Taking Uncertainty into Account in Data Processing and Decision Making: A Brief Overview

Vladik Kreinovich
Center for Theoretical Research and
Its Applications in Computer Science
(TRACS)
University of Texas at El Paso
El Paso, TX 79968, USA
vladik@utep.edu

1. How to Propagate Uncertainty

- How uncertainty with which we know inputs $x_1, \ldots, x_n$ affects the result $y = f(x_1, \ldots, x_n)$ of data processing?
- There are many *efficient algorithms* for interval and for probabilistic uncertainty.
- However, there are also challenges:
  - What if we only have partial information about probabilities?
  - What if we have imprecise ("fuzzy") expert information?
  - Methods are often slow; how can we speed them up?
  - How to take into account that the model $y = f(x_1, \ldots, x_n)$ is usually also only approximate?

2. Heuristic Methods

- We often have heuristic techniques.
- Some techniques are empirically the best.
- However, since these techniques are not validated:
  - practitioners are reluctant to use them, and
  - it's not clear whether they are the best possible.
- In such situations, it is desirable to:
  - either come up with a theoretical explanation of the empirical success,
  - or to come up with better techniques.
- Often, this can be achieved by using appropriate symmetries: a tool ubiquitous in modern physics.
- Indeed, a general way to predict is to see what happens in similar situations.

3. Example: Robust Statistics

- Standard statistical techniques like least squares $\sum e_i^2 \to \min$ work well for normal distributions.
- However, a single outlier can ruin the estimate: e.g., if we have 999 1s and one 10^6, the average is 1,000.
- What if we know that the distribution is not normal, but we do not its shape?
- Empirical analysis shows that in this case, ℓ^p-methods work the best $\sum |e_i|^p \to \min$ (Huber *et al.*)
- It turns out that they are the only ones invariant w.r.t. change of measuring unit $x_i \to \lambda \cdot x_i$.

4. Example: Neural Networks

- The output y of an artificial neuron depends on the inputs $x_1, \ldots, x_n$ as $y = s_0\left(\sum_{i=1}^{n} w_i \cdot x_i - w_0\right)$.
- Empirically, the best activation function $s_0(z)$ is the sigmoid $s_0(z) = \dfrac{1}{1+\exp(-z)}$; but why?
- It turns out that this is the only function which is shift-invariant in the following sense:
  - if we change the starting point $z \to z + z_0$,
  - then $s(z+z_0)$ can be obtained from $s(z)$ by a "natural" transformation $s(z+z_0) = T_{z_0}(s(z))$,
  - i.e., transformations T_{z_0} belong to a finite-parametric family closed under composition.

5. Decision Making under Uncertainty

- Knowing the world is great, but the ultimate goal is to change it.
- Thus, we need to make decisions: find appropriate configuration, appropriate control, etc.
- In situations of complete knowledge, decision making is simply optimization.
- Under uncertainty, often it is not even clear how to formulate the problem in precise terms.
- Here also, some general symmetry ideas can often help.

6. Example: Hurwicz Criterion

- Often, for each alternative a, we only know the bounds $\underline{u}(a)$ and $\overline{u}(a)$ on its utility $u(a)$.
- How should we then select an alternative?
- Nobelist Leo Hurwicz proposed selecting a that maximizes $u(a) = \alpha \cdot \overline{u}(a) + (1 - \alpha) \cdot \underline{u}(a)$.
- It turns out that this is the only criterion which is invariant w.r.t. both:
  - changing the measuring unit $u \to \lambda \cdot u$ and
  - changing the starting point: $u \to u + u_0$.

7. Example of Applications

- *Environmental sciences:* symmetries help decide where best to place the Eddy covariance tower.
- *Physics:* Einstein's, Maxwell's, Schrödinger's equations can all be derived from symmetries only.
- *Processing interval uncertainty:* symmetries explain the use of Cauchy distributions.
- *Radioastronomy:*
  - How to select the most relevant image?
  - Symmetries explain the success of maximizing entropy (or generalized entropy).