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

Vladik Kreinovich[1], Nitaya Buntao[2], and Olga Kosheleva[1]

[1]University of Texas at El Paso
El Paso, TX 79968, USA
vladik@utep.edu, olgak@utep.edu

[2]Department of Applied Statistics
King Mongkut's University of Technology
North Bangkok
Bangkok 10800 Thailand
taltanot@hotmail.com

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 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 $m \stackrel{\text{def}}{=} E[x]$, and
  - a natural characteristic of the deviation from the average is the variance $V \stackrel{\text{def}}{=} E[(x-m)^2]$.
- 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

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

- Equating derivatives to 0, we get:

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

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. 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
$$\int \rho(x) \cdot u(x)\, dx \to \max.$$
- Alternative, the expected value $\int \rho(x) \cdot U(x)\, dx$ of the *disutility* $U \stackrel{\text{def}}{=} -u$ is the smallest possible.
- If we replace $x \to m \approx x$, there is disutility $U(x-m)$.
- So, we choose m s.t. $\int \rho(x) \cdot U(x-m)\, dx \to \min$.
- The resulting minimum describes the deviation of the values from this "average".

7. Resulting Definitions

- Let $U : \mathbb{R} \to \mathbb{R}_0$ 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 a distribution $\rho(x)$, by a *U-mean*, we mean the value m_U that minimizes $\int \rho(x) \cdot U(x - m)\,dx$.
- By a *U-deviation*, we mean

$$V_U \stackrel{\text{def}}{=} \int \rho(x) \cdot U(x - m_U)\,dx.$$

- 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 $V_U = \int \rho(x) \cdot |x - m_U|\,dx$.

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

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

$$\rho_{xy} = \frac{E[(x - E(x)) \cdot (y - E(y))]}{\sqrt{V_x \cdot V_y}}.$$

- For heavy-tailed distributions, variances are infinite, so this formula cannot be applied.
- *Possibility:* Kendall's tau, the proportion of pairs (x, y) and (x', y') s.t. x and y change in the same direction:

$$\text{either } (x \le x' \,\&\, y \le y') \text{ or } (x' \le x \,\&\, y' \le y).$$

- *Remaining problem:* what if we are interested only in linear dependencies?

9. Proposed Definition

- *Idea:* c describes how much disutility decreases when we use x to help predict y:

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

where

$$V_U(y) \stackrel{\text{def}}{=} \min_m \int \rho(x,y) \cdot U(y-m)\,dx\,dy$$

and

$$V_{U,\mathcal{F}}(y|x) \stackrel{\text{def}}{=} \min_{f\in\mathcal{F}} \int \rho(x,y) \cdot U(y-f(x))\,dx\,dy.$$

- 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, $c = \rho^2$.

10. Discussion

- For normal distributions and linear functions, 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;
  - however, if we know y, we can only reconstruct x modulo sign.

11. How to Estimate the New Characteristics from Observations

- *In the above text:* we defined the desired characteristics in terms of the probability density function (pdf) $\rho(x)$.
- *In practice:* we often do not know the distribution.
- *Instead:* we know the sample values $x_1, \ldots, x_n$.
- *A natural idea:* use the "histogram" distribution, in which each x_i appears with equal probability $\dfrac{1}{n}$.
- *Example:* for $\rho(x) = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} \delta(x - x_i)$, the mean

  $E = \int \rho(x) \cdot x\,dx$ turns into $\widehat{E} = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} x_i$.
- *Similarly:* we get $\widehat{V} = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} \left(x_i - \widehat{V}\right)^2$.

12. Resulting Estimates for m_U and V_U

- For each sample $x_1, \ldots, x_n$, by a *U-estimate*, we mean the value $\widehat{m}_U$ that minimizes $\dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} U(x_i - m)$.
- By an *estimate for U-deviation*, we mean

$$\widehat{V}_U \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - \widehat{m}_U).$$

- When $U(x) = x^2$, $\widehat{m}_U$ is arithmetic mean, and $\widehat{V}_U$ is sample variance.
- When $U(x) = |x|$, $\widehat{m}_U$ is sample median, and $\widehat{V}_U$ is average absolute deviation $\widehat{V}_U = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} |x_i - \widehat{m}_U|$.

13. How to Estimate m_U and V_U

- Once we compute $\widehat{m}_U$, the computation of $\widehat{V}_U = \frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - \widehat{m}_U)$ is straightforward.
- Estimating $\widehat{m}_U$ means optimizing a function of a single variable $\frac{1}{n} \cdot \sum_{i=1}^{n} U(x_i - m) \to \min$.
- This optimization problem is equivalent to the Maximum Likelihood (ML): for $U(x) = -\ln(\rho_0(x))$,

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

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

- Similar algorithms are used in *robust statistics*, as *M-methods*, which are mathematically equivalent to ML.

14. Estimates for U-Correlation

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

$$\widehat{c} \stackrel{\text{def}}{=} \frac{\widehat{V}_U(y) - \widehat{V}_{U,\mathcal{F}}(y|x)}{\widehat{V}_U(y)},$$

  where

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

  and

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

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

15. Need to Take into Account Interval Uncertainty

- In practice, we often know approximate values $\widetilde{x}_i \approx x_i$.
- Sometimes, we know the probabilities of different values of the approximation error $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$.
- Often, we only know the upper bound Δ_i: $|\Delta x_i| \le \Delta_i$.
- So, we only know that $x_i \in \mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.
- For each estimator $C(x_1, \ldots, x_n)$, different $x_i \in \mathbf{x}_i$ lead, in general, to different values $C(x_1, \ldots, x_n)$.
- Thus, we must find the range:

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

- This *interval computations* problem is, in general, NP-hard.

16. Estimating the Heavy-Tailed-Related Deviation Characteristics under Interval Uncertainty

- When we know the exact values of x_i, we know how to compute $\widehat{V}_U = \min\limits_m \frac{1}{n} \cdot \sum\limits_{i=1}^{n} U(x_i - m)$.
- In practice, the values x_i are often only known with interval uncertainty.
- We only know the intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ that contain the unknown values x_i.
- In this case, it is desirable to compute the range $[\underline{\widehat{V}_U}, \overline{\widehat{V}_U}]$ of possible values of $\widehat{V}_U$ when $x_i \in \mathbf{x}_i$. Here:
  - The value $\underline{\widehat{V}_U}$ is the minimum of the function $\widehat{V}_U(x_1, \ldots, x_n)$ when $x_i \in \mathbf{x}_i$.
  - The value $\overline{\widehat{V}_U}$ is the maximum of the function $\widehat{V}_U(x_1, \ldots, x_n)$ when $x_i \in \mathbf{x}_i$.

17. Algorithm for Computing $\underline{\widehat{V}_U}$

- First, sort all $2n$ endpoints $\underline{x}_i$ and $\overline{x}_i$ into an increasing sequence $x_{(1)} \le x_{(2)} \le \ldots \le x_{(2n)}$.
- These values, with $x_{(0)} \stackrel{\text{def}}{=} -\infty$ and $x_{(2n+1)} \stackrel{\text{def}}{=} +\infty$, 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 $\overline{x}_i \le r_{(k)}$, 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 $\widehat{m}_U$; if $m_U \in z$, we compute $\widehat{V}_U$.
- The smallest of thus computed U-deviations is the desired value $\underline{\widehat{V}_U}$.

18. 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 $T_{\rm exact}$ – to compute the U-estimate and U-deviation.
- Thus, the total computation time is equal to

$$O(n\cdot\log(n))+O(n^2)+O(n)\cdot T_{\rm exact} = O(n^2)+O(n)\cdot T_{\rm exact}.$$

- Conclusion:
  - if we can compute $\widehat{V}_U$ for exactly known x_i in polynomial time (e.g., linear), then
  - we can compute $\underline{\widehat{V}_U}$ under interval uncertainty also in polynomial time (e.g., quadratic).

19. Computing $\overline{\widehat{V}_U}$: Analysis of the Problem

- *Fact:* the maximum $\overline{\widehat{V}_U}$ is attained:
  - if $\overline{x}_i \le m$, for $x_i = \underline{x}_i$;
  - if $m \le \underline{x}_i$, for $x_i = \overline{x}_i$;
  - if $\underline{x}_i \le m \le \overline{x}_i$, for $x_i = \underline{x}_i$ or $x_i = \overline{x}_i$.
- *Resulting algorithm:*
  - try all possible combinations of endpoints that satisfy the above conditions, and
  - select the largest of the resulting values $\widehat{V}_U$.
- *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 $\overline{\widehat{V}_U}$ is NP-hard.

20. Case when a Feasible Algorithm Is Possible

- *Reminder:* we consider cases where:
  - if $\overline{x}_i \le m$, for $x_i = \underline{x}_i$;
  - if $m \le \underline{x}_i$, for $x_i = \overline{x}_i$;
  - if $\underline{x}_i \le m \le \overline{x}_i$, for $x_i = \underline{x}_i$ or $x_i = \overline{x}_i$.
- *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 $\le C$ 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, $2^C = O(1)$.

21. Resulting Algorithm for Computing $\overline{\widehat{V}_U}$

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

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

- 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 $\widehat{V}_U$ is the desired value $\overline{\widehat{V}_U}$.
- This algorithm requires time $O(n^2) + O(n) \cdot T_{\text{exact}}$.

22. When a Feasible Algorithm Is Possible

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

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

- One can prove that, for some k, the maximum is attained at a tuple $(\underline{x}_1, \ldots, \underline{x}_k, \overline{x}_{k+1}, \ldots, \overline{x}_n)$.
- There are $n+1$ such tuples, so we have a polynomial-time algorithm.
- Similar arguments can be made when the intervals can be divided into m groups with this property.

23. Resulting Algorithms for Computing $\overline{\widehat{V}_U}$

- *Applicable:* when $[\underline{x}_i, \overline{x}_i] \not\subseteq (\underline{x}_j, \overline{x}_j)$ for all i and j.
- First, we sort all the intervals in lexicographic order

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

- Then, we compute V_U for all $n+1$ tuples of the form $(\underline{x}_1, \ldots, \underline{x}_k, \overline{x}_{k+1}, \ldots, \overline{x}_n)$, with $k = 0, 1, \ldots, n$.
- The largest of thus computed U-deviations is the desired value $\overline{\widehat{V}_U}$.
- This algorithm requires time

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

24. Algorithms for Computing $\overline{\widehat{V}_U}$ (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 $n_j \le n$ elements, we consider $n_j + 1 \le n + 1$ tuples of the form

$$(\underline{x}_1, \ldots, \underline{x}_{k_j}, \overline{x}_{k_j+1}, \ldots, \overline{x}_n).$$

- 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 $\widehat{V}_U$.
- The largest of these $\widehat{V}_U$ is the desired value $\overline{\widehat{V}_U}$.
- This algorithm requires time

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

25. Conclusion

- Uncertainty is usually gauged by using standard statistical characteristics: mean, variance, correlation, etc.
- Then, we use the known values of these characteristics to select a decision.
- Sometimes, we only know bounds, then we use these bounds in decision making.
- 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.

26. Conclusion (cont-d)

- *Reminder:* sometimes, the traditional statistical characteristics do not work well.
- In such situations, a natural idea is to come up with characteristics tailored to specific application areas.
- E.g., a characteristic that maximizes the expected utility of the resulting risk-informed decision making.
- How to estimate these characteristics 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.

27. Acknowledgments

- 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 Office of the Higher Education Commission, Thailand.
- The authors are thankful:
  - to Hung T. Nguyen,
  - to Sa-aat Niwitpong,
  - to Tony Wang, and
  - to the anonymous referees

  for valuable suggestions.