathcal{S}^{\tilde{1}}_{pwq}$-robust entropy estimate

of the probability distributions (pds) of the model parameters and measurement noise components follow from solving the problem:

$$H(\mathbf{p}, W, \mathbf{q}) \Rightarrow \max \quad (36)$$

provided that:

- the probability distributions belong to the class:

$$\mathbb{S} = \mathbb{P} \bigcup \mathbb{W} \bigcup \mathbb{Q} \tag{37}$$

  where $\mathbb{P}, \mathbb{W}$ and $\mathbb{Q}$ designate the classes of the pds of the parameters, "input" and "output" noises, respectively;
- the balance condition holds true for the quasi-average vector, $\tilde{\mathbf{v}}$, and the measured vector, $\mathbf{y}$:

$$\begin{aligned} &\mathbf{F}\left[X + (\eta^- + L_\eta W), (\mathbf{a}^- + L_{\mathbf{a}}\mathbf{p})\right] + \\ &\qquad + \xi^- + L_\xi \mathbf{q} = \mathbf{y} \end{aligned} \tag{38}$$

5. Structural Properties of $\mathcal{S}^1_{PQ}$-Estimates

5.1. Power-Type RSM-PQ

Consider the subclass of power-type RSM-PQ with the following nonlinearities:

$$v(t) = \sum_{h=1}^{R} \sum_{i=1}^{n} a_i^h \, x_i^{(h)}(t) + \xi(t) \tag{39}$$

The vector of parameters, $\mathbf{a}$, turns out to be random with independent components $(a_1, \ldots, a_n)$ possessing values in the intervals $\mathcal{A}_i = [a_i^-, a_i^+], (i = \overline{1,n})$ with the pdfs $p_1(a_1), \ldots, p_n(a_n)$.

The "input" and "output" are observed at instants $t_1, \ldots, t_s$. The measured "input" is characterized by the set of R matrices:

$$X^{(h)} = \begin{pmatrix} x_1^{(h)}(t_1) & \cdots & x_n^{(h)}(t_1) \\ \cdots & \cdots & \cdots \\ x_1^{(h)}(t_s) & \cdots & x_n^{(h)}(t_s) \end{pmatrix} = \begin{pmatrix} x_{11}^{(h)} & \cdots & x_{1,n}^{(h)} \\ \cdots & \cdots & \cdots \\ x_{s,1}^{(h)} & \cdots & x_{s,n}^{(h)} \end{pmatrix}, \; h \in [1, R] \tag{40}$$

Additionally, the observed "output" of the RM is characterized by the random vector $\mathbf{v} = \{v(t_1), \ldots, v(t_s)\}$.

Therefore, using observation results, we rewrite the RSM-PQ as:

$$\mathbf{v} = \sum_{h=1}^{R} X^{(h)} \mathbf{a}^{(h)} + \xi \tag{41}$$

Here, $\mathbf{a}^{(h)} = \{a_1^h, \ldots, a_n^h\}$, $\xi = \{\xi(t_1), \ldots, \xi(t_s)\} = \{\xi_1, \ldots, \xi_s\}$ denotes the "output" measurement noise vector. It has independent components belonging to the intervals $\Xi_j = [\xi_j^-, \xi_j^+], (j = \overline{1,s})$ with the pdfs $q_1(\xi_1), \ldots, q_s(\xi_s)$. Suppose that *a priori* information is absent, *i.e.*, $P^0(\mathbf{a}) = Q^0(\xi) = \text{const}$.

To proceed, we analyze the $\mathcal{S}^1_{PQ}$-robust entropy estimate of the pdfs $P(\mathbf{a}) = \{p_1(a_1), \ldots, p_n(a_n)\}$ and $Q(\xi) = \{q_1(\xi_1), \ldots, q_s(\xi_s)\}$. Moreover, we study some structural properties of the estimate.

Revert to problem Equations (32)–(34). Reexpress these as:

$$\begin{aligned} \mathcal{H}[P(\mathbf{a}), Q(\xi)] &= -\sum_{i=1}^{n} \int_{a_i \in \mathcal{A}_i} p_i(a_i) \ln p_i(a_i) da_i - \\ &- \sum_{j=1}^{s} \int_{\xi_j \in \Xi_j} q_j(\xi_j) \ln q_j(\xi_j) d\xi_j . \Rightarrow \max \end{aligned} \tag{42}$$

subject to the constraints:

$$\begin{aligned} D_i[p_i(a_i)] &= 1 - \int_{a_i \in \mathcal{A}_i} p_i(a_i)da_i = 0, \; i = \overline{1,n} \\ T_j[q_j(\xi_j)] &= 1 - \int_{\xi_j \in \Xi_j} q_j(\xi_j)d\xi_j = 0, \; j = \overline{1,s} \end{aligned} \tag{43}$$

$$\begin{aligned} \Phi_j[p(\mathbf{a}), q(\xi)] &= \sum_{h=1}^{R}\sum_{i=1}^{n} x_{ji}^{(h)} \int_{a_i \in \mathcal{A}_i} a_i^h p_i(a_i)da_i + \\ &+ \int_{\xi_j \in \Xi_j} \xi_j q_j(\xi_j)d\xi_j - y_j = 0, \quad j = \overline{1,s} \end{aligned} \tag{44}$$

The structural properties of the $\mathcal{S}^1_{PQ}$-estimate are understood as a certain class of the robust family of probability density functions of the model parameters and noise. The procedure of $\mathcal{S}^1_{PQ}$-estimation consists in solving problem Equations (32)–(34) for RSM-PWQ (alternatively, problem Equations (32), (33) and (35) for RDM-PQ). Both problems belong to the class of functional entropy-linear programming problems with equality constraints. The standard approach to such variational problems (see [19]) naturally brings the following result. The entropy-optimal pdfs (in the class of continuously differentiable functions) of the parameters and noise components are defined by:

$$\begin{aligned} p_i^*(a_i) &\sim \pi_i \exp(-\sum_{h=1}^{R} \alpha_{ih} a_i^h), \qquad i \in [1,n] \\ q_j^*(\xi_j) &\sim \kappa_j \exp(-\theta_j \xi_j), \qquad j = [1,s]) \end{aligned} \tag{45}$$

where:

$$\alpha_{ih} = \sum_{j=1}^{s} \theta_j x^{(h)_{ji}}, \quad \pi_i = \exp(-1-\gamma_i); \qquad \kappa_j = \exp(-1-\omega_j) \tag{46}$$

In the last formulas, the Lagrange multipliers, γ_i, ω_j, correspond to constraint Equation (43), while the ones of θ_j correspond to constraint Equation (44).

Figures 1–4 demonstrate some examples of estimating the pdfs of the parameters. For a linear RM, the $\mathcal{S}^1_{PQ}$-estimate always represents an exponential function (see Figure 1). The "input" and "output" measurements do not affect the structure of the estimate (yet, they change the shape of the functions).

For nonlinear RMs, the structure of the $\mathcal{S}^1_{PQ}$-estimates varies depending on the "input" and "output" measurements. For instance, Figures 2–4 show the estimates of the pdfs for a quadratic, quadratic-linear and cubic RM-PQ.

Figure 1. $p(a_1, a_2) = \exp(-0,3a_1 + 2,5a_2)$.

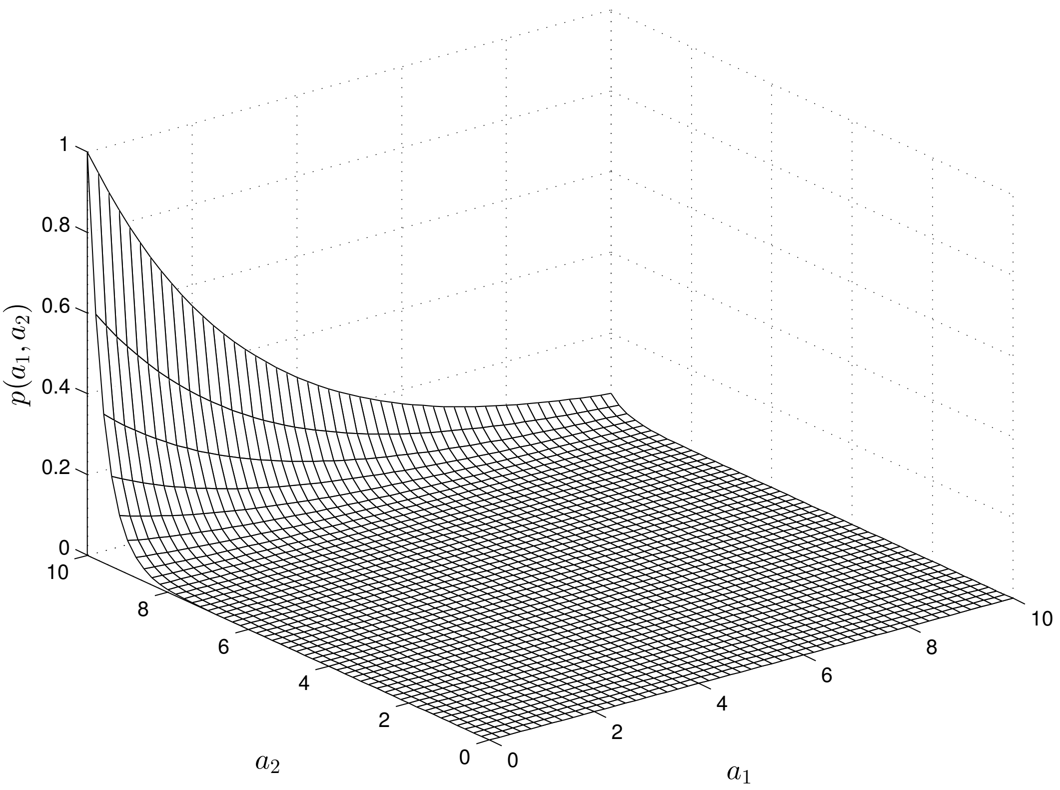


Figure 2. $p(a_1, a_2) = exp(4a_1 - a_1^2 + a_2)$.

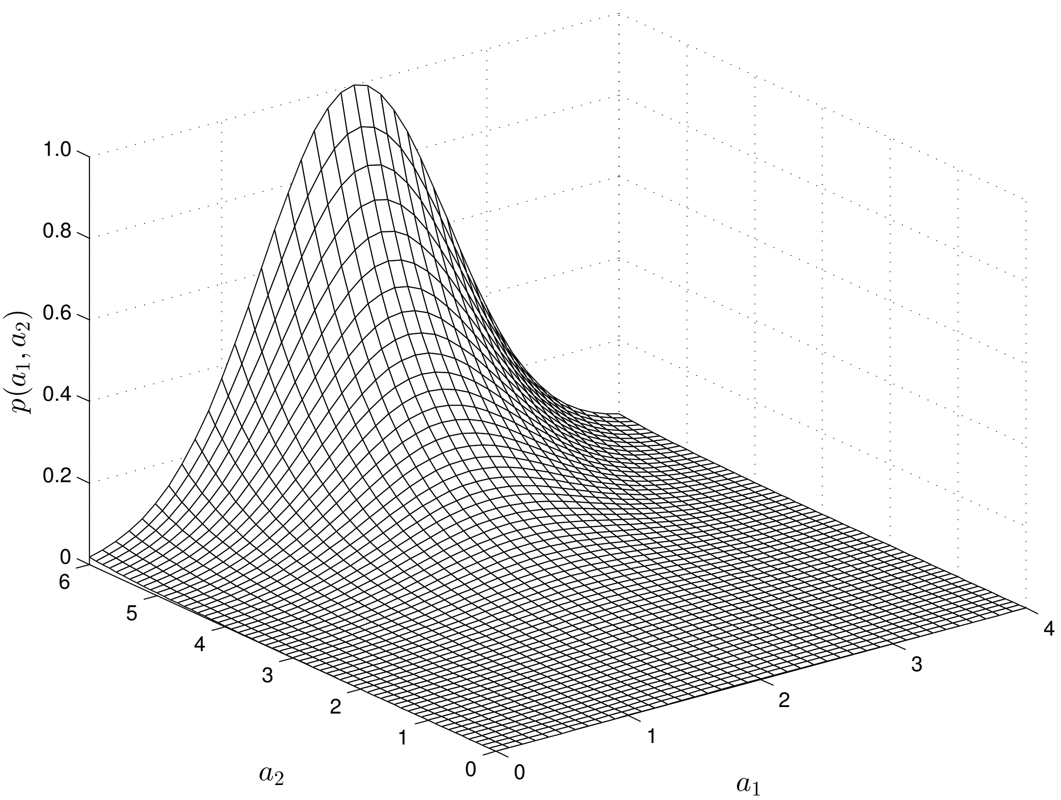


Figure 3. $p(a_1, a_2) = exp(4a_1 - a_1^2 + 6a_2 - a_2^2)$.

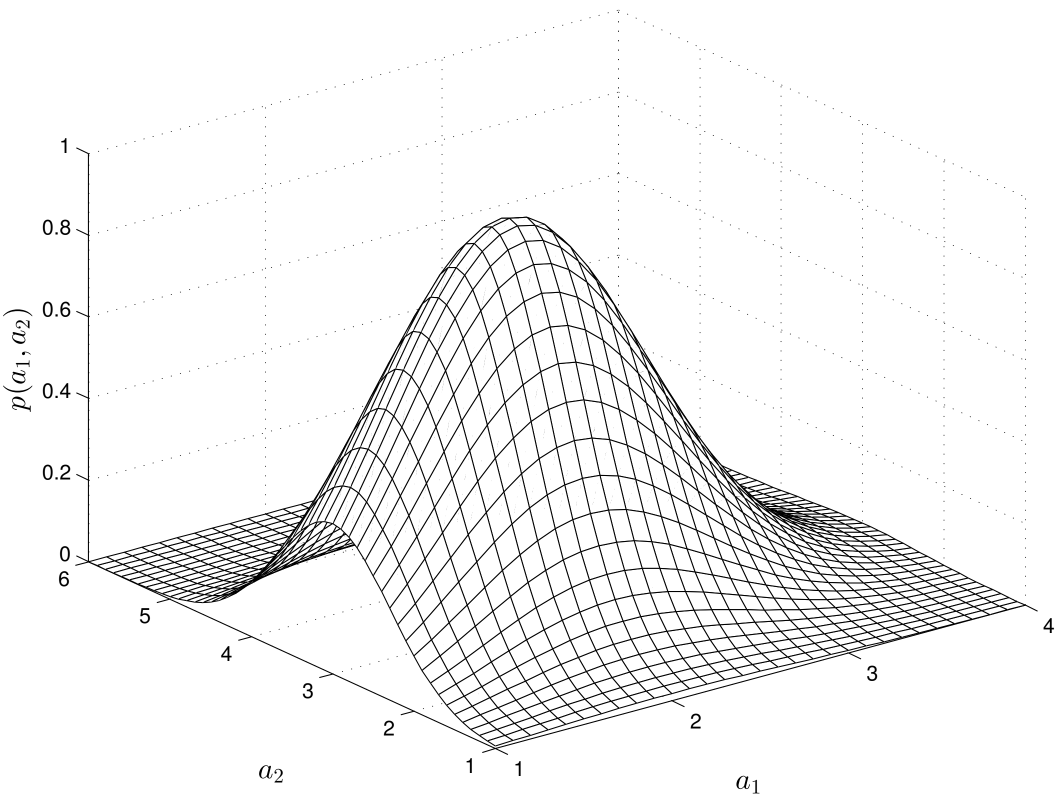
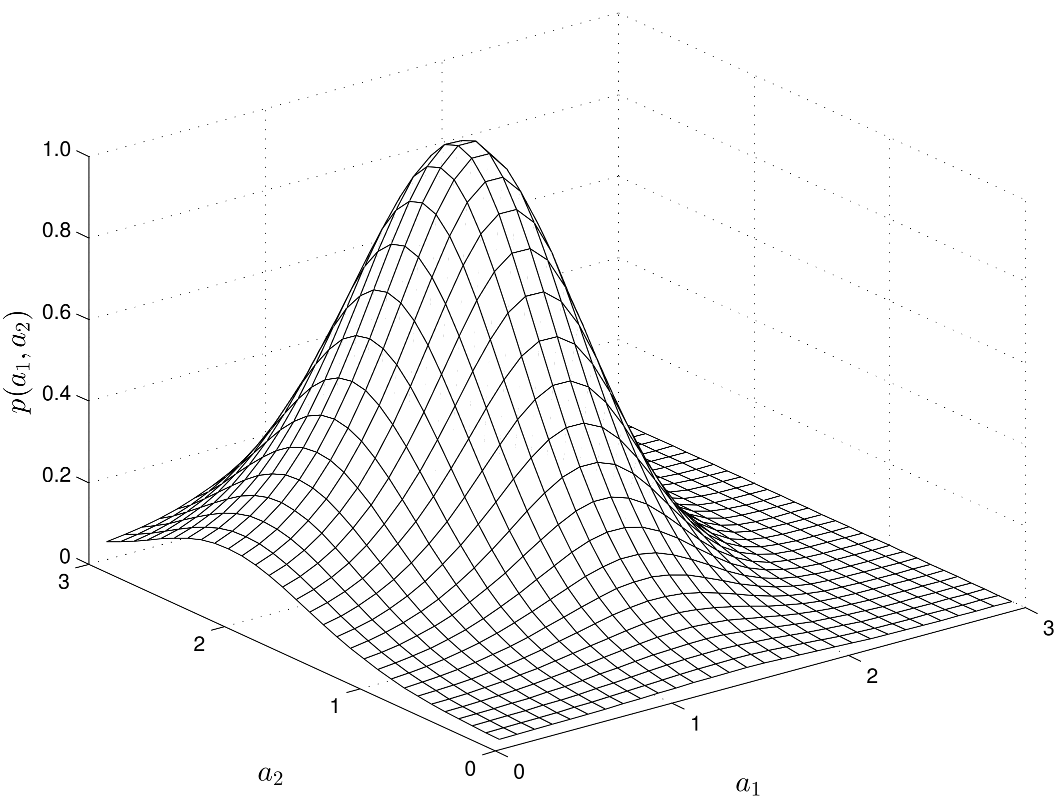


Figure 4. $p(a_1, a_2) = exp(0.5a_1 + 3.5a_2 + 2a_1^2 - a_1^3 - 0.2a_2^2 - 0.2a_2^3)$.



5.2. Power-Type RM-PQ

Address the linear form of RM-PQ Equation (26). In this case, the problem of $\mathcal{S}^{1}_{PQ}$-estimation of the pdfs for parameter Equation (16) and noise Equation (18) acquires the form:

$$\mathcal{H}[P(\mathbf{a}), Q(\xi)] = -\int_{\mathbf{a}\in\mathcal{A}} P(\mathbf{a}) ln P(\mathbf{a}) d\mathbf{a} - \int_{\xi\in Xi} Q(\xi) ln Q(\xi) d\xi \Rightarrow \max \tag{47}$$

subject to the constraints:

$$\begin{aligned} D[P(\mathbf{a})] &= 1 - \int_{\mathbf{a}\in\mathcal{A}} P(\mathbf{a}) d\mathbf{a} = 0 \\ T[Q(\xi)] &= 1 - \int_{\xi\in\Xi} Q(\xi) d\xi = 0 \end{aligned} \tag{48}$$

$$\begin{aligned} \bar{\Phi}[P(\mathbf{a}), Q(\xi)] &= X \int_{\mathbf{a}\in\mathcal{A}} \mathbf{a} P(\mathbf{a}) d\mathbf{a} + \\ &+ \int_{\xi\in\Xi} \xi Q(\xi) d\xi - \mathbf{y} = \mathbf{0} \end{aligned} \tag{49}$$

The vector, $\bar{\Phi}[P(\mathbf{a}), Q(\xi)]$, has the length of $(s+1)$ (the number of measurements).

Problem Equations (47)–(49) represent a modification of functional entropy-linear programming problem Equations (42)–(44), which involves the linear data model. To obtain the solution, again, adopt the standard technique for such variational problems [19]. This yields the following entropy-optimal pdfs of the parameters and noises:

$$\tilde{P}(\mathbf{a}) = \exp\left(-1 - \gamma - \langle\theta, X\mathbf{a}\rangle\right), \qquad \tilde{Q}(\xi) = \exp\left(-1 - \omega - \theta \otimes \xi\right) \tag{50}$$

Here, γ, ω designate the Lagrange multipliers associated with constraint Equation (48). In addition, the vector, θ (whose length equals $(s+1)$), comprises the Lagrange multipliers for constraint Equation (49).

Formula (50) indicates that the entropy-optimal pdfs of the parameters and noises belong to the exponential class, with the parameters representing the Lagrange multipliers. The latter quantities meet a system of equations generated by constraint Equations (48),(49).

6. Normalized and Interval $\mathcal{S}^{\tilde{1}}_{pq}$-Estimates for Linear RSM-pq

Consider the subclass of linear RSM-pq with accurate "input" measurements:

$$\tilde{\mathbf{v}} = XL_{\mathbf{a}}\mathbf{p} + L_{\xi}\mathbf{q} + \mathbf{K}(\mathbf{a}^{-}\xi^{-}) \tag{51}$$

where:

$$\mathbf{K}(\mathbf{a}^{-}, \xi^{-}) = X\mathbf{a}^{-} + \xi^{-} \tag{52}$$

Denote by $\mathbf{p}^0$ and $\mathbf{q}^0$ the *a priori* probabilities of the parameters and noises, respectively.

Let us study the $\mathcal{S}^{\tilde{1}}_{pq}$-estimates for some classes of the vectors, $\mathbf{p}$, $\mathbf{q}$.

Problem 1. Normalized probabilities: the $\mathcal{S}^{\tilde{1}}_{pq}$-estimate results from solving the problem:

$$H(\mathbf{p},\mathbf{q} = -\sum_{i=1}^{n} p_i \ln \frac{p_i}{p_i^0} - -\sum_{j=1}^{s} q_j \ln \frac{q_j}{q_j^0} \Rightarrow \max \tag{53}$$

subject to the probability normalization conditions:

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{j=1}^{s} q_j = 1 \tag{54}$$

and the balance conditions for the real observations, y_j, with the first quasi-moment of the output of RSM-pq:

$$\sum_{i=1}^{n} x_{ji} L^i_{\mathbf{a}} p_i + L^j_{\xi} q_j + K_j = y_j, \qquad j \in [1,s],\ s < n \tag{55}$$

This problem belongs to the class of entropy-linear programming problems [20]. Its solution employs the necessary conditions of optimality in terms of the Lagrange function:

$$\begin{aligned} L(\mathbf{p},\mathbf{q},\lambda,\varepsilon,\theta) &= H(\mathbf{p},\mathbf{q}) + \lambda(1 - \sum_{i=1}^{n} p_i) \\ + \quad \varepsilon(1 - \sum_{j=1}^{s} q_j) &+ \sum_{j=1}^{s} \theta_j [y_j - \sum_{i=1}^{n} x_{ji} L^i_{\mathbf{a}} p_i - L^j_{\xi} q_j - K_j] \end{aligned} \tag{56}$$

Consequently, we derive a system of equations determining the entropy-optimal probabilities, $\mathbf{p}^*, \mathbf{q}^*$:

$$\begin{aligned} 0 \le p_i^*(\theta) &= \frac{p_i^0 \exp\left(-\sum_{j=1}^{s} \theta_j x_{ji} L^i_{\mathbf{a}}\right)}{\sum_{i=1}^{n} p_i^0 \exp\left(-\sum_{j=1}^{s} \theta_j x_{ji} L^i_{\mathbf{a}}\right)} \le 1,\ i = \overline{1,n} \\ 0 \le q_j^*(\theta) &= \frac{q_j^0 \exp\left(-\theta_j L^j_{\xi}\right)}{\sum_{j=1}^{s} q_j^0 \exp\left(-\theta_j L^j_{\xi}\right)} \le 1, \qquad j = \overline{1,s} \end{aligned} \tag{57}$$

and the corresponding Lagrange multipliers $\theta_1, \ldots, \theta_s$:

$$\Phi_j(\theta) \frac{1}{y_j - K_j} \sum_{i=1}^{n} x_{ji} L^i_{\mathbf{a}} p_i^*(\theta) + L^j_{\xi} q_j^*(\theta) = 1, \qquad j = \overline{1,s} \tag{58}$$

Problem 2. The interval probabilities: the $\mathcal{S}^{\tilde{1}}_{pq}$-estimate satisfies the following maximization problem for the generalized informational Fermi–Dirac entropy. The application of the generalized informational Fermi–Dirac entropy provides estimates lying in appropriate intervals [18]:

$$\begin{aligned} H(\mathbf{p},\mathbf{q}) &= -\sum_{i=1}^{n} \left(p_i \ln \frac{p_i}{\varphi_i^0} + (1-p_i)\ln(1-p_i) \right) - \\ &- \sum_{j=1}^{s} \left(q_j \ln \frac{q_j}{\phi_j^0} + (1-q_j)\ln(1-q_j) \right) \Rightarrow \max \end{aligned} \tag{59}$$

subject to the interval probability constraints:

$$0 \le p_i \le 1, \quad i = \overline{1,n}; \qquad 0 \le q_j \le 1, \quad j \in [1,s],\ s < n \tag{60}$$

and the balance conditions for the real observations, y_j, with the first quasi-moment of the output of RM-pq:

$$\sum_{i=1}^{n} x_{ji} L^i_{\mathbf{a}} p_i + L^j_{\xi} q_j + K_j = y_j, \qquad j = \overline{1, s} \tag{61}$$

where:

$$\varphi^0_i = \frac{p^0_i}{1 - p^0_i}, \qquad \phi^0_j = \frac{q^0_j}{1 - q^0_j} \tag{62}$$

Lagrange's method of multipliers gives the following system of equations for the estimates of the probabilities, $\mathbf{p}$, $\mathbf{q}$:

$$\begin{aligned} 0 \le p^*_i(\theta) &= \frac{p^0_i}{p^0 + (1 - p^0_i) \exp\left(\sum_{j=1}^{s} \theta_j x_{ji} L^i_{\mathbf{a}}\right)} \le 1, i = \overline{1, n} \\ 0 \le q^*_j(\theta) &= \frac{q^0_j}{q^0_j + (1 - q^0_j) \exp\left(-\theta_j L^j_{\xi}\right)} \le 1, j = \overline{1, s} \end{aligned} \tag{63}$$

and the Lagrange multipliers $\theta_1, \ldots, \theta_s$:

$$\Phi_j(\theta) = \frac{1}{y_j - K_j} \sum_{i=1}^{n} \theta_j x_{ji} p^*_i(\theta) + L^j_{\xi} q^*_j(\theta) = 1, j = \overline{1, s} \tag{64}$$

So long as the estimates of $\mathbf{p}$ and $\mathbf{q}$ are expressed analytically through the Lagrange multipliers, θ, one should only solve the system of Equation (64). For this, it is possible to use the following multiplicative algorithm [20] for the exponential Lagrange multipliers $z_j = \exp(-\theta_j), \ j \in [1, s]$:

$$z^{k+1}_j = z^k_j \Phi_j(z^k), \qquad , z^0_j > 0, \ j \in [1, s] \tag{65}$$

Under certain conditions, the estimates generated by Problems 1–2 differ in the values of some entropy (e.g., generalized Boltzmann entropy Equation (53)). Let us introduce the following notation:

- the entropy:

$$H(\mathbf{p}, \mathbf{q}) = H(\mathbf{g}), \quad \mathbf{g} = (\mathbf{p}, \mathbf{q}) \in R^{n+s}_+ \tag{66}$$

- the estimate, $\mathbf{g}^*_1$, obtained for normalized probabilities (Problem 1), and the estimate, $\mathbf{g}^*_2$, obtained for interval probabilities (Problem 2);
- the absolute maximum of the entropy $\hat{\mathbf{g}} = \arg\max \ H(\mathbf{g})$;
- the set:

$$\Pi = \{\mathbf{g} : \mathbf{0} \le \mathbf{g} \le \mathbf{1}\} \supset \tilde{\Pi} = \{\mathbf{g} : \langle \mathbf{p}, \mathbf{1} \rangle, \ \langle \mathbf{q}, \mathbf{1} \rangle\} \tag{67}$$

Theorem *Assume that* $\hat{\mathbf{g}} \in (R^{(n+s)}_+ \setminus \tilde{\Pi})$.
Then:

$$H(\mathbf{g}^*_1) \le H(\mathbf{g}^*_2)$$

and the equality takes place iff $\mathbf{g}^*_1 = \hat{\mathbf{g}}$.

Proof. Entropy Equation (53) represents a strictly concave function with a unique absolute maximum at the point, $\hat{\mathbf{g}}$. The value of the entropy at an arbitrary point, $\mathbf{g}$, depends on the distance to the absolute maximum point. Denote by $\varrho(\hat{\mathbf{g}}, \mathbf{g}^*_1)$ and $\varrho(\hat{\mathbf{g}}, \mathbf{g}^*_2)$ the distances between the absolute maximum point, $\hat{\mathbf{g}}$, and the points corresponding to the optimal estimates in Problems (53)–(55) and (59)–(62), respectively. By virtue of the premise above, expression Equation (67) and strict concavity, we have:

$$\varrho(\hat{\mathbf{g}}, \mathbf{g}_1^*) \geq \varrho(\hat{\mathbf{g}}, \mathbf{g}_2^*) \tag{68}$$

These distances coincide only if $\mathbf{g}_1^* = \hat{\mathbf{g}}$.

7. Examples

7.1. Entropy Estimation of the Parameters of Linear RSM-pq under a Small Amount of Data

Consider a linear RM with five random parameter Equations (51) belonging to the same intervals $\mathcal{A} = [0, 10]$ and with the noise vector $\xi = \{\xi_1, \xi_2\}$, whose components lie within the intervals $\Xi_1 = [-3, 3]$ and $\Xi_2 = [-6, 6]$. There are two "output" measurements, $\mathbf{y} = \{y_1, y_2\}$.

The reference model has fixed parameters $\mathbf{a}^0 = \{1, 2, 2, 4, 1\}$. The deviation from these values is described by the relative square error:

$$\varepsilon = \frac{\|\mathbf{a}^0 - \mathbf{a}\|}{\|\mathbf{a}^0\| + \|\mathbf{a}\|}$$

The "input" measurement matrix makes up:

$$X = \begin{pmatrix} 1.8 & 2.1 & 3.3 & 2.0 & 1.5 \\ 4.1 & 3.8 & 3.0 & 2.8 & 1.9 \end{pmatrix}$$

The "output" measurement vector (distorted by noises from appropriate intervals) takes the form:

$$\mathbf{y} = \{21.1; 32.8\}$$

The first quasi-moments for the parameters and noise are defined by:

$$a_i = 10p_i, \quad i = \overline{1, 5}; \qquad \xi_1 = -3 + 6q_1, \;\; \xi_2 = -6 + 12q_2$$

Substituting the last equalities into Equation (51) yields the following RM-pq:

$$T\mathbf{p} + L\mathbf{q} = \mathbf{1}$$

with the matrix:

$$T = XL_{\mathbf{a}} = \begin{pmatrix} 0.75 & 0.87 & 1.37 & 0.83 & 0.62 \\ 1.06 & 0.98 & 0.77 & 0.72 & 0.45 \end{pmatrix}$$

the matrix:

$$L = L_\xi = \begin{pmatrix} 0.25 & 0 \\ 0 & 0.31 \end{pmatrix}$$

and the vector $\mathbf{1} = \{1, 1\}$.

As *a priori* information, we have selected three scenarios for the *a priori* probabilities (see Tables 1 and 2).

Table 1. The *a priori* probabilities, p_i^0, $i \in [1, 5]$.

Scenario/i	**1**	**2**	**3**	**4**	**5**
A	1.0	1.0	1.0	1.0	1.0
B	0.1	0.2	0.3	0.3	0.1
C	0.3	0.4	0.1	0.05	0.15

Table 2. The *a priori* probabilities, q_1^0, q_2^0.

Scenario/i	**1**	**2**
D	0.2	0.8
E	1.0	1.0

Scenario AE corresponds to uniform *a priori* distributions of the parameters and noise. Next, Scenario BD reflects nonuniform *a priori* distributions of the parameters and noise. Finally, CE is the combined scenario (the parameters possess nonidentical *a priori* distributions, whereas the noise has uniform *a priori* distributions).

We study the $\mathcal{S}_{pq}^{\tilde{1}}$-estimates corresponding to Problems 1–2.

Problem 1. Normalized probabilities.

$$H(\mathbf{p}, \mathbf{q}) = -\sum_{i=1}^{5} \left(p_i \ln \frac{p_i}{p_i^0} \right) - \sum_{j=1}^{2} \left(q_j \ln \frac{q_j}{q_j^0} \right) \Rightarrow \max$$

$$\sum_{i=1}^{5} p_i = 1, \quad p_i > 0, \qquad \sum_{j=1}^{2} q_j = 1, \quad q_j > 0$$

$$0.75p_1 + 0.87p_2 + 1.37p_3 + 0.83p_4 + 0.62p_5 + 0.25q_1 = 1$$

$$1.06p_1 + 0.98p_2 + 0.77p_3 + 0.72p_4 + 0.45p_5 + 0.31q_1 = 1$$

Problem 2. Interval probabilities.

$$H(\mathbf{p}, \mathbf{q}) = -\sum_{i=1}^{5} \left(p_i \ln \frac{p_i}{p_i^0} \right) - \sum_{j=1}^{2} \left(q_j \ln \frac{q_j}{q_j^0} \right) \Rightarrow \max$$

$$0 \le \mathbf{p} \le 1, \qquad 0 \le \mathbf{q} \le 1$$

$$0.75p_1 + 0.87p_2 + 1.37p_3 + 0.83p_4 + 0.62p_5 + 0.25q_1 = 1$$

$$1.06p_1 + 0.98p_2 + 0.77p_3 + 0.72p_4 + 0.45p_5 + 0.31q_1 = 1$$

The entropy, H, attains the absolute maximum at the point $\hat{p}_i = 0.36p_i^0$; $\hat{q}_j = 0.36q_j^0$; $i \in [1, 5]$, $j \in [1, 2]$. Tables 3 and 4 present the coordinates of the absolute maximum point for the above scenarios of the *a priori* probabilities (see Tables 1 and 2).

Table 3. The components, $\hat{p}_i,\ i \in [1, 5]$.

Scenario/i	**1**	**2**	**3**	**4**	**5**
A	0.36	0.36	0.36	0.36	0.36
B	0.036	0.072	0.108	0.108	0.036
C	0.108	0.144	0.036	0.018	0.054

Table 4. The components, $\hat{q}_1,\ \hat{q}_2$.

Scenario/i	**1**	**2**
D	0.072	0.288
E	0.36	0.36

Solutions to Problem 1 can be found in Tables 5 and 6.

Table 5. The $\mathcal{S}_{pq}^{\tilde{1}}$-estimates of the probabilities, p, q, in Problem 1.

Scenario	p_1^*	p_2^*	p_3^*	p_4^*	p_5^*	q_1^*	q_2^*	H^*
AD	0.14	0.17	0.28	0.20	0.21	0.28	0.72	1.56
BD	0.08	0.16	0.26	0.34	0.16	0.25	0.75	–0.03
CD	0.17	0.31	0.22	0.06	0.24	0.40	0.60	–0.23
AE	0.23	0.22	0.19	0.19	0.16	0.45	0.55	2.29
BE	0.14	0.24	0.18	0.32	0.12	0.39	0.61	0.63
CE	0.26	0.0.38	0.13	0.05	0.18	0.58	0.42	0.67

Table 6. The first quasi-moments of the parameters and noises in Problem 1.

Scenario	a_1^*	a_2^*	a_3^*	a_4^*	a_5^*	ξ_1^*	ξ_2^*	ε
AD	1.43	1.66	2.78	2.01	2.13	-1.33	2.65	0.13
BD	0.78	1.64	2.62	3.37	1.58	-1.52	3.04	0.03
CD	1.67	3.08	2.25	0.63	2.37	–0.62	1.23	0.30
AE	2.35	2.22	1.91	1.88	1.64	–0.31	0.62	0.15
BE	1.41	2.36	1.84	3.21	1.18	–0.65	1.30	0.02
CE	2.56	3.76	1.30	0.56	1.82	0.46	–0.91	0.36

Solutions to Problem 2 are shown by Tables 7 and 8.

Table 7. The $\mathcal{S}_{pq}^{\tilde{1}}$-estimates of the probabilities, **p**, **q**, in Problem 2.

Scenario	p_1^*	p_2^*	p_3^*	p_4^*	p_5^*	q_1^*	q_2^*	H^*
AD	0.24	0.22	0.18	0.23	0.26	0.0.07	0.0.29	2.05
BD	0.15	0.26	0.27	0.27	0.07	0.07	0.46	0.27
CD	0.23	0.39	0.24	0.05	0.12	0.11	0.26	0.23
AE	0.25	0.22	0.15	0.22	0.25	0.30	0.39	2.37
BE	0.15	0.26	0.23	0.26	0.06	0.34	0.60	0.67
CE	0.26	0.40	0.17	0.05	0.11	0.48	0.37	0.71

The comparative analysis of these computations draws the following conclusions:

- the estimates of the parameters derived for interval probabilities have a larger constrained maximum of the entropy than the ones derived for normalized probabilities (see Theorem 2);
- the reference parameters and *a priori* probabilities appear interconnected: their "successful" choice (Scenario BD) leads to "better" approximation of the reference parameters in terms of the relative squared error than in the case of an "unsuccessful" choice (Scenario CD).

Table 8. The $\tilde{1}$-quasi-moments of the parameters and noise in Problem 2.

Scenario	a_1^*	a_2^*	a_3^*	a_4^*	a_5^*	ξ_1^*	ξ_2^*	ε
AD	2.37	2.26	1.83	2.35	2.64	−2.61	−2.57	0.14
BD	1.50	2.64	2.71	2.74	0.68	−2.58	−0.47	0.06
CD	2.34	3.95	2.44	0.52	1.23	−2.34	−2.92	0.33
AE	2.50	2.24	1.50	2.21	2.50	−1.17	−1.35	0.16
BE	1.52	2.58	2.35	2.60	0.66	−0.97	1.21	0.06
CE	2.62	3.98	1.68	0.47	1.11	−0.12	−1.51	0.36

7.2. The Entropy Estimation of the pdfof Second-Order RDM-PQ Characteristics

Consider the following RDM-PQ:

$$v[k] = \sum_{n_1=0}^{2} w^{(1)}[n_1]x[k-n_1]+$$

$$+\sum_{n_1,n_2=0}^{2} w^{(2)}[n_1,n_2]x[k-n_1]x[k-n_2]+\xi[k],\;\; k\geq 2$$

with the random impulse response:

$$w^{(1)}[n_1] = w^{(2)}[n_1,n_2] = 0, \quad \text{if any of indexes } (n_1,n_2) > 2$$
$$w_2[n_1,n_2] = w_2[n_2,n_1]$$

Therefore, the dynamic RM-PQ in question incorporates nine parameters:

$$\begin{aligned} &a_0 = a_0^{(1)} = w^{(1)}[0],\ a_1 = a_1^{(1)} = w^{(1)}[1],\ a_2 = a_2^{(1)} = w^{(1)}[2] \\ &a_3 = a_0^{(2)} = w^{(2)}[0,0],\ a_4 = a_1^{(2)} = w^{(2)}[0,1] + w^{(2)}[1,0] \\ &a_5 = a_2^{(2)} = w^{(2)}[0,2] + w^{(2)}[2,0],\ a_6 = a_3^{(2)} = w^{(2)}[1,1] \\ &a_7 = a_4^{(2)} = w^{(2)}[1,2] + w^{(2)}[2,1],\ a_8 = a_5^{(2)} = w^{(2)}[2,2] \end{aligned}$$

The values of constants in Equation (8) are combined in Table 9.

Table 9. The values of $\beta_+^{(1,2)}, \beta^{(1,2)-}, \alpha_+^{(1,2)}$ and $\alpha_-^{(1,2)}$.

$\boldsymbol{\beta_+^{(1)}}$	$\boldsymbol{\beta_+^{(2)}}$	$\boldsymbol{\beta_-^{(1)}}$	$\boldsymbol{\beta_-^{(2)}}$	$\boldsymbol{\alpha_+^{(1)}}$	$\boldsymbol{\alpha_+^{(2)}}$	$\boldsymbol{\alpha_-^{(1)}}$	$\boldsymbol{\alpha_-^{(2)}}$
1.0	2.0	0.5	1.0	0.08	0.08	0.08	0.08

Table 10 demonstrates interval Equation (6) covering the components of the vector, **a**.

Table 10. The intervals for the parameters.

j	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
a_j^-	0.50	0.46	0.42	1.00	0.92	0.85	0.85	0.79	0.72
a_j^+	1.00	0.92	0.85	2.00	1.84	1.70	1.70	1.58	1.44

We have two measurements of the "input" and "output." Construct the blocks $X^{(1)}, X^{(2)}$ Equation (11):

$$X^{(1)} = \begin{pmatrix} x[2] & x[1] & x[0] \\ x[3] & x[2] & x[1] \end{pmatrix} = \begin{pmatrix} x_{00} & x_{01} & x_{02} \\ x_{10} & x_{11} & x_{12} \end{pmatrix}$$

$$X^{(2)} = \begin{pmatrix} x^2[2] & x[2]x[1] & x[2]x[0] & x^2[1] & x[1]x[0] & x^2[0] \\ x^2[3] & x[3]x[2] & x[3]x[1] & x^2[2] & x[2]x[1] & x^2[1] \end{pmatrix} =$$

$$= \begin{pmatrix} x_{00} & x_{01} & x_{02} & x_{03} & x_{04} & x_{05} \\ x_{10} & x_{11} & x_{12} & x_{13} & x_{14} & x_{15} \end{pmatrix}$$

The matrix $X = [X^{(1)}, X^{(2)}]$ has the form:

$$X = \begin{pmatrix} 3.9 & 1.9 & 2.8 & 0.9 & 1.6 & 5.2 & 3.6 & 1.9 & 4.2 \\ 9.3 & 3.9 & 1.9 & 3.8 & 8.5 & 4.9 & 0.9 & 1.6 & 2.6 \end{pmatrix}$$

The noise components become $\xi[0] = \xi_0 \in \Xi_0 = [-3,3]$; $\xi[1] = \xi_1 \in \Xi_1 = [-6,6]$. Additionally, the measured values of the "output" are given by $y[2] = y_0 = 18.51$, $y[3] = y_1 = 43.36$.

The $\mathcal{S}_{PQ}^1$-estimate of the pdf of the parameters, $p_j^*(a_j)$, $j \in [0,8]$, and the pdfs of the noise, $q_0^*(\xi_0), q_1^*(\xi_1)$, is defined by:

$$\begin{aligned} &p_j^*(a_j) = \exp[-1 - \gamma_j - \textstyle\sum_{l=0}^1 \theta_l x_{lj} a_j], \quad j \in [0,8] \\ &q_i^*(\xi_i) = \exp[-1 - \omega_i - \theta_i \xi_i], \quad i \in [0,1] \end{aligned}$$

The Lagrange multipliers, α_j, β_i and θ_l, meet the following equations:

$$\Gamma_j(\gamma,\theta) = \int_{a_j^-}^{a_j^+} \exp[-1-\gamma_j - \sum_{l=0}^{1}\theta_l x_{lj} a_j] da_j = 1, \quad j \in [0,8]$$
$$\Omega_i(\omega,\theta) = \int_{\xi_i^-}^{\xi_i^+} \exp[-1-\omega_i - \theta_i \xi_i] d\xi_i = 1, \quad i \in [0,1]$$

$$\Phi_i(\gamma,\omega,\theta) = \sum_{j=0}^{8} x_{ij} \int_{a_j^-}^{a_j^+} a_j \exp[-1-\gamma_j - \sum_{l=0}^{1}\theta_l x_{lj} a_j] da_j +$$
$$+ \int_{\xi_i^-}^{\xi_i^+} \xi_i \exp[-1-\omega_i - \theta_i \xi_i] d\xi_i = y_i, \; i \in [0,1], \quad j \in [0,8]$$

To solve these equations, we have used MATLAB for symbolic transformations and the numerical solution of nonlinear equations, known as the "trust-region dogleg" technique. The computed Lagrange multipliers form Tables 11–13.

Table 11. The Lagrange multipliers, γ_j.

j	**0**	**1**	**2**	**3**	**4**	**5**
γ_j	−10.0475	−6.8421	−13.7688	1.5448	11.2746	−37.2191

Table 12. The Lagrange multipliers, γ_j (continued).

j	**6**	**7**	**8**
γ_j	−32.9907	−15.0797	−29.7189

Table 13. The Lagrange multipliers, ω_i and θ_i.

i	**0**	**1**
ω_i	26.6458	14.7240
θ_i	9.9822	−2.7918

Figure 5 shows the curves of the probability density functions of the parameters, $p_j(a^{(1)_j}),\ j \in [0,2]$ (curves *0,1,2*) and $p_j(a^{(2)_j}),\ j \in [3,8]$ (curves *3–8*). Finally, Figure 6 presents the curves of the probability density functions of the noise, $q_0(\xi_0), q_1(\xi_1)$.

Figure 5. PDFs of the parameters.

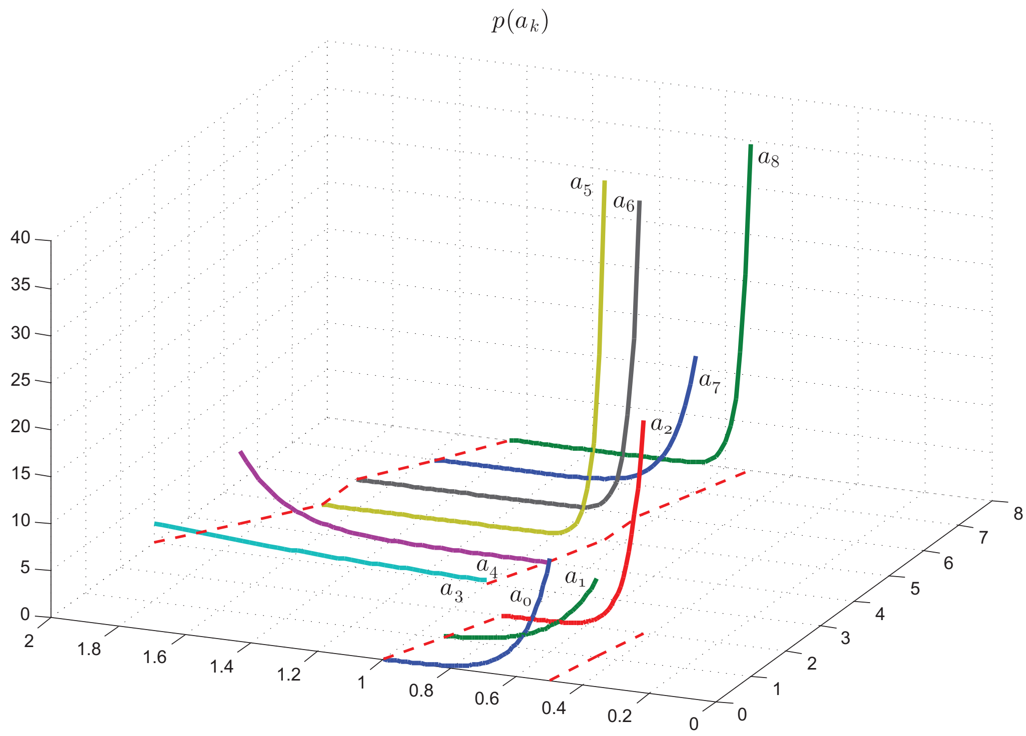
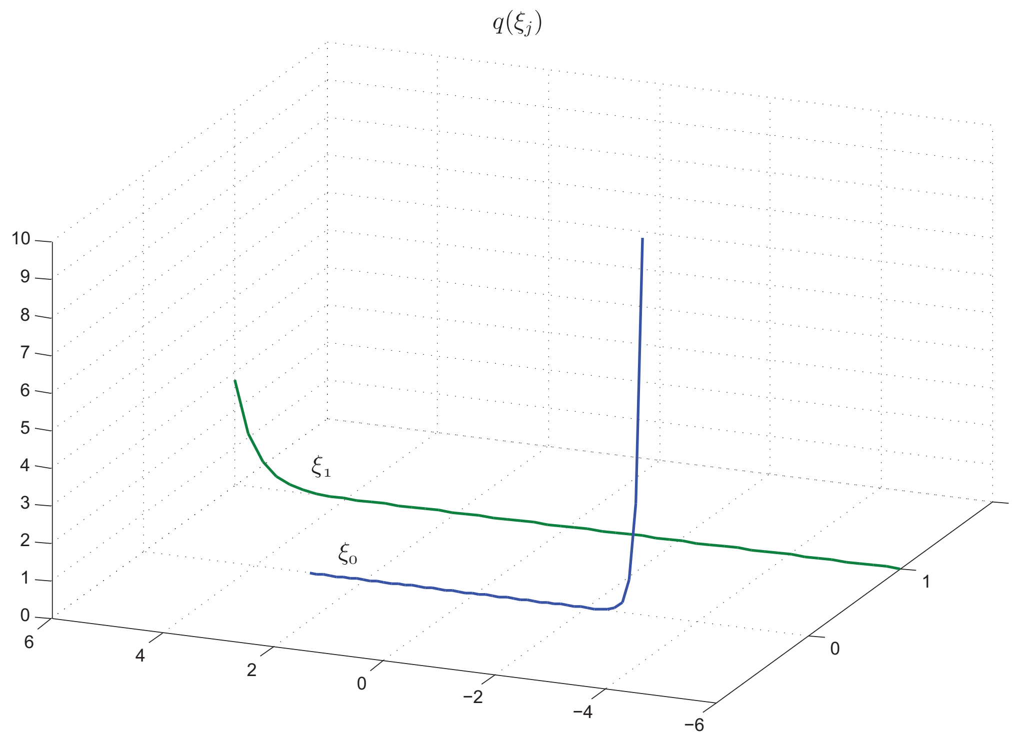


Figure 6. PDFs of the noise.



8. Conclusions

New methods of parametric (probabilities) and non-parametric (probability density functions) estimation of the randomized model characteristics are proposed. These methods are based on entropy functions or entropy functionals maximized under certain constraints. We can interpret the obtained estimate as a robust one, as the entropy function was used for its calculation. These methods are focused on the problems where data is limited and distorted by noises. It is shown that the entropy-robust estimation of the probabilities and probability density functions belong to the exponential class.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Cox, D.R.; Donnelly, C.A. *Principles of Applied Statistics*; Cambridge University Press: New York, NY, USA, 2011.
2. Kendall, M.G.; Stuart, A. *The Advanced Theory of Statistics: Inference and Relationship*; Griffin & Co.: London, UK, 1961; Volume 2.
3. Harrell, F.F., Jr. *Regression Modeling Strategies: With Application to Linear Models, Logistic Regrassion, and Survival Analysis; Springer Series*; Springer-Verlag Inc.: New York, NY, USA, 2001.
4. Cramer, H. *Mathematical Methods of Statistics (PMS-9)*; Vol. 9. Princeton University Press: Princeton, NJ, USA, 1999.
5. Golan, A.; Judge, G.; Miller, D. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*; John Wiley & Sons: New York, NY, USA, 1996.
6. Golan, A. Information and entropy econometrics—A review and synthesis. *Found. Trends Econometr.* **2006**, *2*, 1–145.
7. Racine, J.; Maasoumi, E. A versatile and robust metric entropy test of time-reversibility, and other hypotheses. *J. Econometr.* **2007**, *138*, 547–567.
8. Shiryaev, A.N. *Essentials of Stochastic Finance: Facts, Models, Theory*; World Scientific Publishing Co. Pte. Ltd.: River Edge, NJ, USA, 2000.
9. Del Ruiz, M.C. A new approach to measure volatility in energy markets. *Entropy* **2012**, *14*, 74–91.
10. Polyak, B.T. Robustness Analysis and Multilinear Perturbations. In *Robustness of Dynamic Systems with Parameter Uncertainties*; Mansur, M., Balemi, S., Trueol, W., Eds.; Birkhauser: Basel, Switzerland, 1992; pp. 93–104.
11. Lebiedz, D. Entropy-related extremum principles for model reduction of dissipative dynamical systems. *Entropy* **2010**, *12*, 706–719.
12. Gupta, M.; Srivastava, S. Parametricbayesian estimation of differential entropy and relative entropy. *Entropy* **2010**, *12*, 818–843.
13. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630.
14. Huber, P.J. *Robust Statistics*; John Willey & Sons: New York, NY, USA, 1984.

15. Tsypkin, Y.Z.; Popkov, Y.S. *Theory of Nonlinear Discrete Systems* (in Russian); "Nauka" Publisher: Moscow, Russia, 1973.
16. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.
17. Shannon, C. Communication theory of secrecy systems. *Bell Syst. Tech. J.* **1949**, *28*, 656–715.
18. Popkov, Y.S. *Macrosystems Theory and its Applications: Equilibrium Models, Lecture Notes in Control and Information Sciences*; Springer: London, UK, 1995; Volume 203.
19. Gelfand, I.M.; Fomin, S.V. *Calculus of Variations*; Courier Dover Publications, Inc.: New York, NY, USA, 2000.
20. Popkov, Y.S. New class of multiplicative algorithms for solving of entropy-linear programms. *Eur. J. Oper. Res.* **2006**, *174*, 1368–1379.