Application-Motivated Combinations of Interval and Probability Approaches and their Use in Engineering, Especially in Biomedical Engineering

Vladik Kreinovich

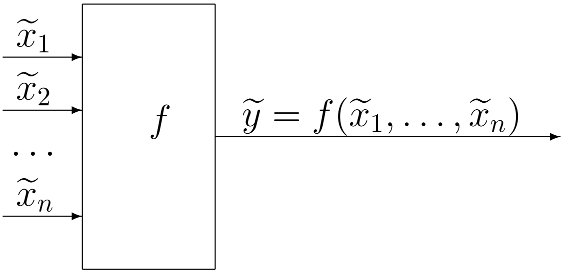
Department of Computer Science,
University of Texas at El Paso,
El Paso, TX 79968, USA
vladik@utep.edu
http://www.cs.utep.edu/vladik

Interval computations website:
http://www.cs.utep.edu/interval-comp

Grant support: NSF grants HRD-0734825 (Cyber-ShARE Center) and DUE-0926721, NIH Grant 1 T36 GM078000-01

1. Data Processing under Uncertainty

- *Problem:* we often need to measure quantities y that are are difficult (or impossible) to measure directly.
- *Usual approach:* measure auxiliary quantities x_i related to y by a known dependence $y = f(x_1, \ldots, x_n)$, then process the measurement results $\widetilde{x}_i$:

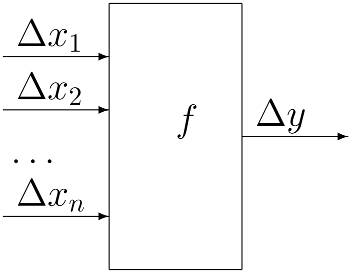


- *Problem:* measurements are never 100% accurate: $\widetilde{x}_i$ differs from the actual (unknown) value x_i:

$$\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i \neq 0 \text{ so } \widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) \neq y = f(x_1, \ldots, x_n).$$

Title Page
◀◀ ▶▶
◀ ▶
Page 2 of 8
Go Back
Full Screen
Close
Quit

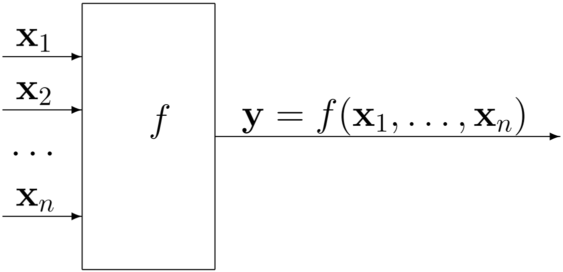
2. Probabilistic Approach to Uncertainty



- *Traditional approach:* we know probability distribution for Δx_i (usually Gaussian).
- *Where it comes from:* calibration using standard MI.
- *Problem:* calibration is often not possible in:
  - cutting-edge research (no standard MI)
  - manufacturing (calibration too expensive)
- *Solution:* we know upper bounds Δ_i on $|\Delta x_i|$ hence

$$x_i \in \mathbf{x}_i \stackrel{\text{def}}{=} [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i].$$

3. Interval Computations



- *Given:* an algorithm $y = f(x_1, \ldots, x_n)$ and n intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$.
- *Compute:* the corresponding range of y:

$$[\underline{y}, \overline{y}] = \{f(x_1, \ldots, x_n) \mid x_1 \in [\underline{x}_1, \overline{x}_1], \ldots, x_n \in [\underline{x}_n, \overline{x}_n]\}.$$

- *Examples of engineering applications*:
  - design of elementary particle colliders: Berz (USA)
  - robotics: Jaulin (France), Neumaier (Austria)
  - chemical engineering: Stadtherr (U. Notre Dame)

4. Need to Go Beyond Intervals: Chip Design

- *Chip design:* one of the main objectives is to decrease the clock cycle.
- *Current approach:* uses worst-case (interval) techniques.
- *Problem:* the probability of the worst-case values is usually very small.
- *Result:* estimates are over-conservative – unnecessary over-design and under-performance of circuits.
- *Difficulty:* we only have *partial* information about the corresponding probability distributions.
- *Objective:* produce estimates valid for all distributions which are consistent with this information.
- *What we do:* provide such estimates for the clock time.

5. Combining Interval and Probabilistic Uncertainty: General Case

- *Problem:* there are many ways to represent a probability distribution.
- *Idea:* look for an objective.
- *Objective:* make decisions $E_x[u(x,a)] \to \max\limits_a$.
- *Case 1:* smooth $u(x)$.
- *Analysis:* we have $u(x) = u(x_0) + (x - x_0) \cdot u'(x_0) + \ldots$
- *Conclusion:* we must know moments to estimate $E[u]$.
- *Case of uncertainty:* interval bounds on moments.
- *Case 2:* threshold-type $u(x)$.
- *Conclusion:* we need cdf $F(x) = \text{Prob}(\xi \le x)$.
- *Case of uncertainty:* p-box $[\underline{F}(x), \overline{F}(x)]$.

6. Second Case Study: Bioinformatics

- *Practical problem:* find genetic difference between cancer cells and healthy cells.
- *Ideal case:* we directly measure concentration c of the gene in cancer cells and h in healthy cells.
- *In reality:* difficult to separate.
- *Solution:* we measure $y_i \approx x_i \cdot c + (1 - x_i) \cdot h$, where x_i is the percentage of cancer cells in i-th sample.
- *Equivalent form:* $a \cdot x_i + h \approx y_i$, where $a \stackrel{\text{def}}{=} c - h$.
- *Interval uncertainty:* experts manually count x_i, and only provide interval bounds $\mathbf{x}_i$, e.g., $x_i \in [0.7, 0.8]$.
- *Problem:* find the range of a and h corresponding to all possible values $x_i \in [\underline{x}_i, \overline{x}_i]$.

7. Selected References

- V. Kreinovich, "Application-motivated combinations of fuzzy, interval, and probability approaches, and their use in geoinformatics, bioinformatics, and engineering", *Int'l J. of Automation and Control (IJAAC)*, 2008, 2(2/3):317–339.
- V. Kreinovich, L. Longpré, S.A. Starks, G. Xiang, J. Beck, R. Kandathi, A. Nayak, S. Ferson, and J. Hajagos, "Interval Versions of Statistical Techniques, with Applications to Environmental Analysis, Bioinformatics, and Privacy in Statistical Databases", *J. Comp. Appl. Math.*, 2007, 199(2):418–423.
- R.L. Cheu, V. Kreinovich, F. Modave, G. Xiang, T. Li, and T. Magoc, "How to Estimate, Take Into Account, and Improve Travel Time Reliability in Transportation Networks", *Proc. Int'l Workshop on Reliable Engineering Computing REC'08*, Savannah, GA, Febr. 20–22, 2008, pp. 289–332.
- M. Orshansky, W.-S. Wang, G. Xiang, and V. Kreinovich, "Interval-Based Robust Statistical Techniques for Non-Negative Convex Functions, with Application to Timing Analysis of Computer Chips", *Proc. Int'l Workshop on Reliable Engineering Computing REC'06*, Savannah, GA, Febr. 22–24, 2006, pp. 197–212.