Auxiliary Problems of Statistical Data Processing: Interval Approach

V. P. Kuznetsov

***Abstract*— In [2], we have considered interval methods of solving the *basic* data statistical data processing problems, in which:**

- **the input statistical information consists of finitely many estimates of generalized moments (i.e., of finitely many inequalities of the type $E(f_i) \in \mathbf{a}_i$); and**
- **the desired output is an (interval) estimate for a given statistical characteristic $E(f)$.**

In many real-life situations, we have statistical problems of different type:

- **we may have an *additional statistical information*; and**
- **we may have an *objective* that is different from finding an estimate for a given statistical characteristics.**

In this paper, we describe how interval methods can be applied to these auxiliary statistical problems.

I. Sometimes, We Encounter Problems That Are Different From Basic Statistical Data Processing

Let us describe possible auxiliary problems (their solutions will be outlined in the forthcoming sections).

A. Problems in which we have additional statistical information

In the basic problems from [2], we consider the case when we have one or several variables x_i that are measured several times, and, as a result of these measurements, we get the interval values for finitely many characteristics. While doing it, we make two implicit assumptions:

- First, this information is usually formulated in terms of several variables x_i, and nothing is known a priori about the relationship between the different variables x_i.
- Second, that the actual values x_i of the physical quantities are not changing from one observations to another (or, to be more precise, that the changes in x_i are negligible). In mathematical terms, this means that x_i is *not* depending on time t.

In real life, we can have an additional information with respect to both possible dependencies:

- First, we can have some additional a priori information about the relationship between x_i.
- Second, it is possible that the values of x_i do change essentially with time.

B. Problems in which we have a different objective

As a basic problem, we considered the problem of estimating the value of a statistical characteristics (namely, of a generalized interval). Instead of that objective, we may have different ones:

- We may want to estimate the value of a more complicated statistical characteristics, or
- Instead of solving an *analysis* problem, we may want to solve a *synthesis* problem, i.e., use the given information to make a certain *decision*.

II. Case of Additional Statistical Information: Dependency Between x_i

A. Formulation of the problem

As we have already mentioned, in the description of a basic problem, we assumed that we only have the information of the type $E(f_i) \in \mathbf{a}_i = [a_i^-, a_i^+]$. This information is obtained by *generalizing* the empirical information:

- We observe that the average value of f_i (estimated from a finite sample) is in between a_i^- and a_i^+; and
- if this finite sample is sufficiently large to make statistical predictions correct, we can then conclude that, crudely speaking, a similar inequality is true for an *arbitrary* sample, i.e., in statistical terms, that $E(f_i) \in \mathbf{a}_i$.

In some cases, we can generalize this information even further: namely, we can, e.g., observe that whenever we estimate the characteristics $f_i(x_i)$ that

The author is with the Dept. of Prognosis, Institute of Complex Systems, Moscow-Murmansk, Russia. The author is thankful to V. Kreinovich for his help in writing and editing this text.

correspond to two different variables, then $E(f_i \cdot f_j) = E(f_i)E(f_j)$. In statistical terms, this means that all experimental results are consistent with the hypothesis that the variables x_i and x_j are *statistically independent.* We can therefore generalize this equality $E(f_i \cdot f_j) = E(f_i)E(f_j)$ (that is confirmed for finitely many f_i that have actually been tested) to *arbitrary* functions of x_i and x_j, i.e., in other words, we can *conclude* that the variables x_i are independent.

In these cases, in addition to the partial statistical information

$$(m_i, f_{i1}, ..., f_{im_i}, \mathbf{a}_{i1}, ..., \mathbf{a}_{m_i})$$

about each of the variables $x_i, 1 \le i \le n$, we know that these variables are *independent.* How to solve statistical data processing problem in this case?

B. Idea

Independence means that

$$E(f(x_i)f'(x_{i'})) = E(f(x_i))E(f'(x_{i'}))$$

whenever $i \neq i'$. Therefore, if we know that $E(f(x_i)) \in \mathbf{a}$, and that $E(f'(x_{i'})) \in \mathbf{a}'$, then we can conclude that

$$E(f(x_i)f'(x_{i'})) = E(f(x_i)E(f'(x_{i'})) \in \mathbf{a} \cdot \mathbf{a}'$$

(here, $\cdot$ indicates the product of the two intervals in the sense of interval computations). In view of the above idea, if we want to estimate the values of some characteristics $g : R^n \to R$, we can use the following statistical information about $\vec{x} = (x_1, ..., x_n)$:

- $E(f_{ij}(x_i)) \in \mathbf{a}_{ij}$ for all $i \le n$ and $j \le n_i$;
- $E(f_{i_1 j_1}(x_{i_1}) f_{i_2 j_2}(x_{i_2})) \in \mathbf{a}_{i_1 j_1} \cdot \mathbf{a}_{i_2 j_2}$;
- ...
- $E(f_{1j_1}(x_1)...f_{nj_n}(x_n)) \in \mathbf{a}_{1j_1} \cdot ... \cdot \mathbf{a}_{nj_n}$.

METHODOLOGY 3. (solving the basic problem of statistical data processing for the case when the error of input variables are independent)

Input:

- `an integer` n `(called the` *number of input variables*`);`
- `for every` $i = 1, ..., n$`, a partial statistical information` $(m_i, f_{i1}, ..., f_{im}, \mathbf{a}_{i1}, ..., \mathbf{a}_{im})$ `on a set` R `(`f_{ij} `are functions of one variable, and` $\mathbf{a}_{ij}$ `are intervals); and`
- `a function` $g : R^n \to R$.

Objective: `To estimate the interval` $\mathbf{a} = [a^-, a^+]$ `of possible values of` $E(g(x_1, ..., x_n))$ `for all distributions in which all` x_i `are independent, and each partial distribution is consistent with the given statistical information.`

Objective in precise mathematical terms: `To estimate the interval of posible values of the mathematical expectation` $E(g(x))$ `for all probability distributions for which all` x_i `are independent, and` $E(f_{ij}(x_i)) \in \mathbf{a}_{ij}$ `for all` i `and` j.

Methodology:

- `Formulate the basic statistical problem (in the sense of [2]) as follows: In this problem,` $X = R^n$`, and the partial statistical information will consists of the following parts:`
  - $E(f_{ij}(x_i)) \in \mathbf{a}_{ij}$ for all $i \le n$ and $j \le n_i$;
  - $E(f_{i_1 j_1}(x_{i_1}) f_{i_2 j_2}(x_{i_2})) \in \mathbf{a}_{i_1 j_1} \cdot \mathbf{a}_{i_2 j_2}$;
  - ...
  - $E(f_{1j_1}(x_1)...f_{nj_n}(x_n)) \in \mathbf{a}_{1j_1} \cdot ... \cdot \mathbf{a}_{nj_n}$.

  `Here, we use interval computations to compute the product of the intervals.`
- `Solve the resulting problem, and return its result as an estimate for the solution of the given problem.`

Comment 1. This method was proposed in Ch. 2 of [1].

Comment 2. This method uses interval computation techniques.

Comment 3. The resulting inequalities do not completely describe the fact that the variables x_i are independent. For example, if we have no conditions on x_i at all, we will thus get no conditions on R^n, but definitely not all probability distributions on R^n correspond to the independent case. We know of no other general method that would estimate the values of an arbitrary statistical characteristics under the condition of independence. The reason is that n independent probability distributions can be described by n probability measures (or density functions ρ_i). Statistical information about each variable x_i can be represented (as above) as a linear inequality in terms of ρ_i. However, the value of the desired characteristics $E(g(x_1, ..., x_n))$ is expressible by a *non-linear* formula $E(f) = \int g(x_1, ..., x_n)\rho(x_1, ..., x_n)\, dx_1...dx_n = \int g(x_1, ..., x_n)\rho_1(x_1) \cdot ... \cdot \rho_n(x_n)\, dx_1...dx_n$. To be more precise, this expression is linear in each of the variables $\rho_i(x_i)$, so, it is *multi-linear.* So, instead of a *linear programming problem,* we get a *multi-linear* optimization problem, that is more difficult to solve.

Comment 4. One important case of this problem is when we have partial information about several independent variables v_i (e.g., we know the intervals to which their moments belong), and we are interested in the characteristics of the sum $v_1 + \ldots + v_n$. This case is important because under some reasonable assumptions, when $n \to \infty$, the resulting distribution tends to Gaussian. The corresponding *central limit theorem* is used in practice to justify the fact that when the error consists of many small independent components, then the resulting error distribution is close to Gaussian. Analogues of central limit theorem for the case of different partial information are considered in Ch. 3 of [1].

III. Case of Additional Statistical Information: Dependency of the Quantities x_i on Time t

In the previous subsections, we considered the case when we are interested in the values of one or several physical quantities. In these case, the set X on which all our probability measures are defined coincides either with R, or with R^n for $n \geq 2$). In these situations, even if we repeat measurements several times, at different moments of time, we are still measuring the same physical quantities. In other words, in these cases, even if in reality the actual values of the measured characteristics change with time, we can safely neglect these changes.

In some real-life cases, however, the change of the values of the desired characteristics from one measurement to another is too large to be neglected. In these cases, the values that we are measuring can no longer be described by one or few values; the adequate description for these characteristics x_i is by *functions* $x_i(t)$ that describe how the values of these characteristics x_i change with time t.

This new *dynamic* feature leads to a possibility of a new type of statistical information. Indeed:

- In *static* case, the only possible type of information consisted of an inequality $E(f_i) \in \mathbf{a}_i$. This information is obtained by *generalizing* the empirical information:
  - We observe that the average value of f_i (estimated from a finite sample) is in between a_i^- and a_i^+; and
  - if this finite sample is sufficiently large to make statistical predictions correct, we can then conclude that, crudely speaking, a similar inequality is true for an *arbitrary* sample, i.e., in statistical terms, that $E(f_i) \in \mathbf{a}_i$.
- In *dynamic* case, we can similarly get estimates, e.g., for $E(x(t))$ for different values $t_1, \ldots, t_n$ of the time t. As a result, we conclude that $E(x(t_i)) \in \mathbf{a}_i = [a^-, a_i^+]$. If the resulting values $a_i^\pm$ are irregularly depending on i, then these n inequality are all the information that we can get. However, in many real-life cases, the bounds $a_i^\pm$ that correspond to the moment of time t_i change rather regularly with time t_i; this regularity means that we can express the dependency of $a_i^\pm$ on t_i by a simple formula: $a_i^\pm = f^\pm(t_i)$. Therefore, as a result of these estimates, we conclude that for the values t_i, $E(x(t_i)) \in [f^-(t_i), f^+(t_i)]$. Now, if this is true for sufficiently many moments of time t_i, then, similarly to the above-described generalization step, we can make a *second generalization*, and conclude that a similar inequality $E(x(t)) \in [f^-(t), f^+(t)]$ is true for all t.

As a result, we have a statistical information that consists of *continuum many* inequalities: in the above example, inequalities that correspond to all possible values of t.

Other examples of this type of statistical information include conditions on correlation $E(x(t)x(t'))$, on moments $E((x(t))^a)$, etc. Methods and results of handling this information are described in Ch. 4 of [1].

IV. Case of a Different Objective: Complicated Statistical Characteristics

A. Formulation of the problem

There are cases when we are interested in possible values of a statistical characteristics that is more complicated than a generalized moment $E(g)$. For example, suppose that we have some information about the general error distribution μ. Now, we have an auxiliary device that enables us to catch errors that exceed a certain number Δ_0. As a result of applying this device, we get only errors that are $\leq \Delta_0$. What are the statistical characteristics of the resulting "curtailed" error distribution? This new distribution can be characterized by the *conditional* mathematical expectations $E(g|B)$ under the condition B (in our case, B stands for $|v| \leq \Delta_0$). This condition mathematical expectation is defined as $E(g \cdot \chi_B)/E(\chi_B)$. How to find the interval of possible values of this fraction? In general, how to find the interval of possible values for the *fraction* $E(N)/E(D)$ of two generalized moments?

B. First approach: naive interval computations

One way to do solve this problem is to compute (or estimate) the intervals of possible values of $E(N)$ and $E(D)$, and then apply interval division to the resulting intervals.

METHODOLOGY 4. (computing conditional statistical characteristics)

Input:

- a set X;
- a partial statistical information $(m, f_1, ..., f_m, \mathbf{a}_1, ..., \mathbf{a}_m)$ on a set X; and
- a function $g : X \to R$:
- a set $B \subset X$.

Objective: To estimate the interval $\mathbf{a} = [a^-, a^+]$ of possible values of the conditional characteristic $E(g(x)|B)$ for all distributions that are consistent with the given statistical information.

Objective in precise mathematical terms: To estimate the interval of posible values of the conditional mathematical expectation $E(g(x)|B)$ for all probability distributions for which $E(f_i(x)) \in \mathbf{a}_i$ for $i = 1, ..., m$.

Methodology:

- Formulate and solve the following two auxiliary statistical problems:
  - For given partial statistical information $(m, f_1, ..., f_m, \mathbf{a}_1, ..., \mathbf{a}_m)$, estimate the interval of possible values of $E(g \cdot \chi_B)$. Denote the resulting interval by $\mathbf{N}$.
  - For given partial statistical information $(m, f_1, ..., f_m, \mathbf{a}_1, ..., \mathbf{a}_m)$, estimate the interval of possible values of $E(\chi_B)$ $(= p(B))$. Denote the resulting interval by $\mathbf{D}$.
- Divide $\mathbf{N}$ by $\mathbf{D}$ (using interval computations), and return the result of this division as the desired interval.

Comment. This idea is similar to naive interval computations, so no wonder that its result is often an overestimate and therefore, a new method is needed to get a better estimate.

C. More complicated techniques

A general methodology for computing

$$E(f \cdot \chi_B)/E(\chi_B)$$

is proposed in [1], Section 1.6. This methodology is based on the following idea:

- First, based on the initial statistical information, we find the interval $[p^-, p^+]$ of possible values of $p(B) = E(\chi_B)$.
- Now, for every $p \in [p^-, p^+]$, we can add the condition $p(B) = p$ (i.e., $E(\chi_B) \in [p, p]$) to the initial set of conditions and find the interval $[a^-(p), a^+(p)]$ of possible values of $E(g \cdot \chi_B)$.
- For every p, the ratio $E(f \cdot \chi_B)/E(\chi_B)$ can thus take any value from $a^-(p)/p$ to $a^+(p)/p$. Therefore, the desired interval can be described as $[a^-, a^+]$, where $a^- = \min(a^-(p)/p)$, $a^+ = \max(a^+(p)/p)$, and p runs from p^- to p^+.

METHODOLOGY 5. (computing conditional statistical characteristics)

Input:

- a set X;
- a partial statistical information $(m, f_1, ..., f_m, \mathbf{a}_1, ..., \mathbf{a}_m)$ on a set X; and
- a function $g : X \to R$:
- a set $B \subset X$.

Objective: To estimate the interval $\mathbf{a} = [a^-, a^+]$ of possible values of the conditional characteristic $E(g(x)|B)$ for all distributions that are consistent with the given statistical information.

Objective in precise mathematical terms: To estimate the interval of posible values of the conditional mathematical expectation $E(g(x)|B)$ for all probability distributions for which

$$E(f_i(x)) \in \mathbf{a}_i$$

for $i = 1, ..., m$.

Methodology:

- Formulate and solve the following auxiliary statistical problem:

  For given partial statistical information $(m, f_1, ..., f_m, \mathbf{a}_1, ..., \mathbf{a}_m)$, estimate the interval of possible values of $E(\chi_B)$ $(= p(B))$.

  Denote the resulting interval by $[p^-, p^+]$.
- For every p from the interval $[p^-, p^+]$, formulate and solve the following auxiliary statistical problem:

  For given statistical information $(m, f_1, ..., f_m, f_{m+1}, \mathbf{a}_1, ..., \mathbf{a}_m, \mathbf{a}_{m+1})$, where

  $$f_{m+1} = \chi_B, \ \ \mathbf{a}_{m+1} = [p, p],$$

  estimate the interval of possible values of $E(g \cdot \chi_B)$.

  Denote the resulting interval by $[a^-(p), a^+(p)]$.
- Estimate

  $$a^- = \min(a^-(p)/p), \ \ a^+ = \max(a^+(p)/p)$$

  (where p runs from p^- to p^+), and return $[a^-, a^+]$ as the desired interval.

This methodology is based on our ability to solve statistical data processing problems (with information of the type $E(f_i) \in \mathbf{a}_i$). Since there is no general algorithm to solve such problems, we cannot expect this method to always work. However, in some cases, it does lead to an algorithm [1], Section 1.6:

- Case when $g = \chi_A$, i.e., when we are interested in *conditional probabilities.*
- Case when $g(x) = |x|$, and $B = [a, b]$, i.e., when we know the average value of $|x|$, and we want to estimate a similar average in case we restrict ourselves only to values form the interval $[a, b]$ (for $[a, b] = [-\Delta_0, \Delta_0]$, this is exactly the case that was our original motivation for considering conditional statistical characteristics).

D. Case study: expert systems

The first case (when we are interested in conditional probabilities) is important for expert systems. One of the main parts of an expert system is its *knowledge base*, that contains experts' statements. One way to describe the experts' uncertainty in their own statements is to assign *probabilities.* Suppose now that we know the probability of a statement A, and that later on, some fact B turned out to be true. This additional information can change the experts' degrees of belief in their statements: instead of the original probability $p(A)$, we can now characterized the new degree of belief by a *conditional* probability $p(A|B)$.

The necessity to update expert systems leads to another statistical problem that is even more complicated than the problem of estimating the conditional probabilities: Namely,

- Conditional probabilities appear when we add to the knowledge base a new fact B that is known to be true.
- However, in many cases, the knowledge that we are adding to knowledge base also comes from an expert, and therefore, comes with a certain probability. In other words, we add B, but B is true only with a certain probability γ.

In [1], Section 1.6, the update problem is formulated and solved for this case as well.

E. Special case: how to estimate the actual value of the desired quantity based on the results of several measurement with random errors

Suppose that we measure the value x of the desired physical quantity in the presence of an additive noise γ. Then, after n measurements, we get n values $\tilde{x}^{(1)} = x+\gamma^{(1)}$, ..., $\tilde{x}^{(n)} = x+\gamma^{(n)}$, where $\gamma^{(i)}$ are the corresponding (unknown) noise values. It is usually assumed that the variables $\gamma^{(i)}$ that correspond to different measurements are independent.

The statistical characteristics of the noise are usually not known precisely; instead, we have a (partial) statistical information about the noise γ. Since the information about γ is statistical, we cannot uniquely determine x: instead, we can only have *statistical* information about the actual value of x. In order to describe this information, we must express the information contained in the measurement results $\tilde{x}^{(i)}$ in terms of the random variables $\gamma^{(i)}$.

We have a description ($\tilde{x}^{(i)} = x + \gamma^{(i)}$), but this description contains an additional (unwelcome) variable x. So, to get the desired description, we must reformulate it without x. This can be done as follows:

PROPOSITION 1. *Let n numbers $\tilde{x}^{(1)}, ..., \tilde{x}_n$ be given. Then, for an arbitrary tuple $(\gamma^{(1)}, ..., \gamma^{(n)})$, the following two conditions are equivalent to each other:*

- *there exists a real number x such that for every i, $x^{(i)} = x + \gamma^{(i)}$;*
- *for every i and j, $\tilde{x}^{(i)} - \gamma^{(i)} = x^{(j)} - \gamma^{(j)}$.*

Due to this proposition, the mathematical expectation of x can be described as a conditional mathematical expectation

$$\bar{x} = E(x^{(1)} - \gamma^{(1)} \,|\, x^{(i)} - \gamma^{(i)} = x^{(j)} - \gamma^{(j)})$$

under the condition that $x^{(i)} - \gamma^{(i)} = x^{(j)} - \gamma^{(j)}$ for all i and j; the standard deviation of x can be described as $\sqrt{E((x^{(1)} - \gamma^{(1)} - \bar{x})^2 \,|\, ...)}$ under the same conditions, etc.

METHODOLOGY 6. (estimating the actual value of the desired quantity based on the results of several measurements with random aditive errors).

Input:

- `an integer` n `(called the` *number of measurements*`);`
- n `real numbers` $\tilde{x}^{(1)}, ..., \tilde{x}^{(n)}$ `(called` *measurement results*`);`
- `for every` $i = 1, ..., n$, `a partial statistical information` $(m_i, f_{i1}, ..., f_{im}, \mathbf{a}_{i1}, ..., \mathbf{a}_{im})$ `on a set` R `(`f_{ij} `are functions of one variable, and` $\mathbf{a}_{ij}$ `are intervals);`
- `a function` $g : R \to R$.

Objective: `To estimate the interval` $\mathbf{a} = [a^-, a^+]$ `of possible values of` $E(g(x))$ `for all distributions` $\gamma^{(i)}$ `in which all` $\gamma^{(i)}$ `are independent,` $\tilde{x}^{(i)} = x + \gamma^{(i)}$, `and each partial distribution is consistent with the given statistical information.`

Objective in precise mathematical terms: To estimate the interval of possible values of the mathematical expectation $E(g(x))$ for all distributions $\gamma^{(i)}$ in which all $\gamma^{(i)}$ are independent, $\tilde{x}^{(i)} = x + \gamma^{(i)}$, and $E(f_{ij}(\gamma^{(i)})) \in \mathbf{a}_{ij}$ for all i and j.

Methodology:

- Formulate the statistical problem (in the sense of [2]) as follows: In this problem, $X = R^n$, and the partial statistical information will consists of the following parts:
  - $E(f_{ij}(\gamma^{(i)})) \in \mathbf{a}_{ij}$ for all $i \le n$ and $j \le n_i$;
  - for all i and j,
  $$\gamma^{(i)} - \gamma^{(j)} = \tilde{x}^{(j)} - \tilde{x}^{(i)};$$
  - all variables $\gamma^{(i)}$ are independent;

  and the goal is to estimate $E(g(x^{(1)} - \gamma^{(1)}))$.
- Solve the formulated problem, and return the resulting estimate as the desired solution.

In [1], Ch. 6, different cases of this problem are considered (e.g., estimating x when we know several moments). Also, in that chapter, several more complicated noises are considered, including the multiplicative noise $\tilde{x}^{(i)} = x \cdot \gamma^{(i)}$. In this case, the possibility to describe the measurement results in terms of the noise values is explained in the following proposition:

PROPOSITION 2. *Let n numbers $\tilde{x}^{(1)}, ..., \tilde{x}_n$ be given. Then, for an arbitrary tuple $(\gamma^{(1)}, ..., \gamma^{(n)})$, the following two conditions are equivalent to each other:*

- *there exists a real number x such that for every i, $x^{(i)} = x \cdot \gamma^{(i)}$;*
- *for every i and j, $\tilde{x}^{(i)}/\gamma^{(i)} = x^{(j)}/\gamma^{(j)}$.*

Due to this proposition, e.g., the mathematical expectation of x is equal to the conditional mathematical expectation of $\tilde{x}^{(i)}/\gamma^{(i)}$ under the condition that $\tilde{x}^{(i)}/\gamma^{(i)} = \tilde{x}^{(j)}/\gamma^{(j)}$, etc.

METHODOLOGY 7. (estimating the actual value of the desired quantity based on the results of several measurements with random multiplicative errors).

Input:

- an integer n (called the *number of measurements*);
- n real numbers $\tilde{x}^{(1)}, ..., \tilde{x}^{(n)}$ (called *measurement results*);
- for every $i = 1, ..., n$, a partial statistical information $(m_i, f_{i1}, ..., f_{im}, \mathbf{a}_{i1}, ..., \mathbf{a}_{im})$ on a set R (f_{ij} are functions of one variable, and $\mathbf{a}_{ij}$ are intervals);
- a function $g : R \to R$.

Objective: To estimate the interval $\mathbf{a} = [a^-, a^+]$ of possible values of $E(g(x))$ for all distributions $\gamma^{(i)}$ in which all $\gamma^{(i)}$ are independent, $\tilde{x}_i = x \cdot \gamma^{(i)}$, and each partial distribution is consistent with the given statistical information.

Objective in precise mathematical terms: To estimate the interval of posible values of the mathematical expectation $E(g(x))$ for all distributions $\gamma^{(i)}$ in which all $\gamma^{(i)}$ are independent, $\tilde{x}^{(i)} = x \cdot \gamma^{(i)}$, and $E(f_{ij}(\gamma^{(i)})) \in \mathbf{a}_{ij}$ for all i and j.

Methodology:

- Formulate the statistical problem (in the sense of [2]) as follows: In this problem, $X = R^n$, and the partial statistical information will consists of the following parts:
  - $E(f_{ij}(\gamma^{(i)})) \in \mathbf{a}_{ij}$ for all $i \le n$ and $j \le n_i$;
  - for all i and j, $\gamma^{(i)}/\gamma^{(j)} = \tilde{x}^{(j)}/\tilde{x}^{(i)}$;
  - all variables $\gamma^{(i)}$ are independent;

  and the goal is to estimate $E(g(x^{(1)}/\gamma^{(1)}))$.
- Solve the formulated problem, and return the resulting estimate as the desired solution.

Comment. In all these cases, for each characteristic $E(g(x))$ of x, we get an *interval* of possible values.

V. Case of a Different Objective: Decision Making

A. *First case of decision making: we know the objective function, and we know the exact values of all the parameters of the system*

Usually, we know what objective we want to achieve, i.e., what numerical criterion J we want to maximize, and how the value of this criterion depends on our decision u and on the parameters $\vec{x} = (x_1, ..., x_n)$ that describe the decision-making situation. In other words, we know the function $J(u, x_1, ..., x_n)$.

In the rare cases when we know the exact values of all the parameters $\vec{x} = (x_1, ..., x_n)$, the choice of a decision u can be described as an optimization problem $J(u, \vec{x}) \to \max$.

B. Second case of decision making: we know the objective function, but we do not know the exact values of the parameters of the system

In real-life situations, however, we do not know the values x_i precisely. Instead, we have a (partial) statistical information about the values of x_i. In many such cases, a reasonable choice is to find a decision u that maximizes the *expected* value of J, i.e., for which $E(J(u,\vec{x})) \to \max$. This choice makes perfect sense if we face a series of repeating similar optimization problems (e.g., in business or manufacturing), then, we want to choose solutions that are in the average most profitable.

Since we only have *partial* statistical information about $\vec{x}$, choosing a decision u does not determine the mathematical expectation $E(J(u,\vec{x}))$ uniquely. Instead, we have an *interval* $[J^-(u), J^+(u)]$ of possible values of $E(J)$. Therefore, we face a problem of choosing a decision when only the *interval* containing an objective function is known. In Ch. 5 of [1], interval decision-making techniques (including pessimism-optimism) are applied to the above-described statistical decision making problem.

Decision is usually characterized by one of several numerical parameters: e.g., how many items of a certain sort to produce, how many to transport, etc. In mathematical terms, the set of all possible decisions is usually *finite-dimensional.* However, there are special situations when the decision is either simpler or more complicated than simply choosing a few parameters. In [1], two such situations are considered:

- *Hypothesis testing*, when we have a *finite* list of hypotheses, and a decision consists of choosing one of them [1], Ch. 7). Here, the set of all possible decisions is finite and therefore, 0–dimensional.
- *Signal reconstruction*, when we try to reconstruct the original signal from the noisy one that we have measured. In this case, a decision consists of choosing a reconstructed signal $x(t)$. Here, the set of decisions coincides with the set of all functions and is, therefore, *infinite-dimensional.* Corresponding algorithms are described in Ch. 5 of [1].

VI. Proof of Proposition 1.

If $x^{(i)} = x - \gamma^{(i)}$, then, subtracting i–th equality from the j–th, we get the second condition; vice versa, if the second condition is true, then we can take this common difference as x. Q.E.D.

References

[1] V. P. Kuznetsov, *Interval statistical models*, Moscow, Radio i Svyaz Publ., 1991 (in Russian).

[2] V. P. Kuznetsov, "Interval Methods For Processing Statistical Characteristics", *These Proceedings.*