Probabilities, Intervals, What Next?

x x x

Extension of Interval Computations to Situations With Partial Information about Probabilities

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

Formulation of the First Problem

- We have n measurement results $x_1, \ldots, x_n$,
- Traditional data processing techniques: compute population parameters, e.g.,

$$\mu = \frac{x_1 + \ldots + x_n}{n},$$

$$\sigma^2 = \frac{(x_1 - \mu)^2 + \ldots + (x_n - \mu)^2}{n} \text{ (or } \sigma = \sqrt{\sigma^2}\text{)}.$$

- Often, we only have intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$.
- *Example:* for measurements, $\mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.
- We need $\mathbf{y} = \{f(x_1, \ldots, x_n) \mid x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}$.
- What are $[\underline{\mu}, \overline{\mu}]$ and $[\underline{\sigma^2}, \overline{\sigma^2}]$?
- For $[\underline{\mu}, \overline{\mu}]$, the answer is easy.
- When $\cap_{i=1}^n \mathbf{x}_i \neq \emptyset$, we have $\underline{\sigma^2} = 0$; else $\underline{\sigma^2} > 0$.
- *Problem* (Walster): what is the total set $[\underline{\sigma^2}, \overline{\sigma^2}]$ of possible values of σ^2?

For this Problem, Straightforward Interval Computations Sometimes Lead to Excess Width

- *Reminder:*
  - parse the function $f(x_1, \ldots, x_n)$, and
  - replace each elementary operation by the corr. operation of interval arithmetic.
- *Example:* for $\mathbf{x}_1 = \mathbf{x}_2 = [0, 1]$.
- *Actual range:* since $\sigma^2 = (x_1 - x_2)^2/4$, the actual range is $[\underline{\sigma^2}, \overline{\sigma^2}] = [0, 0.25]$.
- *Estimate:* $[\underline{\mu}, \overline{\mu}] = [0, 1]$, hence

$$\frac{(\mathbf{x}_1 - [\underline{\mu}, \overline{\mu}])^2 + (\mathbf{x}_2 - [\underline{\mu}, \overline{\mu}])^2}{2} = [0, 1] \supset [0, 0.25].$$

- *Comment:* other formulas also lead to excess width.
- *Explanation:* in each formula for σ^2, each variable occurs several times.

Centered Form Sometimes Leads to Excess Width

- *Reminder:*

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

  where:

  - $\widetilde{x}_i = (\underline{x}_i + \overline{x}_i)/2$ is the interval's midpoint and
  - $\Delta_i = (\underline{x}_i - \overline{x}_i)/2$ is its half-width.

- *Not perfect* (similar to Hertling):
  - it produces an interval centered at $f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$;
  - when all intervals $\mathbf{x}_i$ are equal, all midpoints $\widetilde{x}_i$ are the same;
  - hence the population variance $f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ is 0;
  - so, the estimate's lower bound is < 0, but $\sigma^2 \geq 0$.

First Result: Computing $\underline{\sigma}^2$

The following algorithm always compute $\underline{\sigma}^2$ in $O(n^2)$:

- First, we sort all $2n$ values $\underline{x}_i$, $\overline{x}_i$ into a sequence $x_{(1)} \le x_{(2)} \le \ldots \le x_{(2n)}$.
- Second, we compute $\underline{\mu}$ and $\overline{\mu}$ and select all "small intervals" $[x_{(k)}, x_{(k+1)}]$ that intersect with $[\underline{\mu}, \overline{\mu}]$.
- For each of the selected small intervals $[x_{(k)}, x_{(k+1)}]$, we compute the ratio $r_k = S_k/N_k$, where
$$S_k \stackrel{\text{def}}{=} \sum_{i:\underline{x}_i \ge x_{(k+1)}} \underline{x}_i + \sum_{j:\overline{x}_j \le x_{(k)}} \overline{x}_j,$$
and N_k is the total number of such i's and j's.
- If $r_k \in [x_{(k)}, x_{(k+1)}]$, then we compute
$${\sigma'}_k^2 \stackrel{\text{def}}{=} \frac{1}{n} \cdot \left(\sum_{i:\underline{x}_i \ge x_{(k+1)}} (\underline{x}_i - r_k)^2 + \sum_{j:\overline{x}_j \le x_{(k)}} (\overline{x}_j - r_k)^2 \right).$$
If $N_k = 0$, we take ${\sigma'}_k^2 \stackrel{\text{def}}{=} 0$.
- Finally, we return the smallest of the values ${\sigma'}_k^2$ as $\underline{\sigma}^2$.

Example

- Input: $\mathbf{x}_1 = [2.1, 2.6]$, $\mathbf{x}_2 = [2.0, 2.1]$, $\mathbf{x}_3 = [2.2, 2.9]$, $\mathbf{x}_4 = [2.5, 2.7]$, and $\mathbf{x}_5 = [2.4, 2.8]$.
- "small intervals": $[x_{(1)}, x_{(2)}] = [2.0, 2.1]$, $[2.1, 2.1]$, $[2.1, 2.2]$, $[2.2, 2.4]$, $[2.4, 2.5]$, $[2.5, 2.6]$, $[2.6, 2.7]$, $[2.7, 2.8]$, and $[2.8, 2.9]$.
- Population average $[\underline{\mu}, \overline{\mu}] = [2.24, 2.62]$, so we keep $[2.2, 2.4]$, $[2.4, 2.5]$, $[2.5, 2.6]$, $[2.6, 2.7]$. For these intervals:
  - $S_4 = 7.0$, $N_4 = 3$, so $r_4 = 2.333\ldots$;
  - $S_5 = 4.6$, $N_5 = 2$, so $r_5 = 2.3$;
  - $S_6 = 2.1$, $N_6 = 1$, so $r_6 = 2.1$;
  - $S_7 = 4.7$, $N_7 = 2$, so $r_7 = 2.35$.
- Only r_4 lies within the corresponding small interval.
- Here, ${\sigma'}_4^2 = 0.017333\ldots$, so $\underline{\sigma^2} = 0.017333\ldots$

Second Result: Computing $\overline{\sigma^2}$ is NP-Hard

- **Theorem.** *Computing $\overline{\sigma^2}$ is NP-hard.*
- *Comments:*
  - NP-hard means, crudely speaking, that there are no general ways for solving *all* particular cases of this problem in reasonable time.
  - NP-hardness of computing the range of a quadratic function was proven by Vavasis (1991).
  - By using peeling, we can compute $\overline{\sigma^2}$ in exponential time $O(2^n)$.
- *Natural question:* maybe the difficulty comes from the requirement that the range be computed exactly?
- **Theorem.** *For every $\varepsilon > 0$, the problem of computing $\overline{\sigma^2}$ with accuracy ε is NP-hard.*

Third Result: A Feasible Algorithm that Computes $\overline{\sigma^2}$ in Many Practical Situations

- *Case:* all midpoints ("measured values")

$$\widetilde{x}_i = \frac{\underline{x}_i + \overline{x}_i}{2}$$

  of the intervals

$$\mathbf{x}_i = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$$

  are definitely different from each other.

- *Namely:* the "narrowed" intervals

$$\left[\widetilde{x}_i - \frac{\Delta_i}{n}, \widetilde{x}_i + \frac{\Delta_i}{n}\right]$$

  do not intersect with each other.

- In this case, there exists an algorithm computes $\overline{\sigma^2}$ in quadratic time.

Algorithm

- Sort $2n$ endpoints of narrowed intervals into $x_{(1)} \le x_{(2)} \le \ldots \le x_{(2n)}$.
- Thus, $I\!R$ is divided into $2n+2$ segments ("small intervals") $[x_{(k)}, x_{(k+1)}]$.
- Select only "small intervals" $[x_{(k)}, x_{(k+1)}]$ that intersect with $[\underline{\mu}, \overline{\mu}]$; for each, pick x_i as follows:
  - if $x_{(k+1)} < \widetilde{x}_i - \Delta_i/n$, then we pick $x_i = \overline{x}_i$;
  - if $x_{(k)} > \widetilde{x}_i + \Delta_i/n$, then we pick $x_i = \underline{x}_i$;
  - for all other i, we consider both possible values $x_i = \overline{x}_i$ and $x_i = \underline{x}_i$.
- For each of the sequences x_i, we check whether the average E is indeed within this small interval, and if it is, compute the population variance.
- The largest of these population variances is $\overline{\sigma^2}$.

Third Result (cont-d)

- *Question:* what if two "narrowed" intervals have a common point?
- *Case:* let us fix k and consider all cases C_k in which no more than k "narrowed" intervals can have a common point.
- *Result:* $\forall k$, the above algorithm $\overline{\mathcal{A}}$ computes $\overline{\sigma^2}$ in quadratic time for all problems $\in C_k$.
- *Comments:*
  - Computation time t is quadratic in n.
  - However, t is exponential in k.
  - So, when $k \uparrow$, the algorithm $\overline{\mathcal{A}}$ requires more and more computation time.
  - In our proof of NP-hardness, we use the case when all n narrowed intervals have a common point.

Population Mean, Population Variance: What Next?

- *Population covariance*

$$C = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_x) \cdot (y_i - \mu_y).$$

- *Result:* both computing $\overline{C}$ and computing $\underline{C}$ are NP-hard problems.

- *Population correlation*

$$\rho = \frac{C}{\sigma_x \cdot \sigma_y}.$$

- *Result:* both computing $\overline{\rho}$ and computing $\underline{\rho}$ are NP-hard problems.

- *Open problem:* design feasible algorithms that work in many practical cases.

- *Median:* feasible (since it is monotonic in x_i).

- *Open problem:* analyze other population parameters from this viewpoint.

Bounds for Sample Variance: Variant of the First Problem

- *We know:*
  - measurement results $\widetilde{x}_1, \ldots, \widetilde{x}_n$;
  - the accuracies Δ_i of each measurement;
  - hence, that the actual values x_i are within

$$\mathbf{x}_i \stackrel{\text{def}}{=} [\underline{x}_i, \overline{x}_i] = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i].$$

  - that x_i are normally distributed, w/CDF $F_0\left(\dfrac{x-a}{\sigma}\right)$.
- *Question:* what are the possible values of a and σ?
- *Main idea:* Kolmogorov-Smirnov (KS) inequality implies (with probability $p \ge p_0$) that

$$|F(x) - F_{\text{sample}}(x)| \le \Delta,$$

where $F_{\text{sample}}(x) = \dfrac{i}{n}$ for $x_{(i)} \le x < x_{(i+1)}$.

Bounds for Sample Variance: Solution

- Due to KS, for every i, for some $x_i \in [\underline{x}_i, \overline{x}_i]$:

$$\frac{i}{n} - \Delta \le F_0\left(\frac{x_{(i)} - a}{\sigma}\right) \le \frac{i}{n} + \Delta.$$

- So,

$$\frac{l(x'_i)}{n} - \Delta \le F_0\left(\frac{x'_i - a}{\sigma}\right) \le \frac{u(x'_i)}{n} + \Delta,$$

where $l(x)$ is # of k s.t. $\overline{x}_k \le x$, $u(i)$ is # of k s.t. $\underline{x}_k \le x$, and $x'_i = \underline{x}_i$ or $x'_i = \overline{x}_i$.

- Hence,

$$F_0^{-1}\left(\frac{l(x'_i)}{n} - \Delta\right) \le \frac{x'_i - a}{\sigma} \le \left(\frac{u(x'_i)}{n} + \Delta\right).$$

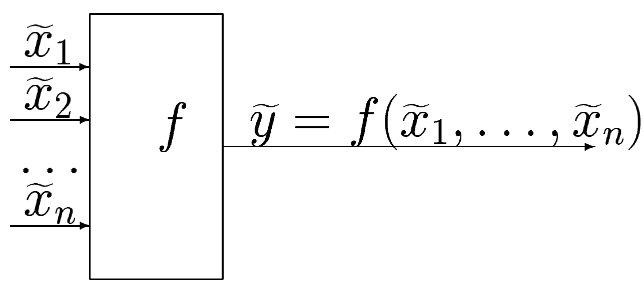
- We get a system of linear inequalities for a and σ:

$$\sigma \cdot F_0^{-1}\left(\frac{l(x_i)}{n} - \Delta\right) \le x_i - a \le \sigma \cdot F_0^{-1}\left(\frac{u(x_i)}{n} + \Delta\right).$$

- So, we can use linear programming to find bounds on a and σ.

Second Problem: Probabilistic Extension of Interval Arithmetic

- *Indirect measurements:* way to measure y that are are impossible or difficult to measure directly.
- *Examples:* distance to a star, the amount of oil in a given well.
- *Idea:* $y = f(x_1, \ldots, x_n)$

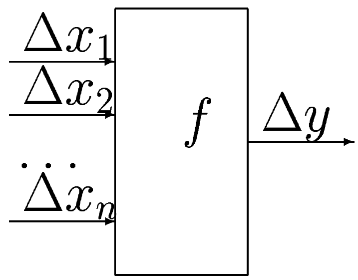


- *Problem:* measurements are never 100% accurate: $\widetilde{x}_i \neq x_i$ ($\Delta x_i \neq 0$) hence

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

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

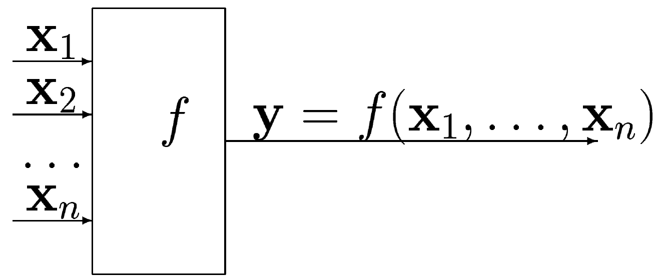
Why Interval Computations: Reminder



- *Traditional approach:* we know probability distribution for Δx_i (usually Gaussian).
- *Problem:* sometimes we do not know the distribution because no "standard" (more accurate) MI is available. Cases:
  - fundamental science
  - manufacturing
- *Solution:* we know upper bounds Δ_i on $|\Delta x_i|$ hence

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

Interval Computations: What? How?



- *What:*

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

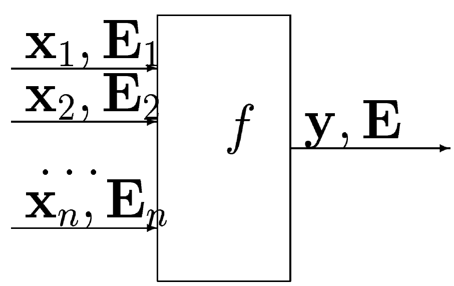
- *How* (straightforward interval computations):
  - parse f into elementary operations $+$, $-$, $\cdot$, $1/x$, min, max;
  - replace each operation by the corresponding operation of interval arithmetic:

$$[\underline{x}_1,\overline{x}_1]+[\underline{x}_2,\overline{x}_1]=[\underline{x}_1+\underline{x}_2,\overline{x}_1+\overline{x}_2];$$

$$[\underline{x}_1,\overline{x}_1]-[\underline{x}_2,\overline{x}_1]=[\underline{x}_1-\overline{x}_2,\overline{x}_1-\underline{x}_2].$$

Adding Moments: Step One

- So far, we have considered two cases:
  - *statistical case:* we know $\text{Prob}(\Delta x_i)$;
  - *interval case:* we know nothing about $\text{Prob}(\Delta x_i)$.
- *Possible:* we have *partial* information about $\text{Prob}(\Delta x_i)$.
- *Example:* we know moments.
- *Simplest case:* we know $E_i \stackrel{\text{def}}{=} E[x_i]$ (or rather $\mathbf{E}_i$).
- *Problem:*



- *Solution:* parse to $+$, $-$, $\cdot$, $1/x$, max, min.

Problem: Formulation, Cases

- *Given:*
  - $[\underline{x}_1, \overline{x}_1]$, $[\underline{E}_1, \overline{E}_1]$,
  - $[\underline{x}_2, \overline{x}_2]$, $[\underline{E}_2, \overline{E}_2]$,
  - an operation $y = x_1 \odot x_2$ ($\odot = +, -, \cdot, 1/x, \max, \min$).
- *Find:* exact bounds on $[\underline{y}, \overline{y}]$ and $[\underline{E}, \overline{E}]$.
- *Comment:* bounds on $[\underline{y}, \overline{y}]$ same.
- *Cases:*
  - we have no info about correlation between x_i;
  - we know that x_i are independent;
  - we know that x_i are maximally + correlated:

$$\exists t \text{ s.t. } x_1(t) \uparrow \,\&\, x_2(t) \uparrow;$$

  - we know that x_i are maximally − correlated:

$$\exists t \text{ s.t. } x_1(t) \uparrow \,\&\, x_2(t) \downarrow .$$

Formulation of the problem in Precise Terms

- *Given:* values $\underline{x}_1$, $\overline{x}_1$, $\underline{x}_2$, $\overline{x}_2$, $\underline{E}_1$, $\overline{E}_1$, $\underline{E}_2$, $\overline{E}_2$, and operation $\odot$.
- *Find:* the values

$$\underline{E} \stackrel{\text{def}}{=} \min\{E(x_1 \odot x_2) \,|\, \text{ all distributions of } (x_1, x_2)$$

$$\text{for which } x_1 \in [\underline{x}_1, \overline{x}_1], x_2 \in [\underline{x}_2, \overline{x}_2],$$

$$E[x_1] \in [\underline{E}_1, \overline{E}_1], E[x_2] \in [\underline{E}_2, \overline{E}_2]\}$$

and

$$\overline{E} \stackrel{\text{def}}{=} \max\{E(x_1 \odot x_2) \,|\, \text{ all distributions of } (x_1, x_2)$$

$$\text{for which } x_1 \in [\underline{x}_1, \overline{x}_1], x_2 \in [\underline{x}_2, \overline{x}_2],$$

$$E[x_1] \in [\underline{E}_1, \overline{E}_1], E[x_2] \in [\underline{E}_2, \overline{E}_2]\}$$

(plus restrictions on the correlation).

Simplest Cases: +, − (All 4 Cases), and Product of Independent x_i

- *Addition:* we know that

$$E[x_1 + x_2] = E[x_1] + E[x_2],$$

so

$$[\underline{E}, \overline{E}] = [\underline{E}_1 + \underline{E}_2, \overline{E}_1 + \overline{E}_2]$$

(in all 4 cases).

- *Subtraction:* similarly,

$$E[x_1 - x_2] = E[x_1] - E[x_2],$$

so

$$[\underline{E}, \overline{E}] = [\underline{E}_1 - \overline{E}_2, \overline{E}_1 - \underline{E}_2].$$

(in all 4 cases).

- *Product, independent x_i:*

here, $E[x_1 \cdot x_2] = E[x_1] \cdot E[x_2]$, hence

$$\mathbf{E} = \mathbf{E}_1 \cdot \mathbf{E}_2.$$

Product – Case When We Have No Info About Correlation: Theorem

Theorem. For multiplication $y = x_1 \cdot x_2$, when we have no information about the correlation,

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

$$\min(p_1, 1 - p_2) \cdot \overline{x}_1 \cdot \underline{x}_2 +$$

$$\min(1 - p_1, p_2) \cdot \underline{x}_1 \cdot \overline{x}_2 +$$

$$\max(1 - p_1 - p_2, 0) \cdot \underline{x}_1 \cdot \underline{x}_2;$$

and

$$\overline{E} = \min(p_1, p_2) \cdot \overline{x}_1 \cdot \overline{x}_2 +$$

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

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

$$\min(1 - p_1, 1 - p_2) \cdot \underline{x}_1 \cdot \underline{x}_2,$$

where $p_i \stackrel{\text{def}}{=} (E_i - \underline{x}_i)/(\overline{x}_i - \underline{x}_i)$.

Meaning of the Theorem

- What are p_i: if we only allow values $\underline{x}_i$ and $\overline{x}_i$, then p_i is $p[\overline{x}_i]$ for which average is E_i; then $p[\underline{x}_i] = 1 - p_i$.
- If we know $p(A)$ and $p(B)$, then $p(A \,\&\, B)$ can take any values:
  - from $\underline{p}(A \,\&\, B) \stackrel{\text{def}}{=} \max(p(A) + p(B) - 1, 0)$
  - to $\overline{p}(A \,\&\, B) \stackrel{\text{def}}{=} \min(p(A), p(B))$;
- Hence,

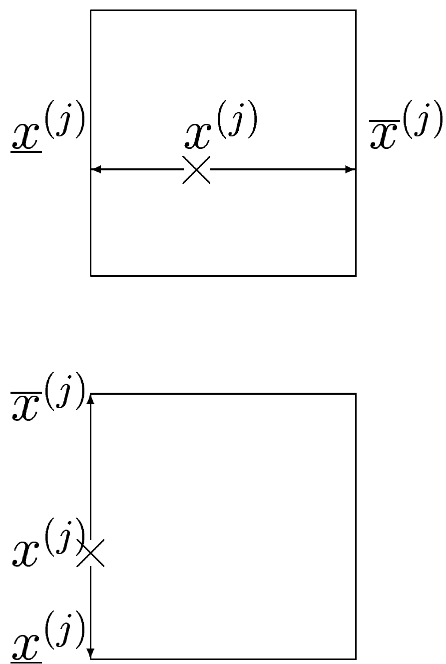
$$\underline{E} = \underline{p}[\overline{x}_1 \,\&\, \overline{x}_2] \cdot \overline{x}_1 \cdot \overline{x}_2 + \overline{p}[\overline{x}_1 \,\&\, \underline{x}_2] \cdot \overline{x}_1 \cdot \underline{x}_2 +$$

$$\overline{p}[\underline{x}_1 \,\&\, \overline{x}_2] \cdot \underline{x}_1 \cdot \overline{x}_2 + \underline{p}[\underline{x}_1 \,\&\, \underline{x}_2] \cdot \underline{x}_1 \cdot \underline{x}_2;$$

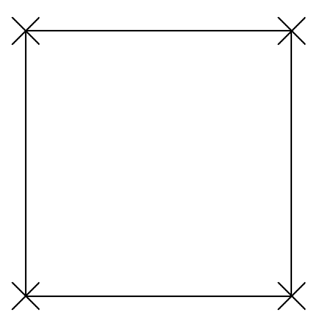
$$\overline{E} = \overline{p}[\overline{x}_1 \,\&\, \overline{x}_2] \cdot \overline{x}_1 \cdot \overline{x}_2 + \underline{p}[\overline{x}_1 \,\&\, \underline{x}_2] \cdot \overline{x}_1 \cdot \underline{x}_2 +$$

$$\underline{p}[\underline{x}_1 \,\&\, \overline{x}_2] \cdot \underline{x}_1 \cdot \overline{x}_2 + \overline{p}[\underline{x}_1 \,\&\, \underline{x}_2] \cdot \underline{x}_1 \cdot \underline{x}_2.$$

Proof: Main Idea



Thus, instead of considering all possible distributions, it is sufficient to consider only distributions for which $x_1 \in \{\underline{x}_1, \overline{x}_1\}$ and $x_2 \in \{\underline{x}_2, \overline{x}_2\}$:



Further Results

- Similar results are given:
  - correlation cases;
  - for the case when we have non-degenerate intervals $\mathbf{E}_i$.
  - for other elementary arithmetic operations ($1/x$, min, max);
- Similar ideas can be used:
  - for more general operations;
  - for the case when we know 2nd moments in addition to the 1st moments.

Acknowledgments

This work was supported in part:

- by NASA under grants NCC5-209 and NCC2-1232;
- by the Air Force Office of Scientific Research grants F30602-00-2-0503 and F49620-00-1-0365;
- by NSF grants CDA-9522207, EAR-0112968, EAR-0225670, and 9710940 Mexico/Conacyt;
- by Small Business Innovation Research grant 9R44CA81741 from the National Institutes of Health (NIH);
- IEEE/ACM SC2002 Minority Serving Institutions Participation Grants;
- by a research grant from Sandia National Laboratories as part of the Department of Energy Accelerated Strategic Computing Initiative (ASCI).