Combining Interval and Probabilistic Uncertainty: What Is Computable?

Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, Texas 79968, USA
vladik@utep.edu

(based on joint work with
Andrzej Pownuk and Olga Kosheleva)

1. Need to Take Uncertainty Into Account When Processing Data

- In practice, we are often interested in a quantity y which is difficult to measure directly.
- *Examples:* distance to a star, amount of oil in the well, tomorrow's weather.
- *Solution:* find easier-to-measure quantities $x_1, \ldots, x_n$ related to y by a known dependence $y = f(x_1, \ldots, x_n)$.
- Then, we measure x_i and use measurement results $\widetilde{x}_i$ to compute an estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$.
- Measurements are never absolutely accurate, so even if the model f is exact, $\widetilde{x}_i \neq x_i$ leads to $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y \neq 0$.
- It is important to use information about measurement errors $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$ to estimate the accuracy Δy.

2. We Often Have Imprecise Probabilities

- *Usual assumption:* we know the probabilities for Δx_i.
- To find them, we measure the same quantities:
  - with our measuring instrument (MI) and
  - with a much more accurate MI, with $\widetilde{x}_i^{\rm st} \approx x_i$.
- In two important cases, this does not work:
  - state-of-the art-measurements, and
  - measurements on the shop floor.
- Then, we have partial information about probabilities.
- Often, all we know is an upper bound $|\Delta x_i| \le \Delta_i$.
- Then, we only know that $x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ and

$$y \in [\underline{y}, \overline{y}] \stackrel{\rm def}{=} \{f(x_1, \ldots, x_n) : x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]\}.$$

- Computing $[\underline{y}, \overline{y}]$ is known as *interval computation.*

3. How Do We Describe Imprecise Probabilities?

- *Ultimate goal of most estimates:* to make decisions.
- *Known:* a rational decision-maker maximizes expected *utility* $E[u(y)]$.
- For smooth $u(y)$, $y \approx \widetilde{y}$ implies that

$$u(y) = u(\widetilde{x}) + (y - \widetilde{y}) \cdot u'(\widetilde{y}) + \frac{1}{2} \cdot (y - \widetilde{y})^2 \cdot u''(\widetilde{y}).$$

- So, to find $E[u(y)]$, we must know moments $E[(y-\widetilde{y})^k]$.
- Often, $u(x)$ abruptly changes: e.g., when pollution level exceeds y_0; then $E[u(y)] \sim F(y) \stackrel{\text{def}}{=} \text{Prob}(y \le y_0)$.
- From $F(y)$, we can estimate moments, so $F(x)$ is enough.
- Imprecise probabilities mean that we know $F(y)$, we only know bounds (*p-box*) $\underline{F}(y) \le F(y) \le \overline{F}(y)$.

4. Imprecise Probabilities: What Is Computable?

- Computations with p-boxes are practically important.
- It is thus desirable to come up with efficient algorithms which are as general as possible.
- It is known that too general problems are often *not* computable.
- To avoid wasting time, it is therefore important to find out what *can* be computed.
- At first glance, this question sounds straightforward:
  - to describe a cdf, we can consider a computable function $F(x)$;
  - to describe a p-box, we consider a computable *function interval* $[\underline{F}(x), \overline{F}(x)]$.
- Often, we can do that, but we will show that sometimes, we need to go *beyond* function intervals.

5. Reminder: What Is Computable?

- A real number x corresponds to a value of a physical quantity.
- We can measure x with higher and higher accuracy.
- So, x is called *computable* if there is an algorithm, that, given k, produces a ration r_k s.t. $|x - r_k| \le 2^{-k}$.
- A *computable function* computes $f(x)$ from x.
- We can only use approximations to x.
- So, an algorithm for computing a function can, given k, request a 2^{-k}-approximation to x.
- Most usual functions are thus computable.
- *Exception:* step-function $f(x) = 0$ for $x < 0$ and $f(x) = 1$ for $x \ge 0$.
- Indeed, no matter how accurately we know $x \approx 0$, from $r_k = 0$, we cannot tell whether $x < 0$ or $x \ge 0$.

6. Consequences for Representing a cdf $F(x)$

- We would like to represent a general probability distribution by its cdf $F(x)$.
- From the purely mathematical viewpoint, this is indeed the most general representation.
- At first glance, it makes sense to consider computable functions $F(x)$.
- For many distributions, e.g., for Gaussian, $F(x)$ is computable.
- However, when $x = 0$ with probability 1, the cdf $F(x)$ is exactly the step-function.
- And we already know that the step-function is not computable.
- Thus, we need to find an alternative way to represent cdf's – beyond computable functions.

7. Back to the Drawing Board

- Each value $F(x)$ is the probability that $X \le x$.
- We cannot empirically find exact probabilities p.
- We can only estimate *frequencies* f based on a sample of size N.
- For large N, the difference $d \stackrel{\text{def}}{=} p - f$ is asymptotically normal, with $\mu = 0$ and $\sigma = \sqrt{\dfrac{p \cdot (1-p)}{N}}$.
- Situations when $|d - \mu| < 6\sigma$ are negligibly rare, so we conclude that $|f - p| \le 6\sigma$.
- For large N, we can get $6\sigma \le \delta$ for any accuracy $\delta > 0$.
- We get a sample $X_1, \ldots, X_N$.
- We don't know the exact values X_i, only measured values $\widetilde{X}_i$ s.t. $|\widetilde{X}_i - X_i| \le \varepsilon$ for some accuracy ε.
- So, what we have is a frequency $f = \text{Freq}(\widetilde{X}_i \le x)$.

8. Resulting Definition

- Here, $X_i \le x - \varepsilon \Rightarrow \widetilde{X}_i \le x \Rightarrow X_i \le x + \varepsilon$, so

$$\text{Freq}(X_i \le x - \varepsilon) \le f = \text{Freq}(\widetilde{X}_i \le x) \le \text{Freq}(X_i \le x + \varepsilon).$$

- Frequencies are δ-close to probabilities, so we arrive at the following:
- *For every x, $\varepsilon > 0$, and $\delta > 0$, we get a rational number f such that $F(x-\varepsilon) - \delta \le f \le F(x+\varepsilon) + \delta$.*
- This is how we define a computable cdf $F(x)$.
- In the computer, to describe a distribution on an interval $[\underline{T}, \overline{T}]$:
  - we select a grid $x_1 = \underline{T}$, $x_2 = \underline{T} + \varepsilon$, ..., and
  - we store the corr. frequencies f_i with accuracy δ.
- A class of possible distribution is represented, for each ε and δ, by a finite list of such approximations.

9. First Equivalent Definition

- *Original:* $\forall x\, \forall \varepsilon_{>0}\, \forall \delta_{>0}$, we get a rational f such that

$$F(x-\varepsilon)-\delta \le f \le F(x+\varepsilon)+\delta.$$

- *Equivalent:* $\forall x\, \forall \varepsilon_{>0}\, \forall \delta_{>0}$, we get a rational f which is δ-close to $F(x')$ for some x' such that $|x'-x| \le \varepsilon$.
- *Proof of equivalence:*
  - We know that $F(x+\varepsilon)-F(x+\varepsilon/3) \to 0$ as $\varepsilon \to 0$.
  - So, for $\varepsilon = 2^{-k}$, $k = 1, 2, \ldots$, we take f and f' s.t.

$$F(x+\varepsilon/3)-\delta/4 \le f \le F(x+(2/3)\cdot\varepsilon)+\delta/4$$

$$F(x+(2/3)\cdot\varepsilon)-\delta/4 \le f' \le F(x+\varepsilon)+\delta/4.$$

  - We stop when f and f' are sufficiently close:

$$|f-f'| \le \delta.$$

  - Thus, we get the desired f.

10. Second Equivalent Definition

- We start with pairs (x_1, f_1), (x_2, f_2), ...
- When $f_{i+1} - f_i > \delta$, we add intermediate pairs

$$(x_i, f_i + \delta), (x_i, f_i + 2\delta), \ldots, (x_i, f_{i+1}).$$

- The resulting set of pairs is (ε, δ)-close to the graph $\{(x, y) : F(x-0) \le y \le F(x)\}$ in Hausdorff metric d_H.
- (x, y) and (x', y') are (ε, δ)-*close* if $|x - x'| \le \varepsilon$ and $|y - y'| \le \delta$.
- The sets S and S' are (ε, δ)-close if:
  - for every $s \in S$, there is a (ε, δ)-close point $s' \in S'$;
  - for every $s' \in S'$, there is a (ε, δ)-close point $s \in S$.
- Compacts with metric d_H form a computable compact.
- So, $F(x)$ is a monotonic computable object in this compact.

11. What Can Be Computed: A Positive Result for the 1D Case

- *Reminder:* we are interested in $F(x)$ and $E_{F(x)}[u(x)]$ for smooth $u(x)$.
- *Reminder:* estimate for $F(x)$ is part of the definition.
- *Question:* computing $E_{F(x)}[u(x)]$ for smooth $u(x)$.
- *Our result:* there is an algorithm that:
  - given a computable cdf $F(x)$,
  - given a computable function $u(x)$, and
  - given accuracy $\delta > 0$,
  - computes $E_{F(x)}[u(x)]$ with accuracy δ.
- For computable classes $\mathcal{F}$ of cdfs, a similar algorithm computes the range of possible values

$$[\underline{u}, \overline{u}] \stackrel{\text{def}}{=} \{E_{F(x)}[u(x)] : F(x) \in \mathcal{F}\}.$$

12. Proof: Main Idea

- Computable functions are computably continuous: for every $\delta > 0$, we can compute $\varepsilon > 0$ s.t.
$$|x - x'| \le \varepsilon \Rightarrow |f(x) - f(x')| \le \delta.$$
- We select ε corr. to $\delta/4$, and take a grid with step $\varepsilon/4$.
- For each x_i, the value f_i is $(\delta/4)$-close to $F(x'_i)$ for some x'_i which is $(\varepsilon/4)$-close to x_i.
- The function $u(x)$ is $(\delta/2)$-close to a piece-wise constant function $u'(x) = u(x_i)$ for $x \in [x'_i, x'_{i+1}]$.
- Thus, $|E[u(x)] - E[u'(x)]| \le \delta/2$.
- Here, $E[u'(x)] = \sum\limits_i u(x_i) \cdot (F(x'_{i+1}) - F(x'_i))$.
- Here, $F(x'_i)$ is close to f_i and $F(x'_{i+1})$ is close to f_{i+1}.
- Thus, $E[u'(x)]$ (and hence, $E[u(x)]$) is computably close to a computable sum $\sum\limits_i u(x_i) \cdot (f_{i+1} - f_i)$.

13. What to Do in a Multi-D Case?

- For each $g(x)$, y, $\varepsilon > 0$, and $\delta > 0$, we can find a frequency f such that:

$$|P(g(x) \le y') - f| \le \varepsilon \text{ for some } y' \text{ s.t. } |y - y'| \le \delta.$$

- We select an ε-net $x_1, \ldots, x_n$ for X. Then,

$$X = \bigcup_i B_\varepsilon(x_i), \text{ where } B_\varepsilon(x) \stackrel{\text{def}}{=} \{x' : d(x, x') \le \varepsilon\}.$$

- We select f_1 which is close to $P(B_{\varepsilon'}(x_1))$ for all ε' from some interval $[\underline{\varepsilon}, \overline{\varepsilon}]$ which is close to ε.
- We then select f_2 which is close to $P(B_{\varepsilon'}(x_1) \cup B_{\varepsilon'}(x_2))$ for all ε' from some subinterval of $[\underline{\varepsilon}, \overline{\varepsilon}]$, etc.
- Then, we get approximations to probabilities of the sets $B_\varepsilon(x_i) - (B_\varepsilon(x_1) \cup \ldots \cup B_\varepsilon(x_{i-1}))$.
- This lets us compute the desired values $E[u(x)]$.

14. Acknowledgments

- The authors are greatly thankful to Dr. Walid Taha and to Rice University for this opportunity.
- This work was supported in part by the National Science Foundation grants:
  - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
  - DUE-0926721.