4-2012

Optimizing Computer Representation and Computer Processing of Epistemic Uncertainty for Risk-Informed Decision Making: Finances etc.

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

Nitaya Buntao
King Mongkut's Institute of Technology North Bangkok

Olga Kosheleva
The University of Texas at El Paso, olgak@utep.edu

Comments:
Technical Report: UTEP-CS-12-04a
Published in *Proceedings of the International Conference on Probabilistic Safety Assessment and Management / European Safety and Reliability PSAM11/ESREL'12*, Helsinki, Finland, June 25-29, 2012.

Optimizing Computer Representation and Computer Processing of Epistemic Uncertainty for Risk-Informed Decision Making: Finances etc.

Vladik Kreinovich[a*], Nitaya Buntao[b], Olga Kosheleva[a]
[a]University of Texas at El Paso, El Paso, Texas 79968, USA
[b]King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

Abstract: Uncertainty is usually gauged by using standard statistical characteristics: mean, variance, correlation, etc. Then, we use the known values of these characteristics (or the known bounds on these values) to select a decision. Sometimes, it becomes clear that the selected characteristics do not always describe a situation well; then other known (or new) characteristics are proposed. A good example is description of volatility in finance: it started with variance, and now many descriptions are competing, all with their own advantages and limitations.

In such situations, a natural idea is to come up with characteristics tailored to specific application areas: e.g., select the characteristic that maximize the expected utility of the resulting risk-informed decision making.

With the new characteristics, comes the need to estimate them when the sample values are only known with interval uncertainty. Algorithms originally developed for estimating traditional characteristics can often be modified to cover new characteristics.

Keywords: Uncertainty, application-tailored characteristics, interval uncertainty

1. NEED FOR APPLICATION-TAILORED STATISTICAL CHARACTERISTICS

To make a proper decision, we need to know the state of the world, i.e., we need to know the values of the physical quantities that characterize this state. These values usually come from measurements. Measurement are never absolutely accurate; see, e.g., [30]. As a consequence, the measurement result $\widetilde{x}$ is, in general, different from the actual (unknown) value x of the desired physical quantity. In other words, in practice, we only know the values of the desired quantities with *uncertainty*.

Once we know the measurement result $\widetilde{x}$, we need to know the upper bound Δ on the (absolute value of) measurement error $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$. Indeed, if no such upper bound is known, this means that the difference between the actual value and the measurement result can be arbitrarily large – this is not a measurement, this is a wild guess. For the procedure to be called "measurement", we need to know this upper bound.

Once we know the upper bound Δx (for which $|\Delta x| \le \Delta$), we can conclude that the measurement error Δx belongs to the interval $[-\Delta, \Delta]$ and thus, that the actual value x belongs to the interval $[\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.

In practice, different values $\Delta x \in [-\Delta, \Delta]$ occur with different frequency. It is therefore desirable, in addition to the upper bound Δ, to also know the probability of different values $\Delta x \in [-\Delta, \Delta]$. In other words, to describe uncertainty, it is desirable to describe a probability distribution for the measurement error.

In many practical situations, the measurement error is a result of many independent factors, each of which is reasonably small. It is known (see, e.g., [35]), that, under reasonable assumption, the sum of a large number of small independent random variables tends to a normal distribution. This result

is known as a Central Limit Theorem. Thus, in such situations, the probability distribution of the measurement error is close to normal.

A normal distribution is uniquely determined by its mean and standard deviation. Because of this, to describe the probability distribution of a measurement error, we traditionally use mean and standard deviation. The mean is called a *systematic error*, and the standard deviation is said to characterize the *random error*; see, e.g., [30].

To describe a normal distribution of several variables, in addition to their means and standard deviations, we also need to know their covariances or, alternatively, their pairwise correlations. Thus, these covariances (or correlations) are traditionally used in the description of *epistemic* uncertainty (i.e., uncertainty with which we know the actual values).

To make a proper decision, we need to know not only the *current* state of the world, we also need to be able to predict the *future* state of the world, the state that will occur as a result of different actions. Often, the future state depends not only on the current state, but also on many difficult-to-measure factors. For example, in weather prediction, the future temperature depends not only on the current meteorological characteristics such as temperature, humidity, etc., at the sensor locations, it also depends on the physical processes in the oceans, in the Arctic, and in many other places where sensors are rare.

In such situations, even if we know the exact (or at least very accurate) values of the physical quantities that characterize the current state of the world, we cannot exactly predict their future values. In different situations, we may have slightly different future values. Once we have collected a large number of such situations, we can talk about the frequencies (probabilities) of different future states. These *aleatoric* (relative-frequency) probabilities reflect the probability distribution of the un-measured factors that affect the transition from the current to the future state.

Similarly to the case of epistemic uncertainty, the aleatoric uncertainty is also often caused by a joint effect of a large number of small independent factors. In such situations, due to the Central Limit Theorem, the corresponding probability distributions are close to normal. As a result, these distributions are characterized by mean, standard deviation, and covariance (or correlation).

In many practical situations, distributions, e.g., distributions of measurement errors, are indeed close to normal. However, there are many other practical situations in which the probability distribution is drastically different from normal. In many such situations, 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 [21] that large-scale fluctuations follow the Pareto power-law distribution, with the probability density function $\rho(x) = A \cdot x^{-\alpha}$ for $x \geq x_0$, for some constants $\alpha \approx 2.7$ and x_0. 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., [22, 23].

Since then, similar heavy-tailed distributions have been empirically found in other financial situations [3, 4, 5, 10, 24, 26, 31, 34, 38, 39, 40], and in many other application areas [1, 12, 22, 25, 33].

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.

This situation is typical in financial and economic applications, where this deviation is known as volatility. At first, economists followed a natural idea to use standard deviation as a quantitative measure of volatility. However, since the empirical distribution is heavy-tailed, its standard deviation is infinite, so other characteristics of volatility are needed.

In the following text, we will describe different characteristics of deviation from the average, and we will describe how we can estimate these characteristics – both when we know the exact sample values and when the sample values are only known with interval uncertainty.

2. HOW TO DESCRIBE DEVIATION FROM THE "AVERAGE" FOR GENERAL HEAVY-TAILED DISTRIBUTIONS

Of course, there are many possible mathematical definitions. Since our main motivation is to make decisions, it is desirable 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, 17, 20, 32]. According to decision theory, a user prefers an alternative for which the expected value $E[u] = \int \rho(x) \cdot u(x)\, dx$ of the utility is the largest possible. Alternative, we can say that the expected value $E[U] = \int \rho(x) \cdot U(x)\, dx$ of the *disutility* $U \stackrel{\text{def}}{=} -u$ is the smallest possible.

In our case, instead of considering all possible values x, we consider a single value m (e.g., the measurement result). Since we are replacing each original value x with a new value m which is only an approximation to x, there is some resulting disutility. For example, if we dress based on the expected average temperature m and the actual temperature is $x \neq m$, 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 $x \neq m$, 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, the larger the disutility. Let $U(d)$ denote the disutility cause by the difference $d = x - m$. When x 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 x, the disutility is equal to $U(x - m)$. Thus, the expected value of the disutility is equal to

$$\int \rho(x) \cdot U(x - m)\, dx. \tag{1}$$

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.

Definition 1. *By a* disutility function*, we mean a function $U(d) \ge 0$ from real numbers to non-negative real numbers for which $U(0) = 0$, $U(d)$ is (non-strictly) increasing for $d \ge 0$, and $U(d)$ is (non-strictly) decreasing for $d \le 0$.*

Definition 2. *For each probability distribution $\rho(x)$ and a disutility function $U(d)$, by a* U-mean*, we mean the value m_U that minimizes the expression (1). By a* U-deviation*, we mean the value $V_U \stackrel{\text{def}}{=} \min\limits_m \int \rho(x) \cdot U(x - m)\, dx$.*

Comment. Because of the definition of m_U, the value V_U takes the form $V_U = \int \rho(x) \cdot U(x - m_U)\, dx$.

Examples. When $U(x) = x^2$, the expression (1) turns into the expression $\int \rho(x) \cdot (x - m)^2\, dx$ for which minimization leads to the mean $m = \int \rho(x) \cdot x\, dx$. For this mean, the expression V_U is the usual variance.

When $U(x) = |x|$, the expression (1) turns into the expression $\int \rho(x) \cdot |x - m|\, dx$ for which minimization leads to the median. For the median m_U, the expression V_U is the *average absolute deviation* $V_U = \int \rho(x) \cdot |x - m_U|\, dx$.

3. HOW TO DESCRIBE DEPENDENCE FOR GENERAL HEAVY-TAILED DISTRIBUTIONS

How can we describe dependence for general heavy-tailed distributions? In the traditional statistics, a reasonable measure of dependence is the correlation, which is defined as $\rho_{xy} = \dfrac{E[(x - E(x)) \cdot (y - E(y))]}{\sqrt{V_x \cdot V_y}}$. 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.

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., [35]) – which can be estimated as the proportion of pairs of tuples (x, y) and (x', y') for which x and y change in the same direction, i.e., either $x \le x'$ and $y \le y'$, or $x' \le x$ and $y' \le y$. Kendall's tau can be applied (and has been applied) to heavy-tailed distributions as well.

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).

Let us again take into account disutility. The above measure of deviation estimates the disutility of replacing all the values x with a single value m_x, and the disutility of replacing all the values y with a single value m_y. Dependence means that if we know x, we can get a better approximation for y 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, instead of using a single value m_y (and selecting the value for which the expected disutility is the smallest), we use the value $f(x)$ 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:

Definition 3. *Let us assume that we have a random 2-D vector (x, y). Let $U(d) \ge 0$ be a disutility 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 $E[U(y - f(x))] = \int \rho(x, y) \cdot U(y - f(x))\, dx\, dy$ 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. It is now reasonable to define correlation as the proportion of how much the average disutility decreases when we use x to help predict the values y.

Definition 4. *By a $(U, \mathcal{F})$-correlation c, we mean the value $c \stackrel{\text{def}}{=} \dfrac{V_U(y) - V_{U,\mathcal{F}}(y|x)}{V_U(y)}$, where $V_U(y) \stackrel{\text{def}}{=} \min\limits_m \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} U(y_i - m)$ and $V_{U,\mathcal{F}}(y|x) \stackrel{\text{def}}{=} \min\limits_{f \in \mathcal{F}} \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} U(y_i - f(x_i))$.*

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

For normal distributions, correlation is symmetric: if we can reconstruct y from x, then we can reconstruct x from y. Our definition is, in general, not symmetric. This asymmetry make perfect

sense. For example, suppose that $y = x^2$. Then, if we know x, then we can uniquely reconstruct y, so the reconstruction of y from x is perfect. However, if we know y, we can only reconstruct x modulo sign, so the reconstruction of x from y is not perfect.

4. HOW TO ESTIMATE THE NEW CHARACTERISTICS FROM OBSERVATIONS

In the above text, we defined the desired characteristics in terms of the corresponding probability density functions $\rho(x)$ and $\rho(x, y)$. In practice, we often do not know the actual distribution, i.e., we do not know the probability density $\rho(x)$ (or, for two variables, $\rho(x, y)$). Instead, we know the sample values $x_1, \ldots, x_n$. How do we estimate the above characteristics based on the sample values?

A natural idea is to use the "histogram" distribution, i.e., the distribution in which each of n observed values $x_1, \ldots, x_n$ appears with equal probability $\dfrac{1}{n}$. This idea is behind the usual estimates for the mean and for the variance. Indeed, if we plug in the corresponding distribution $\rho(x) = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} \delta(x - x_i)$ into the definitions of the mean $E = \int \rho(x) \cdot x \, dx$ and variance $V = \int \rho(x) \cdot (x - E)^2 \, dx$, we get the usual estimates $\widehat{E} = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} x_i$ and $\widehat{V} = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} \left(x_i - \widehat{E}\right)^2$.

Applying this idea to the above formulas for the deviation from the mean, we get the following estimates.

Definition 5. *Let $U(d) \geq 0$ be a disutility function. For each sample $x_1, \ldots, x_n$, by an* estimate for the U-mean*, we mean the value $\widehat{m}_U$ that minimizes the expression*

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

By an estimate for U-deviation*, we mean the value $\widehat{V}_U \stackrel{\text{def}}{=} \min\limits_{m} \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} U(x_i - m)$.*

Comment. Because of the definition of $\widehat{m}_U$, the value $\widehat{V}_U$ takes the form $\widehat{V}_U = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} U\left(x_i - \widehat{m}_U\right)$.

Let us give two examples. When $U(x) = x^2$, the expression (2) 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 $\widehat{m} = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} x_i$. For this arithmetic average, the expression $\widehat{V}_U$ is the usual sample 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 sample median. For the sample median $\widehat{m}_U$, the expression $\widehat{V}_U$ is the *average absolute deviation* $\widehat{V}_U = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} \left|x_i - \widehat{m}_U\right|$.

How to estimate $\widehat{V}_U$? Once we compute $\widehat{m}_U$, the computation of $\widehat{V}_U$ is straightforward: we just apply the function $U(d)$ n times and compute the corresponding expression.

Estimating $\widehat{m}_U$ 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 $\psi \stackrel{\text{def}}{=} -\ln(L)=\sum\limits_{i=1}^{n}U(x_i-m)$, where we denoted $U(x)\stackrel{\text{def}}{=}-\ln(\rho_0(x))$. Minimizing this value is equivalent to minimizing the value (2); thus, this value is exactly our estimate $\widehat{m}_U$.

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 [36].

How to estimate measures of dependence? 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.

Definition 6. *Let $x_1,\ldots,x_n$ and $y_1,\ldots,y_n$ be two tuples, let $U(d)\ge 0$ be a disutility function, and let $\mathcal{F}$ be a class of functions from real numbers to real numbers. By an* estimate for the $\mathcal{F}$-regression, *we mean a function $\widehat{f}\in\mathcal{F}$ for which the value $\dfrac{1}{n}\cdot\sum\limits_{i=1}^{n}U\left(y_i-\widehat{f}(x_i)\right)$ 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.

Definition 7. *By an* estimate for the $(U,\mathcal{F})$-correlation c, *we mean the value $\widehat{c}\stackrel{\text{def}}{=}\dfrac{\widehat{V}_U(y)-\widehat{V}_{U,\mathcal{F}}(y|x)}{V_U(y)}$, where $\widehat{V}_U(y)\stackrel{\text{def}}{=}\min\limits_{m}\dfrac{1}{n}\cdot\sum\limits_{i=1}^{n}U(y_i-m)$ and $\widehat{V}_{U,\mathcal{F}}(y|x)\stackrel{\text{def}}{=}\min\limits_{f\in\mathcal{F}}\dfrac{1}{n}\cdot\sum\limits_{i=1}^{n}U(y_i-f(x_i))$.*

5. ESTIMATING THE NEW CHARACTERISTICS UNDER 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., [16, 15, 27].

In spite of the simplicity of the problem's formulation, in general, the interval computations problem is NP-hard (computationally intensive [29]); see, e.g., [19].

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 $\widehat{V}(x_1,\ldots,x_n)=\frac{1}{n}\cdot\sum_{i=1}^{n}x_i^2-\left(\frac{1}{n}\cdot\sum_{i=1}^{n}x_i\right)^2$ that describes the sample variance [6, 7].

Let us describe an algorithm for computing the lower endpoint $\underline{\widehat{V}_U}$ of the range $\left[\underline{\widehat{V}_U},\overline{\widehat{V}_U}\right]$ of possible values of $\widehat{V}_U$. In order to find $\underline{\widehat{V}_U}$, 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-mean and – if this U-mean is in the zone – take m equal to this U-mean and compute the corresponding U-deviation.

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

Proof of correctness of this algorithm is similar to the algorithm for variance; see, e.g., [28] and [2].

Computation time for this algorithm is $O(n^2)+O(n)\cdot T_{\text{exact}}$, where T_{exact} is the time to compute the U-mean and U-deviation for the given values x_i.

This formula leads us to the following conclusion: If we can compute V_U for exactly known x_i in polynomial time, then we can compute $\underline{\widehat{V}_U}$ 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 $O(n^3)$.

In contrast to the lower endpoint, computing $\overline{\widehat{V}_U}$ is, in general, NP-hard: indeed, it is NP-hard already for the usual variance. We can, however, have efficient algorithms for some practically important cases. Let us describe three such cases.

Our first efficient algorithm for computing $\overline{\widehat{V}_U}$ is applicable to the case when for some integer C, every group of $>C$ different 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 $\overline{\widehat{V}_U}$.

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

Our second efficient algorithm for computing $\overline{\widehat{V}_U}$ 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 $\overline{\widehat{V}_U}$.

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

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

6. CONCLUSION

Uncertainty is usually gauged by using standard statistical characteristics: mean, variance, correlation, etc. Then, we use the known values of these characteristics (or the known bounds on these values) to select a decision. Sometimes, it becomes clear that the selected characteristics do not always describe a situation well; then other known (or new) characteristics are proposed. A good example is description of volatility in finance: it started with variance, and now many descriptions are competing, all with their own advantages and limitations.

In such situations, a natural idea is to come up with characteristics tailored to specific application areas: e.g., select the characteristic that maximize the expected utility of the resulting risk-informed decision making.

With the new characteristics, comes the need to estimate them when the sample values are only known with interval uncertainty. We show that algorithms originally developed for estimating traditional characteristics can often be modified to cover new characteristics.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation grants HRD-0734825 and DUE-0926721 and by Grant 1 T36 GM078000-01 from the National Institutes of Health. 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, Sa-aat Niwitpong, and Tony Wang for valuable discussions, and to the anonymous referees for important suggestions.

REFERENCES

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

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

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

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

[5] Farmer J D and Lux T (eds.). *Applications of statistical physics in economics and finance*, a special issue of the *Journal of Economic Dynamics and Control*, 32(1):1–320, 2008.

[6] Ferson S, Ginzburg L, Kreinovich V, Longpré L, and Aviles M. Computing Variance for Interval Data is NP-Hard, *ACM SIGACT News*, 33(2):108–118, 2002.

[7] Ferson S, Ginzburg L, Kreinovich V, Longpré L, and Aviles M. Exact Bounds on Finite Populations of Interval Data, *Reliable Computing*, 11(3):207–233, 2005.

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

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

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

[11] Gabaix X, Parameswaran G, Vasiliki P, and Stanley H E. A theory of power-law distributions in financial market fluctuations, *Nature*, 423(6937):267–270, 2003.

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

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

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

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

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

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

[18] Kinsella S and O'Brien F. 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.

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

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

[21] Mandelbrot B. The variation of certain speculative prices, *J. Business*, 36:394–419, 1963.

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

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

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

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

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

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

[28] Nguyen H T, Kreinovich V, Wu B, and Xiang G. *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.

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

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

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

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

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

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

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

[36] Sriboonchitta S, Nguyen H T, and Kreinovich V. How to Relate Spectral Risk Measures and Utilities, *International Journal of Intelligent Technologies and Applied Statistics*, 3(2):141–158, 2010.

[37] Stanley H E. Econophysics and the current economic turmoil, *American Physical Society News*, 17(11):8, 2008.

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

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

[40] Vasiliki P and Stanley H E. Stock return distributions: tests of scaling and universality from three distinct stock markets, *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 77(3), Pt. 2, Publ. 037101, 2008.