Estimating Variance under Interval and Fuzzy Uncertainty: Parallel Algorithms

Karen Villaverde
Department of Computer Science
New Mexico State University
Las Cruces, NM 88003, USA
email kvillave@cs.nmsu.edu

Gang Xiang
Philips Healthcare Informatics
El Paso, Texas
email gxiang@acm.org

Estimating Variance . . .
Case of Measurement . . .
Case of Expert . . .
Estimating Mean . . .
Estimating Mean . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Acknowledgments

1. Estimating Variance under Uncertainty

- *Computing statistics is important:* traditional data processing starts with computing population mean and population variance:

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

- *Traditional approach:* assumes that we know the *exact* values x_i.
- *In practice:* these values come either from measurements or from expert estimates.
- *Uncertainty:* in both cases, we get only *approximations* $\widetilde{x}_i$ to the actual (unknown) values x_i.
- *Result:* we only get approximate valued $\widetilde{E}$ and $\widetilde{V}$.
- *Question:* what is the accuracy of these approximations?

Estimating Variance . . .
Case of Measurement . . .
Case of Expert . . .
Estimating Mean . . .
Estimating Mean . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Acknowledgments

2. Case of Measurement Uncertainty

- The result $\widetilde{x}$ of the measurement is, in general, different from the (unknown) actual value x: $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x \neq 0$.
- *Upper bound* Δ is usually supplied by the manufacturer: $|\Delta x| \leq \Delta$.
- *Interval uncertainty:* $x \in [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.
- *Probabilistic approach:* often, we know probabilities of different values of Δx.
- *How these probabilities are determined:* by comparing with standard measuring instrument (SMI).
- *Cases when we do not know probabilities:*
  - cutting-edge measurements;
  - manufacturing.
- *Resulting problem:* find the ranges $\mathbf{E}$ and $\mathbf{V}$ of E and V.

Estimating Variance . . .
Case of Measurement . . .
Case of Expert . . .
Estimating Mean . . .
Estimating Mean . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Acknowledgments

3. Case of Expert Uncertainty

- *Situation:* an expert use natural language.
- *Example:* "most probably, the value of the quantity is between 6 and 7, but it is somewhat possible to have values between 5 and 8".
- *Natural formalization:* for every i, a fuzzy set $\mu_i(x_i)$.
- *Resulting problem:* given fuzzy numbers x_i, find the fuzzy numbers for E and V.
- *Reduction to interval case:* the α-cut for $C(x_1, \ldots, x_n)$ is equal to the range of C when x_i are in the corresponding α-cuts: $x_i \in \mathbf{x}_i(\alpha)$.
- *Conclusion:* for each characteristic $C(x_1, \ldots, x_n)$, it is sufficient to be able to compute the range

$$C(\mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{def}}{=} \{C(x_1, \ldots, x_n) \mid x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}.$$

Estimating Variance . . .
Case of Measurement . . .
Case of Expert . . .
Estimating Mean . . .
Estimating Mean . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Acknowledgments

4. Estimating Mean under Interval Uncertainty: What Is Known

- *Fact:* the arithmetic average $E(x_1, \ldots, x_n)$ is an increasing function of $x_1, \ldots, x_n$.
- *Conclusions:*
  - the smallest possible value $\underline{E}$ of E is attained when each value x_i is the smallest possible ($x_i = \underline{x}_i$);
  - the largest possible value $\overline{E}$ of E is attained when $x_i = \overline{x}_i$ for all i.
- *Resulting formulas:* the range $\mathbf{E}$ of E is equal to

$$[E(\underline{x}_1, \ldots, \underline{x}_n), E(\overline{x}_1, \ldots, \overline{x}_n)],$$

i.e., to

$$\mathbf{E} = [\underline{E}, \overline{E}] = \left[\frac{1}{n} \cdot (\underline{x}_1 + \ldots + \underline{x}_n), \frac{1}{n} \cdot (\overline{x}_1 + \ldots + \overline{x}_n)\right].$$

Estimating Variance . . .
Case of Measurement . . .
Case of Expert . . .
Estimating Mean . . .
Estimating Mean . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Acknowledgments

5. Estimating Mean under Interval Uncertainty: Parallelization

- *General problem:* for large n, the corresponding algorithm may require a large computation time.
- *Possible solution:* if we have several (p) processors, we may speed up computations by parallelization.
- *Case of the mean – reminder:*

$$\mathbf{E} = [\underline{E}, \overline{E}] = \left[\frac{1}{n} \cdot (\underline{x}_1 + \ldots + \underline{x}_n), \frac{1}{n} \cdot (\overline{x}_1 + \ldots + \overline{x}_n)\right].$$

- *Parallelization:*
  - divide n elements into p groups of n/p elements;
  - each of p processors computes the sum of all the elements from the corresponding group in time $O(n/p)$;
  - then, we add p subsums.
- *Resulting computation time:* $O(n/p + \log(p))$.

Estimating Variance . . .
Case of Measurement . . .
Case of Expert . . .
Estimating Mean . . .
Estimating Mean . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Acknowledgments

6. Estimating Variance under Interval Uncertainty: What is Known

- *Problem:* compute the range $\mathbf{V} = [\underline{V}, \overline{V}]$ of the variance V over interval data $x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.
- *Known:* there is a polynomial-time algorithm for computing $\underline{V}$.
- *In general:* computing $\overline{V}$ is NP-hard.
- *In many practical situations:* there are efficient algorithms for computing $\overline{V}$.
- *Example:* consider narrowed intervals $[x_i^-, x_i^+]$, where $x_i^- \stackrel{\text{def}}{=} \widetilde{x}_i - \dfrac{\Delta_i}{n}$ and $x_i^+ \stackrel{\text{def}}{=} \widetilde{x}_i + \dfrac{\Delta_i}{n}$.
- *Case:* no two narrowed intervals are proper subsets of one another, i.e., $[x_i^-, x_i^+] \not\subseteq (x_j^-, x_j^+)$ for all i and j.
- *For this case:* there exists an $O(n \cdot \log(n))$ time algorithm for computing $\overline{V}$.

Estimating Variance . . .
Case of Measurement . . .
Case of Expert . . .
Estimating Mean . . .
Estimating Mean . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Acknowledgments

7. Estimating Variance Under Interval Uncertainty: Main Idea

- *Reminder:* $V = M - E^2$, where $M = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2$ and $E = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$.
- *Main lemma:* if we sort the narrowed intervals in lexicographic order, then $\overline{V}$ is attained at one of the vectors

$$x = (\underline{x}_1, \ldots, \underline{x}_k, \overline{x}_{k+1}, \ldots, \overline{x}_n).$$

- *Conclusion:* for some k, we have $\overline{V} = M_k - E_k^2$, where $M_k = \underline{M}_k + \overline{M}_k$, $E_k = \underline{E}_k + \overline{E}_k$, $\underline{M}_k = \frac{1}{n} \cdot \sum_{i=1}^{k} (\underline{x}_i)^2$, $\overline{M}_k = \frac{1}{n} \cdot \sum_{i=k+1}^{n} (\overline{x}_i)^2$, $\underline{E}_k = \frac{1}{n} \cdot \sum_{i=1}^{k} \underline{x}_i$, $\overline{E}_k = \frac{1}{n} \cdot \sum_{i=k+1}^{n} \overline{x}_i$.

Estimating Variance . . .
Case of Measurement . . .
Case of Expert . . .
Estimating Mean . . .
Estimating Mean . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Acknowledgments

8. Estimating Variance Under Interval Uncertainty: Resulting Algorithm

- First, we sort the intervals; this takes time $O(n \cdot \log(n))$.
- Then, for every k, we compute $\underline{M}_k$, $\overline{M}_k$, $\underline{E}_k$, $\overline{E}_k$, and

$$V_k = (\underline{M}_k + \overline{M}_k) - (\underline{E}_k + \overline{E}_k)^2 :$$

  - computing the values for $k = 0$ takes linear time $O(n)$;
  - then, we update in $O(1)$ steps for each k, e.g., $\underline{M}_{k+1} = \underline{M}_k + \frac{1}{n} \cdot (\underline{x}_{k+1})^2$.
- Finally, we find the largest of the values $V_0, \ldots, V_{n+1}$; this takes $O(n)$ time.
- Total time $O(n \cdot \log(n)) + O(n) + n \cdot O(1) + O(n) = O(n \cdot \log(n))$.

Estimating Variance . . .
Case of Measurement . . .
Case of Expert . . .
Estimating Mean . . .
Estimating Mean . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Acknowledgments

9. Estimating Variance Under Interval Uncertainty: Parallel Algorithm

If we have $p < n$ processors, then we can:

- on Stage 1, sort n values in time $O\left(\dfrac{n \cdot \log(n)}{p} + \log(n)\right)$;
- on Stage 2, compute all $n+1$ values $\underline{M}_k$, $\overline{M}_k$, $\underline{E}_k$, $\overline{E}_k$ (and hence V_k) in time in $O(n/p + \log(p))$;
- on Stage 3, compute the maximum of $V_0, \ldots, V_n$ in time $O(n/p + \log(p))$.

Overall, we thus need time (since $p \le n$):

$$O\left(\frac{n \cdot \log(n)}{p} + \log(n)\right) + O\left(\frac{n}{p} + \log(p)\right) + O\left(\frac{n}{p} + \log(p)\right) =$$

$$O\left(\frac{n \cdot \log(n)}{p} + \log(n)\right).$$

Estimating Variance . . .
Case of Measurement . . .
Case of Expert . . .
Estimating Mean . . .
Estimating Mean . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Acknowledgments

10. Acknowledgments

This work was supported in part

- by an internal grant from New Mexico State University,
- by NSF grants HRD-0734825, EAR-0225670, and EIA-0080940, and
- by Texas Department of Transportation contract No. 0-5453.

The authors are thankful to the anonymous referees for valuable suggestions.

Estimating Variance . . .
Case of Measurement . . .
Case of Expert . . .
Estimating Mean . . .
Estimating Mean . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Estimating Variance . . .
Acknowledgments