Estimating Risk of Extreme and Catastrophic Events under Interval Uncertainty

Nitaya Buntao[1] and Vladik Kreinovich[2]

[1]Department of Applied Statistics
King Mongkut's University of Technology North Bangkok
1518 Piboonsongkhram Road, Bangsue, Bangkok 10800 Thailand
Email: taltanot@hotmail.com
[2]Department of Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA, vladik@utep.edu

Abstract

In many application areas, we encounter heavy-tail distributions – for example, such distributions are ubiquitous in financial applications. These distributions are often described by Pareto law. There exist techniques for estimating the parameters of such the corresponding Pareto distributions based on the sample $x_1, \ldots, x_n$. In practice, we often only know the values x_i with interval uncertainty. In this paper, we show how to estimate the parameters of the Pareto distribution under such uncertainty and how to describe deviation and dependence for general heavy-tailed distributions.

The results of this paper partly appeared in [2, 3].

1 Formulation of the General Problem: Estimating Parameters of the Heavy-Tailed Distributions under Interval Uncertainty

Need for heavy-tailed distributions. To estimate the risk of extreme and catastrophic events, we usually perform a statistical analysis of the existing data, and use the results of this analysis to predict the probabilities (and risk) of extreme and catastrophic events. For example, based on the observations of small earthquakes recorded in a certain area, we estimate the probability of a future potentially catastrophic earthquake; based on the observed variations in the economic characteristics, we predict the probability of catastrophic (crisis) changes, etc.

In most applications of statistical methods to science and engineering, researchers use either the normal distribution or distributions related to normal – such as lognormal; see, e.g., [31, 36]. In these distributions, the probability of a value decrease exponentially with this value. As a result, large deviations are practically impossible.

For example, for a normal distribution with mean a and standard deviation σ, the probability of a value x to be outside the "three sigma" interval $[a - 3\sigma, a + 3\sigma]$ is approximately 0.1%, and the probability to be outside the "six sigma" interval $[a - 6\sigma, a + 6\sigma]$ is approximately 10^{-8}.

In practice, however, we often encounter random processes in which large deviations are possible. In many such distributions, the variance is infinite; such distributions are called *heavy-tailed*. These distributions surfaced in the 1960s, when Benoit Mandelbrot, the author of fractal theory, empirically studied the fluctuations and showed [23] that larger-scale fluctuations follow the Pareto power-law distribution, with the probability density function $\rho(x) = A \cdot x^{-\alpha}$, for some constant $\alpha \approx 2.7$. For this distribution, variance is infinite. The above empirical result, together with similar empirical discovery of heavy-tailed laws in other application areas, has led to the formulation of *fractal theory*; see, e.g., [24, 25].

Since then, similar heavy-tailed distributions have been empirically found in other financial situations [4, 5, 7, 12, 26, 28, 32, 35, 39, 40, 41], and in many other application areas [1, 14, 24, 27, 34].

Need to take into account interval uncertainty. In practice, we rarely know the exact values of x_i. For example, in financial situations, we can take, as x_i, the price of the financial instrument at the i-th moment of time – e.g., on the i-th day. However, the price does not remain stable during the day – it fluctuates. Of course, we can always arbitrarily select a value, but it is more reasonable to consider the whole range $[\underline{x}_i, \overline{x}_i]$ of the daily prices instead of a single value x_i.

Different values x_i from the corresponding intervals lead, in general, to different estimates $f(x_1, \ldots, x_n)$ for the parameters of the heavy-tailed distribution. To get a good understanding of the corresponding risk, it is therefore desirable to compute not just a *single* value of each characteristic, but rather the *range*

$$\mathbf{y} = \{f(x_1, \ldots, x_n) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}$$

of possible values of this characteristic when each x_i takes different values from the corresponding interval $\mathbf{x}_i$. It is therefore desirable to find the range of all resulting values of $f(x_1, \ldots, x_n)$.

Due to the ubiquity of interval uncertainty, the need to estimate a range of a given function $f(x_1, \ldots, x_n)$ over given intervals $\mathbf{x}_1, \ldots, \mathbf{x}_n$ occurs in many other application areas. The problem of computing this range is known as the main problem of *interval computations*; see, e.g., [18, 17, 29].

In spite of the simplicity of the problem's formulation, in general, the interval computations problem is NP-hard (computationally intensive [30]); see, e.g., [21].

It is even NP-hard if we restrict ourselves to simple functions: e.g., to quadratic ones. Moreover, the problem is NP-hard even for the simplest statistically meaningful quadratic function: the function

$$V(x_1, \ldots, x_n) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i\right)^2$$

that describes the sample variance [8, 9].

2 First Problem: Estimating Parameters of the Pareto Distribution under Interval Uncertainty

Estimating parameters of the Pareto distribution. Large deviations describe crises, so their analysis is very important. To get accurate predictions of the possible large deviations, we must get accurate estimates of the parameters x_0 and α based on the observed data values $x_1, \ldots, x_n$.

In [20], it was shown that by applying the Maximum Likelihood techniques to the Pareto distribution, we get the following estimates:

$$\hat{x}_0 = \min(x_1, \ldots, x_n); \tag{1}$$

and

$$\hat{\alpha} = n \cdot \left(\sum_{i=1}^{n} \ln\left(\frac{x_i}{\min(x_1, \ldots, x_n)}\right)\right)^{-1}. \tag{2}$$

Need to take into account interval uncertainty. As we have mentioned, it is desirable to compute the ranges of possible values of these characteristics x_0 and α when each x_i takes different values from the corresponding interval $\mathbf{x}_i$.

3 First Result: Estimating x_0 under Interval Uncertainty

Problem: reminder. Let us first estimate the range of the estimate $x_0 = \min(x_1, \ldots, x_n)$ when $x_1 \in [\underline{x}_1, \overline{x}_1]$, …, $x_n \in [\underline{x}_n, \overline{x}_n]$.

How we can solve this problem. The function $x_0 = \min(x_1, \ldots, x_n)$ is a (non-strictly) increasing function of each of its variables.

Thus, the largest possible value of this function is attained when each of the variables x_i attains its largest possible value $x_i = \overline{x}_i$. So, the largest possible value of x_0 is equal to $\min(\overline{x}_1, \ldots, \overline{x}_n)$.

Similarly, the smallest possible value of this function is attained when each of the variables x_i attains its smallest possible value $x_i = \underline{x}_i$. Thus, the smallest possible value of x_0 is equal to $\min(\underline{x}_1, \ldots, \underline{x}_n)$.

So, we arrive at the following result.

The interval of possible values of x_0: result. The interval $[\underline{x}_0, \overline{x}_0]$ of possible value of the parameter x_0 can be computed as follows:

$$\underline{x}_0 = \min(\underline{x}_1, \ldots, \underline{x}_n); \tag{3}$$

$$\overline{x}_0 = \min(\overline{x}_1, \ldots, \overline{x}_n). \tag{4}$$

4 Estimating α under Interval Uncertainty: Analysis of the Problem

Reducing the problem to simpler ones: idea. We want to find the range $[\underline{\alpha}, \overline{\alpha}]$ of the estimate α, as described by the formula (2). To simplify our analysis, let us reduce this problem to several simpler ones.

Reducing the problem to simpler ones: first step. First, according to the description of the estimate α (formula (2)), this estimate has the form

$$\alpha = \frac{n}{r}, \tag{5}$$

where we denoted

$$r = \sum_{i=1}^{n} \ln\left(\frac{x_i}{\min(x_1, \ldots, x_n)}\right). \tag{6}$$

Since the function $\dfrac{n}{r}$ is decreasing,

- the largest possible value $\overline{\alpha}$ of $\alpha = \dfrac{n}{r}$ is attained when r takes the smallest possible value, and
- the smallest possible value $\underline{\alpha}$ of $\alpha = \dfrac{n}{r}$ is attained when r takes the largest possible value.

So, if we can find the range $[\underline{r}, \overline{r}]$ of possible values of r, we can then find the range $[\underline{\alpha}, \overline{\alpha}]$ for α as follows:

$$\underline{\alpha} = \frac{n}{\overline{r}}; \quad \overline{\alpha} = \frac{n}{\underline{r}}. \tag{7}$$

Thus, the original problem of computing the range of a complex expression (2) can be reduced to the auxiliary problem of computing the range of a somewhat simpler expression (6).

Reducing the problem to simpler ones: second step. To reduce the problem further, let us further simplify the expression (6). For this simplification, we can use the fact that r is the sum of several logarithms, and the sum of the logarithms is equal to the logarithm of the product:

$$r = \ln(S), \tag{8}$$

where we denoted

$$S \stackrel{\text{def}}{=} \prod_{i=1}^{n} \frac{x_i}{\min(x_1, \ldots, x_n)} = \frac{\prod\limits_{i=1}^{n} x_i}{(\min(x_1, \ldots, x_n))^n}. \quad (9)$$

Since the function $\ln(S)$ is increasing,

- the largest possible value $\overline{r}$ of $r = \ln(S)$ is attained when S takes the largest possible value, and
- the smallest possible value $\underline{r}$ of $r = \ln(S)$ is attained when S takes the smallest possible value.

So, if we can find the range $[\underline{S}, \overline{S}]$ of possible values of S, we can then find the range $[\underline{r}, \overline{r}]$ for r as follows:

$$\underline{r} = \ln(\underline{S}); \quad \overline{r} = \ln(\overline{S}). \quad (10)$$

Thus, the problem of computing the range of a complex expression (6) can be reduced to the auxiliary problem of computing the range of a somewhat simpler expression (9).

Further reduction. When we know that x_j is the smallest of n values $x_1, \ldots, x_n$, then the expression (9) can be simplified even further:

$$S = \frac{\prod\limits_{i=1}^{n} x_i}{x_j^n}. \quad (11)$$

By canceling the terms x_j in the numerator and in the denominator, we can further simplify this expression into

$$S = \frac{\prod\limits_{i \neq j} x_i}{x_j^{n-1}}. \quad (12)$$

Let us show how after this reduction, we can explicitly compute both bounds $\underline{S}$ and $\overline{S}$.

Computing $\overline{S}$: analysis. The expression (12) is increasing as a function of all the variables x_i with $i \neq j$ and decreasing as a function of the remaining variable x_j. Thus, its largest possible value is attained when:

- all the variables x_i with $i \neq j$ attains its largest possible value $\overline{x}_i$, while
- the variable x_j attains its smallest possible value $\underline{x}_j$.

The corresponding expression is equal to

$$S_j = \frac{\prod\limits_{i \neq j} \overline{x}_i}{\underline{x}_j^{n-1}}. \quad (13)$$

Multiplying both numerator and denominator by $\overline{x}_j$, we conclude that

$$S_j = \frac{\prod\limits_{i=1}^{n} \overline{x}_i}{\overline{x}_j \cdot \underline{x}_j^{n-1}}. \tag{14}$$

This expression is only possible when $x_j \le x_i$ for all $i \ne j$, i.e., when $\underline{x}_j \le \overline{x}_i$ for all i. A number is smaller than several numbers if it is smaller than the smallest of them, i.e., if

$$\underline{x}_j \le \min(\overline{x}_1, \ldots, \overline{x}_n). \tag{15}$$

It should be mentioned that the right-hand side of this inequality has already appeared in this text – as the upper endpoint $\overline{x}_0$ for the parameter x_0.

Among the values S_j corresponding to all such j, we need to choose the largest one. According to (14), each of the values S_j is the result of dividing the same product $\prod\limits_{i=1}^{n} \overline{x}_i$ by the value $\overline{x}_j \cdot \underline{x}_j^{n-1}$. Thus, the largest possible value S_j corresponds to the smallest possible value of the product $\overline{x}_j \cdot \underline{x}_j^{n-1}$.

The largest value $\overline{S}$ of S corresponds to the largest value $\overline{r}$ of r and thus, to the smallest value $\underline{\alpha}$ of α. Thus, we arrive at the algorithm for computing $\underline{\alpha}$ that is described in the next section.

Computing $\underline{S}$: analysis. Let $x_1, \ldots, x_n$ be the values at which the function S attains its minimum, and let x_j be the smallest of these values.

If all the values x_i are equal to each other, then we get $S = 1$. In this case, we can increase all the values until we reach the upper endpoint $\overline{x}_i$ of one of the intervals. Then, we get $x_i = \overline{x}_i$, and for every other k, we get $\overline{x}_i = x_i = x_k \le \overline{x}_k$, hence $\overline{x}_i \le \overline{x}_k$ for all k, and $\overline{x}_i = \min(\overline{x}_1, \ldots, \overline{x}_n)$ $(= \overline{x}_0)$.

For this i, we have $x_i = \overline{x}_i$, and for all other $k \ne i$, we get $x_k = \max(\overline{x}_i, \underline{x}_k)$. Let us show that a similar formula holds when not all the coordinates of the optimizing vector $(x_1, \ldots, x_n)$ are equal to each other.

Indeed, the expression (12) is increasing as a function of all the variables x_i with $i \ne j$ and decreasing as a function of the remaining variable x_j.

Thus, if we could increase x_j without changing all other values x_i – and still preserve the conditions $x_j \le \overline{x}_j$ and the inequalities $x_j \le x_i$ – we would be able to further decrease the value (12). Since we started with the values for which S attains its minimum, such a increase in x_j is impossible. The fact that we cannot increase x_j without violating the constraints $x_j \le \overline{x}_j$ and $x_j \le x_i$ means that at least in one of the constraints, we have equality. Thus:

- we either have $x_j = \overline{x}_j$,
- or we have $x_j < \overline{x}_j$ and $x_j = x_i$ for some $i \ne j$.

Let us consider the second case. In the second case, we may have several values x_i for which $x_j = x_i$. If for all these values, we have $x_i < \overline{x}_i$, then we can increase this common value $x_j = x_i = \ldots$ and thus, further decrease S. Thus, the fact that we have selected the minimizing vector implies that at least for one i, we have $x_j = x_i = \overline{x}_i$.

Thus, for the minimizing vector, the smallest value $\min(x_1, \ldots, x_n)$ is attained at one of the upper endpoints $\overline{x}_i$. Since this value $\overline{x}_i$ is the smallest, we get $\overline{x}_i \le x_k$ for all $k \ne i$, and since $x_k \le \overline{x}_k$, we conclude that $\overline{x}_i \le \overline{x}_k$ for all k. Thus, the minimal value $\overline{x}_i = \min(x_1, \ldots, x_n)$ is the smallest of n upper endpoints:

$$x_j = \min(\overline{x}_1, \ldots, \overline{x}_n) = \overline{x}_0. \tag{16}$$

For every $k \ne i$, we select the smallest possible value $x_k \in [\underline{x}_k, \overline{x}_k]$ for which $x_k \ge \overline{x}_i$, i.e., the value $x_k = \max(\overline{x}_i, \underline{x}_k)$.

The smallest value $\underline{S}$ of S corresponds to the smallest value $\underline{r}$ of r and thus, to the largest value $\overline{\alpha}$ of α. Thus, we arrive at the algorithm for computing $\overline{\alpha}$ that is described in the corresponding section.

5 Algorithm for Computing $\underline{\alpha}$

First stage. To find $\underline{\alpha}$, first, we compute the value

$$\overline{x}_0 = \min(\overline{x}_1, \ldots, \overline{x}_n). \tag{17}$$

Comment. If we have already computed the range $[\underline{x}_0, \overline{x}_0]$, for x_0, then we do not need to compute anything: we just borrow the corresponding value $\overline{x}_0$.

Second stage. We test $j = 1, \ldots, n$, and among all the indices j for which $\underline{x}_j \le \overline{x}_0$, we select the one with the smallest possible value of the product $\overline{x}_j \cdot \underline{x}_j^{n-1}$.

Final formula. The smallest possible value of α is attained when x_j takes the value $\underline{x}_j$ while all other variables take the values $\overline{x}_i$. For these values, $\min(x_1, \ldots, x_n) = \underline{x}_j$, hence the j-th term in the sum (2) disappears, and the expression (2) takes the form

$$\underline{\alpha} = n \cdot \left(\sum_{i \ne j} \ln\left(\frac{\overline{x}_i}{\underline{x}_j} \right) \right)^{-1}. \tag{18}$$

Computation time. At each stage, this algorithm takes the linear number of steps, i.e., the number of steps bounded by the number of variables n. Thus, overall, we have a linear-time algorithm.

This computation time is asymptotically optimal. Indeed, we need to take into account each of the intervals $[\underline{x}_i, \overline{x}_i]$. We need at least one computation step to read each of these values. Thus, the overall number of computation steps cannot be smaller than n. So, our algorithm that takes times $\le \text{const} \cdot n$ is asymptotically optimal.

6 Algorithm for Computing $\overline{\alpha}$

First stage. To find $\overline{\alpha}$, first, we compute the value

$$\overline{x}_0 = \min(\overline{x}_1, \ldots, \overline{x}_n). \tag{19}$$

Comment. If we have already computed the range $[\underline{x}_0, \overline{x}_0]$, for x_0, then we do not need to compute anything: we just borrow the corresponding value $\overline{x}_0$.

Second stage. For each $k = 1, \ldots, n$, we take $x_k = \max(\overline{x}_0, \underline{x}_k)$, and then compute the corresponding value α as

$$\overline{\alpha} = n \cdot \left(\sum_{k=1}^{n} \ln \left(\frac{\max(\overline{x}_0, \underline{x}_k)}{\overline{x}_0} \right) \right)^{-1}. \tag{20}$$

Computation time. This algorithm also takes linear time and is, thus, also asymptotically optimal.

7 Second Problem: How to Describe Deviation from the "Average" for General Heavy-Tailed Distributions

Formulation of the problem. For heavy-tailed distributions, variance is infinite, so we cannot use variance to describe the deviation from the "average". Thus, we need to come up with other characteristics for describing this deviation.

In the following text, we will describe such characteristics, and we will describe how we can estimate these characteristics under interval uncertainty.

Analysis of the problem. In this section, we handle the first problem: how to characterize deviation from the "average" for heavy-tailed distributions. Of course, there are many possible mathematical definitions, our objective is to select a definition that best reflects the user's preferences.

A standard way to describe preferences of a decision maker is to use the notion of *utility* u; see, e.g., [10, 11, 19, 22, 33]. According to decision theory, a user prefers an alternative for which the expected value $\sum\limits_{i=1}^{n} p_i \cdot u_i$ of the utility is the largest possible. Alternative, we can say that the expected value $\sum\limits_{i=1}^{n} p_i \cdot U_i$ of the *disutility* $U \stackrel{\text{def}}{=} -u$ is the smallest possible.

In our case, instead of considering n different values $x_1, \ldots, x_n$, we consider a single value m. Since we are replacing each original value x_i with a new value m which is only an approximation to x_i, there is some resulting disutility. For example, if we dress based on the expected average temperature m and the actual temperature is $x_i \neq m$, then we may feel somewhat warm or somewhat cold. Similarly, if the heating and cooling system of the campus buildings is programmed based on the assumption that the outside temperature is m and the actual temperature is $x_i \neq m$, the system does not work perfectly well, and we may need to spend extra resources (and extra heaters and/or ventilators) to make the temperature in the offices most comfortable.

The further away the approximate value m from the actual one x_i, the larger the disutility. Let $U(d)$ denote the disutility cause by the difference $d = x_i - m$. When x_i coincides with m, there is no disutiluty, i.e., $U(0) = 0$. If this difference d is positive, then, the larger d, the larger the disutility: $d_1 \le d_2$ implies $U(d_1) \le U(d_2)$. Similarly, if the difference d is negative, the smaller d, the larger the disutility: $d_1 \le d_2$ implies $U(d_1) \ge U(d_2)$.

Under this notation, for each i, the disutility is equal to $U(x_i - m)$. In the sample, we have n estimates with equal probability $p_i = \frac{1}{n}$; thus, the expected value of the disutility is equal to

$$\frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - m). \tag{21}$$

It is therefore reasonable to select, as the "average" m, the value for which this disutility attains the smallest possible value. The resulting value of expected disutility can then be used as the desired characteristic of the deviation of the values from the average. Thus, we arrive at the following definitions.

Resulting definitions. Let $U(d) \ge 0$ be a function from real numbers to non-negative real numbers such that $U(0) = 0$, $U(d)$ is (non-strictly) increasing for $d \ge 0$, and $U(d)$ is (non-strictly) decreasing for $d \le 0$.

For each sample $x_1, \ldots, x_n$, by a *U-estimate*, we mean the value m_U that minimizes the expression (21). By a *U-deviation*, we mean the value

$$V_U \stackrel{\text{def}}{=} \min_m \frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - m). \tag{22}$$

Comment. Because of the definition of m_U, the value V_U takes the form

$$V_U = \frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - m_U). \tag{23}$$

Examples. When $U(x) = x^2$, the expression (21) turns into the expression $\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - m)^2$ for which minimization leads to the arithmetic average $m = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$. For this arithmetic average, the expression V_U is the usual variance.

When $U(x) = |x|$, the expression turns into the expression $\frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - m|$ for which minimization leads to the median. For the median m_U, the expression V_U is the *average absolute deviation*

$$V_U = \frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - m_U|.$$

How to estimate m_U and V_U. Once we compute m_U, the computation of V_U is straightforward: we just apply the function $U(d)$ n times and compute the corresponding expression.

Estimating m_U means optimizing a function of a single variable. This particular optimization problem is well-known and actively used in statistics, because, as we will show, it is equivalent to the Maximum Likelihood approach to the following problem. Let us assume that we know the shape $\rho_0(x)$ of the actual distribution but not the starting point, i.e., we know that the actual distribution has the form $\rho_0(x - m)$ for some unknown value m. To estimate this value m based on the sample $x_1, \ldots, x_n$, we can use the maximum likelihood method, i.e., find m for which the probability density

$$L = \rho_0(x_1 - m) \cdot \ldots \cdot \rho_0(x_n - m)$$

attains the largest possible value. Maximizing this probability is equivalent to minimizing the value

$$\psi \stackrel{\text{def}}{=} -\ln(L) = \sum_{i=1}^{n} U(x_i - m),$$

where we denoted $U(x) \stackrel{\text{def}}{=} -\ln(\rho_0(x))$. Minimizing this value is equivalent to minimizing the value (21); thus, this value is exactly our estimate m_U.

Similar algorithms are also used in *robust statistics* – an area of statistics in which we need to make statistical estimates under partial information about the probability distribution.

In robust statistics (see, e.g., [16]), there are several different types of techniques for estimating a shift-type parameter a based on a sample $x_1, \ldots, x_n$. The most widely used methods are *M-methods*, methods which are mathematically equivalent to the maximum likelihood approach from the traditional (non-robust) statistics.

Comment. The relation between utilities, maximum likelihood methods, and robust statistics was analyzed in [37].

8 Estimating the Heavy-Tailed-Related Deviation Characteristics under Interval Uncertainty: Analysis of the Problem

What we want. In the previous section, we described how to define the deviation V_U in the heavy-tailed case, and how to estimate the value of the deviation when we know the exact values $x_1, \ldots, x_n$. As we have mentioned, in practice, the values x_i are often only known with interval uncertainty, i.e., we only know the intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ that contain the unknown values x_i. In this case, it is desirable to compute the range $\mathbf{V}_U = [\underline{V}_U, \overline{V}_U]$ of possible values of V_U when $x_i \in \mathbf{x}_i$.

The value $\underline{V}_U$ is the minimum of the function $V_U(x_1, \ldots, x_n)$ when $x_i \in \mathbf{x}_i$, and the value $\overline{V}_U$ is the maximum of the function $V_U(x_1, \ldots, x_n)$ when $x_i \in \mathbf{x}_i$. So, to estimate these values, let us recall when a function attains its minimum and maximum.

When does a function attains its minimum and maximum on an interval: a general reminder. Let us start with functions of one variable $f(x)$ defined on an interval $[\underline{x}, \overline{x}]$. A continuous function always attains its smallest possible value at some point $x \in [\underline{x}, \overline{x}]$. This point can be:

- either inside the interval $\underline{x} < x < \overline{x}$;
- or the left endpoint $x = \underline{x}$,
- or at the right endpoint $x = \overline{x}$.

It is well known, from calculus, that if a function $f(x)$ attains its minimum or maximum at some point x inside the interval, then at this point, the derivative of f is equal to 0: $\dfrac{df}{dx} = 0$.

If the minimum is attained at the left endpoint $x = \underline{x}$, then at this point, we cannot have $\dfrac{df}{dx} < 0$, because otherwise, for small $\Delta x > 0$, we would have

$$f(\underline{x} + \Delta x) = f(\underline{x}) + \Delta x \cdot \frac{df}{dx} + o(\Delta x) < f(\underline{x}),$$

which contradicts our assumption that $f(\underline{x})$ is the smallest value of $f(x)$ on the given interval. Thus, in this case, we must have $\dfrac{df}{dx} \ge 0$.

Similarly, if the minimum is attained at the right endpoint $x = \overline{x}$, we must have $\dfrac{df}{dx} \le 0$. For maximum:

- if the maximum is attained at the left endpoint $x = \underline{x}$, we must have $\dfrac{df}{dx} \le 0$;
- if the maximum is attained at the right endpoint $x = \overline{x}$, we must have $\dfrac{df}{dx} \ge 0$.

Thus, for minimum, we have one of the following three option:

- either the minimum is attained for $x = \underline{x}$ and $\dfrac{df}{dx} \ge 0$;
- or the minimum is attained for $x = \overline{x}$ and $\dfrac{df}{dx} \le 0$;
- the minimum is attained strictly inside the interval $[\underline{x}, \overline{x}]$, and $\dfrac{df}{dx} = 0$.

When does a function of several variables attains its minimum and its maximum? For a function of several variables, a similar conclusion can be reached for each of these variables. Thus, if $(x_1, \ldots, x_n)$ denotes the tuple at which the tuple attains its minimum, then for every i, we have one of three following options:

- either $x_i = \underline{x}_i$ and $\dfrac{\partial f}{\partial x_i} \ge 0$;
- or $x_i = \overline{x}_i$ and $\dfrac{\partial f}{\partial x_i} \le 0$;
- or $x_i \in (\underline{x}_i, \overline{x}_i)$ and $\dfrac{\partial f}{\partial x_i} = 0$.

Similarly, if $(x_1, \ldots, x_n)$ denotes the tuple at which the tuple attains its *maximum*, then for every i, we have one of three following options:

- either $x_i = \underline{x}_i$ and $\dfrac{\partial f}{\partial x_i} \le 0$;
- or $x_i = \overline{x}_i$ and $\dfrac{\partial f}{\partial x_i} \ge 0$;
- or $x_i \in (\underline{x}_i, \overline{x}_i)$ and $\dfrac{\partial f}{\partial x_i} = 0$.

Applying the general conclusions about minima and maxima to our problem. Let us apply these conclusions to the function $V_U(x_1, \ldots, x_n)$. From the fact that the value m_U corresponds to the minimum of the expression (21), we conclude that for this value, the derivative of the expression (21) with respect to m is equal to 0, i.e., that

$$-\frac{1}{n} \cdot \sum_{i=1}^{n} U'(x_i - m) = 0, \tag{24}$$

where $U'(d)$ denotes the derivative of the function $U(d)$. Differentiating the expression (23) with respect to x_i and taking into account that m_U also depends on x_i, we conclude that

$$\frac{\partial V_U}{\partial x_i} = U'(x_i - m) - \left(\frac{1}{n} \cdot \sum_{i=1}^{n} U'(x_i - m)\right) \cdot \frac{\partial m_V}{\partial x_i}.$$

Due to (24), the expression in parentheses is equal to 0 and thus,

$$\frac{\partial V_U}{\partial x_i} = U'(x_i - m). \tag{25}$$

By definition of the function $U(d)$, we have $U'(x_i - m) > 0$ only for $x_i > m$ and $U'(x_i - m) < 0$ only for $x_i < m$.

Thus, when the function V_U attains its minimum, we have:

- either $x_i = \underline{x}_i$ and $x_i \geq m$,
- or $x_i = \overline{x}_i$ and $x_i \leq m$,
- or $x_i \in (\underline{x}_i, \overline{x}_i)$, and $x_i = m$.

If $\overline{x}_i < m$, then the i-th interval is fully to the left of the value m, i.e., $x_i < m$ for all $x_i \in [\underline{x}_i, \overline{x}_i]$. In this case, we cannot have $x_i \in (\underline{x}_i, \overline{x}_i)$ – otherwise we would have $x_i = m$, and we know that $x_i < m$. Similarly, we cannot have $x_i = \underline{x}_i$ because otherwise, we will have $x_i \geq m$, and we know that $x_i < m$. Thus, the only remaining option is $x_i = \overline{x}_i$.

Similarly, when $m < \underline{x}_i$, then the i-th interval is right to the left of the value m, i.e., $x_i > m$ for all $x_i \in [\underline{x}_i, \overline{x}_i]$. In this case, the only possible option is $x_i = \underline{x}_i$.

Finally, when $\underline{x}_i \leq m \leq \overline{x}_i$, the only remaining option is $x_i = m$.

Comment. For simplicity, in our analysis, we ignored the fact that it is possible to have $U'(d) = 0$ for $d > 0$; if we take this possibility into account, then, strictly speaking, we can no longer argue that *every* tuple for which the deviation measure V_U attains its minimum has the above type, we can still argue that *there is* a tuple of this type for which V_U attains its minimum. Crudely speaking, if the minimum is attained for the value x_i at which $U'(x_i - m) = 0$, we can still modify x_i without changing the value V until we can no longer do that – i.e., until we either get the endpoint or the value m.

Thus, once we know where m is with respect to all the bounds $\underline{x}_i$ and $\overline{x}_i$, we can uniquely determine where the minimum of V_U is attained under this restriction on m:

- if $\overline{x}_i \leq m$, then we have $x_i = \overline{x}_i$;
- if $m \leq \underline{x}_i$, then we have $x_i = \underline{x}_i$;
- if $\underline{x}_i \leq m \leq \overline{x}_i$, then $x_i = m$.

In all three cases, x_i is the closest value to m on the interval $[\underline{x}_i, \overline{x}_i]$.

The value m can now be determined by the requirement that for this m, the sum (21) take the smallest possible value. Since for $x_i = m$, we have $U(x_i - m) = U(0) = 0$, it is sufficient to consider only the intervals i for which $x_i \neq m$. Thus, m is equal to the U-average of such values x_i. So, we arrive at the following algorithm.

9 Algorithm for Computing $\underline{V}_U$

Algorithm. In order to find $\underline{V}_U$, let us first sort all $2n$ endpoints $\underline{x}_i$ and $\overline{x}_i$ into an increasing sequence

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(2n)}.$$

To these values, we add $x_{(0)} \stackrel{\text{def}}{=} -\infty$ and $x_{(2n+1)} \stackrel{\text{def}}{=} +\infty$, then we get

$$-\infty = x_{(0)} \le x_{(1)} \le x_{(2)} \le \ldots \le x_{(2n)} \le x_{(2n+1)} = +\infty.$$

The resulting values divide the real line into $2n+1$ zones $[x_{(k)}, x_{(k+1)}]$, $k = 0, 1, \ldots, 2n$. For each zone, we select the values $x_1, \ldots, x_n$ as follows: for some value m (to be determined),

- if $\overline{x}_i \le r_{(k)}$, then we select $x_i = \overline{x}_i$;
- if $r_{(k+1)} \le \underline{x}_i$, then we select $x_i = \underline{x}_i$;
- for all other i, we select $x_i = m$.

Then, we take only the values for which $x_i \ne m$, and find their U-estimate and – if this U-estimate is in the zone – compute the corresponding U-deviation.

The smallest of thus computed U-deviations is the desired value $\underline{V}_U$.

Computation time for this algorithm. Sorting takes

$$O(n \cdot \log(n))$$

steps; see, e.g., [6]. After that, for each of $2n = O(n)$ zones, we need $O(n)$ steps to perform the computations and the time – that we will denote by T_{exact} – to compute the U-estimate and U-deviation. Thus, the total computation time is equal to $O(n \cdot \log(n)) + O(n^2) + O(n) \cdot T_{\text{exact}}$. Since $O(n \cdot \log(n)) + O(n^2) = O(n^2)$, we thus conclude that the algorithm takes time

$$O(n^2) + O(n) \cdot T_{\text{exact}}.$$

Conclusion. If we can compute V_U for exactly known x_i in polynomial time, then we can compute $\underline{V}_U$ under interval uncertainty also in polynomial time. For example:

- if we could compute V_U for exact x_i in linear time $O(n)$, then we can compute $\underline{V}_U$ for interval x_i in quadratic time $O(n^2)$;
- if we could compute V_U for exact x_i in quadratic time $O(n^2)$, then we can compute $\underline{V}_U$ for interval x_i in cubic time $O(n^3)$.

10 Computing $\overline{V}_U$: Analysis of the Problem

Where does the function V_U attains its maximum? Similar analysis of the problem of computing the maximum $\overline{V}_U$ of the function (23) leads to the following conclusion:

- if $\overline{x}_i \le m$, then we have $x_i = \underline{x}_i$;
- if $m \le \underline{x}_i$, then we have $x_i = \overline{x}_i$;
- if $\underline{x}_i \le m \le \overline{x}_i$, then we can have both $x_i = \underline{x}_i$ and $x_i = \overline{x}_i$.

Resulting algorithm is not feasible for large n. So, in principle, we can find $\overline{V}_U$ by trying all possible combinations of endpoints that satisfy the above conditions, and selecting the largest of the appropriate values V_U.

The problem with this idea is that, in general, we have two possibilities for each i, so overall, we may have an exponential number 2^n of combinations. Even for reasonable-size n, e.g., for $n = 300$, the number of combinations exceeds the number of particles in the Universe and thus, cannot be feasibly computed.

This is in line with the above fact that even for the case when $U(d) = d^2$, the problem of computing $\overline{V}_U$ is NP-hard.

Cases when a feasible algorithm is possible. However, there are practically important cases when we *can* compute $\overline{V}_U$ in polynomial time.

First case. The first case is when there is a constant C such that every group of $> C$ intervals has an empty intersection.

In this case, for each zone, there are $\le C$ intervals for which $\underline{x}_i \le m \le \overline{x}_i$, so we need to check $\le 2^C$ combinations for each zone. Since C is a constant, this means $O(1)$ and not affecting the asymptotic computation time.

Second case. The second case is when no interval is a proper subinterval of another, i.e., when $[\underline{x}_i, \overline{x}_i] \not\subseteq (\underline{x}_j, \overline{x}_j)$ for all i and j.

This happens, e.g., when all the measurements are made by the same measuring instrument. A measuring instrument can have different accuracy at different parts of the scale, e.g., it may lead to a narrower interval $[0.59, 0.61]$ in one part of the scale and wider interval $[1.2, 1.4]$ at another part. However, it is not realistic to expect two intervals $[0.59, 0.61]$ and $[0.1, 1.2] \supseteq [0.59, 0.61]$ produced by the same measuring instrument.

Under this no-subinterval property, as one can check, lexicographic order

$$[\underline{x}_i, \overline{x}_i] \le [\underline{x}_j, \overline{x}_j] \Leftrightarrow ((\underline{x}_i < \underline{x}_j) \vee (\underline{x}_i = \underline{x}_j \,\&\, \overline{x}_i < \overline{x}_j))$$

sorts the intervals by both the left- and the right endpoints:

$$\underline{x}_1 \le \underline{x}_2 \le \ldots \le \underline{x}_n; \quad \overline{x}_1 \le \overline{x}_2 \le \ldots \le \overline{x}_n.$$

In this case, for all intervals for which $\overline{x}_i \le m$, we have $x_i = \underline{x}_i$, and for all intervals for which $m < \underline{x}_i$, we have $x_i = \overline{x}_i$, For intermediate intervals, we may have both $x_i = \underline{x}_i$ and $x_i = \overline{x}_i$.

Let us show that among all the tuples on which the maximum is attained, there is always a tuple of the type $(\underline{x}_1, \ldots, \underline{x}_k, \overline{x}_{k+1}, \ldots, \overline{x}_n)$, i.e., a tuple in which we first have only lower endpoints, and then all upper endpoints.

Indeed, let us assume that the maximum is attained on some tuple for which $x_i = \overline{x}_i$ and $x_j = \underline{x}_j$ for some $j > i$. If the two intervals coincide, then we can swap them and eliminate this problem. Thus, it is sufficient to consider the case when the intervals are different.

In this case, we cannot have $\overline{x}_i < m$ because then, we would have $x_i = \underline{x}_i$, so $m \le \overline{x}_i$. Similarly, we cannot have $m < \underline{x}_i$ because then, due to the above ordering property, we would have $m < \underline{x}_i \le \underline{x}_j$ hence $m < \underline{x}_j$ and $x_j = \overline{x}_j$, Thus, we have $\underline{x}_i \le m \le \overline{x}_i$. Similarly, we can prove that in this case, $\underline{x}_j \le m \le \overline{x}_j$, i.e., that

$$\underline{x}_i \le \underline{x}_j \le m \le \overline{x}_i \le \overline{x}_j.$$

The maximum is attained when $x_i = \overline{x}_i$ and $x_j = \underline{x}_j$. Here, both values $\overline{x}_i$ and $\underline{x}_j$ belong to both intervals $[\underline{x}_i, \overline{x}_i]$ and $[\underline{x}_j, \overline{x}_j]$. The value V_U does not change if we simply swap two values x_i and x_j, i.e., take $x_i = \underline{x}_j$ and $x_j = \overline{x}_i$. Since the intervals are different, we cannot have both $x_i = \underline{x}_i$ and $x_j = \overline{x}_j$, so either $x_i > \underline{x}_i$ or $x_j < \overline{x}_j$. We already know that in this case, maximum cannot be attained.

Thus, it is sufficient to check only the tuples of the type $(\underline{x}_1, \ldots, \underline{x}_k, \overline{x}_{k+1}, \ldots, \overline{x}_n)$. There are $n + 1$ such tuples, so we have a polynomial-time algorithm.

Third case. Similar arguments can be made when the intervals can be divided into a fixed number m of groups within each of which there is a no-subinterval property. This can happen, e.g., when all the measurements are made by m different measuring instruments.

In this case, we can similarly sort intervals corresponding to each group (i.e., each measuring instrument), so it is sufficient to pick a transition point k_j for each of the groups $j = 1, \ldots, m$.

Thus, we arrive at the following algorithms.

11 Efficient Algorithms for Computing $\overline{V}_U$

First algorithm. This algorithm is applicable to the case when for some integer C, every subset of $> C$ intervals $[\underline{x}_i, \overline{x}_i]$ has an empty intersection. The algorithm is as follows.

First, we sort all $2n$ endpoints $\underline{x}_i$ and $\overline{x}_i$ into an increasing sequence, and add the values $x_{(0)} = -\infty$ and $x_{(2n+1)} = +\infty$, resulting in:

$$-\infty = x_{(0)} \le x_{(1)} \le x_{(2)} \le \ldots \le x_{(2n)} \le x_{(2n+1)} = +\infty.$$

For each zone $[x_{(k)}, x_{(k+1)}]$, we do the following:

- if $\overline{x}_i \le r_{(k)}$, then we select $x_i = \underline{x}_i$;
- if $r_{(k+1)} \le \underline{x}_i$, then we select $x_i = \overline{x}_i$;

- for all other i, we select either $x_i = \underline{x}_i$ or $x_i = \overline{x}_i$.

For each zone, we have $\le C$ indices i that allow two selections, so we thus get $\le 2^C$ selections. For each of these selections, we compute the U-deviation. The largest of these U-deviations is the desired value $\overline{V}_U$.

This algorithm requires time $O(n^2) + O(n) \cdot T_{\text{exact}}$.

Second algorithm. This algorithm is applicable to the case when no two intervals are proper subintervals of each other, i.e., when $[\underline{x}_i, \overline{x}_i] \not\subseteq (\underline{x}_j, \overline{x}_j)$ for all i and j.

In this case, first, we sort all the intervals in lexicographic order, i.e., by the order

$$[\underline{x}_i, \overline{x}_i] \le [\underline{x}_j, \overline{x}_j] \Leftrightarrow ((\underline{x}_i < \underline{x}_j) \vee (\underline{x}_i = \underline{x}_j \,\&\, \overline{x}_i < \overline{x}_j)).$$

We then consider all $n+1$ tuples of the form $(\underline{x}_1, \ldots, \underline{x}_k, \overline{x}_{k+1}, \ldots, \overline{x}_n)$, with $k = 0, 1, \ldots, n$. For each of these tuples, we compute the U-deviation. The largest of these U-deviations is the desired value $\overline{V}_U$.

This algorithm requires time $O(n \cdot \log(n)) + O(n) \cdot T_{\text{exact}}$.

Third algorithm. This algorithm is applicable if for some m, all the intervals can be divided into m groups each of which satisfies the above no-subinterval property. In this case, we sort all intervals within each group in lexicographic order. For each group $j = 1, \ldots, m$, with $n_j \le n$ elements, we consider $n_j + 1 \le n + 1$ tuples of the form $(\underline{x}_1, \ldots, \underline{x}_{k_j}, \overline{x}_{k_j+1}, \ldots, \overline{x}_n)$, and we consider all possible combinations of such tuples corresponding to all possible vectors $(k_1, \ldots, k_m)$. For each of these $\le n^m$ vectors, we compute the U-deviation. The largest of these U-deviations is the desired value $\overline{V}_U$.

This algorithm requires time $O(n \cdot \log(n)) + O(n^m) \cdot T_{\text{exact}}$.

12 What Are the Reasonable Measures of Dependence for Heavy-Tailed Distributions?

Formulation of the problem. If we have several possibly related samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, then, in addition to knowing how much each sample deviates from its "average", it is also desirable to know how much they depend on each other.

In the traditional statistics, a reasonable measure of dependence is the correlation, which is defined as

$$\rho_{xy} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - m_x) \cdot (y_i - m_y)}{\sqrt{V_x \cdot V_y}}.$$

This correlation describes linear dependencies.

For heavy-tailed distributions, as we have mentioned, variances are infinite, so this formula cannot be applied. Thus, we need to come up with a numerical characteristic for describing dependence.

One possibility: use Kendall's tau. The traditional correlation only describes linear dependence.

To describe possibly non-linear monotonic dependencies, we can use, e.g., Kendall's tau (see, e.g., [36]) – which can be estimated as the proportion of pairs (i, j) for which x and y change in the same direction, i.e.

- either $x_i \le x_j$ and $y_i \le y_j$
- or $x_j \le x_i$ and $y_j \le y_i$.

Kendall's tau can be applied (and has been applied) to heavy-tailed distributions as well.

Remaining problem. But what is we are interested not in all possible monotonic dependencies, but only in linear ones, or, more generally, only in dependencies $y = f(x)$ belonging to a certain class of functions $\mathcal{F}$ (e.g., all quadratic functions, or all fractionally linear functions).

Our idea. Let us again take into account disutility. The above measure of deviation estimates the disutility of replacing all the values x_i with a single value m_x, and the disutility of replacing all the values y_i with a single value m_y. Dependence means that if we know x_i, we can get a better approximation for y_i than m_y.

For example, if we want to predict temperature in El Paso, then we approximate this temperature by an average value and get some deviation. However, we know that there is a correlation between the temperature in El Paso and the temperature in the nearby city of Las Cruces. Thus means that if we know the temperature in Las Cruces, we can predict the temperature in El Paso better than by simply taking the average of El Paso temperatures.

In general, to approximate the values y_i,

- instead of using a single value m_y (and selecting the value for which the expected disutility is the smallest),
- we use the value $f(x_i)$ for an appropriate function $f \in \mathcal{F}$ – and we select the function f for which the expected disutility is the smallest possible.

Thus, we arrive at the following definitions:

Resulting definitions. Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be two tuples, let $U(d) \ge 0$ be a utility function, and let $\mathcal{F}$ be a class of functions from real numbers to real numbers.

By an *$\mathcal{F}$-regression*, we mean a function $f \in \mathcal{F}$ for which the value

$$\frac{1}{n} \cdot \sum_{i=1}^{n} U(y_i - f(x_i)) \tag{26}$$

is the smallest possible.

In particular, when $\mathcal{F}$ is the class of all constant functions, we get the U-estimate. When $U(d) = d^2$ and $\mathcal{F}$ is the class of all linear functions, we get the usual linear regression.

By a $(U, \mathcal{F})$-*correlation* c, we mean the proportion of how much the average disutility decreases when we use x_i to help predict the values y_i, i.e.,

$$c \stackrel{\text{def}}{=} \frac{V_U(y) - V_{U,\mathcal{F}}(y|x)}{V_U(y)},$$

where

$$V_U(y) \stackrel{\text{def}}{=} \min_m \frac{1}{n} \cdot \sum_{i=1}^n U(y_i - m)$$

and

$$V_{U,\mathcal{F}}(y|x) \stackrel{\text{def}}{=} \min_{f \in \mathcal{F}} \frac{1}{n} \cdot \sum_{i=1}^n U(y_i - f(x_i)).$$

Observation. For the class of linear functions $\mathcal{F}$ and for $U(d) = d^2$, the resulting value c coincides with the square ρ^2 of the usual correlation.

Discussion. For normal distributions, correlation is symmetric: if we can reconstruct y_i from x_i, then we can reconstruct x_i from y_i. Our definition is, in general, not symmetric. This asymmetry make perfect sense. For example, suppose that $y_i = x_i^2$.

- Then, if we know x_i, then we can uniquely reconstruct y_i, so the reconstruction of y_i from x_i is perfect.
- However, if we know y_i, we can only reconstruct x_i modulo sign, so the reconstruction of x_i from y_i is not perfect.

Remaining open problem. It is desirable to come up with efficient algorithms that would estimate the above measures of dependence under interval uncertainty.

References

[1] J. Beirlant, Y. Goegevuer, J. Teugels, and J. Segers, *Statistics of Extremes: Theory and Applications*, Wiley, Chichester, 2004.

[2] N. Buntao, "Estimating parameters of Pareto distribution under interval and fuzzy uncertainty", *Proceedings of the 30th Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2011*, El Paso, Texas, March 18–20, 2011, to appear.

[3] N. Buntao and V. Kreinovich, "Measures of Deviation (and Dependence) for Heavy-Tailed Distributions and their Estimation under Interval and Fuzzy Uncertainty", submitted to *World Conference on Soft Computing*, San Francisco, CA, May 23–26, 2011, to appear.

[4] B. K. Chakrabarti, A. Chakraborti, and A. Chatterjee, *Econophysics and Sociophysics: Trends and Perspectives*, Wiley-VCH, Berlin, 2006.

[5] A. Chatterjee, S. Yarlagadda, B. K. Chakrabarti, *Econophysics of Wealth Distributions*, Springer-Verlag Italia, Milan, 2005.

[6] C. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, Boston, Massachusetts, 2009.

[7] J. D. Farmer and T. Lux (eds.), *Applications of statistical physics in economics and finance*, a special issue of the *Journal of Economic Dynamics and Control*, 2008, Vol. 32, No. 1, pp. 1–320.

[8] S. Ferson, L. Ginzburg, V. Kreinovich, L. Longpré, and M. Aviles, "Computing Variance for Interval Data is NP-Hard", *ACM SIGACT News*, 2002, Vol. 33, No. 2, pp. 108–118.

[9] S. Ferson, L. Ginzburg, V. Kreinovich, L. Longpré, and M. Aviles, "Exact Bounds on Finite Populations of Interval Data", *Reliable Computing*, 2005, Vol. 11, No. 3, pp. 207–233.

[10] P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.

[11] P. C. Fishburn, *Nonlinear Preference and Utility Theory*, The John Hopkins Press, Baltimore, Maryland, 1988.

[12] X. Gabaix, G. Parameswaran, P. Vasiliki, and H. E. Stanley, "Understanding the cubic and half-cubic laws of financial fluctuations", *Physica A*, 2003, Vol. 324, pp. 1–5.

[13] X. Gabaix, G. Parameswaran, P. Vasiliki, and H. E. Stanley, "A theory of power-law distributions in financial market fluctuations", *Nature*, 2003, Vol. 423, No. 6937, pp. 267–270.

[14] C. P. Gomez and D. B. Shmoys, "Approximations and Randomization to Boost CSP Techniques', *Annals of Operations Research*, 2004, Vol. 130, pp. 117–141.

[15] C. Hu, R. B. Kearfott, A. de Korvin, and V. Kreinovich (eds.), *Knowledge Processing with Interval and Soft Computing*, Springer Verlag, London, 2008.

[16] P. J. Huber, *Robust Statistics*, Wiley, Hoboken, New Jersey, 2004.

[17] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, London, 2001.

[18] R. B. Kearfott and V. Kreinovich (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 1996.

[19] R. L. Keeney and H. Raiffa, *Decisions with Multiple Objectives*, John Wiley and Sons, New York, 1976.

[20] S. Kinsella and F. O'Brien, "Maximum likelihood estimation of stable Paretian distribution applied to index and option data", *Proceedings of the INFINITI Conference on International Finance*, Dublin, Ireland, June 8–9, 2009.

[21] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1998.

[22] R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

[23] B. Mandelbrot, "The variation of certain speculative prices", *J. Business*, 1963, Vol. 36, pp. 394–419.

[24] B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, California, 1983.

[25] B. Mandelbrot and R. L. Hudson, *The (Mis)behavior of Markets: A Fractal View of Financial Turbulence*, Basic Books, 2006.

[26] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, Cambridge, Massachusetts, 1999.

[27] N. Markovich (ed.), *Nonparametric Analysis of Univariate Heavy-Tailed Data: Research and Practice*, Wiley, Chichester, 2007.

[28] J. McCauley, *Dynamics of Markets, Econophysics and Finance*, Cambridge University Press, Cambridge, Massachusetts, 2004.

[29] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM Press, Philadelphia, Pennsylvania, 2009.

[30] C. Papadimitriou, *Computational Complexity*, Addison Welsey, Reading, Massachusetts, 1994.

[31] S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer Verlag, New York, 2005.

[32] S. T. Rachev and S. Mittnik, *Stable Paretian Models in Finance*, Wiley Publishers, New York, 2000.

[33] H. Raiffa, *Decision Analysis*, Addison-Wesley, Reading, Massachusetts, 1970.

[34] S. I. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer-Varlag, New York, 2007.

[35] B. Roehner, *Patterns of Speculation - A Study in Observational Econophysics*, Cambridge University Press, Cambridge, Massachusetts, 2002.

[36] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2007.

[37] S. Sriboonchitta, H. T. Nguyen, and V. Kreinovich, "How to Relate Spectral Risk Measures and Utilities", *International Journal of Intelligent Technologies and Applied Statistics*, 2010, Vol. 3, No. 2, pp. 141–158.

[38] H. E. Stanley, "Econophysics and the current economic turmoil", *American Physical Society News*, 2008, Vol. 17, No. 11, p. 8.

[39] H. E. Stanley, L. A. N. Amaral, P. Gopikrishnan, and V. Plerou, "Scale invariance and universality of economic fluctuations", *Physica A*, 2000, Vol. 283, pp. 31–41.

[40] S. V. Stoyanov, B. Racheva-Iotova, S. T. Rachev, and F. J. Fabozzi, "Stochastic models for risk estimation in volatile markets: a survey", *Annals of Operations Research*, 2010, Vol. 176, pp. 293–309.

[41] P. Vasiliki and H. E. Stanley, "Stock return distributions: tests of scaling and universality from three distinct stock markets", *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2008, Vol. 77, No. 3, Pt. 2, Publ. 037101.