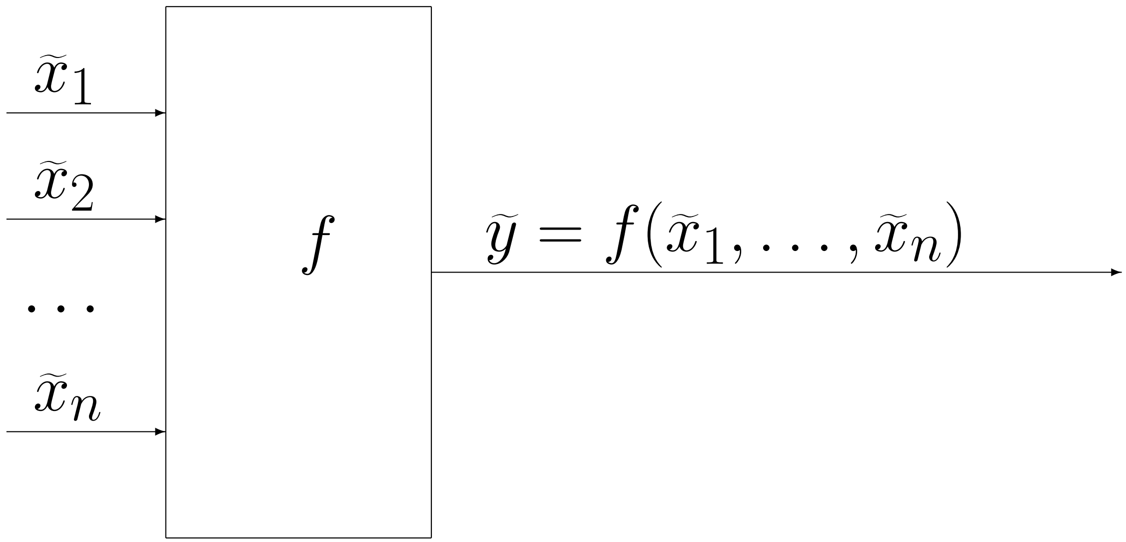
How to Fuse Interval and Probabilistic Uncertainty: From Practice to Theory and Back

Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso TX 79968, USA
vladik@utep.edu
http://www.cs.utep.edu/vladik

1. General Problem of Data Processing under Uncertainty

- *Indirect measurements:* way to measure y that are difficult (or even impossible) to measure directly.
- *Idea:* $y = f(x_1, \ldots, x_n)$

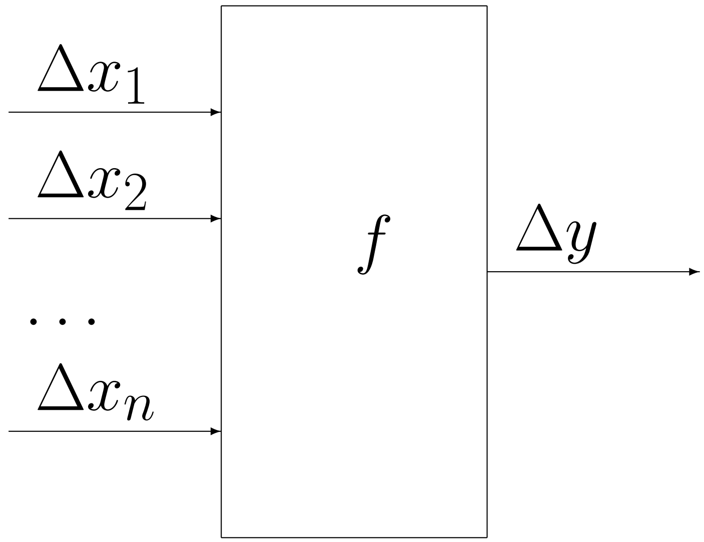


- *Problem:* measurements are never 100% accurate: $\widetilde{x}_i \neq x_i$ ($\Delta x_i \neq 0$) hence

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

What are bounds on $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$?

2. Probabilistic and Interval Uncertainty



- *Traditional approach:* we know probability distribution for Δx_i (usually Gaussian).
- *Where it comes from:* calibration using standard MI.
- *Problem:* calibration is not possible in:
  - fundamental science
  - manufacturing
- *Solution:* we know upper bounds Δ_i on $|\Delta x_i|$ hence

$$x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i].$$

3. Interval Arithmetic: Foundations of Interval Techniques

- *Problem:* compute the range

$$[\underline{y}, \overline{y}] = \{f(x_1, \ldots, x_n) \,|\, x_1 \in [\underline{x}_1, \overline{x}_1], \ldots, x_n \in [\underline{x}_n, \overline{x}_n]\}.$$

- *Interval arithmetic:* for arithmetic operations $f(x_1, x_2)$ (and for elementary functions), we have explicit formulas for the range.

- *Examples:* when $x_1 \in \mathbf{x}_1 = [\underline{x}_1, \overline{x}_1]$ and $x_2 \in \mathbf{x}_2 = [\underline{x}_2, \overline{x}_2]$, then:
  - The range $\mathbf{x}_1 + \mathbf{x}_2$ for $x_1 + x_2$ is $[\underline{x}_1 + \underline{x}_2, \overline{x}_1 + \overline{x}_2]$.
  - The range $\mathbf{x}_1 - \mathbf{x}_2$ for $x_1 - x_2$ is $[\underline{x}_1 - \overline{x}_2, \overline{x}_1 - \underline{x}_2]$.
  - The range $\mathbf{x}_1 \cdot \mathbf{x}_2$ for $x_1 \cdot x_2$ is $[\underline{y}, \overline{y}]$, where

$$\underline{y} = \min(\underline{x}_1 \cdot \underline{x}_2, \underline{x}_1 \cdot \overline{x}_2, \overline{x}_1 \cdot \underline{x}_2, \overline{x}_1 \cdot \overline{x}_2);$$

$$\overline{y} = \max(\underline{x}_1 \cdot \underline{x}_2, \underline{x}_1 \cdot \overline{x}_2, \overline{x}_1 \cdot \underline{x}_2, \overline{x}_1 \cdot \overline{x}_2).$$

- The range $1/\mathbf{x}_1$ for $1/x_1$ is $[1/\overline{x}_1, 1/\underline{x}_1]$ (if $0 \notin \mathbf{x}_1$).

4. Interval Techniques: General Idea

- First, we check for monotonicity, by using straightforward interval computations to find the range of the partial derivatives.
- If the function is increasing in one of the variables, its max is when $x_i = \overline{x}_i$ and min when $x_i = \underline{x}_i$.
- Thus, we reduce the range computation problem for a function of n variables to two functions on $n-1$ variables.
- For non-monotonic case, we can use centered form $f(x_1, \ldots, x_n) \in \mathbf{Y}$, where

$$\mathbf{Y} = \widetilde{y} + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \cdot [-\Delta_i, \Delta_i].$$

- If we want more accurate estimates, we bisect.

5. Need for Fusing Interval and Probabilistic Uncertainty

- In many practical situations, we need to fuse data corresponding to different types of uncertainty.
- In some cases, we know the probability distribution of the corresponding measurement error.
- Then, we have *probabilistic uncertainty.*
- In other cases, we only know upper bound Δ on the measurement error.
- Then, we have *interval uncertainty.*
- Indeed, if the measurement result is $\widetilde{x}$, then all we know about the actual (unknown) value x is that $x \in [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.
- In yet other cases, we have partial information about the corresponding probabilities.
- To make a decision, we need to combine (fuse) data corresponding to different types of uncertainty.

6. Fusing Interval and Probabilistic Uncertainty: Problem

- Many semi-heuristic techniques have been proposed for fusing interval and probabilistic uncertainty:
  - p-boxes,
  - Dempster-Shafer techniques,
  - maximum entropy ideas, etc.
- The problem is that often, several seemingly reasonable techniques lead to different results.
- Which of them should we choose?
- To make such a choice, it is desirable to go from practice to theory, i.e., to consider the problem from the fundamental viewpoint.
- The hope is that this analysis will help us go back to practice.

7. How Can We Decide Which Technique Is Better?

- The ultimate goal of all this analysis is to enable practitioners to make reasonable decisions.
- To get a proper combination of different types of uncertainty, we thus need to start by analyzing what types of uncertainty are most appropriate for decision making.
- For this, we need to take into account the users' preferences.
- Traditional approach is based on an assumption that for each two alternatives A' and A'', a user can tell:
  - whether the first alternative is better for him/her; we will denote this by $A'' < A'$;
  - or the second alternative is better; we will denote this by $A' < A''$;
  - or the two given alternatives are of equal value to the user; we will denote this by $A' = A''$.

8. The Notion of Utility

- Under the above assumption, we can form a natural numerical scale for describing preferences.
- Let us select a very bad alternative A_0 and a very good alternative A_1.
- Then, most other alternatives are better than A_0 but worse than A_1.
- For every prob. $p \in [0, 1]$, we can form a lottery $L(p)$ in which we get A_1 w/prob. p and A_0 w/prob. $1 - p$.
- When $p = 0$, this lottery simply coincides with the alternative A_0:

$$L(0) = A_0.$$

- The larger the probability p of the positive outcome increases, the better the result:

$$p' < p'' \text{ implies } L(p') < L(p'').$$

9. The Notion of Utility (cont-d)

- Finally, for $p=1$, the lottery coincides with the alternative A_1:

$$L(1)=A_1.$$

- Thus, we have a continuous scale of alternatives $L(p)$ that monotonically goes from $L(0)=A_0$ to $L(1)=A_1$.
- Due to monotonicity, when p increases, we first have $L(p)<A$, then we have $L(p)>A$.
- The threshold value is called the *utility* of the alternative A:

$$u(A)\stackrel{\text{def}}{=}\sup\{p:L(p)<A\}=\inf\{p:L(p)>A\}.$$

- Then, for every $\varepsilon>0$, we have

$$L(u(A)-\varepsilon)<A<L(u(A)+\varepsilon).$$

- We will describe such (almost) equivalence by $\equiv$, i.e., we will write that

$$A\equiv L(u(A)).$$

10. Fast Iterative Process for Determining $u(A)$

- *Initially:* we know the values $\underline{u} = 0$ and $\overline{u} = 1$ such that $A \equiv L(u(A))$ for some $u(A) \in [\underline{u}, \overline{u}]$.
- *What we do:* we compute the midpoint $u_{\rm mid}$ of the interval $[\underline{u}, \overline{u}]$ and compare A with $L(u_{\rm mid})$.
- *Possibilities:* $A \le L(u_{\rm mid})$ and $L(u_{\rm mid}) \le A$.
- *Case 1:* if $A \le L(u_{\rm mid})$, then $u(A) \le u_{\rm mid}$, so
$$u \in [\underline{u}, u_{\rm mid}].$$
- *Case 2:* if $L(u_{\rm mid}) \le A$, then $u_{\rm mid} \le u(A)$, so
$$u \in [u_{\rm mid}, \overline{u}].$$
- After each iteration, we decrease the width of the interval $[\underline{u}, \overline{u}]$ by half.
- After k iterations, we get an interval of width 2^{-k} which contains $u(A)$ – i.e., we get $u(A)$ w/accuracy 2^{-k}.

11. How to Make a Decision Based on Utility Values

- Suppose that we have found the utilities $u(A')$, $u(A'')$, $\ldots$, of the alternatives A', A'', $\ldots$
- Which of these alternatives should we choose?
- By definition of utility, we have:
  - $A \equiv L(u(A))$ for every alternative A, and
  - $L(p') < L(p'')$ if and only if $p' < p''$.
- We can thus conclude that A' is preferable to A'' if and only if

$$u(A') > u(A'').$$

- In other words, we should always select an alternative with the largest possible value of utility.
- Interval techniques can help in finding the optimizing decision.

12. How to Estimate Utility of an Action

- For each action, we usually know possible outcomes $S_1, \ldots, S_n$.
- We can often estimate the prob. $p_1, \ldots, p_n$ of these outcomes.
- By definition of utility, each situation S_i is equivalent to a lottery $L(u(S_i))$ in which we get:
  - A_1 with probability $u(S_i)$ and
  - A_0 with the remaining probability $1 - u(S_i)$.
- Thus, the action is equivalent to a complex lottery in which:
  - first, we select one of the situations S_i with probability p_i: $P(S_i) = p_i$;
  - then, depending on S_i, we get A_1 with probability $P(A_1 \mid S_i) = u(S_i)$ and A_0 with probability $1 - u(S_i)$.

13. How to Estimate Utility of an Action (cont-d)

- *Reminder:*
  - first, we select one of the situations S_i with probability p_i: $P(S_i) = p_i$;
  - then, depending on S_i, we get A_1 with probability $P(A_1 \mid S_i) = u(S_i)$ and A_0 w/probability $1 - u(S_i)$.
- The probability of getting A_1 in this complex lottery is:

$$P(A_1) = \sum_{i=1}^{n} P(A_1 \mid S_i) \cdot P(S_i) = \sum_{i=1}^{n} u(S_i) \cdot p_i.$$

- In the complex lottery, we get:
  - A_1 with probability $u = \sum_{i=1}^{n} p_i \cdot u(S_i)$, and
  - A_0 w/probability $1 - u$.
- So, we should select the action with the largest value of expected utility

$$u = \sum p_i \cdot u(S_i).$$

14. First Important Comment: Utility Is Different from Money

- Empirical data shows that utility u is proportional to square root of money x:
  - when $x > 0$, we have $u(x) = c_+ \cdot \sqrt{x}$;
  - when $x < 0$, we have $u(x) = -c_- \cdot \sqrt{|x|}$.
- This explains why most people are risk-averse.
- Indeed, let us consider two cases:
  - getting \$50, and
  - getting \$100 with probability 0.5.
- In both cases, the expected amount is the same, but:
  - in the first case, $u(x) = c_+ \cdot \sqrt{50} \approx 7 \cdot c_+$;
  - in the second case, the expected utility is

$$0.5 \cdot c_+ \cdot \sqrt{100} + 0.5 \cdot c_+ \cdot \sqrt{0} = 5 \cdot c_+ \ll 7 \cdot c_+.$$

15. Second Important Comment: Non-Uniqueness of Utility

- The above definition of utility u depends on A_0, A_1.
- What if we use different alternatives A'_0 and A'_1?
- Every A is equivalent to a lottery $L(u(A))$ in which we get A_1 with probability $u(A)$ and A_0 with probability $1 - u(A)$.
- For simplicity, let us assume that $A'_0 < A_0 < A_1 < A'_1$.
- Then, $A_0 \equiv L'(u'(A_0))$ and $A_1 \equiv L'(u'(A_1))$.
- So, A is equivalent to a complex lottery in which:

  1) we select A_1 with probability $u(A)$ and A_0 with probability $1-u(A)$;
  2) depending on A_i, we get A'_1 with probability $u'(A_i)$ and A'_0 with probability $1 - u'(A_i)$.

- In this complex lottery, we get A'_1 with probability

$$u'(A) = u(A) \cdot (u'(A_1) - u'(A_0)) + u'(A_0).$$

- So, in general, utility is defined modulo an (increasing) linear transformation $u' = a \cdot u + b$, with $a > 0$.

16. Third Important Comment: Subjective Probabilities

- In practice, we often do not know the probabilities p_i of different outcomes.
- For each event E, a natural way to estimate its subjective probability is to fix a prize (e.g., \$1) and compare:
  - the lottery ℓ_E in which we get the fixed prize if the event E occurs and 0 is it does not occur, with
  - a lottery $\ell(p)$ in which we get the same amount with probability p.
- Here, similarly to the utility case, we get a value $ps(E)$ for which, for every $\varepsilon > 0$:

$$\ell(ps(E) - \varepsilon) < \ell_E < \ell(ps(E) + \varepsilon).$$

- Then, the utility of an action with possible outcomes $S_1, \ldots, S_n$ is equal to $u = \sum_{i=1}^{n} ps(E_i) \cdot u(S_i)$.

17. Decision Theory: Summary

- According to decision theory, decision making of a rational person can be described by a special function called utility.
- Specifically, a rational person should select an alternative that maximizes the expected value of his/her utility.
- When we have partial information about the probabilities, we thus have partial information about the corresponding expected values.
- What type of information we get depends on the corresponding utility function.

18. Describing Partial Info about Probabilities: Decision Making Viewpoint

- *Problem:* there are many ways to represent a probability distribution.
- *Idea:* look for an objective.
- *Objective:* make decisions $E_x[u(x,a)] \to \max\limits_a$.
- In most practical situations, the utility function smoothly depends on the corresponding parameters.
- This is how utility depends on money, comfort on room temperature, etc.
- A small change in the corresponding parameters leads to a small change in our degree of satisfaction.
- In such cases, we can expand the utility function in Taylor series and keep the few first terms in this expansion:

$$u(x) = u(x_0) + (x - x_0) \cdot u'(x_0) + \ldots$$

19. Describing Partial Info about Probabilities: Smooth Case

- We consider the case when the utility is smooth:

$$u(x) = u(x_0) + (x - x_0) \cdot u'(x_0) + \ldots$$

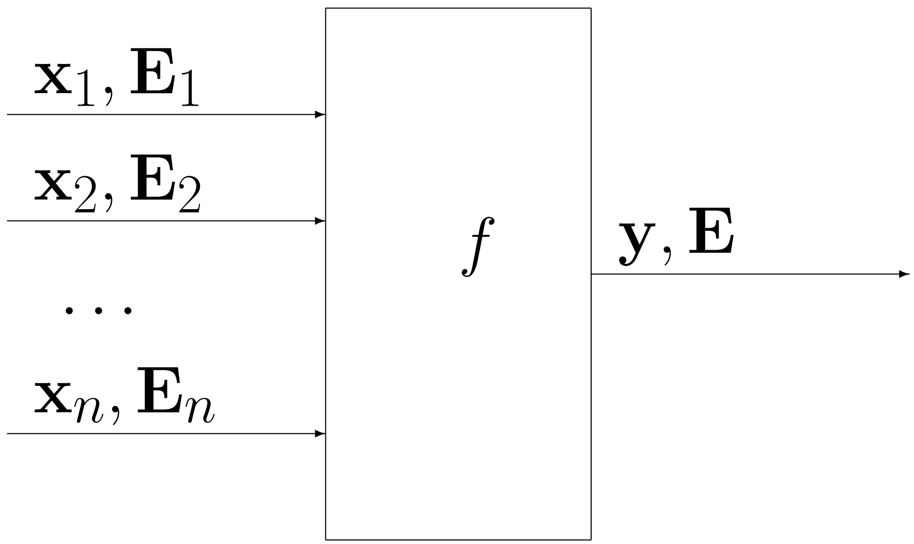
- The resulting expected value of the utility is thus a linear combination of the distribution's moments.
- In this case, partial information means that:
  - instead of the exact values of the moments,
  - we only have approximate values known with some uncertainty,
  - i.e., we know intervals of possible values.
- Thus, we need to have *interval bounds on moments.*

20. Describing Partial Info about Probabilities (cont-d)

- In other practical situations, there is a threshold causing an abrupt change in utility.
- For example, a chemical plant gets a heavy fine if the concentration of unwanted chemicals in the air exceed a regulatory threshold.
- In such situations, expected utility is proportional to the corresponding value of the cumulative distribution function (cdf) $F(x) = \text{Prob}(\xi \le x)$.
- Thus, partial information means that:
  - instead of knowing the exact values of the cdf,
  - we only have an approximate value,
  - i.e., in effect, that for each x, we only know the interval $[\underline{F}(x), \overline{F}(x)]$ of possible values of $F(x)$.
- This is what Scott Ferson calls a *p-box*.

21. Extension of Interval Arithmetic to Probabilistic Case: Successes

- *General solution:* parse to elementary operations $+$, $-$, $\cdot$, $1/x$, max, min.
- Explicit formulas for arithmetic operations known for intervals, for p-boxes $\mathbf{F}(x) = [\underline{F}(x), \overline{F}(x)]$, for intervals + 1st moments $E_i \stackrel{\text{def}}{=} E[x_i]$:



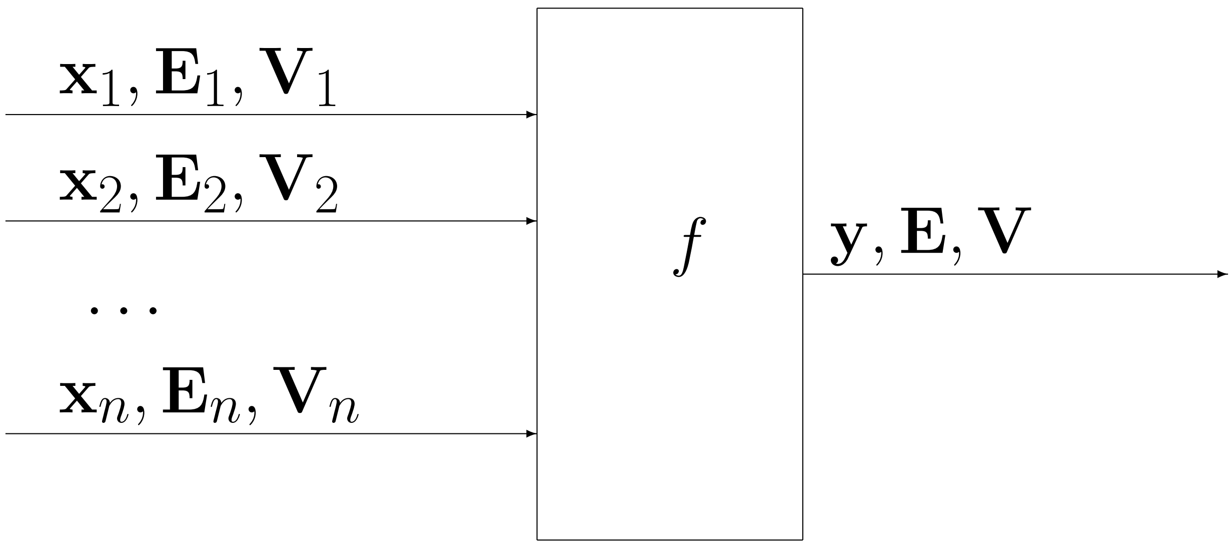
22. Successes (cont-d)

- *Easy cases:* $+$, $-$, product of independent x_i.
- *Example of a non-trivial case:* multiplication $y = x_1 \cdot x_2$, when we have no information about the correlation:
  - $\underline{E} = \max(p_1 + p_2 - 1, 0) \cdot \overline{x}_1 \cdot \overline{x}_2 + \min(p_1, 1 - p_2) \cdot \overline{x}_1 \cdot \underline{x}_2 +$ $\min(1 - p_1, p_2) \cdot \underline{x}_1 \cdot \overline{x}_2 + \max(1 - p_1 - p_2, 0) \cdot \underline{x}_1 \cdot \underline{x}_2$;
  - $\overline{E} = \min(p_1, p_2) \cdot \overline{x}_1 \cdot \overline{x}_2 + \max(p_1 - p_2, 0) \cdot \overline{x}_1 \cdot \underline{x}_2 +$ $\max(p_2 - p_1, 0) \cdot \underline{x}_1 \cdot \overline{x}_2 + \min(1 - p_1, 1 - p_2) \cdot \underline{x}_1 \cdot \underline{x}_2$,

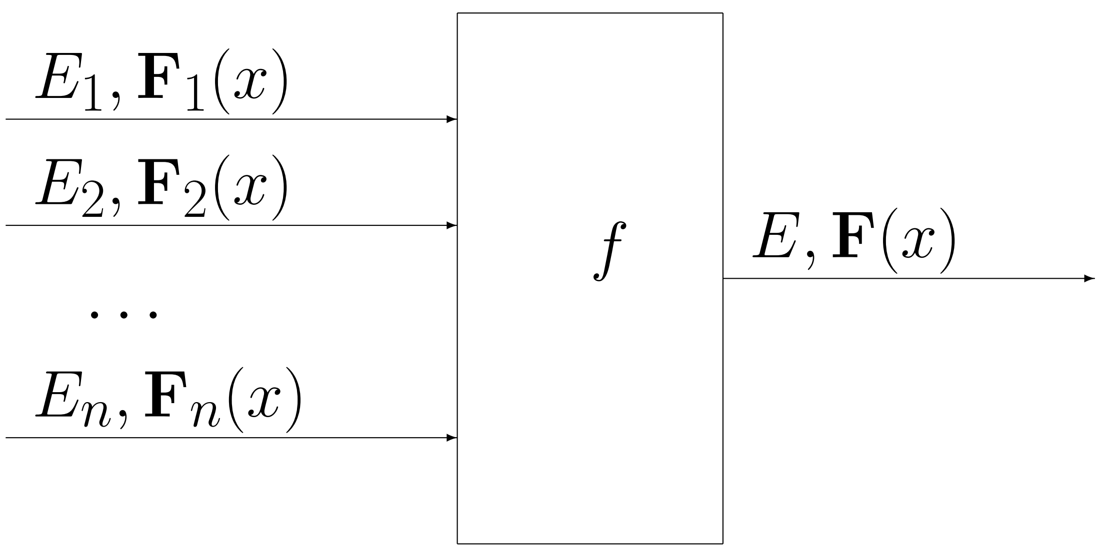
  where $p_i \stackrel{\text{def}}{=} (E_i - \underline{x}_i)/(\overline{x}_i - \underline{x}_i)$.
- *Interesting:* the corresponding formulas are very similar to the formulas of fuzzy logic, mixing two pairs of "and"- and "or"-operations:
  - $\min(p_1, p_2)$ and $\max(p_1, p_2)$, and
  - $\max(p_1 + p + 2 - 1, 0)$ and $\min(p_1 + p + 2, 1)$.

23. Challenges

- How to extend this to general functions?
- intervals + 2nd moments:



- moments + p-boxes; e.g.:



Appendix A: Basic Interval Techniques
A1. Interval Arithmetic: Reminder

- *Problem:* compute the range

$$[\underline{y}, \overline{y}] = \{f(x_1, \ldots, x_n) \mid x_1 \in [\underline{x}_1, \overline{x}_1], \ldots, x_n \in [\underline{x}_n, \overline{x}_n]\}.$$

- *Interval arithmetic:* for arithmetic operations $f(x_1, x_2)$ (and for elementary functions), we have explicit formulas for the range.
- *Examples:* when $x_1 \in \mathbf{x}_1 = [\underline{x}_1, \overline{x}_1]$ and $x_2 \in \mathbf{x}_2 = [\underline{x}_2, \overline{x}_2]$, then:
  - The range $\mathbf{x}_1 + \mathbf{x}_2$ for $x_1 + x_2$ is $[\underline{x}_1 + \underline{x}_2, \overline{x}_1 + \overline{x}_2]$.
  - The range $\mathbf{x}_1 - \mathbf{x}_2$ for $x_1 - x_2$ is $[\underline{x}_1 - \overline{x}_2, \overline{x}_1 - \underline{x}_2]$.
  - The range $\mathbf{x}_1 \cdot \mathbf{x}_2$ for $x_1 \cdot x_2$ is $[\underline{y}, \overline{y}]$, where

$$\underline{y} = \min(\underline{x}_1 \cdot \underline{x}_2, \underline{x}_1 \cdot \overline{x}_2, \overline{x}_1 \cdot \underline{x}_2, \overline{x}_1 \cdot \overline{x}_2);$$

$$\overline{y} = \max(\underline{x}_1 \cdot \underline{x}_2, \underline{x}_1 \cdot \overline{x}_2, \overline{x}_1 \cdot \underline{x}_2, \overline{x}_1 \cdot \overline{x}_2).$$

- The range $1/\mathbf{x}_1$ for $1/x_1$ is $[1/\overline{x}_1, 1/\underline{x}_1]$ (if $0 \notin \mathbf{x}_1$).

A2. Straightforward Interval Computations: Example

- *Example:* $f(x) = (x-2)\cdot(x+2)$, $x \in [1,2]$.
- How will the computer compute it?
  - $r_1 := x - 2;$
  - $r_2 := x + 2;$
  - $r_3 := r_1 \cdot r_2.$
- *Main idea:* perform the same operations, but with *intervals* instead of *numbers*:
  - $\mathbf{r}_1 := [1,2] - [2,2] = [-1,0];$
  - $\mathbf{r}_2 := [1,2] + [2,2] = [3,4];$
  - $\mathbf{r}_3 := [-1,0] \cdot [3,4] = [-4,0].$
- *Actual range:* $f(\mathbf{x}) = [-3,0]$.
- *Comment:* this is just a toy example, there are more efficient ways of computing an enclosure $\mathbf{Y} \supseteq \mathbf{y}$.

A3. First Idea: Use of Monotonicity

- *Reminder:* for arithmetic, we had exact ranges.
- *Reason:* $+$, $-$, $\cdot$ are monotonic in each variable.
- *How monotonicity helps:* if $f(x_1, \ldots, x_n)$ is (non-strictly) increasing ($f\uparrow$) in each x_i, then

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) = [f(\underline{x}_1, \ldots, \underline{x}_n), f(\overline{x}_1, \ldots, \overline{x}_n)].$$

- *Similarly:* if $f\uparrow$ for some x_i and $f\downarrow$ for other x_j.
- *Fact:* $f\uparrow$ in x_i if $\dfrac{\partial f}{\partial x_i} \geq 0$.
- *Checking monotonicity:* check that the range $[\underline{r}_i, \overline{r}_i]$ of $\dfrac{\partial f}{\partial x_i}$ on $\mathbf{x}_i$ has $\underline{r}_i \geq 0$.
- *Differentiation:* by Automatic Differentiation (AD) tools.
- *Estimating ranges of* $\dfrac{\partial f}{\partial x_i}$*:* straightforward interval comp.

A4. Monotonicity: Example

- *Idea:* if the range $[\underline{r}_i, \overline{r}_i]$ of each $\dfrac{\partial f}{\partial x_i}$ on $\mathbf{x}_i$ has $\underline{r}_i \geq 0$, then
$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) = [f(\underline{x}_1, \ldots, \underline{x}_n), f(\overline{x}_1, \ldots, \overline{x}_n)].$$
- *Example:* $f(x) = (x-2) \cdot (x+2)$, $\mathbf{x} = [1, 2]$.
- *Case $n = 1$:* if the range $[\underline{r}, \overline{r}]$ of $\dfrac{df}{dx}$ on $\mathbf{x}$ has $\underline{r} \geq 0$, then
$$f(\mathbf{x}) = [f(\underline{x}), f(\overline{x})].$$
- *AD:* $\dfrac{df}{dx} = 1 \cdot (x+2) + (x-2) \cdot 1 = 2x.$
- *Checking:* $[\underline{r}, \overline{r}] = [2, 4]$, with $2 \geq 0$.
- *Result:* $f([1, 2]) = [f(1), f(2)] = [-3, 0]$.
- *Comparison:* this is the exact range.

A5. Non-Monotonic Example

- *Example:* $f(x) = x \cdot (1 - x)$, $x \in [0, 1]$.
- How will the computer compute it?
  - $r_1 := 1 - x$;
  - $r_2 := x \cdot r_1$.
- *Straightforward interval computations:*
  - $\mathbf{r}_1 := [1, 1] - [0, 1] = [0, 1]$;
  - $\mathbf{r}_2 := [0, 1] \cdot [0, 1] = [0, 1]$.
- *Actual range:* min, max of f at $\underline{x}$, $\overline{x}$, or when $\dfrac{df}{dx} = 0$.
- Here, $\dfrac{df}{dx} = 1 - 2x = 0$ for $x = 0.5$, thus we:
  - compute $f(0) = 0$, $f(0.5) = 0.25$, and $f(1) = 0$, so
  - $\underline{y} = \min(0, 0.25, 0) = 0$, $\overline{y} = \max(0, 0.25, 0) = 0.25$.
- *Resulting range:* $f(\mathbf{x}) = [0, 0.25]$.

A6. Second Idea: Centered Form

- *Main idea:* Intermediate Value Theorem

$$f(x_1, \ldots, x_n) = f(\tilde{x}_1, \ldots, \tilde{x}_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\chi) \cdot (x_i - \tilde{x}_i)$$

for some $\chi_i \in \mathbf{x}_i$.

- *Corollary:* $f(x_1, \ldots, x_n) \in \mathbf{Y}$, where

$$\mathbf{Y} = \tilde{y} + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \cdot [-\Delta_i, \Delta_i].$$

- *Differentiation:* by Automatic Differentiation (AD) tools.
- *Estimating the ranges of derivatives:*
  - if appropriate, by monotonicity, or
  - by straightforward interval computations, or
  - by centered form (more time but more accurate).

A7. Centered Form: Example

- *General formula:*

$$\mathbf{Y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \cdot [-\Delta_i, \Delta_i].$$

- *Example:* $f(x) = x \cdot (1 - x)$, $\mathbf{x} = [0, 1]$.
- Here, $\mathbf{x} = [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$, with $\widetilde{x} = 0.5$ and $\Delta = 0.5$.
- *Case $n = 1$:* $\mathbf{Y} = f(\widetilde{x}) + \dfrac{df}{dx}(\mathbf{x}) \cdot [-\Delta, \Delta]$.
- *AD:* $\dfrac{df}{dx} = 1 \cdot (1 - x) + x \cdot (-1) = 1 - 2x$.
- *Estimation:* we have $\dfrac{df}{dx}(\mathbf{x}) = 1 - 2 \cdot [0, 1] = [-1, 1]$.
- *Result:* $\mathbf{Y} = 0.5 \cdot (1 - 0.5) + [-1, 1] \cdot [-0.5, 0.5] = 0.25 + [-0.5, 0.5] = [-0.25, 0.75]$.
- *Comparison:* actual range $[0, 0.25]$, straightforward $[0, 1]$.

A8. Third Idea: Bisection

- *Known:* accuracy $O(\Delta_i^2)$ of first order formula

$$f(x_1, \ldots, x_n) = f(\tilde{x}_1, \ldots, \tilde{x}_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\chi) \cdot (x_i - \tilde{x}_i).$$

- *Idea:* if the intervals are too wide, we:

  – split one of them in half ($\Delta_i^2 \to \Delta_i^2/4$); and

  – take the union of the resulting ranges.

- *Example:* $f(x) = x \cdot (1 - x)$, where $x \in \mathbf{x} = [0, 1]$.
- *Split:* take $\mathbf{x}' = [0, 0.5]$ and $\mathbf{x}'' = [0.5, 1]$.
- *1st range:* $1 - 2 \cdot \mathbf{x} = 1 - 2 \cdot [0, 0.5] = [0, 1]$, so $f \uparrow$ and $f(\mathbf{x}') = [f(0), f(0.5)] = [0, 0.25]$.
- *2nd range:* $1 - 2 \cdot \mathbf{x} = 1 - 2 \cdot [0.5, 1] = [-1, 0]$, so $f \downarrow$ and $f(\mathbf{x}'') = [f(1), f(0.5)] = [0, 0.25]$.
- *Result:* $f(\mathbf{x}') \cup f(\mathbf{x}'') = [0, 0.25]$ – exact.

Appendix B. Case Study

B1. Chip Design

- *Chip design:* one of the main objectives is to decrease the clock cycle.
- *Current approach:* uses worst-case (interval) techniques.
- *Problem:* the probability of the worst-case values is usually very small.
- *Result:* estimates are over-conservative – unnecessary over-design and under-performance of circuits.
- *Difficulty:* we only have *partial* information about the corresponding probability distributions.
- *Objective:* produce estimates valid for all distributions which are consistent with this information.
- *What we do:* provide such estimates for the clock time.

B2. Estimating Clock Cycle: a Practical Problem

- *Objective:* estimate the clock cycle on the design stage.
- The clock cycle of a chip is constrained by the maximum path delay over all the circuit paths

$$D \stackrel{\text{def}}{=} \max(D_1, \ldots, D_N).$$

- The path delay D_i along the i-th path is the sum of the delays corresponding to the gates and wires along this path.
- Each of these delays, in turn, depends on several factors such as:
  - the variation caused by the current design practices,
  - environmental design characteristics (e.g., variations in temperature and in supply voltage), etc.

B3. Traditional (Interval) Approach to Estimating the Clock Cycle

- *Traditional approach:* assume that each factor takes the worst possible value.
- *Result:* time delay when all the factors are at their worst.
- *Problem:*
  - different factors are usually independent;
  - combination of worst cases is improbable.
- *Computational result:* current estimates are 30% above the observed clock time.
- *Practical result:* the clock time is set too high – chips are over-designed and under-performing.

B4. Robust Statistical Methods Are Needed

- *Ideal case:* we know probability distributions.
- *Solution:* Monte-Carlo simulations.
- *In practice:* we only have *partial* information about the distributions of some of the parameters; usually:
  - the mean, and
  - some characteristic of the deviation from the mean – e.g., the interval that is guaranteed to contain possible values of this parameter.
- *Possible approach:* Monte-Carlo with several possible distributions.
- *Problem:* no guarantee that the result is a valid bound for all possible distributions.
- *Objective:* provide *robust* bounds, i.e., bounds that work for all possible distributions.

B5. Towards a Mathematical Formulation of the Problem

- *General case:* each gate delay d depends on the difference $x_1, \ldots, x_n$ between the actual and the nominal values of the parameters.
- *Main assumption:* these differences are usually small.
- Each path delay D_i is the sum of gate delays.
- *Conclusion:* D_i is a linear function: $D_i = a_i + \sum\limits_{j=1}^{n} a_{ij} \cdot x_j$ for some a_i and a_{ij}.
- The desired maximum delay $D = \max\limits_i D_i$ has the form

$$D = F(x_1, \ldots, x_n) \stackrel{\text{def}}{=} \max_i \left(a_i + \sum_{j=1}^{n} a_{ij} \cdot x_j \right).$$

B6. Towards a Mathematical Formulation of the Problem (cont-d)

- *Known:* maxima of linear function are exactly convex functions:

$$F(\alpha \cdot x + (1-\alpha) \cdot y) \le \alpha \cdot F(x) + (1-\alpha) \cdot F(y)$$

for all x, y and for all $\alpha \in [0, 1]$;

- *We know:* factors x_i are independent;

– we know distribution of some of the factors;

– for others, we know ranges $[\underline{x}_j, \overline{x}_j]$ and means E_j.

- *Given:* a convex function $F \ge 0$ and a number $\varepsilon > 0$.
- *Objective:* find the smallest y_0 s.t. for all possible distributions, we have $y \le y_0$ with the probability $\ge 1 - \varepsilon$.

B7. Additional Property: Dependency is Non-Degenerate

- *Fact:* sometimes, we learn additional information about one of the factors x_j.
- *Example:* we learn that x_j actually belongs to a proper subinterval of the original interval $[\underline{x}_j, \overline{x}_j]$.
- *Consequence:* the class $\mathcal{P}$ of possible distributions is replaced with $\mathcal{P}' \subset \mathcal{P}$.
- *Result:* the new value y_0' can only decrease: $y_0' \le y_0$.
- *Fact:* if x_j is irrelevant for y, then $y_0' = y_0$.
- *Assumption:* irrelevant variables been weeded out.
- *Formalization:* if we narrow down one of the intervals $[\underline{x}_j, \overline{x}_j]$, the resulting value y_0 decreases: $y_0' < y_0$.

B8. Formulation of the Problem

GIVEN:

- n, $k \le n$, $\varepsilon > 0$;
- a convex function $y = F(x_1, \ldots, x_n) \ge 0$;
- $n - k$ cdfs $F_j(x)$, $k + 1 \le j \le n$;
- intervals $\mathbf{x}_1, \ldots, \mathbf{x}_k$, values $E_1, \ldots, E_k$,

TAKE: all joint probability distributions on R^n for which:

- all x_i are independent,
- $x_j \in \mathbf{x}_j$, $E[x_j] = E_j$ for $j \le k$, and
- x_j have distribution $F_j(x)$ for $j > k$.

FIND: the smallest y_0 such that for all such distributions, $F(x_1, \ldots, x_n) \le y_0$ with probability $\ge 1 - \varepsilon$.

WHEN: the problem is *non-degenerate* – if we narrow down one of the intervals $\mathbf{x}_j$, y_0 decreases.

B9. Main Result and How We Can Use It

- *Result:* y_0 is attained when for each j from 1 to k,
  - $x_j = \underline{x}_j$ with probability $\underline{p}_j \stackrel{\text{def}}{=} \dfrac{\overline{x}_j - E_j}{\overline{x}_j - \underline{x}_j}$, and
  - $x_j = \overline{x}_j$ with probability $\overline{p}_j \stackrel{\text{def}}{=} \dfrac{E_j - \underline{x}_j}{\overline{x}_j - \underline{x}_j}$.
- *Algorithm:*
  - simulate these distributions for x_j, $j < k$;
  - simulate known distributions for $j > k$;
  - use the simulated values $x_j^{(s)}$ to find

$$y^{(s)} = F(x_1^{(s)}, \ldots, x_n^{(s)});$$

  - sort N values $y^{(s)}$: $y_{(1)} \le y_{(2)} \le \ldots \le y_{(N_i)}$;
  - take $y_{(N_i \cdot (1-\varepsilon))}$ as y_0.

B10. Comment about Monte-Carlo Techniques

- *Traditional belief:* Monte-Carlo methods are inferior to analytical:
  - they are approximate;
  - they require large computation time;
  - simulations for *several* distributions, may mis-calculate the (desired) maximum over *all* distributions.
- *We proved:* the value corresponding to the selected distributions indeed provide the desired maximum value y_0.
- *General comment:*
  - justified Monte-Carlo methods often lead to *faster* computations than analytical techniques;
  - example: multi-D integration – where Monte-Carlo methods were originally invented.

B11. Comment about Non-Linear Terms

- *Reminder:* in the above formula $D_i = a_i + \sum_{j=1}^{n} a_{ij} \cdot x_j$, we ignored quadratic and higher order terms in the dependence of each path time D_i on parameters x_j.
- *In reality:* we may need to take into account some quadratic terms.
- *Idea behind possible solution:* it is known that the max $D = \max_i D_i$ of convex functions D_i is convex.
- *Condition when this idea works:* when each dependence $D_i(x_1, \ldots, x_k, \ldots)$ is still convex.
- *Solution:* in this case,
  - the function function D is still convex,
  - hence, our algorithm will work.

B12. Conclusions

- *Problem of chip design:* decrease the clock cycle.
- *How this problem is solved now:* by using worst-case (interval) techniques.
- *Limitations of this solution:* the probability of the worst-case values is usually very small.
- *Consequence:* estimates are over-conservative, hence over-design and under-performance of circuits.
- *Objective:* find the clock time as y_0 s.t. for the actual delay y, we have $\text{Prob}(y > y_0) \le \varepsilon$ for given $\varepsilon > 0$.
- *Difficulty:* we only have *partial* information about the corresponding distributions.
- *What we have described:* a general technique that allows us, in particular, to compute y_0.

Appendix C. Second Case Study:
C1. Bioinformatics

- *Practical problem:* find genetic difference between cancer cells and healthy cells.
- *Ideal case:* we directly measure concentration c of the gene in cancer cells and h in healthy cells.
- *In reality:* difficult to separate.
- *Solution:* we measure $y_i \approx x_i \cdot c + (1 - x_i) \cdot h$, where x_i is the percentage of cancer cells in i-th sample.
- *Equivalent form:* $a \cdot x_i + h \approx y_i$, where $a \stackrel{\text{def}}{=} c - h$.

C2. Case Study: Bioinformatics (cont-d)

- *If we know x_i exactly:* Least Squares Method

$$\sum_{i=1}^{n}(a\cdot x_i+h-y_i)^2 \to \min_{a,h}, \text{ hence } a=\frac{C(x,y)}{V(x)} \text{ and } h=E(y)-a\cdot E(x),$$

where $E(x)=\frac{1}{n}\cdot\sum_{i=1}^{n} x_i$,

$$V(x)=\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i-E(x))^2,$$

$$C(x,y)=\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i-E(x))\cdot(y_i-E(y)).$$

- *Interval uncertainty:* experts manually count x_i, and only provide interval bounds $\mathbf{x}_i$, e.g., $x_i\in[0.7,0.8]$.
- *Problem:* find the range of a and h corresponding to all possible values $x_i\in[\underline{x}_i,\overline{x}_i]$.

C3. General Problem

- *General problem:*
  - we know intervals $\mathbf{x}_1 = [\underline{x}_1, \overline{x}_1], \ldots, \mathbf{x}_n = [\underline{x}_n, \overline{x}_n]$,
  - compute the range of $E(x) = \frac{1}{n} \sum_{i=1}^{n} x_i$, population variance $V = \frac{1}{n} \sum_{i=1}^{n} (x_i - E(x))^2$, etc.
- *Difficulty:* NP-hard even for variance.
- *Known:*
  - efficient algorithms for $\underline{V}$,
  - efficient algorithms for $\overline{V}$ and $C(x, y)$ for reasonable situations.
- *Bioinformatics case:* find intervals for $C(x, y)$ and for $V(x)$ and divide.