Processing Measurement Uncertainty: From Intervals and p-Boxes to Probabilistic Nested Intervals

Konstantin K. Semenov[1], Gennady N. Solopchenko[1], and Vladik Kreinovich[2]

[1]Measur. Info. Techn. Dept., Saint-Petersburg State Polytechnic Univ., Russia
[2]Computer Science Dept., University of Texas, El Paso, Texas 79968, USA
semenov.k.k@gmail.com, g.n.solopchenko@mail.ru, vladik@utep.edu

Keywords: measurement uncertainty, interval computations, p-boxes, probabilistic nested intervals

Due to measurement errors, the result $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ of processing measurement outcomes is, in general, different from the desired result $y = f(x_1, \ldots, x_n)$ of processing actual (unknown) values x_i. It is desirable to estimate the difference $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$ [4].

When we only know the bounds Δ_i on measurement errors $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$, the only information that we have about y is that $y \in \boldsymbol{y} \stackrel{\text{def}}{=} \{f(x_1, \ldots, x_n) : x_1 \in \boldsymbol{x}_1, \ldots, x_n \in \boldsymbol{x}_n\}$, where $\boldsymbol{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$. Computing such a range $\boldsymbol{y}$ is one of the main problems solved by *interval computations* [2].

Often, in addition to the bounds Δ_i, we have partial information about the probability of different values Δx_i. A general probability distribution can be described by the cumulative distribution function (cdf) $F(x) \stackrel{\text{def}}{=} \text{Prob}(\eta \le x)$. Partial information means that instead of knowing the exact values $F(x)$, we only know *bounds* $\underline{F}(x) \le F(x) \le \overline{F}(x)$. The corresponding "interval-valued" cdf $[\underline{F}(x), \overline{F}(x)]$ is known as a *probability box*, or *p-box*, for short [1].

P-boxes are useful in decision making, where the objective is often to satisfy a given inequality-type constraint, and p-boxes provide the probability of satisfying this constraint. In many practical situations (e.g., in control applications), the objective is to find how far is the actual value y from our estimate $\widetilde{y}$. We know that the desired probability $p \stackrel{\text{def}}{=} \text{Prob}(-\Delta \le \eta \le \Delta)$ is equal to $F(\Delta) - F(-\Delta)$, so based on the known p-boxes, we can conclude that $p \le \widetilde{p} \stackrel{\text{def}}{=} \overline{F}(\Delta) - \underline{F}(-\Delta)$. However, often, this $\widetilde{p}$ is an overestimation: e.g., for $\Delta = 0$, we have $p = 0$, while for p-boxes of finite width w, we have $\widetilde{p} = 2w$.

To get better bounds for p, we use *probabilistic nested intervals*: 1-parametric families of confidence intervals $\boldsymbol{x}_i(\alpha)$ for which $\text{Prob}(x_i \in \boldsymbol{x}_i(\alpha)) \geq 1-\alpha$ and $\boldsymbol{x}_i(\alpha) \subseteq \boldsymbol{x}_i(\alpha')$ when $\alpha' < \alpha$. E.g., when we have a systematic error component with known bounds $[-\Delta_{si}, \Delta_{si}]$ and a normally distributed random error component with a known σ_i, the confidence intervals are obtained by adding the usual Gaussian confidence interval to the interval $[\widetilde{x}_i - \Delta_{si}, \widetilde{x}_i + \Delta_{si}]$.

Probabilistic nested intervals are a particular case of nested intervals [3]. However, [3] focused on expert estimates, where it was reasonable to assume that when we know that $x_i \in \boldsymbol{x}_i(\alpha)$ with confidence $1-\alpha$, then $y = f(x_1, \ldots, x_n) \in f(\boldsymbol{x}_1(\alpha), \ldots, \boldsymbol{x}_n(\alpha))$ with the same confidence $1-\alpha$. This assumption led to explicit formulas for propagating expert-related nested intervals through computations.

In contrast, it is usually assumed that random errors of different measurements are independent [4]; in this case, when for each i, we have $x_i \in \boldsymbol{x}_i(\alpha)$ with probability $\geq 1-\alpha$, then we can only conclude that $(x_1, \ldots, x_n) \in \boldsymbol{x}_1(\alpha) \times \ldots \times \boldsymbol{x}_n(\alpha)$ (and thus, that $y = f(x_1, \ldots, x_n) \in f(\boldsymbol{x}_1(\alpha), \ldots, \boldsymbol{x}_n(\alpha))$) with probability $\leq (1-\alpha)^n \ll 1-\alpha$. So, we need *new formulas* for propagating probabilistic nested intervals. Such formulas will be described in the talk.

When measurement errors Δx_i are small, we can safely ignore terms quadratic (and of higher order) in Δx_i. For this linearized case, we can use *automatic differentiation* to design efficient algorithms. We can further speed up computations because in practice, inputs are usually known with 5-10% accuracy. In such situations, the result can only be computed with a similar 1-digit accuracy, so there is no need to perform iterations that improve the 2nd digit. A practical example of such a speed-up will be presented.

References:

[1] S. Ferson, *RAMAS Risk Calc 4.0: Risk Assessment with Uncertain Numbers*, CRC Press, Boca Raton, Florida, 2002.

[2] R.E. Moore, R.B. Kearfott, M.J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.

[3] H.T. Nguyen, V. Kreinovich, Nested intervals and sets: concepts, relations to fuzzy sets, and applications, In: R.B. Kearfott et al. (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 1996, pp. 245–290.

[4] S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer, New York, 2005.