How to Gauge a Combination of Uncertainties of Different Type: General Foundations

Ingo Neumann[1], Vladik Kreinovich[2], and
Thach Ngoc Nguyen[3]

[1]Geodetic Institute, Leibniz University of Hannover
30167 Hannover, Germany, neumann@gih.uni-hannover.de
[2]University of Texas at El Paso
El Paso, Texas 79968, USA, vladik@utep.edu
[3]Banking University of Ho Chi Minh City
Ho Chi Minh City, Vietnam, Thachnn@buh.edu.vn

1. Need to Gauge Uncertainty

- Measurements are never absolutely accurate.
- The measurement result $\widetilde{x}$ is, in general, different from the actual (unknown) value x of the quantity.
- To understand how accurate is the measurement, we need to gauge the corresponding uncertainty.
- In other words, we must provide a number describing the measurement error $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$.
- For different types of uncertainty, it is natural to use different characteristics.
- For probabilistic uncertainty, we know the probability distribution of the corresponding measurement error.
- Then, a natural measure of deviation is the standard deviation σ.

2. Need to Gauge Uncertainty (cont-d)

- For interval uncertainty, we only know the upper bound Δ on the absolute value of the measurement error.
- This upper bound is an appropriate measure of uncertainty.
- It is reasonable to select a characteristic that is described in the same unit as the measured quantity.
- In this case, if we change the measuring unit to the one which is λ times smaller, then:
  - not only all numerical value x should multiply by λ ($x \to x' = \lambda \cdot x$), but also
  - the corresponding characteristic of uncertainty should change the same way: $u \to u' = \lambda \cdot c$.

3. Need to Gauge Uncertainty (cont-d)

- For many quantities like coordinate or charge, changing the sign does not change its physical sense; so:
  - if we simply change the sign of the quantity,
  - then the corresponding characteristic of uncertainty should not change: $u' = u$.
- Thus, in general:
  - when we go from x to $x' = c \cdot x$,
  - then the corresponding characteristic of uncertainty should change as $u' = |c| \cdot u$.

4. Need to Combine Uncertainty and to Gauge the Combined Uncertainty

- The measurement error often consists of several components: $\Delta x = \Delta x_1 + \ldots + \Delta x_k$.
- For each of these components Δx_i, we usually know the corresponding characteristic of uncertainty u_i.
- Based on these characteristics, we need to estimate the characteristic u of the overall uncertainty Δx.
- A similar problem occurs when we process data:
  - we have the results $\widetilde{x}_i$ of measuring quantities x_i;
  - we are interested in a quantity y related to x_i by a known dependence $y = f(x_1, \ldots, x_n)$.
- To estimate y, we use the measurement results $\widetilde{x}_i$ and thus, come up with an estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$.

5. Need to Gauge Combined Uncertainty (cont-d)

- We need to estimate the resulting approximation error

$$\Delta y = \widetilde{y} - y = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(x_1, \ldots, x_n) =$$

$$f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n).$$

- Measurements are usually reasonably accurate, so the measurement errors Δx_i are small; thus:
  - we can safely ignore terms which are quadratic or of higher order in terms of Δx_i, and
  - consider only linear terms in the Taylor expansion.
- Then, $\Delta y = \sum\limits_{i=1}^{k} c_i \cdot \Delta x_i$, where $c_i \stackrel{\text{def}}{=} \dfrac{\partial f}{\partial x_i}$ (computed as the point $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$).

6. Need to Gauge Combined Uncertainty (cont-d)

- Thus:
  - once we know the uncertainty characteristics u_i of each measurement error Δx_i,
  - we can find the uncertainty characteristics $U_i = |c_i| \cdot u_i$ of each term $X_i = c_i \cdot \Delta x_i$.
- Based on these characteristics, we need to estimate the uncertainty characteristic of the sum

$$\Delta y = X_1 + \ldots + X_n.$$

- Why do we want to characterize the joint uncertainty by a single number?
- Because we want to select a single less uncertain option.

7. Example: Portfolio Allocation

- A good example is the traditional Markowitz's portfolio allocation problem.
- In this problem, we have full information about all the probabilities.
- The objective is to find:
  - among all portfolios with the given value of expected rate of return,
  - the one with the smallest possible standard deviation (= smallest risk).

8. Need to Gauge Combined Uncertainty (cont-d)

- In many practical situations, we know probabilities only with some uncertainty.
- As a result, for each portfolio, in addition to the random uncertainty, we have an additional uncertainty.
- It caused by the fact that we only have partial knowledge about the probabilities.
- If we minimize the random component, we risk missing a huge interval component.
- If we minimize the interval component, we risk missing a huge random component.
- It is more adequate to minimize the appropriate combination of both uncertainties.
- This will make sure that none of the components become too large.

9. How Uncertainty Is Combined and Gauged Now

- In the case of probabilistic uncertainty, if we know the standard deviations σ_i of each component Δx_i and
  - we have no information about their correlation,
  - then a natural idea is to assume that the error components are independent.
- The same conclusion can be made if we use the Maximum Entropy approach.
- This approach selects:
  - among all possible joint distributions,
  - the distribution with the largest possible value of entropy.
- *Known:* the variance of the sum of independent random variables is equal to the sum of their variances.

10. How Uncertainty Is Gauged Now (cont-d)

- So for the variance σ^2 of the sum Δx we have

$$\sigma^2 = \sigma_1^2 + \ldots + \sigma_n^2 \text{ and } \sigma = \sqrt{\sigma_1^2 + \ldots + \sigma_n^2}.$$

- Sometimes, we only know that each component Δx_i can take any value from the interval $[-\Delta_i, \Delta_i]$.
- Then the largest possible value Δ of the sum is attained when:
  - each of the components Δx_i
  - attains its largest possible value Δ_i; then:

$$\Delta = \Delta_1 + \ldots + \Delta_n.$$

11. How Uncertainty Is Gauged Now (cont-d)

- In both probabilistic and interval cases, we know the uncertainty characteristics u_i of the components.
- But we have two different formulas for combining uncertainty:
  - in the probabilistic case, $u = \sqrt{u_1^2 + \ldots + u_n^2}$;
  - in the interval case, $u = u_1 + \ldots + u_n$.

12. What Is a General Case?

- What are the general properties of the binary combination operation $u * u'$?
- Adding 0 should not change anything, including the accuracy: $u * 0 = u$.
- The sum does not depend on the order in which we add the components.
- So the result of combination should also not depend on the order in which we combine the components.
- So, we should have

$$u * u' = u' * u \text{ (commutativity) and}$$

$$u * (u' * u'') = (u * u') * u'' \text{ (associativity).}$$

13. What Is a General Case (cont-d)

- Monotonicity:
  - if we replace one of the components with a less accurate one (with larger u),
  - the result cannot become more accurate:

$$\text{if } u_1 \le u_2 \text{ and } u_1' \le u_2', \text{ then } u_1 * u_1' \le u_2 * u_2'.$$

- It turns out that under these conditions, every combination operation has:
  - either the from $u*u' = (u^p+(u')^p)^{1/p}$ for some $p > 0$,
  - or the form $u * u' = \max(u, u')$ (corresponding to the limit case $p \to \infty$).

14. Remaining Problem: What If We Combine Uncertainties Of Different Type?

- In many cases, we have different information about the uncertainty of different components.
- For example, the measurement error is often represented as the sum of:
  - a systematic error (the mean value) and
  - the remaining part – known as random error.
- About the random error component, we usually know the standard deviation.
- So it can be a viewed as a probabilistic uncertainty.
- However, about the systematic error component, we only know the error bound: interval uncertainty.
- How should we gauge the result of combining uncertainties of different type?

15. How the Combination of Uncertainties of Different Type Is Gauged Now?

- There are several ways to gauge the combination of probabilistic and interval uncertainty.
- The first way takes into account that, in practice, probability distributions are often:
  - either Gaussian (normal)
  - or close to Gaussian.
- This empirical fact is easy to explain.
- Indeed, in many cases, the measurement error is a result of a large number of independent small factors.
- *Known:* the distribution of the sum of many small independent random variables is close to Gaussian.
- It tends to Gaussian when the number of components tends to ∞.

16. Combination of Uncertainties (cont-d)

- This fact is known as the *Central Limit Theorem*.
- Strictly speaking, a normally distributed random variable with 0 mean can take arbitrarily large value.
- Indeed, its probability density function $\rho(x)$ remains positive for all values x.
- However, from the practical viewpoint, the probabilities of very large values are very small.
- So, for all practical purposes, such values can be safely ignored.
- Thus, in practice, we assume that:
  - all the values of a normal random variable with 0 mean and standard deviation σ
  - are located in an interval $[-k_0 \cdot \sigma, k_0 \cdot \sigma]$.

17. Combination of Uncertainties (cont-d)

- Here, k_0 depends on how small the probability we can ignore.
- Usually, people take k_0 equal to 2 (corresponding to 5%), 3 (0.1%) and 6 (10^{-6}%).
- So, a random error component with standard deviation σ implies that this component lies in the interval

$$[-k_0 \cdot \sigma, k_0 \cdot \sigma].$$

- Thus, all we have to do to combine it with the interval uncertainty $[-\Delta, \Delta]$ is to combine the two intervals.
- As a result, we get $\Delta + k_0 \cdot \sigma$.

18. Combination of Uncertainties (cont-d)

- Another frequently used approach is based on the Maximum Entropy idea:
  - if we do not know the exact distribution,
  - then, out of all possible probability distributions, we should select the one whose entropy is the largest.
- For example, assume that all we know that the systematic error is located on the interval $[-\Delta, \Delta]$; then:
  - out of all possible probability distributions on this interval,
  - we should select the distribution whose entropy is the largest.
- This turns out to be the uniform distribution on this interval.

19. Combination of Uncertainties (cont-d)

- For the uniform distribution on the interval $[-\Delta, \Delta]$, the standard deviation is equal to $\dfrac{\Delta}{\sqrt{3}}$; thus:
  - to combine it with the random error component with known standard deviation σ,
  - it is sufficient to use the general formula for combining standard deviations:

$$\sqrt{\frac{\Delta^2}{2} + \sigma^2}.$$

- So what is the general formula?
- This is a problem to which, in this talk, we provide an answer.

20. Definitions

- Let us assume that $\mathcal{T}$ is the set of possible types of uncertainty with T elements.
- For simplicity, let us enumerate the types, i.e., let us identify $\mathcal{T}$ with the set $\{1, 2, \ldots, T\}$.
- By combining uncertainties from some subset $S \subseteq \mathcal{T}$, we get, in effect, a new type of uncertainty.
- Thus, we has, in effect, as many types of uncertainty as there are nonempty subsets $S \subseteq \mathcal{T}$.
- Since uncertainties can be of different type, in order to properly combine them, we need to know the type.
- Thus, an uncertainty is described not just by a number, but also by a type.

21. Definitions (cont-d)

- **Definition.** *Let a finite set $\mathcal{T}$ be given. By an* uncertainty, *we mean a pair (u, S), where:*
  - *u is a non-negative real number and*
  - *S is a non-empty subset of the set $\mathcal{T}$.*
- Let us consider binary operations $*$ that map two uncertainties (u, S) and (u', S') into a new uncertainty

$$(u'', S \cup S').$$

22. Definitions (cont-d)

Definition. *By a* combination operation, *we means a binary operation* $*$ *that has the following properties:*

- *the operation* $*$ *is commutative and associative;*
- *the operation* $*$ *is monotonic:*
  - *if* $u_1 \le u_2$, $u'_1 \le u'_2$, $(u_1, S) * (u'_1, S') = (u''_1, S \cup S')$ *and* $(u_2, S) * (u'_2, S') = (u''_2, S \cup S')$,
  - *then* $u''_1 \le u''_2$;
- *scale-invariance: for every* $\lambda > 0$, *if* $(u, S) * (u', S') = (u'', S \cup S')$, *then*

$$(\lambda \cdot u, S) * (\lambda \cdot u', S') = (\lambda \cdot u'', S \cup S');$$

- *zero-property: for each set* S, *we have* $(u, S) * (0, S) = (u, S)$; *and*
- *non-zero property: if* $u > 0$ *and* $(u, S) * (u', S') = (u'', S \cup S')$, *then* $u'' > 0$.

23. Main Result

Proposition. *For each combination operation, there exist positive values* $c_1, \ldots, c_T$ *such that:*

- *either* $(u_1, \{1\}) * (u_2, \{2\}) * \ldots * (u_T, \{T\}) = ((c_1^p \cdot u_1^p + \ldots + c_T^p \cdot u_T^p)^{1/p}, \mathcal{T})$ *for all* u_i,
- *or* $(u_1, \{1\}) * (u_2, \{2\}) * \ldots * (u_T, \{T\}) = (\max(c_1 \cdot u_1, \ldots, c_T \cdot u_T), \mathcal{T})$ *for all* u_i.

24. Proof

- Due to the zero property, we have

$$(u, \mathcal{T}) \stackrel{\text{def}}{=} (u_1, \{1\}) * (u_2, \{2\}) * \ldots * (u_T, \{T\}) =$$

$$(u_1, \{1\}) * (0, \{1\}) * \ldots * (0, \{1\}) * \ldots *$$
$$(u_T, \{T\}) * (0, \{T\}) * \ldots * (0, \{T\}).$$

- Here, each term $(0, \{t\})$ is repeated T times.
- Due to associativity and commutativity, we can rearrange the terms and get

$$(u, \mathcal{T}) = (u'_1(u_1), \mathcal{T}) * \ldots * (u'_T(u_T), \mathcal{T}), \text{ where}$$

$$(u'_t(u_t), \mathcal{T}) \stackrel{\text{def}}{=} (u_t, \{t\}) * (0, \{1\}) * \ldots * (0, \{T\}).$$

- Due to non-zero property, if $u_t = 1$, then $u'_t(1) \neq 0$.
- Let us denote $c_t \stackrel{\text{def}}{=} u'_t(1) > 0$.

25. Proof (cont-d)

- Due to scale-invariance, $u'_t(u_t) = c_t \cdot u_t$, so:

$$(u, \mathcal{T}) = (c_1 \cdot u_1, \mathcal{T}) * \ldots * (c_T \cdot u_T, \mathcal{T}).$$

- For the values of type $\mathcal{T}$, we get the usual properties of the combination operation.
- So, for uncertainties of this type:
  - either $(u, \mathcal{T}) * (u', \mathcal{T}) = ((u^p + (u')^p)^{1/p}, \mathcal{T})$
  - or $(u, \mathcal{T}) * (u', \mathcal{T}) = (\max(u, u'), \mathcal{T})$.
- In both cases, we get exactly the formulas from the proposition.
- The proposition is thus proven.
- *Comment.* Both existing methods for combining uncertainty are particular cases of this general approach.
- They correspond to $p = 1$ and $p = 2$.

26. Acknowledgments

This work was supported:

- by the Institute of Geodesy, Leibniz University of Hannover, and
- by the US National Science Foundation grants:
  - 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science),
  - and HRD-1242122 (Cyber-ShARE Center of Excellence).