Faster Algorithms for Estimating the Mean of a Quadratic Expression under Uncertainty

Martine Ceberio[1], Vladik Kreinovich[1], Olga Kosheleva[2], and Lev Ginzburg[3]
[1,2]Departments of [1]Computer Science and [2]Teacher Education
University of Texas at El Paso, El Paso, Texas 79968, USA
mceberio@utep.edu, vladik@utep.edu, olgak@utep.edu
[3]Department of Ecology and Evolution, Stony Brook University
Stony Brook, New York 11794, USA, levrginzburg@gmail.com

1. Need for uncertainty propagation

- In many practical situations:
  - we are interested in the value of a physical quantity y
  - which is difficult – or even impossible – to measure directly.
- For example, we may want to estimate tomorrow's temperature, the amount of oil in a given oilfield, or a distance to a faraway star.
- Since we cannot measure the quantity y directly, a natural idea is to measure this quantity indirectly, i.e.:
  - to measure easier-to-measure quantities $x_1, \ldots, x_n$ that are related to y by a known dependence $y = f(x_1, \ldots, x_n)$,
  - and then to use the results $\widetilde{x}_i$ of these measurements to provide the estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ for y.
- The problem is that measurements are never absolutely accurate.
- The measurement result $\widetilde{x}$ is, in general, different from the actual (unknown) value of the corresponding quantity.

2. Need for uncertainty propagation (cont-d)

- In other words, there is, usually, a non-zero difference $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$ known as a *measurement error*.

- As a result, the estimate $\widetilde{y}$ is, in general, different from the desired value y.

- It is therefore desirable:

  – to use available information about measurement errors Δx_i

  – to gauge the difference $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$ between the estimate and the actual value.

- Estimating this difference is known as *uncertainty propagation*.

3. Comment

- In many practical situations:
  - even if we know the exact values of the measured quantities $x_1, \ldots, x_n$,
  - we can only approximately determine the value y.
- For example, in weather prediction:
  - even if we know the exact values of the temperature, atmospheric pressure, etc. at all the locations where we placed the sensors,
  - we will still retain some uncertainty, since tomorrow's temperature is also affected by temperature at other locations.
- In such cases, the difference $\delta y \stackrel{\text{def}}{=} y - f(x_1, \ldots, x_n)$ between the actual value of the quantity y and the expression $f(x_1, \ldots, x_n)$ is non-zero.
- This difference – known as *approximation error* – is unpredictable, i.e., random.

4. Comment (cont-d)

- In other words, instead of the above equation, we have a more accurate formula $y = f(x_1, \ldots, x_n) + \delta y$.
- From the previous experience of predicting the value y, we can get a sample of the values of the approximation error.
- Based in this sample, we can determine the probability distribution of the approximation error.
- It may turn out that the mean value m of the approximation error is different from 0, i.e., that the approximate expression f has a bias.
- Then practitioners usually compensate for this bias by replacing the original function $f(x_1, \ldots, x_n)$ with a new function
$$f_{\text{new}}(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) + m.$$
- For this new function, the mean value of approximation error is 0.
- Thus, without loss of generality, we can safely assume that the mean value of the approximation error is 0.

5. Possibility of linear and quadratic approximations

- By definition of a measurement error, we have $x_i = \widetilde{x}_i - \Delta x_i$.
- Thus, the estimation error Δy can be represented as

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

- Measurement errors are usually reasonably small.
- Thus, terms which are quadratic in Δx_i are much smaller than linear terms, and terms which are cubic in Δx_i are even smaller.
- For example, for a 20% measurement error, its square is 4%, and its cube is 0.8%.
- In such situations, from the practical viewpoint:
  - cubic (and higher order) terms can be safely ignored
  - in comparison with linear and quadratic terms.

6. Possibility of linear and quadratic approximations (cont-d)

- So, to simplify computations, we can expand f and only keep linear and quadratic terms in this expansion.
- In this case, we have the expression

$$\Delta y = \sum_{i=1}^{n} a_i \cdot \Delta x_i + \sum_{i=1}^{n} a_{ii} \cdot (\Delta x_i)^2 + \sum_{i=1}^{n} \sum_{j \neq i} a_{ij} \cdot \Delta x_i \cdot \Delta x_j.$$

7. Probabilistic, interval, and fuzzy uncertainty

- In the ideal case:
  - we should know the probability distribution of each measurement error,
  - and, since the measurement errors can be correlated, the joint probability distribution of these measurement errors.
- However, determining these probability distributions is very time-consuming and thus, rarely done in practice.
- In many practical situations, all we know is the upper bound Δ on the absolute value of the measurement error Δx: $|\Delta x| \le \Delta$.
- In such situations, after we learn the measurement result $\widetilde{x}$:
  - the only information that we have about the actual (unknown) value of the corresponding quantity x
  - is that this value is contained in the interval $[\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.
- This situation is known as *interval uncertainty*.

8. Probabilistic, interval, and fuzzy uncertainty (cont-d)

- In some cases, all we know is the expert estimates on the possible values of the measurement error.
- These estimates are often expressed by using imprecise ("fuzzy") words from natural language, such as "small", "approximately 0.1".
- To describe the resulting knowledge in precise terms, it is reasonable to use fuzzy techniques.
- These techniques were specifically designed for such a description.
- In these techniques, in effect, for each level of confidence α, we provide a bound $\Delta(\alpha)$ that is satisfied with this degree of confidence.
- So, from the computational viewpoint, this is similar to interval uncertainty.
- The main difference that for each quantity, we have several different intervals corresponding to different levels of α.

9. Probabilistic, interval, and fuzzy uncertainty (cont-d)

- In view of this similarity, in the following text, we will only talk about interval uncertainty.
- However, all the results are also applied to the case when we have fuzzy uncertainty (expressed by the corresponding intervals).
- In addition to the upper bounds on the measurement errors Δx_i, we often have partial information on the probabilities of different values

$$\Delta x \stackrel{\text{def}}{=} (\Delta x_1, \ldots, \Delta x_n).$$

10. What is the typical case of partial information about the probabilities?

- In addition to the interval range of each variable Δx_i, we often know the mean ΔE_i of Δx_i.
- We get it, e.g., from the results of the testing the measuring instrument.
- Then, the mean is estimated as the average of measurement errors.
- The more tests we undertake:
  - the more information we get about the probability distribution, and
  - the more characteristics of the probability distribution we can determine.
- In the ideal situation, we can perform as many tests as necessary to determine the probability distribution of Δx_i.

11. What is the typical case of partial information about the probabilities (cont-d)

- In many real-life situations, however, we can only afford to determine one (or two) characteristics.
- In such situations, a natural choice is to determine the mean (and, if possible, the standard deviation).
- In measurement terms, the mean value of the measurement error is called the *systematic error* of the measurement procedure.
- In measurements, it is a common practice to *calibrate* the measuring instrument so that the systematic error (bias) is eliminated.
- Calibration means that:
  - instead of the original measured value $\widetilde{x}_i$ of the desired property,
  - we return the value $E_i \stackrel{\text{def}}{=} \widetilde{x}_i - \Delta E_i$.
- The mean value of the re-calibrated measurement error $\Delta x_i' \stackrel{\text{def}}{=} E_i - x_i$ is exactly 0.

12. What is the typical case of partial information about the probabilities (cont-d)

- The original measurement error Δx_i can attain any value from the interval $[-\Delta_i, \Delta_i]$.
- As a result, the re-calibrated measurement error Δ'_i can take all possible values from the interval $[-\Delta_i^-, \Delta_i^+]$, where

$$\Delta_i^- \stackrel{\text{def}}{=} \Delta_i + \Delta E_i \text{ and } \Delta_i^+ \stackrel{\text{def}}{=} \Delta_i - \Delta E_i.$$

- For example:
  - if we know that $\Delta x_i \in [-0.1, 0.1]$, and its mean is $\Delta E_i = 0.05$,
  - then for $\Delta x'_i = \Delta x_i - \Delta E_i$, the mean is 0, and the interval of possible values is $[-0.15, 0.05]$.
- The mean is the only information that we have about *each* measurement error.
- What do we know about the dependence *between* the corresponding random variables?

13. What is the typical case of partial information about the probabilities (cont-d)

- In many applications, we know that the same source of noise contributes to the errors of different measurements.
- So these errors Δx_i are not independent.
- We do not have enough statistics to get any information about each distribution except for its mean.
- Thus, we also cannot determine the correlation between Δx_i.
- So, if we are interested in guaranteed estimates, we must consider:
  - all possible n-dimensional distributions,
  - with all possible correlations.

14. Formulation of the problem

- In the case of interval uncertainty:
  - we know the intervals of possible values of the measurement errors Δx_i, and
  - we want to find the interval of possible values of the desired quantity $y = f(x_1, \ldots, x_n)$.
- We consider situations when:
  - in addition to the interval of possible values of measurement error,
  - we also know the mean of the measurement errors.
- In this case:
  - in addition to knowing the interval of possible values for the result y of data processing,
  - it is desirable to also know the interval of possible values of the mean value $E \stackrel{\text{def}}{=} E[\Delta y]$ of the quantity Δy.

15. Formulation of the problem (cont-d)

- In this paper, we consider the case when the dependence $f(x_1, \ldots, x_n)$ can be safely described by a quadratic function.
- So, the problem is:
  - we know the intervals $[-\Delta_i, \Delta_i]$ of possible values of Δx_i, and we know the means $\Delta E_i = E[\Delta x_i]$;
  - based on this information, we want to estimate the mean $E[\Delta y]$ of the quadratic expression.

16. Formulation of the problem (cont-d)

- Alternatively:
  - if we re-calibrate all the measuring instruments and take the re-calibrated measured values $E_i = \widetilde{x}_i - \Delta E_i$,
  - then we can expand f around these new values E_i:

$$\Delta y = a'_0 + \sum_{i=1}^{n} a'_i \cdot \Delta x'_i + \sum_{i=1}^{n} a'_{ii} \cdot (\Delta x'_i)^2 + \sum_{i=1}^{n} \sum_{j \neq i} a'_{ij} \cdot \Delta x'_i \cdot \Delta x'_j.$$

- After this calibration, the problem takes the following form:
  - we know the intervals $[-\Delta_i^-, \Delta_i^+]$ of possible values of $\Delta x'_i$, and we know that $E[\Delta x'_i] = 0$ for all i;
  - based on this information, we want to find the interval of possible values of $E[\Delta y]$.

17. Comment

- What if the dependence f is only approximate?
- Then, we get a similar formula, but with the additional term δy describing the approximation error:

$$\Delta y = a'_0 + \sum_{i=1}^{n} a'_i \cdot \Delta x'_i + \sum_{i=1}^{n} a'_{ii} \cdot (\Delta x'_i)^2 + \sum_{i=1}^{n} \sum_{j \neq i} a'_{ij} \cdot \Delta x'_i \cdot \Delta x'_j + \delta y.$$

- As we have mentioned earlier, the mean value of the approximation error is 0; so:
  - the expected value of the new expression is
  - exactly the same as the expected value of the original expression – that ignores the approximation error.
- Thus, in the problem of estimating the mean value of a quadratic expression, we can safely ignore the approximation error.

18. What is currently known

- A general quadratic expression is a linear combination of linear and quadratic terms.
- These terms are either Δx_i and $\Delta x_i \cdot \Delta x_j$ or $\Delta x'_i$ and $\Delta x'_i \cdot \Delta x'_j$.
- The expected value of a linear combination is equal to the linear combination of the corresponding expected values.
- The expected values of Δx_i and $\Delta x'_i$ are known; so:
  - to be able to estimate the expected value of y,
  - it is sufficient to be able to estimate the expected value of a product $\Delta x_i \cdot \Delta x_j$.
- In our previous papers, we showed that:
  - if we have two random variables v_1 and v_2 with known ranges $[\underline{v}_i, \overline{v}_i]$ and known means E_i,
  - then the interval $[\underline{E}, \overline{E}]$ of possible values of $E \stackrel{\text{def}}{=} E[v_1 \cdot v_2]$ can be computed as follows.

19. What is currently known (cont-d)

- First, we compute the auxiliary values $p_i \stackrel{\text{def}}{=} (E_i - \underline{v}_i)/(\overline{v}_i - \underline{v}_i)$.
- Then, we compute

$$\underline{E} = \min(p_1 + p_2 - 1, 0) \cdot \overline{v}_1 \cdot \overline{v}_2 + \min(p_1, 1 - p_2) \cdot \overline{v}_1 \cdot \underline{v}_2 +$$

$$\min(1 - p_1, p_2) \cdot \underline{v}_1 \cdot \overline{v}_2 + \max(1 - p_1 - p_2, 0) \cdot \underline{v}_1 \cdot \underline{v}_2;$$

$$\overline{E} = \min(1 - p_1, 1 - p_2) \cdot \underline{v}_1 \cdot \underline{v}_2 + \max(p_1 - p_2, 0) \cdot \overline{v}_1 \cdot \underline{v}_2 +$$

$$\max(p_2 - p_1, 0) \cdot \underline{v}_1 \cdot \overline{v}_2 + \min(p_1, p_2) \cdot \overline{v}_1 \cdot \overline{v}_2.$$

- In principle, we can use these formulas to estimate $E[\Delta y]$.

20. Remaining problem

- In the previous papers, our objective was:
  - to come up with an expression
  - for a *single* exact range of $E[v_1 \cdot v_2]$ for two variables v_1 and v_2.
- The fact that we found an explicit analytical expression makes it easy to compute the range.
- In our new problem, however, we need to estimate *many* ($\sim n^2$) such ranges.
- Indeed, n may be large.
- Each range computation requires two divisions (to compute p_1 and p_2) and several multiplications.
- Division is known to take longer to compute.
- We need to repeat these computations $\sim n^2$ times.

21. Remaining problem (cont-d)

- So, it is desirable to look for simpler expressions for $\underline{E}$ and $\overline{E}$.
- Ideally:
  - these expressions would hopefully avoid division altogether and require fewer multiplications, and
  - thus, will lead to faster computations.

22. Our result

- The main result of this paper is that such faster-to-compute expressions are indeed possible:
- Suppose that we have two random variables v_1 and v_2 with known means E_i and known ranges $[E_i - \Delta_i^-, E_i + \Delta_i^+]$.
- Then the interval $[\underline{E}, \overline{E}]$ of possible values of $E = E[v_1 \cdot v_2]$ is equal to

  $$[E_1 \cdot E_2 - \min(\Delta_1^- \cdot \Delta_2^-, \Delta_1^+ \cdot \Delta_2^+), E_1 \cdot E_2 + \min(\Delta_1^- \cdot \Delta_2^+, \Delta_1^+ \cdot \Delta_2^-)].$$

- For the expression v_i^2, the range can be computed by using the known general techniques.

23. Auxiliary result

- Suppose that we have a random variable v_i with a known mean E_i and a known range $[E_i - \Delta_i^-, E_i + \Delta_i^+]$.
- Then, the interval $[\underline{M}_i, \overline{M}_i]$ of possible values of $M_i = E[v_i^2]$ is equal to $[E_i^2, E_i^2 + \Delta_i^- \cdot \Delta_i^+]$.

24. Let us apply these results

- We want to apply these results to the variables $v_i = \Delta x'_i$ for which $E_i = E[\Delta x'_i] = 0$.
- As a result, the enclosure $\mathbf{E}_{\text{orig}}$ for the range $[\underline{E}, \overline{E}]$ of $E = E[\Delta y]$ takes the following form:

$$\mathbf{E}_{\text{orig}} = a'_0 - \sum_i a'_{ii} \cdot [0, \Delta_i^- \cdot \Delta_i^+] +$$

$$\sum_{i=1}^n \sum_{j \neq i} a'_{ij} \cdot [-\min(\Delta_i^- \cdot \Delta_j^+, \Delta_i^+ \cdot \Delta_j^-), \min(\Delta_i^- \cdot \Delta_j^-, \Delta_i^+ \cdot \Delta_j^+)].$$

25. Comment

- Alternatively, we can represent the matrix a'_{ij} in terms of its eigen-values λ'_k and the corresponding unit eigenvectors $e'_k = (e'_{k1}, \ldots, e'_{kn})$.
- In this case,

$$\Delta y = a'_0 + \sum_{i=1}^{n} a'_i \cdot \Delta x'_i + \sum_{k=1}^{n} \lambda'_k \cdot \left(\sum_{i=1}^{n} e'_{ki} \cdot \Delta x'_i \right).$$

- Each variable $\Delta x'_i$ has 0 mean and range $[-\Delta_i^-, \Delta_i^+]$.
- Thus, for the linear combination $\sum\limits_{i=1}^{n} e'_{ki} \cdot \Delta x'_i$:
  - the mean is equal to 0, and
  - the range is

$$\sum_{i=1}^{n} e'_{ki} \cdot [-\Delta_i^-, \Delta_i^+].$$

26. Comment (cont-d)

- In other words, the range is equal to $[-\delta_k^-, \delta_k^+]$, where

$$\delta_k^- = \sum_{i:e'_{ki}<0} |e'_{ki}| \cdot \Delta_i^+ + \sum_{i:e'_{ki}>0} e'_{ki} \cdot \Delta_i^-;$$

$$\delta_k^+ = \sum_{i:e'_{ki}<0} |e'_{ki}| \cdot \Delta_i^- + \sum_{i:e'_{ki}>0} e'_{ki} \cdot \Delta_i^+.$$

- As a result, we get a different enclosure $\mathbf{E}_{\text{new}}$ for the range $[\underline{E}, \overline{E}]$ of $E = E[\Delta y]$:

$$\mathbf{E}_{\text{new}} = a'_0 + \sum_{k=1}^{n} \lambda_k \cdot [0, \delta_k^- \cdot \delta_k^+].$$

27. Even faster computations are possible

- According to our result, in general:
  - to compute the range for E, it is sufficient to perform 5 multiplications:
  - one to compute $E_1 \cdot E_2$ and 4 to compute 4 products $\Delta_1^{\pm} \cdot \Delta_2^{\pm}$.
- In particular, for $v_i = \Delta x_i'$, since we have $E_i = 0$, we only need 4 multiplications.
- It is possible:
  - to follow the ideas behind the fast algorithm for interval multiplication and
  - to reduce the number of multiplications by one.
- For this, first, we compare Δ_1^- with Δ_1^+ and Δ_2^- with Δ_2^+.
- As a result, we get $2 \times 2 = 4$ different comparison results.

28. Even faster computations are possible (cont-d)

- In all 4 cases, we can avoid at least one multiplication in the above formula.
- Indeed:
  - if $\Delta_1^- \le \Delta_1^+$ and $\Delta_2^- \le \Delta_2^+$, then $\min(\Delta_1^-\cdot\Delta_2^-, \Delta_1^+\cdot\Delta_2^+) = \Delta_1^-\cdot\Delta_2^-$;
  - if $\Delta_1^- \le \Delta_1^+$ and $\Delta_2^+ \le \Delta_2^-$, then $\min(\Delta_1^-\cdot\Delta_2^+, \Delta_1^+\cdot\Delta_2^-) = \Delta_1^-\cdot\Delta_2^+$;
  - if $\Delta_1^+ \le \Delta_1^-$ and $\Delta_2^- \le \Delta_2^+$, then $\min(\Delta_1^-\cdot\Delta_2^+, \Delta_1^+\cdot\Delta_2^-) = \Delta_1^+\cdot\Delta_2^-$;
  - if $\Delta_1^+ \le \Delta_1^-$ and $\Delta_2^+ \le \Delta_2^-$, then $\min(\Delta_1^-\cdot\Delta_2^-, \Delta_1^+\cdot\Delta_2^+) = \Delta_1^+\cdot\Delta_2^+$.

29. Proof of our result

- Let us start with the known expression for $\overline{E}$.
- When $p_1 \ge p_2$, this expression takes the following simplified form:

$$\overline{E} = p_2 \cdot \overline{v}_1 \cdot \overline{v}_2 + (p_1 - p_2) \cdot \overline{v}_1 \cdot \underline{v}_2 + (1 - p_1) \cdot \underline{v}_1 \cdot \underline{v}_2.$$

- Grouping together terms proportional to p_1 and p_2, we conclude that

$$\overline{E} = p_2 \cdot \overline{v}_1 \cdot (\overline{v}_2 - \underline{v}_2) + p_1 \cdot (\overline{v}_1 - \underline{v}_1) \cdot \underline{v}_2 + \underline{v}_1 \cdot \underline{v}_2.$$

- Substituting the expression defining p_i into this formula, we conclude that

$$\overline{E} = \overline{v}_1 \cdot (E_2 - \underline{v}_2) + (E_1 - \underline{v}_1) \cdot \underline{v}_2 + \underline{v}_1 \cdot \underline{v}_2.$$

- Grouping the last two terms, we get

$$\overline{E} = \overline{v}_1 \cdot (E_2 - \underline{v}_2) + E_1 \cdot \underline{v}_2.$$

30. Proof of our result (cont-d)

- Finally, substituting $\underline{v}_i = E_i - \Delta_i^-$ and $\overline{v}_i = E_i + \Delta_i^+$, we conclude that

$$\overline{E} = (E_1 + \Delta_1^+) \cdot \Delta_2^- + E_1 \cdot (E_2 - \Delta_2^-) = E_1 \cdot \Delta_2^- + \Delta_1^+ \cdot \Delta_2^- + E_1 \cdot E_2 - E_1 \cdot \Delta_2^-,$$

$$\overline{E} = E_1 \cdot E_2 + \Delta_1^+ \cdot \Delta_2^-. \tag{9}$$

- Similarly, if $p_1 \le p_2$, we get

$$\overline{E} = E_1 \cdot E_2 + \Delta_1^- \cdot \Delta_2^+. \tag{10}$$

- The condition $p_1 \ge p_2$ means that

$$(E_1 - \underline{v}_1)/(\overline{v}_1 - \underline{v}_1) \ge (E_2 - \underline{v}_2)/(\overline{v}_2 - \underline{v}_2).$$

- This condition can be equivalently described as

$$(E_1 - \underline{v}_1) \cdot (\overline{v}_2 - \underline{v}_2) \ge (E_2 - \underline{v}_2) \cdot (\overline{v}_1 - \underline{v}_1).$$

- In terms of E_i and $\Delta_i^\pm$ it takes the form

$$\Delta_1^- \cdot (\Delta_2^- + \Delta_2^+) \ge \Delta_2^- \cdot (\Delta_1^- + \Delta_1^+).$$

31. Proof of our result (cont-d)

- This is equivalent to

$$\Delta_1^- \cdot \Delta_2^+ \geq \Delta_1^+ \cdot \Delta_2^-.$$

- This condition determines whether we have an expression corresponding to $p_1 \leq p_2$ or to $p_2 \leq p_1$.
- We thus get the desired formula for $\overline{E}$.
- For $\underline{E}$, we similarly consider two cases: $p_1 + p_2 \geq 1$ and $p_1 + p_2 < 1$.
- In the first case, we have

$$\underline{E} = (p_1 + p_2 - 1) \cdot \overline{v}_1 \cdot \overline{v}_2 + (1 - p_2) \cdot \overline{v}_1 \cdot \underline{v}_2 + (1 - p_1) \cdot \underline{v}_1 \cdot \overline{v}_2.$$

- This leads to

$$\underline{E} = p_1 \cdot (\overline{v}_1 - \underline{v}_1) \cdot \overline{v}_2 + p_2 \cdot \overline{v}_1 \cdot (\overline{v}_2 - \underline{v}_2) +$$

$$\underline{v}_1 \cdot \overline{v}_2 + \overline{v}_1 \cdot \underline{v}_2 - \overline{v}_1 \cdot \overline{v}_2.$$

32. Proof of our result (cont-d)

- Substituting the expressions for p_i, we conclude that

$$\underline{E} = (E_1 - \underline{v}_1) \cdot \overline{v}_2 + \overline{v}_1 \cdot (E_2 - \underline{v}_2) + \underline{v}_1 \cdot \overline{v}_2 + \overline{v}_1 \cdot \underline{v}_2 - \overline{v}_1 \cdot \overline{v}_2.$$

- So, we have:

$$\underline{E} = E_1 \cdot \overline{v}_2 + \overline{v}_1 \cdot E_2 - \overline{v}_1 \cdot \overline{v}_2.$$

- Finally, substituting the expressions $\underline{v}_i = E_i - \Delta_i^-$ and $\overline{v}_i = E_i + \Delta_i^+$, we conclude that

$$\underline{E} = E_1 \cdot E_2 - \Delta_1^+ \cdot \Delta_2^+.$$

- Similarly, if $p_1 + p_2 < 1$, then

$$\underline{E} = p_1 \cdot \overline{v}_1 \cdot \underline{v}_2 + p_2 \cdot \underline{v}_1 \cdot \overline{v}_2 + (1 - p_1 - p_2) \cdot \underline{v}_1 \cdot \underline{v}_2.$$

- This leads to

$$\underline{E} = p_1 \cdot (\overline{v}_1 - \underline{v}_1) \cdot \underline{v}_2 + p_2 \cdot \underline{v}_1 \cdot (\overline{v}_2 - \underline{c}_2) + \underline{v}_1 \cdot \underline{v}_2.$$

33. Proof of our result (cont-d)

- Substituting the expressions for p_i, we conclude that

$$\underline{E} = (E_1 - \underline{v}_1) \cdot \underline{v}_2 + \underline{v}_1 \cdot (E_2 - \underline{v}_2) + \underline{v}_1 \cdot \underline{v}_2.$$

- This is equivalent to

$$\underline{E} = E_1 \cdot \underline{v}_2 + \underline{v}_1 \cdot (E_2 - \underline{v}_2).$$

- Finally, substituting the expressions $\underline{v}_i = E_i - \Delta_i^-$ and $\overline{v}_i = E_i + \Delta_i^+$, we conclude that

$$\underline{E} = E_1 \cdot E_2 - \Delta_1^- \cdot \Delta_2^-.$$

- The inequality $p_1 + p_2 \geq 1$ can be reformulated as

$$(E_1 - \underline{v}_1) \cdot (\overline{v}_2 - \underline{v}_2) + (E_2 - \underline{v}_2) \cdot (\overline{v}_1 - \underline{v}_1) \geq (\overline{v}_1 - \underline{v}_1) \cdot (\overline{v}_2 - \underline{v}_2).$$

- Subtracting the first product from both sides, we get"

$$(E_2 - \underline{v}_2) \cdot (\overline{v}_1 - \underline{v}_1) \geq (\overline{v}_1 - E_1) \cdot (\overline{v}_2 - \underline{v}_2).$$

34. Proof of our result (cont-d)

- In terms of $\Delta_i^{\pm}$, we get:

$$\Delta_2^- \cdot (\Delta_1^- + \Delta_1^+) \geq \Delta_1^+ \cdot (\Delta_2^- + \Delta_2^+).$$

- This is equivalent to

$$\Delta_1^- \cdot \Delta_2^- \geq \Delta_1^+ \cdot \Delta_2^+.$$

- This condition determines whether we have one of the two above expressions.
- We thus get the desired formula for $\overline{E}$.
- The proposition is proven.

35. Proof of the auxiliary result

- It is known that the second moment M_i is equal to $V_i + E_i^2$, where V_i is the variance.
- The variance is the second moment of the auxiliary variable

$$v \stackrel{\text{def}}{=} v_i - E_i.$$

- For v, $E[v] = 0$ and the range of possible values is equal to $[-\Delta_i^-, \Delta_i^+]$.
- Thus, to prove the result, it is sufficient to prove that:
  - if we have a random variable v with a known mean $E[v] = 0$ and a known range $[-\Delta^-, \Delta^+]$,
  - then the interval $[\underline{M}, \overline{M}]$ of possible values of $M = E[v^2]$ is equal to $[0, \Delta^- \cdot \Delta^+]$.
- The second moment is always non-negative, so $\underline{M} \geq 0$.
- It is possible that v is identically 0, in which case $M = 0$; thus, $\underline{M} = 0$.

36. Proof of the auxiliary result (cont-d)

- We consider all possible distributions with 0 mean located on the interval $[-\Delta^-, \Delta^+]$.
- Among such distributions, we want to find a distribution for which the second moment is the largest.
- For discrete distributions:
  - that attains values $x_1, \ldots, x_n$ with probabilities $p_1, \ldots, p_n$,
  - this means that we must maximize the expression $\sum p_i \cdot x_i^2$ under the constraints $p_i \geq 0$, $\sum p_i = 1$, and $\sum p_i \cdot x_i = 1$.
- Once the values x_i are fixed, this constraint optimization problem becomes a linear programming problem with 2 equality constraints.
- According to the general properties of linear programming problems problems:
  - the maximum is attained
  - when at most 2 of the variables p_i are non-zero.

37. Proof of the auxiliary result (cont-d)

- Thus, to find the value $\overline{M}$, it is sufficient to consider distributions located at only two points x_1 and x_2.
- Since the average is 0, one of these values should be negative, and another positive.
- Let us denote the negative value by $-x^-$ and the positive one by x^+, and the corresponding probabilities by p^- and p^+.
- Since $p^- + p^+ = 1$, we get $p^+ = 1 - p^-$.
- Thus, the condition becomes $p^- \cdot (-x^-) + (1 - p^-) \cdot x^+ = 0$, i.e., equivalently, $p^-(x^- + x^+) = x^+$.
- We conclude that $p^- = x^+/(x^- + x^+)$ – and hence, that
$$p^+ = 1 - p^- = x^-/(x^- + x^+).$$

38. Proof of the auxiliary result (cont-d)

- Therefore,
$$M = p^- \cdot (x^-)^2 + p^+ \cdot (x^+)^2 =$$
$$\frac{x^+}{x^- + x^+} \cdot (x^-)^2 + \frac{x^-}{x^- + x^+} \cdot (x^+)^2 = \frac{x^- \cdot x^+ \cdot (x^- + x^+)}{x^- + x^+} = x^- \cdot x^+.$$
- This expression is a strictly increasing function of both its variables $x^-, x^+ \geq 0$.
- Thus, its maximum under the constraints $x^- \leq \Delta^-$ and $x^+ \leq \Delta^+$ is attained when $x^- = \Delta^-$ and $x^+ = \Delta^+$.
- So, the corresponding value is exactly the one we described.
- The proposition is proven.

39. What if measurements are actually independent?

- In the above text, we assumed that:
  - the measurement errors Δx_i can be correlated, and
  - we have no information about possible correlations.
- In some cases, however, we do know that the variables Δx_i are independent.
- Under this additional assumption of independence, what is the possible range of $E[\Delta y]$?
- This question is the easiest to answer under the formula using $\Delta x'_i$.
- Indeed, in this case, $E[\Delta x'_i] = 0$, and for $i \neq j$, we have $E[\Delta x'_i \cdot \Delta x'_j] = E[\Delta x'_i] \cdot E[\Delta x'_j] = 0 \cdot 0 = 0$.
- So, we conclude that

$$E = E[\Delta y] = a'_0 + \sum_{i=1}^{n} a'_{ii} \cdot E[(\Delta x'_i)^2].$$

40. What if measurements are actually independent (cont-d)

- From the auxiliary result, we already know how to compute the range of possible values of $E[(\Delta x'_i)^2]$ for $i = 1, \ldots, n$.
- So, in the case of independence, the interval of possible values of E is equal to

$$[\underline{E}, \overline{E}] = a'_0 + \sum_{i=1}^{n} a'_{ii} \cdot [0, \Delta_i^- \cdot \Delta_i^+].$$

41. Remaining open problem

- In practice:
  - in addition to the first moments $E[\Delta x_i]$,
  - we often know the second moments $E[\Delta x_i \cdot \Delta x_j]$ of the corresponding distributions.
- If we know the second moments, then, of course, computing the first moment of the quadratic expression y is easy.
- Indeed, y is a linear combination of terms Δx_i and $\Delta x_i \cdot \Delta x_j$.
- However, in this case, a natural next question remains open: what can we say about the second moment of y?

42. Acknowledgments

This work was supported in part by:

- National Science Foundation grants 1623190, HRD-1834620, HRD-2034030, and EAR-2225395;
- AT&T Fellowship in Information Technology;
- program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
- a grant from the Hungarian National Research, Development and Innovation Office (NRDI).