1-2011

# Measures of Deviation (and Dependence) for Heavy-Tailed Distributions and heir Estimation under Interval and Fuzzy Uncertainty

Nitaya Buntao

Vladik Kreinovich
*The University of Texas at El Paso*, vladik@utep.edu



Comments:
Technical Report: UTEP-CS-11-07
Published in: Ronald R. Yager, Marek Z. Reformat, Shahnaz N. Shahbazova, and Sergei Ovchinnikov (eds.), *Proceedings of the World Conference on Soft Computing*, San Francisco, CA, May 23-26, 2011.

# Measures of Deviation (and Dependence) for Heavy-Tailed Distributions and their Estimation under Interval and Fuzzy Uncertainty

Nitaya Buntao and Vladik Kreinovich

*Abstract*— **Traditionally, in science and engineering, most statistical techniques are based on the assumption that the random variables are normally distributed. For such distributions, a natural characteristic of the "average" value is the mean, and a natural characteristic of the deviation from the average is the variance. However, in many practical situations, e.g., in economics and finance, we encounter probability distributions for which the variance is infinite; such distributions are called *heavy-tailed*. For such distributions, we describe which characteristics can be used to describe the average and the deviation from the average, and how to estimate these characteristics under interval and fuzzy uncertainty. We also discuss what are the reasonable analogues of correlation for such heavy-tailed distributions.**

## I. Introduction to the Problem.

**Normal distributions are most widely used.** Traditionally, in science and engineering, most statistical techniques are based on the assumption that the random variables are normally distributed, with the probability density

$$\rho(x) = \frac{1}{\sqrt{2\pi \cdot V}} \cdot \exp\left(-\frac{(x-m)^2}{2V}\right);$$

see, e.g., [33].

For such distributions, a natural characteristic of the "average" value is the mean $m \stackrel{\rm def}{=} E[x]$, and a natural characteristic of the deviation from the average is the variance $V \stackrel{\rm def}{=} E[(x-m)^2]$.

In principle, we can think of other possible characteristics such as mode or median. However, it is known that a normal distribution is uniquely determined by its first two moments $m = E[x]$ and $M \stackrel{\rm def}{=} E[x^2]$; thus, each characteristic is uniquely determined by $m$ and $M$. Since it is known that $M = V + m^2$, we can thus conclude that every characteristic can be described in terms of $m$ and $V$.

Nitaya Buntao is with the Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, 1518 Piboonsongkhram Road, Bangsue, Bangkok 10800 Thailand (email: taltanot@hotmail.com). Vladik Kreinovich is with the Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, TX9968, USA (email: vladik@utep.edu).

This work was supported in part by the National Science Foundation grants HRD-0734825 and DUE-0926721, by Grant 1 T36 GM078000-01 from the National Institutes of Health, by Grant MSM 6198898701 from MŠMT of Czech Republic, and by Grant 5015 "Application of fuzzy logic with operators in the knowledge based systems" from the Science and Technology Centre in Ukraine (STCU), funded by European Union. The work of N. Buntao was supported by a grant from the Office of the Higher Education Commission, Thailand, under the Strategic Scholarships for Frontier Research Network. The authors are thankful to Hung T. Nguyen and Sa-aat Niwitpong for valuable discussions.

**Estimating the value of the characteristics: case of normal distributions.** For the case of normal distributions, once we have a sample consisting of the values $x_1, \ldots, x_n$, we can use the Maximum Likelihood Method to find the estimates for $m$ and $V$. According to this method, we find the values $m$ and $V$ for which the corresponding probability density

$$L = \rho(x_1) \cdot \ldots \cdot \rho(x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \cdot V}} \cdot \exp\left(-\frac{(x_i-m)^2}{2V}\right).$$

Maximizing this probability is equivalent to minimizing the value

$$\psi \stackrel{\rm def}{=} -\ln(L) = \sum_{i=1}^{n} \left[\frac{1}{2} \cdot \ln(2\pi \cdot V) + \frac{(x_i-m)^2}{2V}\right].$$

It is known that if we differentiate this expression by $m$ and $V$ and equate the corresponding derivatives to 0, then we get the following formulas:

$$m = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i; \quad V = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - m)^2.$$

**In many practical situations, we encounter heavy-tailed distributions.** In many practical situations, e.g., in economics and finance, we encounter probability distributions for which the variance is infinite; such distributions are called *heavy-tailed*. These distributions surfaced In the 1960s, Benoit Mandelbrot, the author of fractal theory, empirically studied the fluctuations and showed [20] that larger-scale fluctuations follow the 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., [21], [22].

Since then, similar heavy-tailed distributions have been empirically found in other financial situations [2], [3], [5], [10], [23], [25], [32], [36], [37], [38], and in many other application areas [1], [12], [21], [24], [31].

**First problem: how to characterize such distributions?** For such distributions, we cannot use variance to describe the deviation from the "average". Thus, we need to come up with other characteristics for describing this deviation.

We will describe such characteristics in the first part of this paper. We will also describe how we can estimate these characteristics.

**Need to take into account interval uncertainty.** The above estimators for $m$ and $V$ are based on the simplifying assumption that the sample values $x_i$ are known exactly.

In practice, we often know the values $x_i$ only approximately. In other words, instead of the *exact* value of $x_i$, we only know the *approximate* estimation $\widetilde{x}_i$. We also have some information about the approximation error $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$.

In some cases, we know the probability distribution of different values of the approximation error. However, in many practical situations, we only know the upper bound $\Delta_i$ on this error, i.e., the value for which $|\Delta x_i| \le \Delta_i$; see, e.g., [29].

In such situations, the only information that we have about the actual (unknown) value $x_i$ is that $x_i$ belongs to the interval $\mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$. Because of this, such uncertainty is also known as an *interval uncertainty*.

Interval uncertainty also naturally appears in the analysis of financial data; see, e.g., [13] and references therein. For example, in the analysis of stock market data, each sample value $x_i$ may represent the price of a certain stock on the $i$-th day. In reality, the price of each stock slightly fluctuates during the day.

Usually, practitioners take, as $x_i$, the average price or the price at a certain specific time. The problem is that there are several possibilities of select a single day price, and different selections lead to (slightly) different results. It is therefore reasonable, instead of artificially picking one number $x_i$, to consider the entire interval $[\underline{x}_i, \overline{x}_i]$ of all possible prices offered during the $i$-th day.

As shown in [13], not only this approach more reasonable – the resulting use of the additional information about daily variances of stock prices leads to a better predictions of future stock values.

For each estimator $C(x_1, \ldots, x_n)$, different combinations of values $x_i \in \mathbf{x}_i$ lead, in general, to different values of $C(x_1, \ldots, x_n)$.

It is therefore desirable to find the range $[\underline{C}, \overline{C}]$ of possible values of $C$:

$$\mathbf{C} = [\underline{C}, \overline{C}] = \{C(x_1, \ldots, x_n) : x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}.$$

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., [15], [26].

In spite of the simplicity of the problem's formulation, in general, the interval computations problem is NP-hard (computationally intensive [28]); see, e.g., [18].

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 above function $V(x_1, \ldots, x_n)$ that describes the sample variance [6], [7].

**Case of fuzzy uncertainty.** Not all the values within the interval $[\underline{x}_i, \overline{x}_i]$ may be equally reasonable to consider. Some of these values may be flukes caused by accidental errors. While it is difficult to decide for sure, financial experts can usually tell to what extent the corresponding values are possible. This extent is usually formulated not in precise terms, but by using words from a natural language. For example, an expert may say that some values are most probably flukes, while some other values are most probably reasonable.

To describe these natural-language statements, it is reasonable to use *fuzzy logic*, a formalism specifically designed to formalize such statements; see, e.g., [17], [27], [39]. Based on the information about the possibility of different values $x_i \in [\underline{x}_i, \overline{x}_i]$, it is desirable to conclude what is the degree of possibility of different values $C(x_1, \ldots, x_n)$ from the corresponding intervals.

An alternative way to describe a membership function $\mu_i(x_i)$ is to describe, for each possible value $\alpha \in [0, 1]$, the set of all values $x_i$ for which the degree of possibility is at least $\alpha$. This set $\{x_i : \mu_i(x_i) \ge \alpha\}$ is called an *alpha-cut* and is denoted by $X_i(\alpha)$.

It is known (see, e.g., [17], [27]), that for alpha-cuts, Zadeh's extension principle takes the following form: for every $\alpha$, we have

$$R(\alpha) = \{R(x_1, \ldots, x_n) : x_i \in X_i(\alpha)\}.$$

Thus, for every $\alpha$, finding the alpha-cut of the resulting membership function $\mu(R)$ is equivalent to applying interval computations to the corresponding intervals $X_1(\alpha)$, …, $X_n(\alpha)$.

Because of this reduction, in the following text, we will only consider the case of interval uncertainty. So, we arrive at the following problem.

**Second problem.** How can we estimate the values of the heavy-tailed deviation characteristic under interval and fuzzy uncertainty? For characteristics described in the first part of this paper, in the second part, we describe how to compute them under this uncertainty.

## II. How to Describe Deviation from the "Average" for Heavy-Tailed Distributions

**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., [8], [9], [16], [19], [30]. According to decision theory, a user prefers an alternative for which the expected value $\sum_{i=1}^{n} p_i \cdot u_i$ of the utility is the largest possible. Alternative, we can say that the expected value $\sum_{i=1}^{n} p_i \cdot U_i$ 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: $d_1 \leq d_2$ implies $U(d_1) \leq U(d_2)$. Similarly, if the difference $d$ is negative, the smaller $d$, the larger the disutility: $d_1 \leq d_2$ implies $U(d_1) \geq U(d_2)$.

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 $p_i = \dfrac{1}{n}$; thus, the expected value of the disutility is equal to

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

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

For each sample $x_1, \ldots, x_n$, by a *U-estimate*, we mean the value $m_U$ that minimizes the expression (1). By a *U-deviation*, we mean the value

$$V_U \stackrel{\text{def}}{=} \min_m \frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - m). \tag{2}$$

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

$$V_U = \frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - m_U). \tag{3}$$

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

When $U(x) = |x|$, the expression turns into the expression $\dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} |x_i - m|$ 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 (1); 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., [14]), 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 [34].

## III. 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 $\dfrac{df}{dx} \geq 0$.

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

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

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

- either the minimum is attained for $x = \underline{x}$ and $\dfrac{df}{dx} \geq 0$;
- or the minimum is attained for $x = \overline{x}$ and $\dfrac{df}{dx} \leq 0$;
- 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 $\dfrac{\partial f}{\partial x_i} \geq 0$;
- or $x_i = \overline{x}_i$ and $\dfrac{\partial f}{\partial x_i} \leq 0$;
- 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 $\dfrac{\partial f}{\partial x_i} \leq 0$;
- or $x_i = \overline{x}_i$ and $\dfrac{\partial f}{\partial x_i} \geq 0$;
- 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 (1), we conclude that for this value, the derivative of the expression (1) with respect to $m$ is equal to 0, i.e., that

$$-\frac{1}{n} \cdot \sum_{i=1}^{n} U'(x_i - m) = 0, \tag{4}$$

where $U'(d)$ denotes the derivative of the function $U(d)$. Differentiating the expression (3) 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 (4), the expression in parentheses is equal to 0 and thus,

$$\frac{\partial V_U}{\partial x_i} = U'(x_i - m). \tag{5}$$

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 $\overline{x}_i \le m$, then we have $x_i = \overline{x}_i$;
- if $m \le \underline{x}_i$, then we have $x_i = \underline{x}_i$;
- if $\underline{x}_i \le m \le \overline{x}_i$, 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 (1) 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 $x_i \ne m$. Thus, $m$ is equal to the $U$-average of such values $x_i$. So, we arrive at the following algorithm.

## IV. 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., [4]. 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 $T_{\text{exact}}$ – 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 (hence fuzzy) 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 $O(n^3)$.

## V. 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 (3) 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.

## VI. 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 $O(n \cdot \log(n)) + O(n^m) \cdot T_{\text{exact}}$.

## VII. 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., [33]) – 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

$$\frac{1}{n} \cdot \sum_{i=1}^{n} U(y_i - f(x_i)) \tag{6}$$

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 $\rho^2$ 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 and fuzzy uncertainty.

## REFERENCES


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

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

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

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

[5] 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.

[6] 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.

[7] 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.

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

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

[10] 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.

[11] 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.

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

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

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

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

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

[17] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, Upper Saddle River, New Jersey, 1995.

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

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

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

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

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

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

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

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

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

[27] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Boca Raton, Florida: Chapman & Hall/CRC, 2006.

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

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

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

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

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

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

[34] 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.

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

[36] 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.

[37] 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.

[38] 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.

[39] L. A. Zadeh, "Fuzzy sets", *Information and control*, vol. 8, pp. 338–353, 1965.