Need for Techniques Intermediate Between Interval and Probabilistic Ones

Olga Kosheleva[1] and Vladik Kreinovich[2]

Departments of [1]Teacher Education and [2]Computer Science
University of Texas at El Paso, El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

1. Need to take uncertainty into account in high-performance computing

- One of the main applications of high performance computing is estimating the values of some quantities y based on the inputs $x_1, \ldots, x_n$.
- For example, in weather prediction:
  - we estimate tomorrow's temperature y at some location
  - based on the results x_i of meteorological measurements in the vicinity of this location.
- The problem is that:
  - even when the data processing algorithm $y = f(x_1, \ldots, x_n)$ describes the exact relation between y and x_i,
  - the value $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ – that we obtain by processing measurement results $\widetilde{x}_i$ – is not exact:
  - since the measurement results $\widetilde{x}_i$ are, in general, different from the actual (unknown) values x_i of the corresponding quantities.

2. Need to take uncertainty into account (cont-d)

- Because of the measurement errors $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$, the result $\widetilde{y}$ of data processing is, in general, different from the desired value y.
- It is important to provide an estimate for the resulting uncertainty

$$\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y.$$

3. What do we usually know and what we usually do not know about the measurement errors Δx_i

- For each measuring instrument, we know the upper bound Δ_i on the absolute value of the measurement error: $|\Delta x_i| \le \Delta_i$.
- Indeed, if no such bound is guaranteed:
  - this would mean that for any measurement result, the actual value can be anything,
  - this would be a wild guess, not a measuring instrument.
- In many practical applications, each measuring instrument is calibrated:
- Before using this instrument, we several times compare its results with the results of a much more accurate instrument.

4. What do we usually do (cont-d)

- Thus, if the mean value of the measurement error was not 0:
  - we can find this mean value (known as *bias*) and
  - correct for it by subtracting this mean value from all the measurement results.
- Thus, we can safely assume that for each instrument, the mean value of the measurement error is 0.
- In most applications, we can also safely assume that the measurement errors are relatively small.
- So we can safely ignore terms which are quadratic or higher order in terms of these errors.
- For example, even if the relative measurement error is 10%, its square is 1%, which can be safely ignored in comparison with 10%.
- This is often all we know.

5. What do we usually do (cont-d)

- Ideally, we should also know the probability distributions of all the measurement errors and all the correlations between them.
- In simple computations:
  - when the number n of inputs is small,
  - it is possible to extract this information for all n instruments and all $n^2/2$ pairs of instruments.
- So, for simple computations, this information is sometimes available.
- However, for high-performance computing, when n is large, it is not feasible to extract all this information.
- So this information is usually not available.

6. Possibility of linearization

- By definition of the measurement errors, we have $x_i = \widetilde{x}_i - \Delta x_i$, thus

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

- Since the measurement errors Δx_i are small, we can:
  - expand the expression $f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n)$ in Taylor series in terms of Δx_i and
  - keep only linear terms in this expansion.
- As a result, we get $\Delta y = \sum\limits_{i=1}^{n} c_i \cdot \Delta x_i$, where $c_i \stackrel{\text{def}}{=} \dfrac{\partial f}{\partial x_i}_{|x_1=\tilde{x}_1,\ldots,x_n=\tilde{x}_n}$.

7. How Δy is estimated now: interval technique

- Since we have no information about the correlation between the measurement errors, a natural idea is to consider all possible correlations.
- In general, since $|a+b| \le |a|+|b|$ and $|a \cdot b| = |a| \cdot |b|$, from the above formula, we get $|\Delta y| \le \sum\limits_{i=1}^{n} |c_i| \cdot |\Delta x_i|$.
- Since $|\Delta x_i| \le \Delta_i$, we get $|\Delta y| \le \Delta_{\text{int}} \stackrel{\text{def}}{=} \sum\limits_{i=1}^{n} |c_i| \cdot \Delta_i$.
- This value Δ_{int} is the exact upper bound, in the sense that it is possible to have $|\Delta y| = \Delta_{\text{int}}$ with probability 1.
- Indeed, this happens when:
  - with probability 1/2, we have $\Delta x_i = \Delta_i \cdot \text{sign}(c_i)$, where, as usual, $\text{sign}(x) = +1$ for $x > 0$ and $\text{sign}(x) = -1$ for $x < 0$; and
  - with probability 1/2, we have $\Delta x_i = -\Delta_i \cdot \text{sign}(c_i)$.

8. How Δy is estimated now: interval technique (cont-d)

- In this case:
  - with probability 1/2, we have $\Delta y = \Delta_{\rm int}$, and
  - with probability 1/2, we have $\Delta y = -\Delta_{\rm int}$.
- This worst-case estimate is known as the *interval estimate*, since:
  - this is the only estimate that we can guarantee based on the available information,
  - that all measurement errors Δx_i are located within the corresponding interval $[-\Delta_i, \Delta_i]$.

9. Interval technique: efficient algorithms

- How can we compute the value Δ_{int}?
- A natural idea is to explicitly use the above expression.
- When the function $f(x_1, \ldots, x_n)$ is given by an *explicit expression*, we can simply differentiate it and compute c_i.
- Often, the algorithm $f(x_1, \ldots, x_n)$ is only given as a proprietary *black box*.
- We do not know the exact algorithm, so we cannot differentiate this function.
- In this case, to find the values c_i, we can use numerical differentiation techniques.
- For example, we can take into account that, in general, the derivative is the limit of the ratios

$$c_i = \lim_{h_i \to 0} \frac{f(\widetilde{x}_1, \ldots, \widetilde{x}_{i-1}, \widetilde{x}_i + h_i, \widetilde{x}_{i+1}, \ldots, \widetilde{x}_n) - \widetilde{y}}{h_i}.$$

10. Interval technique: efficient algorithms (cont-d)

- Thus, we can estimate c_i as the value of this ratio for some small h_i.
- The limitation of this natural idea is that for large n, it requires applying the algorithm $f(x_1, \ldots, x_n)$ $n+1$ times:
  - one time to compute $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ and
  - n times to compute n values $c_1, \ldots, c_n$.
- In practice, the algorithm $f(x_1, \ldots, x_n)$ is often time-consuming – it may require several hours on a high-performance computer.
- The number of inputs n may be in the thousands.
- In this case, the overall time needed to apply this natural idea is unrealistically large.

11. Interval technique: efficient algorithms (cont-d)

- In such situations, we can use a faster algorithm based on using Cauchy distribution, with probability density proportional to

$$\frac{1}{1+\left(\frac{x}{\Delta}\right)^2}.$$

- This algorithm is based on the fact that:
  - if we have n independent random variables δ_i each of which is Cauchy-distributed with parameter Δ_i,
  - then their linear combination $\sum c_i \cdot \delta_i$ is also Cauchy-distributed, with parameter $\Delta = \sum |c_i| \cdot \Delta_i$ – which is exactly Δ_{int}.
- Thus, to estimate Δ_{int}, we, several times $k = 1, \ldots, K$, simulate the Cauchy-distributed values $\delta_i^{(k)}$.
- Then, we compute $\delta^{(k)} = f(\widetilde{x}_1 + \delta_1^{(k)}, \ldots, \widetilde{x}_n + \delta_n^{(k)}) - \widetilde{y}$.
- Due to linearization, we have $\delta^{(k)} = \sum\limits_{i=1}^{n} c_i \cdot \delta_i^{(k)}$.

12. Interval technique: efficient algorithms (cont-d)

- Then we use the resulting Cauchy-distributed sample $\delta^{(1)}, \ldots, \delta^{(K)}$ to estimate the desired parameter $\Delta = \Delta_{\text{int}}$.
- The advantage of this Cauchy-based method is that, as with all Monte-Carlo simulation methods:
  - the number of simulations – and thus, the number of times we apply the time-consuming algorithm $f(x_1, \ldots, x_n)$
  - depends only on the desired accuracy and does not depend on the number of inputs n.
- In general, the accuracy of a statistical estimate based on a sample of size K is approximately equal to $1/\sqrt{K}$.
- So, to estimate Δ_{int} with accuracy 20%, it is sufficient to apply the algorithm $f(x_1, \ldots, x_n)$ $K \approx 25$ times.
- This is much smaller than the thousands times needed for a direct estimation.

13. Comment

- The largest value of the expression Δ_{int} is attained at the endpoints of the corresponding intervals.
- So, a seemingly natural idea may be to apply the Monte-Carlo idea directly:
  - select, for each i, one of the endpoints $\Delta x_i = \Delta_i$ or $\Delta x_i = -\Delta_i$ with equal probability $1/2$,
  - repeat this several (K) times and
  - take the largest of the resulting values $\Delta y^{(1)}, \ldots, \Delta y^{(K)}$.
- Unfortunately, this simple idea leads to a drastic underestimation, even in the simplest case:
  - when $f(x_1, \ldots, x_n) = x_1 + \ldots + x_n$ and
  - when all the values x_i are measured with the same accuracy

$$\Delta_i = \Delta_1.$$

- Indeed, in this case, the above formula leads to $\Delta_{\text{int}} = n \cdot \Delta_1$.

14. Comment (cont-d)

- As we have mentioned, it is possible that this value is actually attained.
- On the other hand, when we use the above seemingly natural idea, then for large n, according to the Central Limit Theorem:
  - the distribution of the sum is close to Gaussian,
  - with mean 0 (equal to the sum of the means) and variance $V = \sigma^2$ equal to the sum of the variances $V = n \cdot \delta_1^2$.
- Thus, e.g., with confidence 99.9%, we can conclude that the resulting values $\Delta y^{(k)}$ are:
  - within the 3-sigma interval, i.e.,
  - smaller that $3\sigma = 3\sqrt{n} \cdot \Delta_1$.
- For large n, we have $3\sqrt{n} \ll n$.
- So indeed, this seemingly natural idea can lead to a drastic underestimation.

15. Interval technique: limitation

- The main problem with this approach is that the resulting worst-case estimates are too pessimistic.
- In most practical situations, the actual value Δy is much smaller than this estimate Δ_{int}.

16. How can we explain this limitation

- The above limitation can be easily explained.
- Indeed, in the arrangement that leads to $\Delta y = \Delta_{\text{int}}$, all measurement errors are highly correlated, with correlation coefficients ± 1.
- In practice, it is possible that common factors affect several measurement instruments.
- However, there are also usually other factors which affect only one measuring instrument.
- So the correlation is usually larger than -1 and smaller than 1.

17. How Δy is estimated now: probabilistic technique

- Another idea is that:
  - since we have no reason to prefer negative or positive correlation,
  - it is reasonable to assume that the correlation is 0, and,
  - more generally, that different measurement errors are independent.
- This is also what follows from the Maximum Entropy approach, when:
  - out of all possible joint distributions $\rho(\Delta x_1, \ldots, \Delta x_n)$ for which mean of each x_i is 0 and which are located on $[-\Delta_i, \Delta_i]$,
  - we select the distribution with the largest value of entropy

$$S \stackrel{\text{def}}{=} -\int \rho(\Delta x_1, \ldots, \Delta x_n) \cdot \ln(\rho(\Delta x_1, \ldots, \Delta x_n))\, d\Delta x_1 \ldots d\Delta x_n.$$

- Independence implies that $E[\Delta x_i \cdot \Delta x_j] = E[\Delta x_i] \cdot E[\Delta x_j]$.
- The mean value of each measurement error is 0, so $E[\Delta x_i \cdot \Delta x_j] = 0$.

18. How Δy is estimated now: probabilistic technique (cont-d)

- In this case, the expected value of $(\Delta y)^2$ is $E[(\Delta y)^2] = \sum\limits_{i=1}^{n} c_i^2 \cdot V_i$.
- Here by $V_i \stackrel{\text{def}}{=} E[(\Delta x_i - E[\Delta x_i])^2] = E[(\Delta x_i)^2]$, we denoted the variance of the i-th measurement error.
- As is well known in statistics, for large n, the deviation from this average is small – since:
  - this deviation grows with n as $\sqrt{n}$,
  - while the expected value itself grows as n.
- So we conclude that the actual value $(\Delta y)^2$ is, with high accuracy, equal to this expected value: $(\Delta y)^2 \approx \sum\limits_{i=1}^{n} c_i^2 \cdot V_i$.
- We do not know the variances V_i, but, since $|\Delta x_i| \le \Delta_i$, we have $(\Delta x_i)^2 \le \Delta_i^2$.

19. How Δy is estimated now: probabilistic technique (cont-d)

- Thus, the expected value V_i of the square $(\Delta x_i)^2$ is also bounded by the same bound Δ_i^2: $V_i \le \Delta_i^2$.
- This upper bound on the variance V_i is the best we can have – it is attained if:
  - we have $\Delta x_i = \Delta_i$ with probability 1/2, and
  - we have $\Delta x_i = -\Delta_i$ with probability 1/2.
- Thus, we conclude that $(\Delta y)^2 \le \sum\limits_{i=1}^{n} c_i^2 \cdot \Delta_i^2$, i.e., that:

$$|\Delta y| \le \Delta_{\text{prob}} \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{n} c_i^2 \cdot \Delta_i^2}.$$

20. Probabilistic technique: efficient algorithms

- How can we estimate the value $\Delta_{\rm prob}$?
- A natural idea is to explicitly use the above expression.
- When describing interval techniques, we have already mentioned how we can estimate the values c_i.
- The limitation of this natural idea is the same as for the similar interval idea:
  - for large n, it requires applying the algorithm $f(x_1,\ldots,x_n)$ too many times and
  - thus, the overall time needed to apply this natural idea is sometimes unrealistically large.

21. Probabilistic technique: efficient algorithms (cont-d)

- To speed up computations, we can use the fact that:
  - if we have n independent random variables δ_i each of which is normally distributed with mean 0 and standard deviation Δ_i,
  - then their linear combination $\sum c_i \cdot \delta_i$ is also normally distributed, with the standard deviation equal to the desired expression.
- Thus, to estimate Δ_{prob}, we:
  - several times $k = 1, \ldots, K$, simulate the normally distributed values $\delta_i^{(k)}$,
  - compute $\delta^{(k)} = f(\widetilde{x}_1 + \delta_1^{(k)}, \ldots, \widetilde{x}_n + \delta_n^{(k)}) - \widetilde{y}$,
  - due to linearization, we have $\delta^{(k)} = \sum\limits_{i=1}^{n} c_i \cdot \delta_i^{(k)}$,
  - then use the resulting normally distributed sample $\delta^{(1)}, \ldots, \delta^{(K)}$ to estimate the desired standard deviation Δ_{prob}.

22. Probabilistic technique: efficient algorithms (cont-d)

- This method has the same advantage as the Cauchy method – that when we use this method:
  - the number of times we apply the time-consuming algorithm $f(x_1, \ldots, x_n)$ depends only on the desired accuracy and
  - does not depend on the number of inputs n.

23. Probabilistic technique: limitation

- The main problem with this probabilistic technique is that it is too optimistic.
- It often drastically decreases the approximation error Δy.
- We had an example of such a drastic underestimation when we explained why:
  - a seemingly natural Monte-Carlo algorithm
  - does not lead to a reasonable estimate for interval uncertainty.

24. How can we explain this limitation

- The above limitation can be easily explained.
- Indeed, this technique assumes that all the measurement errors are independent.
- However, as we have mentioned, in reality, there may be common factors affecting several instruments, and thus, there is correlation.

25. Need for intermediate techniques

- The interval techniques are too pessimistic and the probability techniques are too optimistic.
- It is desirable to have intermediate techniques that would provide more realistic estimates.
- The main objective of this talk is to provide such estimates.

26. Main idea

- As we have mentioned, the problem with the interval technique is that it assumes that the absolute value of the correlation can be 1.
- In practice, it is always smaller than 1.
- Similarly, the problem with the probabilistic technique is that it assumes that all correlations are 0s.
- In practice, they can take non-zero values.
- So, a natural idea is to assume that:
  - there is some number b between 0 and 1
  - that provides an upper bound for absolute values $|r_{ij}|$ of all the correlations $r_{ij} \stackrel{\text{def}}{=} \dfrac{E[\Delta x_i \cdot \Delta x_j]}{\sigma_i \cdot \sigma_j}$, where $\sigma_i \stackrel{\text{def}}{=} \sqrt{V_i}$: $|r_{ij}| \le b$.

27. Main idea (cont-d)

- The value b can be determined empirically, by:
  - computing absolute value of the correlation for several randomly selected pairs of measuring instruments and
  - selecting the largest of these values.

28. Comments

- In principle, the experimental determination of the correlations is possible.
- However, it is not easy to do in the field.
- Hopefully:
  - when several different sensors are produced by the same manufacturer,
  - this manufacturer will be able to provide these correlation values.
- Our estimates are based on on computing the largest absolute value of the observed correlations.
- We can instead make a usual reasonable assumption that correlations are normally distributed.
- Then, based on the observed correlation values, we can find the sample mean μ and the sample standard deviation σ.

29. Comments (cont-d)

- We can then conclude, with some confidence, that the actual correlation values are within the interval $[\mu - k_0 \cdot \sigma, \mu + k_0 \cdot \sigma]$.
- Here, as usual, $k_0 = 2$ corresponds to reliability 95%, $k_0 = 3$ to 99.9%, and $k_0 = 6$ to reliability $1 - 10^{-8}$.

30. From the idea to the resulting formula

- From linearized formula, we conclude that

$$(\Delta y)^2 = \sum_{i=1} c_i^2 \cdot (\Delta x_i)^2 + \sum_{i \neq j} c_i \cdot c_j \cdot \Delta x_i \cdot \Delta x_j, \text{ hence}$$

$$E[(\Delta y)^2] = \sum_{i=1}^{n} c_i^2 \cdot E[(\Delta x_i)^2] + \sum_{i \neq j} c_i \cdot c_j \cdot E[\Delta x_i \cdot \Delta x_j], \text{ i.e.,}$$

$$E[(\Delta y)^2] = \sum_{i=1}^{n} c_i^2 \cdot V_i + \sum_{i \neq j} c_i \cdot c_j \cdot r_{ij} \cdot \sigma_i \cdot \sigma_j.$$

- We know that $(\Delta y)^2 \approx E[(\Delta y)^2]$, we know that $|r_{ij}| \leq b$, so:

$$(\Delta y)^2 \leq \sum_{i=1}^{n} c_i^2 \cdot \sigma_i^2 + \sum_{i \neq j} |c_i| \cdot |c_j| \cdot b \cdot \sigma_i \cdot \sigma_j.$$

- We have mentioned that $\sigma_i \leq \Delta_i$, thus

$$(\Delta y)^2 \leq \sum_{i=1}^{n} c_i^2 \cdot \Delta_i^2 + \sum_{i \neq j} |c_i| \cdot |c_j| \cdot b \cdot \Delta_i \cdot \Delta_j.$$

31. From the idea to the resulting formula (cont-d)

- Here, $\Delta_{\text{int}}^2=\left(\sum\limits_{i=1}^n |c_i|\cdot\Delta_i\right)^2=\sum\limits_{i=1}^n c_i^2\cdot\Delta_i^2+\sum\limits_{i\ne j}|c_i|\cdot|c_j|\cdot\Delta_i\cdot\Delta_j.$
- So, the above formula the formula takes the form

$$(\Delta y)^2\le b\cdot\Delta_{\text{int}}^2+(1-b)\cdot\left(\sum_{i=1}^n c_i^2\cdot\Delta_i^2\right),\ \text{i.e.,}$$

$$(\Delta y)^2\le b\cdot\Delta_{\text{int}}^2+(1-b)\cdot\Delta_{\text{prob}}^2,\ \text{and}$$

$$|\Delta y|\le\Delta_b\stackrel{\text{def}}{=}\sqrt{b\cdot\Delta_{\text{int}}^2+(1-b)\cdot\Delta_{\text{prob}}^2}.$$

32. How to compute this estimate

- As we have mentioned earlier, there exist efficient algorithms:
  - for computing Δ_{prob} – based on Monte-Carlo simulation of normally distributed measurement errors – and
  - for computing Δ_{int} – based on using Cauchy distribution.
- In both algorithms, the number of simulations depend only on the desired accuracy and does not depends on the number n of inputs.
- By using these algorithms, we can efficiently compute the new estimate.

33. Future work

- How realistic is this new estimate?
- How close is it to the actual error Δy?
- To answer these questions, it is necessary to test this method on real-life examples.

34. Acknowledgments

- This work was supported in part by the National Science Foundation grants:
  - 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and
  - HRD-1834620 and HRD-2034030 (CAHSI Includes).
- It was also supported by the AT&T Fellowship in Information Technology.
- It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.