Soft Probability of Large Deviations

D.A. Molodtsov
Dorodnitsyn Computing Center, Russian Academy of Sciences, Moscow, Russia

Abstract
A concept of soft probability is presented. An analogue of Chebyshev's inequality for soft probability is proved. Soft large deviation probabilities for a nonnegative random variable under a mean hypothesis is calculated.
Keywords soft probability, hypothesis replacement, soft Chebyshev inequality, large deviations.

1 What is Soft Probability?

In modern textbooks of probability theory, the exposition usually begins with a discussion of the subject matter of this science. All phenomena are divided into three types. Phenomena of the first type are those characterized by deterministic regularity; this means that a given set of circumstances always leads to the same outcome.

Phenomena of the second and the third type are not deterministic regular. The second type consists of statistically regular phenomena, and the third type, of the remaining phenomena.

By statistical regularity the statistical stability of outcome frequencies is usually understood. As a rule, this notion is associated with the example of coin tossing. Some authors outline more constructive ways of interpretation [1], but a final formalization of statistical stability has never been proposed, and its verification is always left to the reader's judgment and intuition.

Thus, the most important question of the applicability of the theory to real phenomena remains essentially unanswered.

Soft probability [2-7] is merely the logical completion of the construction of statistical regularity, whose approximate description is usually contained in probability theory textbooks.

Our approach is based on examine the conclusions to which the logical development of the notion of statistical regularity leads.

First, we mention at once that the verification of statistical regularity suggested here requires a certain set of trial outcomes, that is, a statistical database. If there are no trial results, then there is no object of examination.

2 Statistical Database

First, we introduce the base outcome space Ω. Each trial is associated with an element of the set Ω, namely, the outcome of this trial. A statistical database is merely a finite sequence of outcomes: $Base = \{\omega_1, ..., \omega_n\}, \omega_i \in \Omega$.

By an event A we mean a subset of $\Omega : A \subseteq \Omega$. We say that an event occurs under a certain trial if the outcome of this trial belongs to the given event. We have to define the statistical regularity of the occurrence of an event A. In effect, statistical regularity means the closeness of frequencies for a given set of samples. Thus, in formalizing this concept, we must begin with specifying a set of samples for a statistical database.

We identify a sample I with the positions in the sequence $Base = \{\omega_1, ..., \omega_n\}$ occupied by its elements; in other words, a sample is a subset $I \subseteq Ind(Base)$ of the set $Ind(Base) = \{1, ..., n\}$.

In specifying a set of samples, it is natural to first constrain their size; thus, we introduce a parameter determining the size of a sample. As admissible samples we consider any samples of size m which consists of consecutive elements of the set $Ind(Base)$ and are not too "old". We denote the set of such samples by $S(Base, m, \tau)$:

$$S(Base, m, \tau) = \{(i, i+1, ..., i+m-1), i = \tau, ..., N-m+1\}.$$

Apparently, the set $S(Base, m, \tau)$ is a minimal set of samples which are natural to consider in defining the notion of statistical regularity. Of course, other definitions of admissible samples are also possible, which lead to different definitions of statistical regularity; it is important that the set of admissible samples be precisely specified.

We define the frequency of occurrence of an event A in a sample I as

$$\mu(Base, \chi(A, \cdot), I) = \frac{1}{|I|} \sum_{i \in I} \chi(A, \omega_i).$$

Here $|I|$ denotes the cardinality of the set I and $\chi(A, \omega) = \{^{1, \omega \in A}_{0, \omega \notin A}$.

3 Statistical Regularity

Now, it is natural to understand the statistical regularity of the occurrence of an event A as the closeness of the frequencies of occurrence of A in any admissible samples from the set $S(Base, m, \tau)$. Below we give a more formal definition of this notion.

Definition 1. An event A is said to be *statistically* (m, τ, δ) $-$ *regular on a database Base* if

$$|\mu(Base, \chi(A, \cdot), I) - \mu(Base, \chi(A, \cdot), J)| \leq \delta.$$

for any samples $I, J \in S(Base, m, \tau)$.

We see that the definition of statistical regularity involves several parameters. Apparently, it is for this reason that this notion has not been used, because the dependence of statistical regularity on parameters inevitably makes probability dependent on parameters as well, while probability is traditionally thought of as a single real number in the unit interval. However, the parameterization of probability only means the detailing of the description of the situation under consideration. In different situations, probabilities corresponding to different parameters are important. The description using the same number as probability in all situations is coarser than that using parameterized probability.

Parameterized families are used very extensively in theory and practice. Thus, different methods of measuring physical quantities yield different results, which naturally leads to parameterized families. A striking example of such a family is the description of mine deposits in geology.

In computational mathematics, it often happens that only approximate solutions can be found numerically; such solutions form parameterized families too [2,8].

For dealing with such objects, the notion of a soft set was introduced and the theory of soft sets was developed, which has found numerous applications in various areas of mathematics [2-7, 9-24]. For this reason, the alternative to classical probability considered in this paper is called soft probability.

Let us contemplate Definition 1. The statistical database consists of the outcomes of events which have already occurred, while the main purpose of the theory is to produce informative statements concerning future events. Thus, the only interesting aspect of the notion of statistical regularity is its use as a hypothesis on the future behavior of trial outcomes.

Accordingly, the application of soft probability is divided into two processes:

- Verifying hypotheses at each step.
- Given a set of accepted hypotheses, constructing estimates, predictions, etc.

Note that, in classical probability theory, the former process is virtually absent; the statistical regularity hypothesis is accepted only once, before launching the apparatus of probability theory, after which this hypothesis is neither controlled nor verified anew. Moreover, in classical probability theory, the statistical regularity hypothesis is accepted for all possible events together. It remains unclear how to handle situations in which some of the events are statistically regular and the other events are not.

The definition of statistical regularity is easy to generalize to a random function.

By a random function f we mean any real-valued function defined on the set Ω, that is, any function $f : \Omega \to E$, where E is the set of real numbers. An example of such a function is the characteristic function $\chi(A, \cdot)$ of a set.

The mean value of a random function f on a sample I is defined as

$$\mu(Base, f, I) = \frac{1}{|I|} \sum_{i \in I} f(\omega_i).$$

Definition 2. A random function f is said to be *statistically* (m, τ, δ) $-$ *regular on a database* $Base$ if

$$|\mu(Base, f, I) - \mu(Base, f, J)| \leq \delta.$$

for any samples $I, J \in S(Base, m, \tau)$.

Thus, the statistical regularity of an event means simply that the characteristic function of this event is statistically regular.

The condition that a random function is statistically regular can be written in a different equivalent form as follows. We set

$$2a = \max_{I \in S(Base,m,\tau)} \mu(Base, f, I) + \min_{I \in S(Base,m,\tau)} \mu(Base, f, I),$$

and

$$2b = \max_{I \in S(Base,m,\tau)} \mu(Base, f, I) - \min_{I \in S(Base,m,\tau)} \mu(Base, f, I).$$

It is easy to see that the statistical regularity of f is equivalent to the inequality $2b \leq \delta$.

For any $I \in S(Base, m, \tau)$, we have $|\mu(Base, f, I) - a| \leq b$. This readily implies the equivalence of the statistical regularity of f to the fulfillment of the inequality $|\mu(Base, f, I) - a| \leq \delta/2$, or the inclusion $\mu(Base, f, I) \in [a - \delta/2, a + \delta/2]$, for any $I \in S(Base, m, \tau)$. The inclusion is also equivalent to

$$[\min_{I \in S(Base,m,\tau)} \mu(Base, f, I), \max_{I \in S(Base,m,\tau)} \mu(Base, f, I)] \subseteq [a - \delta/2, a + \delta/2].$$

This suggests the following natural definition.

Definition 3. The interval

$$\lambda(f, Base, m, \tau) = [\min_{I \in S(Base,m,\tau)} \mu(Base, f, I), \max_{I \in S(Base,m,\tau)} \mu(Base, f, I)]$$

is called the $(m, \tau)-$ *approximate mean value* of the random function f on the database $Base$.

We denote the left and right endpoints of this interval by an underscore and an overscore, respectively:

$$\lambda(f, Base, m, \tau) = [\underline{\lambda}(f, Base, m, \tau), \overline{\lambda}(f, Base, m, \tau)].$$

4 Hypotheses on the Behavior of a Random Function

On the basis of the notion of the statistical regularity of a random function, we can formulate hypotheses of two types on the future behavior of the values of a random function. In fact, these are hypotheses on the future values of the statistical database.

Definition 4. A database $Base$ is said to be *statistically* $(m, \delta) - regular\ with$ respect to a random function f if

$$|\mu(Base, f, I) - \mu(Base, f, J)| \leq \delta.$$

for any samples $I, J \in S(Base, m, 1)$.

Definition 5. A database $Base$ is said to be $statistically\ significantly$ $(m, \delta, a) - regular$ with respect to a random function f if

$$|\mu(Base, f, I) - a| \leq \delta.$$

for any samples $I \in S(Base, m, 1)$.

The difference between Definitions 4 and 5 is in that Definition 4 supposes only the closeness of mean values on any admissible samples, while Definition 5 specifies the number to which these mean values must be close with a given accuracy. In other words, the mean values must belong to the interval $[a - \delta.a + \delta]$, or the approximate mean values of the random function under consideration must belong to $[a - \delta.a + \delta]$.

These definitions can also be regarded as hypotheses on certain properties of the approximate means of a random function. Definition 4 specifies only the length of an interval containing the approximate mean, and Definition 5 specifies the boundaries of the entire range of this mean.

An analysis shows that these hypotheses are often insufficient for obtaining instructive results. In addition to hypotheses on approximate mathematical expectation, hypotheses describing the deviation of the random function under consideration from its mathematical expectation (that is, hypotheses similar to that of the existence of variance) are very useful. Such hypotheses can be formulated in various forms. We give only one version.

Together with a random function f, consider the random function equal to the absolute value of the deviation of f from the interval $[a - \delta.a + \delta]$, that is, defined by

$$g(\omega) = max\{|f(\omega) - a| - \delta, 0\}.$$

Definition 6. We say that a database $Base$ satisfies the $(m, \delta, a, \triangle) - variance\ hypothesis$ with respect to a random function f if $Base$ is statistically significantly $(m, \triangle, 0) - regular$ with respect to the random function g, i.e.

$$\mu(Base, g, I) \leq \triangle.$$

for any samples $I \in S(Base, m, 1)$.

The approximate mathematical expectation and approximate variance hypotheses describe a variable oscillating about a certain interval. It is also natural to consider other hypotheses, which describe the growth, decline, periodicity, and other properties of random variables [6-7].

After classes of hypotheses are chosen, dealing with hypotheses is a dynamical step process. At each step, new realization from the base space (trial outcomes) appear. Thus, at each step, it is required to form a list of accepted hypotheses and perform calculations on the basis of this list.

5 Properties of the Approximate Mean Value of a Random Function

Properties of an approximate mean are similar to those of mathematical expectation. Inequalities involving intervals are assumed to hold componentwise, that is, at each endpoint of the interval.

1). $\lambda(c, Base, m, \tau) = [c, c], c \in E$.
2). if $f(\omega) \geq g(\omega)$ for any $\omega \in \Omega$, then $\lambda(f, Base, m, \tau) \geq \lambda(g, Base, m, \tau)$.
3). $\lambda(cf, Base, m, \tau) = c\lambda(f, Base, m, \tau), c \in E, c \geq 0$.
4). $\lambda(f + c, Base, m, \tau) = \lambda(f, Base, m, \tau) + c, c \in E$.
5). $\lambda(-f, Base, m, \tau) = -\lambda(f, Base, m, \tau) = [-\overline{\lambda}(f, Base, m, \tau), -\underline{\lambda}(f, Base, m, \tau)]$.
6). $\lambda(f + g, Base, m, \tau) \subseteq \lambda(f, Base, m, \tau) + \lambda(g, Base, m, \tau)$.
7). $\lambda(f, Base, m, \tau) \subseteq \lambda(f, Base, m, t), \tau \geq t$.

6 An Approximate Variance of a Random Variable

In classical probability theory, the variance of a random variable is defined as the mathematical expectation of the squared difference between this variable and its expectation. The mathematical expectation of a random function is merely a number.

In the case under consideration, the approximate mean is a parametric family of intervals which characterizes the random variable on a certain database. Hypotheses involving the notion of statistically significant regularity impose constraints on approximate means. The intervals describing the constraints are not required to equal the corresponding approximate means. Thus, it is convenient to define approximate variance as the measure of deviation of a random variable from a certain interval rather than from the corresponding approximate mean.

Definition 7. The *approximate* $(m, \tau, a, \delta) -$ *variance* of a random function f on a database $Base$ is the $(m, \tau) -$ *approximate* mean of the random function $max\{|f(\omega) - a| - \delta, 0\}$, that is, the interval

$$D(f, Base, m, \tau, a, \delta) = \lambda(max\{|f(\cdot) - a| - \delta, 0\}, Base, m, \tau).$$

The simplest properties of approximate variance are as follows.

1). $D(f, Base, m, \tau, a, \delta) \geq 0$.

2). $D(cf, Base, m, \tau, ca, c\delta) = cD(f, Base, m, \tau, a, \delta), c \in E, c \geq 0$.

3). $D(-f, Base, m, \tau, -a, \delta) = D(f, Base, m, \tau, a, \delta)$.

4). $\overline{D}(f+g, Base, m, \tau, a+b, \delta+\gamma) \leq \overline{D}(f, Base, m, \tau, a, \delta)+\overline{D}(g, Base, m, \tau, b, \gamma)$.

5). $\underline{D}(f+g, Base, m, \tau, a+b, \delta+\gamma) \leq \underline{D}(f, Base, m, \tau, a, \delta)+\overline{D}(g, Base, m, \tau, b, \gamma)$.

6). $D(f, Base, m, \tau, a, \delta) \geq D(f, Base, m, \tau, a, \gamma), \delta \leq \gamma$.

7). $D(f, Base, m, \tau, a, \delta) \supseteq D(f, Base, m, \gamma, a, \delta), \tau \leq \gamma$.

7 Chebyshev's Inequality for Soft Probability

Statement 1 (Chebyshev's inequality). If a random function f is nonnegative everywhere on Ω and $\varepsilon > 0$, then

$$\lambda(\chi(\{\omega' \in \Omega | f(\omega') \geq \varepsilon\}, \cdot), Base, m, \tau) \leq \frac{1}{\varepsilon}\lambda(f, Base, m, \tau).$$

Proof. If a random function f is nonnegative everywhere on Ω, then, as is easy to see, we have

$$f(\omega) \geq \varepsilon\chi(\{\omega' \in \Omega | f(\omega') \geq \varepsilon\}, \omega)$$

for any $\varepsilon > 0$ and any $\omega \in \Omega$. Property 2 of approximate mean implies

$$\lambda(f, Base, m, \tau) \geq \varepsilon\lambda(\chi(\{\omega' \in \Omega | f(\omega') \geq \varepsilon\}, \omega), Base, m, \tau).$$

The approximate means of the characteristic function of a set equals the soft probability of this set; therefore, the soft probability of the set $\{\omega' \in \Omega | f(\omega') \geq \varepsilon\}$ is estimated as

$$\lambda(\chi(\{\omega' \in \Omega | f(\omega') \geq \varepsilon\}, \cdot), Base, m, \tau) \leq \frac{1}{\varepsilon}\lambda(f, Base, m, \tau)$$

(recall that the inequality is interval). This completes the proof of the statement.

Now, let f be an arbitrary random function. For the function $|f|$, we have

$$\lambda(\chi(\{\omega' \in \Omega |\ |f(\omega')| \geq \varepsilon\}, \cdot), Base, m, \tau) \leq \frac{1}{\varepsilon}\lambda(|f|, Base, m, \tau).$$

Since the inequality $|f(\omega')| \geq \varepsilon$ is equivalent to $|f^s(\omega')| \geq \varepsilon^s$, where $s > 0$, it follows that

$$\lambda(\chi(\{\omega' \in \Omega |\ |f(\omega')| \geq \varepsilon\}, \cdot), Base, m, \tau) \leq \frac{1}{\varepsilon^s}\lambda(|f|^s, Base, m, \tau).$$

therefore,

$$\lambda(\chi(\{\omega' \in \Omega |\ |f(\omega')| \geq \varepsilon\}, \cdot), Base, m, \tau) \leq \inf_{s>0} \frac{1}{\varepsilon^s}\lambda(|f|^s, Base, m, \tau).$$

Now, consider the nonnegative random function $g(\omega) = max\{|f(\omega) - a| - \delta, 0\}$, where f is an arbitrary random function. Applying Chebyshev's inequality to g, we obtain

$$\lambda(\chi(\{\omega' \in \Omega |\, max\{|f(\omega') - a| - \delta, 0\} \geq \varepsilon\}, \cdot), Base, m, \tau) \leq \frac{1}{\varepsilon} D(f, Base, m, \tau, a, \delta).$$

Elementary transformations yield

$$\lambda(\chi(\{\omega' \in \Omega || f(\omega') - a| \geq \delta + \varepsilon\}, \cdot), Base, m, \tau) \leq \frac{1}{\varepsilon} D(f, Base, m, \tau, a, \delta).$$

This allows us to write the inequality in the equivalent form

$$\lambda(\chi(\{\omega' \in \Omega || f(\omega') - a| \geq \delta\}, \cdot), Base, m, \tau) \leq \inf_{0<\varepsilon<\delta} \frac{1}{\varepsilon} D(f, Base, m, \tau, a, \delta - \varepsilon).$$

Thus, Chebyshev's inequality gives an estimate of the soft probability of "large" deviations in terms of approximate mean or approximate variance. If the database is known, then this information is of little value, because it is easy to directly calculate the exact values of any probabilistic characteristics, including the probabilities of large deviations. Apparently, it is of more interest to apply this inequality to estimating the probability of a random function on a future database. Naturally, this requires hypotheses on the approximate mean or the approximate variance of the function under consideration. However, in the presence of hypotheses, of interest are sharp bounds for the probability of large deviations.

8 Soft Probability of Large Deviations for a Nonnegative Random Function Under an Approximate Mean Hypothesis

Suppose that a database $Base$ is statistically significantly $(m, \delta, a) - regular$ with respect to a nonnegative random function f, i.e., given any sample $I \in S(Base, m, 1)$, we have

$$|\mu(Base, f, I) - a| \leq \delta$$

We assume that $a \geq 0$.

We define a large deviation as the event $A(f, \varepsilon) = \{\omega \in \Omega | f(\omega) \geq \varepsilon\}$, where $\varepsilon > 0$. We are interested in the range of values of the soft probability of this event under the above hypothesis. Clearly, the solution essentially depends on the range of the function f. Consider the case where $f(\Omega) = E_+ = \{x \in E | x \geq 0\}$.

Let $f(\omega_i) = x_i \in E_+$. Then the constraints on the database can be written as constraints on the vector $x = \{x_1, ...x_n\} \in X(n, m, a, \delta)$, where

$$X(n, m, a, \delta) = \{x \in E_+^n | \, |\sum_{i=j}^{j+m-1} x_i - am| \leq \delta m, j = 1, ..., n - m + 1\}.$$

The boundaries of the range of the soft probability are

$$u^*(n,m,a,k,\tau,\delta,\varepsilon) = \sup_{x\in X(n,m,a,\delta)} \max_{\tau\le j\le n-k+1} \frac{1}{k}\sum_{i=j}^{j+k-1} \chi(\{y|y\ge\varepsilon\},x_i)\}.$$

and

$$u_*(n,m,a,k,\tau,\delta,\varepsilon) = \inf_{x\in X(n,m,a,\delta)} \min_{\tau\le j\le n-k+1} \frac{1}{k}\sum_{i=j}^{j+k-1} \chi(\{y|y\ge\varepsilon\},x_i)\}.$$

Note that, at $n = m$, the set $X(n, m, a, \delta)$ takes the form

$$X(m,m,a,\delta) = \{x\in E_+^m|\ |\sum_{i=1}^{m} x_i - am| \le \delta m\}.$$

In the case $k \le m$, the evaluation of u^* and u_* is based on the following assertion.

Statement 2. If $x \in E_+^m$ and $(x_j, x_{j+1}, ..., x_{j+m-1}) \in X(m, m, a, \delta)$, then there exists a vector $y \in X(n, m, a, \delta)$ such that $y_i = x_i$ for $i = j, j+1, ..., j+m-1$.

Proof. Let $y_i = x_{j+(i-j) mod m}$ for $i = 1, ..., n$. Then $\sum_{i=l}^{l+m-1} y_i = \sum_{i=j}^{j+m-1} x_i$ for any $l = 1, ..., n-m+1$. Therefore, $y \in X(n, m, a, \delta)$. This proves the required assertion.

The u^* and u_* problems can be formulated as

$$u^*(n,m,a,k,\tau,\delta,\varepsilon) = \max_{\tau\le j\le n-k+1} \sup_{x\in X(n,m,a,\delta)} \frac{1}{k}\sum_{i=j}^{j+k-1} \chi(\{y|y\ge\varepsilon\},x_i)\}.$$

and

$$u_*(n,m,a,k,\tau,\delta,\varepsilon) = \min_{\tau\le j\le n-k+1} \inf_{x\in X(n,m,a,\delta)} \frac{1}{k}\sum_{i=j}^{j+k-1} \chi(\{y|y\ge\varepsilon\},x_i)\}.$$

At $k \le m$, in the problems

$$\sup_{x\in X(n,m,a,\delta)} \frac{1}{k}\sum_{i=j}^{j+k-1} \chi(\{y|y\ge\varepsilon\},x_i) \text{ and } \inf_{x\in X(n,m,a,\delta)} \frac{1}{k}\sum_{i=j}^{j+k-1} \chi(\{y|y\ge\varepsilon\},x_i)$$

the function to be optimized depends only on the variables $x^j = (x_j, x_{j+1}, x_{j+m-1})$; hence, we can perform optimization over the projection of the set $X(n, m, a, \delta)$

on the corresponding coordinates rather over the entire set. It follows from Statement 2 that this projection coincides with the set $X(m,m,a,\delta)$; thus, at $k \leq m$, the u^* and u_* problems take the forms

$$u^*(n,m,a,k,\tau,\delta,\varepsilon) = \sup_{x \in X(m,m,a,\delta)} \frac{1}{k}\sum_{i=1}^{k}\chi(\{y|y \geq \varepsilon\}, x_i)\}.$$

and

$$u_*(n,m,a,k,\tau,\delta,\varepsilon) = \inf_{x \in X(m,m,a,\delta)} \frac{1}{k}\sum_{i=1}^{k}\chi(\{y|y \geq \varepsilon\}, x_i)\}.$$

For $x \in E_+^m$, we set $\pi(x,m,\varepsilon) = |\{i \in \{1,...,m\}|x_i \geq \varepsilon\}|$; this is the number of components greater than or equal to ε. We have

$$u^*(n,m,a,k,\tau,\delta,\varepsilon) = \sup_{x \in X(m,m,a,\delta)} \frac{min\{\pi(x,m,\varepsilon),k\}}{k}$$
$$= \frac{1}{k} min\{\sup_{x \in X(m,m,a,\delta)} \pi(x,m,\varepsilon),k\}.$$

and

$$u_*(n,m,a,k,\tau,\delta,\varepsilon) = \frac{1}{k} max\{k + \inf_{x \in X(m,m,a,\delta)} \pi(x,m,\varepsilon) - m, 0\}.$$

Thus, it is required to find the maximum and the minimum value of the function $\pi(x,m,\varepsilon)$ on $X(m,m,a,\delta)$. Let us introduce the set

$$\Pi(m,\varepsilon,p) = \{x \in E_+^m | \pi(x,m,\varepsilon) = p\}.$$

It is easy to see that the image of the function $\sum_{i=1}^{m} x_i$ on the set $\Pi(m,\varepsilon,p)$ equals

$$\begin{cases} [\varepsilon p, +\infty), p > 0 \\ [0, m\varepsilon), p = 0, \end{cases}$$

Therefore, the function $\pi(x,m,\varepsilon)$ takes the value p on the set $X(m,m,a,\delta)$ if and only if

$$\Pi(m,\varepsilon,p) \cap X(m,m,a,\delta) \neq \emptyset.$$

that is, $1 \leq p \leq \frac{ma+m\delta}{\varepsilon}$ or $a - \delta < \varepsilon$ at $p = 0$.
Let $[x]$ denote the largest integer not exceeding x. Then the condition on those positive values p which the function $\pi(x,m,\varepsilon)$ can take on $X(m,m,a,\delta)$ can be written in the form

$$1 \leq p \leq min\{m, [\frac{m(a+\delta)}{\varepsilon}]\}.$$

The condition that $\pi(x,m,\varepsilon)$ vanishes on $X(m,m,a,\delta)$ has the form $a<\delta+\varepsilon$. Thus, we have proved the following assertion.

Statement 3. Let $k\le m$.

1). If $0\le a<\delta+\varepsilon$, then $u_*(n,m,a,k,\tau,\delta,\varepsilon)=0$.

2). If $a>\delta+\varepsilon$, then $u_*(n,m,a,k,\tau,\delta,\varepsilon)=\frac{max\{k+1-m,0\}}{k}=\{^{1/m,k=m}_{0,k<m}$.

3). If $m(a+\delta)\ge\varepsilon$, then $u^*(n,m,a,k,\tau,\delta,\varepsilon)=\frac{min\{[\frac{m(a+\delta)}{\varepsilon}],k\}}{k}$.

4). If $m(a+\delta)<\varepsilon$, then $u^*(n,m,a,k,\tau,\delta,\varepsilon)=0$.

Note that the solution has the useful feature that, for each of the u^* and u_* problems, there exists the same vector, no matter at what parameter the solution is attained.

Now, consider the case $k>m$. The function $\sum_{i=j}^{j+k-1}\chi(\{y|y\ge\varepsilon\},x_i)$ depends only on those components of the vector x whose numbers belong to $\{j,j+1,...,j+k-1\}$. Thus, we introduce the set

$$X(k,m,a,\delta)=\{x\in E_+^k|\ |\sum_{i=j}^{j+m-1}x_i-am|\le\delta m, j=1,...,k-m+1\}.$$

For this set, an assertion similar to Statement 2 is valid.

Statement 4. If $x\in E_+^n$ and $(x_j,x_{j+1},...,x_{j+k-1})\in X(k,m,a,\delta)$, then there exists a vector $y\in X(n,m,a,\delta)$ such that $y_i=x_i$ for $i=j,j+1,...,j+k-1$.

Proof. We set

- $y_i=x_{j+(i-j)mod\ m}$ for $i=1,...,j-1$,
- $y_i=x_i$ for $i=j,j+1,...,j+k-1$,
- $y_i=x_{j+k-m+(i-j-k+m)mod\ m}$ for $i=j+k,...,n$.

We have

- $\sum_{i=l}^{l+m-1}y_i=\sum_{i=j}^{j+m-1}x_i$ for $l=1,...,j-1$,
- $\sum_{i=l}^{l+m-1}y_i=\sum_{i=l}^{l+m-1}x_i$ for $l=j,j+1,...,j+k-m$,
- $\sum_{i=l}^{l+m-1}y_i=\sum_{i=j+k-m}^{j+k-1}x_i$ for $l=j+k-m,...,n-m+1$.

Therefore, $y\in X(n,m,a,\delta)$. This completes the proof of the statement.

Now, the problems for u^* and u_* with $k>m$ take the forms

$$u^*(n,m,a,k,\tau,\delta,\varepsilon)=\frac{1}{k}\underset{x\in X(k,m,a,\delta)}{sup}\pi(x,k,\varepsilon).$$

and

$$u_*(n,m,a,k,\tau,\delta,\varepsilon)=\frac{1}{k}\underset{x\in X(k,m,a,\delta)}{inf}\pi(x,k,\varepsilon).$$

Let us divide k by m with a remainder, that is, write $k=mq+r, m>r\ge0$.

Take an arbitrary vector $x\in X(k,m,a,\delta)$ and consider its decomposition into the parts

$$x^0=(x_1,...,x_m)\in E^m\ and\ x^j=(x_{r+(j-1)m+1},x_{r+(j-1)m+m})\in E^m, j=1,...,q.$$

It is assumed that $r > 0$; if $r = 0$, then the vector x^0 is absent.

Obviously,

$$\pi(x,k,\varepsilon) = \pi(x^0,r,\varepsilon) + \sum_{j=1}^{q} \pi(x^j,m,\varepsilon) \le \min\{\pi(x^0,m,\varepsilon), r\} + \sum_{j=1}^{q} \pi(x^j,m,\varepsilon).$$

Note that the condition $x \in X(k,m,a,\delta)$ implies that $x^j \in X(m,m,a,\delta)$ for any $j = 0,...,q$. Hence, we have

$$\sup_{x\in X(k,m,a,\delta)} \pi(x,k,\varepsilon) \le min\{ \max_{y\in X(m,m,a,\delta)} \pi(y,m,\varepsilon), r\} + q \max_{y\in X(m,m,a,\delta)} \pi(y,m,\varepsilon).$$

Let us show that this inequality is, in fact, an equality. We take a vector $y^* \in X(m,m,a,\delta)$ at which $\max_{y\in X(m,m,a,\delta)} \pi(y,m,\varepsilon)$ is attained and place all components of this vector greater than or equal to ε at the first positions. Let $x^* \in E_+^k$ be the vector formed by as many copies of y^* written one after another as needed to achieve the required dimension. It is easy to see that $x^* \in X(k,m,a,\delta)$ and

$$\sup_{x\in X(k,m,a,\delta)} \pi(x,k,\varepsilon) \ge \pi(x^*,k,\varepsilon) = min\{ \max_{y\in X(m,m,a,\delta)} \pi(y,m,\varepsilon), r\} + q \max_{y\in X(m,m,a,\delta)} \pi(y,m,\varepsilon).$$

Thus, we have proved the following assertion.

Statement 5. Suppose that $n \ge k > m$ and $k = mq + r, \quad m > r \ge 0$.

1). If $m(a+\delta) \ge \varepsilon$, then $u^*(n,m,a,k,\tau,\delta,\varepsilon) = \frac{min\{[\frac{m(a+\delta)}{\varepsilon}],r\}+q\ min\{[\frac{m(a+\delta)}{\varepsilon}],m\}}{k}$.

2). If $m(a+\delta) \le \varepsilon$, then $u^*(n,m,a,k,\tau,\delta,\varepsilon) = 0$.

It is easy to see that Statement 5 is also valid for $n \ge m \ge k > 0$.

Now, consider the u_* problem. For the arbitrary vector $x \in X(k,m,a,\delta)$ under consideration and its partition constructed above, we have

$$\pi(x,k,\varepsilon) = \pi(x^0,r,\varepsilon) + \sum_{j=1}^{q} \pi(x^j,m,\varepsilon) \ge max\{k-m+\pi(x^0,m,\varepsilon), 0\} + \sum_{j=1}^{q} \pi(x^j,m,\varepsilon).$$

Since $x^j \in X(m,m,a,\delta)$ for any $j = 0,...,q$, it follows that

$$\inf_{x\in X(k,m,a,\delta)} \pi(x,k,\varepsilon) \ge max\{r-m+ \inf_{y\in X(m,m,a,\delta)} \pi(y,m,\varepsilon), 0\} + q \inf_{y\in X(m,m,a,\delta)} \pi(y,m,\varepsilon).$$

As above, this inequality is, in fact, an equality. To show this, we take a vector $y_* \in X(m,m,a,\delta)$ at which $\min_{y\in X(m,m,a,\delta)} \pi(y,m,\varepsilon)$ is attained and place all components of this vector which are greater than or equal to ε at the last positions.

Consider the vector $x_* \in E_+^m$ consisting of copies of y_* written one after another. It is easy to see that $x_* \in X(k,m,a,\delta)$ and

$$\inf_{x \in X(k,m,a,\delta)} \pi(x,k,\varepsilon) \le \pi(x_*,k,\varepsilon)$$

$$= \max\{r - m + \min_{y \in X(m,m,a,\delta)} \pi(y,m,\varepsilon), 0\} + q \min_{y \in X(m,m,a,\delta)} \pi(y,m,\varepsilon).$$

Thus, we have proved the following assertion.

Statement 6. Suppose that $n \ge k > m$ and $k = mq + r, m > r \ge 0$.

1). If $0 \le a < \delta + \varepsilon$, then $u_*(n,m,a,k,\tau,\delta,\varepsilon) = 0$.

2). If $a \ge \delta + \varepsilon$, then $u_*(n,m,a,k,\tau,\delta,\varepsilon) = \frac{q}{k}$.

9 Conclusion

The exact boundaries of the range of the soft probability of large deviations under a single mean hypothesis, which were found in this paper, show (although, for a very simple example) that it is quite possible to deal with soft probabilities, in spite of the presence of parameters and the interval form of soft probability. The next goal is to solve more complicated problems on evaluating various probabilities and other characteristics in the presence of several hypotheses, preferably of different types.

Of special interest is the application of the ideas and results presented in this paper to a real statistical problem, which would make it possible to verify the effectiveness of the approach for real data. All readers interested in such a practical experiment are kindly requested to send their suggestions to the author at dmitri_molodtsov@mail.ru.

References

[1] V. N. Tutubalin. (1972), *Probability Theory: A Short Course and Scientific-Methodological Notes*, Izd. Moskov. Univ, Moscow, Russian.

[2] D. A. Molodtsov. (2004), *Theory of Soft Sets*, URSS, Moscow, Russian.

[3] D. A. Molodtsov. (2007), "Portfolio control using soft probability", Vestn. Nats. Assots. Uchastnikov Fond. Rynka, No.7-8.

[4] D. A. Molodtsov. (2007), "Portfolio control using soft probability (short positions)", *Vestn. Nats. Assots. Uchastnikov Fond. Rynka*, No.10.

[5] D. A. Molodtsov. (2011), "Soft sets and prediction", *Nechetkie Sist. Myagkie Vychisl*, Vol.6, No.1.

[6] D. A. Molodtsov. (2010), "Finite frequency stability and probability", *Nechetkie Sist. Myagkie Vychisl*, Vol.5, No.1.

[7] D. A. Molodtsov. (2008), "Soft uncertainty and probability", *Nechetkie Sist. Myagkie Vychisl*, Vol.3, No.1.

[8] I. S. Berezin and N. P. Zhidkov. (1962), *Computing Methods*, Pergamon Press, Vol.1/2.

[9] H. Aktas and N. Cagman N. (2007), "Soft sets and soft groups", *Inf. Sci*, Vol.177, pp.2726C2735.

[10] M. I. Ali et al. (2009), "On some new operations in soft set theory", *Computers and Math. With Appl*, Vol.57, pp.1547-1553.

[11] D. Chen. ((2005)), "The parameterization reduction of soft sets and its applications", *Comput. Math. Appl*, Vol.49, pp.757-763.

[12] F. Feng et al. (2008), "Soft semirings", *Computers and Math. With Appl*, Vol.56, pp.2621-2628.

[13] F. Feng et al. (2010), "Soft sets combined with fuzzy sets and rough sets: a tentative approach", *Soft Computing*, Vol.14, pp.899-911.

[14] Y. B. Jun. (2008), "Soft BCK/BCI-algebras", *Computers and Math. With Appl*, Vol.56, pp.1408-1413.

[15] Y. B. Jun and C. H. Park. (2008), "Applications of soft sets in ideal theory of BCK/BCI-algebras", *Inf. Sci*, Vol.178, pp.2466-2475.

[16] Z. Kong et al. (2008), "The normal parameter reduction of soft sets and its algorithm", *J.Comp. Appl. Math*, Vol.56, pp.3029-3037.

[17] P. K. Maji et al. (2002), "An application of soft sets in a decision making problem", *Comput. Math. Appl*, Vol.44, pp.1077-1083.

[18] P. K. Maji et al. (2003), "Soft set theory", *Comput. Math. Appl*, Vol.45, pp.555-562.

[19] P. Majumdar and S. K. Samanta. (2010), "On soft mappings", *Computers and Math. With Appl*, Vol.60, pp.2666-2672.

[20] D. Molodtsov. (1999), "Soft set theory first results", *Comput. Math. Appl*, Vol.37, pp.19-31.

[21] D. Pie and D. Miao. (2005), "From soft sets to information systems, granular computing", *IEEE Iinter. Conf*, No.2, pp.617-621.

[22] M. Shabir et al. (2009), "Soft ideals and generalized fuzzy ideals in semigroups", *New Math. Nat. Comput*, No.5, pp.599-615.

[23] M. Shabir et al. (2011), "On soft topological spaces", *Computers and Math. With Appl*, Vol.61, pp.1786-1799.

[24] Zou Yan et al. (2008), "Data analysis approaches of soft sets under incomplete information", *Knowl.-Based Syst*, Vol.21, pp.941-945.

Corresponding author
D.A Molodtsov can be contacted at: dmitri_molodtsov@mail.ru