Towards Optimal Representation and Processing of Uncertainty for Decision Making, on the Example of Economics-Related Heavy-Tailed Distributions

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA, vladik@utep.edu
1. Formulation of the Problem

- Traditionally, most statistical techniques assume 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.

- In practice, we encounter heavy-tailed distributions, with infinite variance; what are analogs of:
  - “average” and deviation from average?
  - correlation?
  - how to take into account interval and fuzzy uncertainty?
2. Normal Distributions Are Most Widely Used

- Most statistical techniques assume that the random variables are normally distributed:

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

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

- It is known that a normal distribution is uniquely determined by $m$ and $V$.

- Thus, each characteristic (mode, median, etc.) is uniquely determined by $m$ and $V$. 
3. Estimating the Values of the Characteristics: Case of Normal Distributions

- We have a sample consisting of the values $x_1, \ldots, x_n$.
- We can use the Maximum Likelihood Method: $m$ and $V$ maximizing

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

- Maximizing $L$ is equivalent to minimizing

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

- Equating derivatives to 0, we get:

$$m = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i; \quad V = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - m)^2.$$
4. In Many Practical Situations, We Encounter Heavy-Tailed Distributions

- In the 1960s, Benoit Mandelbrot empirically studied fluctuations.
- He showed that larger-scale fluctuations follow the power-law distribution $\rho(x) = A \cdot x^{-\alpha}$, with $\alpha \approx 2.7$.
- For this distribution, variance is infinite.
- Such distributions are called heavy-tailed.
- Similar heavy-tailed laws were empirically discovered in other application areas.
- These result led to the formulation of fractal theory.
- Since then, similar heavy-tailed distributions have been empirically found:
  - in other financial situations and
  - in many other application areas.
5. First Problem: How to Characterize Such Distributions?

- Usually, *variance* is used to describe deviation from the average.
- 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.
- We will *describe* such characteristics in the first part of this talk.
- We will also describe *how we can estimate* these characteristics.
6. Need to Take into Account Interval Uncertainty

- **Reminder:** \( m = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i, \quad V = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - m)^2. \)

- In practice, we often know approximate values \( \tilde{x}_i \approx x_i. \)

- Sometimes, we know the probabilities of different values of the approximation error \( \Delta x_i \overset{\text{def}}{=} \tilde{x}_i - x_i. \)

- Often, we only know the upper bound \( \Delta_i: \quad |\Delta x_i| \leq \Delta_i. \)

- So, we only know that \( x_i \in x_i = [\tilde{x}_i - \Delta_i, \tilde{x}_i + \Delta_i]. \)

- For each estimator \( C(x_1, \ldots, x_n), \) different \( x_i \in x_i \) lead, in general, to different values \( C(x_1, \ldots, x_n). \)

- Thus, we must find the range:
  \[
  C = [\underline{C}, \overline{C}] = \{ C(x_1, \ldots, x_n) : x_1 \in x_1, \ldots, x_n \in x_n \}.
  \]

- This *interval computations* problem is, in general, NP-hard.
7. How to Describe Deviation from the “Average” for Heavy-Tailed Distributions: Analysis

- A standard way to describe preferences of a decision maker is to use the notion of utility $u$.
- According to decision theory, a user prefers an alternative for which the expected value $\sum_{i=1}^{n} p_i \cdot u_i \to \text{max}$.
- Alternative, the expected value $\sum_{i=1}^{n} p_i \cdot U_i$ of the disutility $U \overset{\text{def}}{=} -u$ is the smallest possible.
- If we replace $x_i \to m \approx x_i$, there is disutility $U(x_i - m)$.
- So, we choose $m$ s.t. $\frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - m) \to \text{min}$.
- The resulting minimum describes the deviation of the values from this “average”.
8. Resulting Definitions

• Let $U : \mathbb{R} \rightarrow \mathbb{R}_0$ be a function for which:
  • $U(0) = 0$,
  • $U(d)$ 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
  $$\frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - m).$$

• By a $U$-deviation, we mean
  $$V_U \overset{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - m_U).$$

• When $U(x) = x^2$, $m_U$ is mean, and $V_U$ is variance.
• When $U(x) = |x|$, $m_U$ is median, and $V_U$ is average absolute deviation (AAD)
  $$V_U = \frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - m_U|.$$
9. These Are Estimates: What Statistical Characteristics Are They Estimating?

- Let $U : \mathbb{R} \to \mathbb{R}_0$ be a f-n for which $U(0) = 0$, and:
  - $U(d)$ is (non-strictly) increasing for $d \geq 0$, and
  - $U(d)$ is (non-strictly) decreasing for $d \leq 0$.
- For each distribution $\rho(x)$, by a $U$-mean, we mean the value $m_U$ that minimizes the expected disutility
  \[ E[U(x - m)] = \int U(x - m) \cdot \rho(x) \, dx. \]
- By a $U$-deviation, we mean this expected disutility value $E[U(x - m_U)]$.
- When $U(x) = x^2$, $m_U$ is the expected value mean, and $V_U$ is variance.
- When $U(x) = |x|$, $m_U$ is the median, and $V_U$ is MAD:
  \[ V_U = E[|x - m_U|] = \int |x - m_U| \cdot \rho(x) \, dx. \]
10. **How to Estimate** $m_U$ and $V_U$

- Once we compute $m_U$, the computation of 
  \[ V_U = \frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - m_U) \] 
  is straightforward.

- Estimating $m_U$ means optimizing a function of a single variable 
  \[ \frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - m) \rightarrow \min. \]

- This optimization problem is equivalent to the Maximum Likelihood (ML): for $U(x) = -\ln(\rho_0(x))$, 
  \[ L = \rho_0(x_1 - m) \cdot \ldots \cdot \rho_0(x_n - m) \rightarrow \max \iff \psi \overset{\text{def}}{=} -\ln(L) = \sum_{i=1}^{n} U(x_i - m) \rightarrow \min. \]

- Similar algorithms are used in *robust statistics*, as *M-methods*, which are mathematically equivalent to ML.
11. Estimating the Heavy-Tailed-Related Deviation Characteristics under Interval Uncertainty: Analysis of the Problem

- When we know the exact values of $x_i$, we know how to compute
  $$V_U = \min_{m} \frac{1}{n} \sum_{i=1}^{n} U(x_i - m).$$

- In practice, the values $x_i$ are often only known with interval uncertainty.

- We only know the intervals $x_i = [x_i, \bar{x}_i]$ that contain the unknown values $x_i$.

- In this case, it is desirable to compute the range $V_U = [V_U, \overline{V}_U]$ of possible values of $V_U$ when $x_i \in x_i$. Here:
  - The value $V_U$ is the minimum of the function $V_U(x_1, \ldots, x_n)$ when $x_i \in x_i$.
  - The value $\overline{V}_U$ is the maximum of the function $V_U(x_1, \ldots, x_n)$ when $x_i \in x_i$. 
12. Algorithm for Computing $V_U$

- First, sort all $2n$ endpoints $x_i$ and $\bar{x}_i$ into an increasing sequence $x(1) \leq x(2) \leq \ldots \leq x(2n)$.

- These values, with $x(0) \overset{\text{def}}{=} -\infty$ and $x(2n+1) \overset{\text{def}}{=} +\infty$, divide the real line into zones $[x(k), x(k+1)], k = 0, 1, \ldots, 2n$.

- For each zone $z$, we select the values $x_1, \ldots, x_n$ as follows: for some value $m$ (to be determined),
  - if $\bar{x}_i \leq r(k)$, then we select $x_i = \bar{x}_i$;
  - if $r(k+1) \leq x_i$, then we select $x_i = x_i$;
  - for all other $i$, we select $x_i = m$.

- Then, we take only the values for which $x_i \neq m$, and find their $U$-estimate $m_U$; if $m_U \in z$, we compute $V_U$.

- The smallest of thus computed $U$-deviations is the desired value $V_U$. 
13. Computation Time for This Algorithm

- Sorting takes $O(n \cdot \log(n))$ steps.
- 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
  
  \[ O(n \cdot \log(n)) + O(n^2) + O(n) \cdot T_{\text{exact}} = O(n^2) + O(n) \cdot T_{\text{exact}}. \]
- Conclusion:
  - if we can compute $V_U$ for exactly known $x_i$ in polynomial time (e.g., linear), then
  - we can compute $V_U$ under interval (hence fuzzy) uncertainty also in polynomial time (e.g., quadratic).
14. Computing $\overline{V}_U$: Analysis of the Problem

- **Fact:** the maximum $\overline{V}_U$ is attained:
  - if $\bar{x}_i \leq m$, for $x_i = \bar{x}_i$;
  - if $m \leq x_i$, for $x_i = \bar{x}_i$;
  - if $x_i \leq m \leq \bar{x}_i$, for $x_i = \bar{x}_i$ or $x_i = \bar{x}_i$.

- **Resulting algorithm:**
  - try all possible combinations of endpoints that satisfy the above conditions, and
  - select the largest of the resulting values $V_U$.

- **Problem:** we may need $2^n$ combinations, too long already for $n \approx 300$.

- **Explanation:** even for $U(d) = d^2$, the problem of computing $\overline{V}_U$ is NP-hard.
15. Case when a Feasible Algorithm Is Possible

• **Reminder:** we consider cases where:
  
  − if $\bar{x}_i \leq m$, for $x_i = \bar{x}_i$;
  
  − if $m \leq \bar{x}_i$, for $x_i = \bar{x}_i$;
  
  − if $\bar{x}_i \leq m \leq \bar{x}_i$, for $x_i = \bar{x}_i$ or $x_i = \bar{x}_i$.

• **Situation.** For some $C$, every group of $> C$ intervals has an empty intersection.

• **Algorithm:** for each zone $z$, we consider case $m \in z$.

• For each zone, there are $\leq C$ intervals for which

$$x_i \leq m \leq \bar{x}_i.$$ 

• So we need to check $\leq 2^C$ combinations for each zone.

• Since $C$ is a constant, $2^C = O(1)$. 
16. Resulting Algorithm for Computing $\overline{V}_U$

- First, we sort all endpoints $x_i$ and $\overline{x}_i$ into an increasing sequence, and add $x(0) = -\infty$ and $x(2n+1) = +\infty$:
  
  $$-\infty = x(0) \leq x(1) \leq x(2) \leq \ldots \leq x(2n) \leq x(2n+1) = +\infty.$$ 

- For each zone $[x(k), x(k+1)]$, we do the following:
  - if $\overline{x}_i \leq r(k)$, then we select $x_i = \overline{x}_i$;
  - if $r(k+1) \leq \overline{x}_i$, then we select $x_i = \overline{x}_i$;
  - for all other $i$, we select either $x_i = \overline{x}_i$ or $x_i = x_i$.

- For each zone, we have $\leq C$ indices $i$ that allow two selections, so we thus get $\leq 2^C$ selections.

- For each of these selections, we compute the $U$-deviation.

- The largest of these $V_U$ is the desired value $\overline{V}_U$.

- This algorithm requires time $O(n^2) + O(n) \cdot T_{\text{exact}}$. 
17. When a Feasible Algorithm Is Possible

- **2nd case:** no interval is a proper subinterval of another: $[x_i, \bar{x}_i] \not\subseteq (x_j, \bar{x}_j)$ for all $i$ and $j$.
- **Example:** measurements made by the same instrument.
- Under this property, lexicographic order

$$[x_i, \bar{x}_i] \leq [x_j, \bar{x}_j] \iff ((x_i < x_j) \lor (x_i = x_j \land \bar{x}_i < \bar{x}_j))$$

sorts the intervals by both endpoints:

$$\bar{x}_1 \leq \bar{x}_2 \leq \ldots \leq \bar{x}_n; \quad \bar{x}_1 \leq \bar{x}_2 \leq \ldots \leq \bar{x}_n.$$ 

- One can prove that, for some $k$, the maximum is attained at a tuple $(x_1, \ldots, x_k, \bar{x}_{k+1}, \ldots, \bar{x}_n)$.
- There are $n + 1$ such tuples, so we have a polynomial-time algorithm.
- Similar arguments can be made when the intervals can be divided into $m$ groups with this property.
18. Resulting Algorithms for Computing $\overline{V}_U$

- **Applicable:** when $[x_i, \overline{x}_i] \not\subseteq (x_j, \overline{x}_j)$ for all $i$ and $j$.

- First, we sort all the intervals in lexicographic order

  $$[x_i, \overline{x}_i] \leq [x_j, \overline{x}_j] \iff ((x_i < x_j) \lor (x_i = x_j \land \overline{x}_i < \overline{x}_j)).$$

- Then, we compute $V_U$ for all $n + 1$ tuples of the form

  $$(x_1, \ldots, x_k, \overline{x}_{k+1}, \ldots, \overline{x}_n),$$

  with $k = 0, 1, \ldots, n$.

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

- This algorithm requires time

  $$O(n \cdot \log(n)) + O(n) \cdot T_{\text{exact}}.$$
19. **Algorithms for Computing $\overline{V}_U$ (cont-d)**

- **Applicable:** all intervals can be divided into $m$ groups each of which satisfies the no-subinterval property.
- We sort all intervals within each group in lexicographic order.
- For each group $j = 1, \ldots, m$, with $n_j \leq n$ elements, we consider $n_j + 1 \leq n + 1$ tuples of the form
  $$(\bar{x}_1, \ldots, \bar{x}_{k_j}, \bar{x}_{k_j+1}, \ldots, \bar{x}_n).$$
- We consider all possible combinations of such tuples corresponding to all possible vectors $(k_1, \ldots, k_m)$.
- For each of these $\leq n^m$ vectors, we compute $V_U$.
- The largest of these $V_U$ is the desired value $\overline{V}_U$.
- This algorithm requires time
  $$O(n \cdot \log(n)) + O(n^m) \cdot T_{\text{exact}}.$$
20. What Are the Reasonable Measures of Dependence for Heavy-Tailed Distributions?

- In the traditional statistics, a reasonable measure of dependence is the correlation

\[ \rho_{xy} = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - m_x) \cdot (y_i - m_y) \]

\[ \sqrt{V_x \cdot V_y} \]

- For heavy-tailed distributions, variances are infinite, so this formula cannot be applied.

- **Possibility:** use Kendall’s tau, the proportion of pairs \((i, j)\) for which \(x\) and \(y\) change in the same direction:

  either \((x_i \leq x_j \& y_i \leq y_j)\) or \((x_j \leq x_i \& y_j \leq y_i)\).

- **Remaining problem:** what if we are interested only in linear dependencies?
21. Proposed Definition

- **Idea:** $c$ describes how much disutility decreases when we use $x_i$ to help predict $y_i$:

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

where

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

and

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

- The function $f$ at which the minimum is attained is called $\mathcal{F}$-regression.
- When $U(d) = d^2$ and $\mathcal{F}$ is the class of all linear functions, $c = \rho^2$. 
22. Discussion

- For normal distributions and linear functions, 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 makes perfect sense.
- For example, suppose that $y_i = x_i^2$:
  - then, if we know $x_i$, then we can uniquely reconstruct $y_i$;
  - however, if we know $y_i$, we can only reconstruct $x_i$ modulo sign.
- **Remaining open problem:** estimate the above measures of dependence under interval uncertainty.
23. Acknowledgments

- The 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 . . .” funded by European Union.
- The author is thankful to Hung T. Nguyen for valuable discussions.