Departmental Technical Reports (CS)   Computer Science

12-2008

Estimating Risk under Interval Uncertainty: Sequential and Parallel Algorithms

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Hung T. Nguyen

Songsak Sriboonchitta

Recommended Citation

Estimating Risk under Interval Uncertainty: Sequential and Parallel Algorithms

Vladik Kreinovich[1], Hung T. Nguyen[2], and Songsak Sriboonchita[3]

[1]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
vladik@utep.edu

[2]Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
hunguyen@nmsu.edu

[3]Faculty of Economics
Chiang Mai University
Chiang Mai 50200 Thailand
songsak@econ.cmu.ac.th

Abstract

In traditional econometrics, the quality of an individual investment – and of the investment portfolio – is characterized by its expected return and its risk (variance). For an individual investment or portfolio, we can estimate the future expected return and a future risk by tracing the returns $x_1, \ldots, x_n$ of this investment (and/or similar investments) over the past years, and computing the statistical characteristics based on these returns. The return (per unit investment) is defined as the selling of the corresponding financial instrument at the ends of, e.g., a one-year period, divided by the buying price of this instrument at the beginning of this period. It is usually assumed that we know the exact return values $x_1, \ldots, x_n$. In practice, however, both the selling and the buying prices unpredictably fluctuate from day to day – and even within a single day. These minute-by-minute fluctuations are rarely recorded; what we usually have recorded is the daily range of prices. As a result, we can only find the range $[\underline{x}_i, \overline{x}_i]$ of possible values of the return x_i. In this case, different

possible values of x_i lead, in general, to different values of the expected return E and of the risk V. In such situations, we are interested in producing the intervals of possible values of E and V.

In the paper, we describe algorithms for producing such interval estimates. The corresponding sequential algorithms, however, are reasonably complex and time-consuming. In financial applications, it is often very important to produce the result as fast as possible. One way to speed up computations is to perform these algorithms in parallel on several processors, and thus, to speed up computations. In this paper, we show how the algorithms for estimating variance under interval uncertainty can be parallelized.

1 Computing statistics is important

In traditional econometrics, the quality of an individual investment – and of the investment portfolio – is characterized by its expected return and its risk (variance). For an individual investment or portfolio, we can estimate the future expected return and a future risk by tracing the returns $x_1, \ldots, x_n$ of this investment (and/or similar investments) over the past years, and computing the statistical characteristics based on these returns: the expected return

$$E = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \tag{1}$$

and the risk

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

2 Additional problem: interval uncertainty

The return (per unit investment) is defined as the selling price of the corresponding financial instrument at the end of, e.g., a one-year period, divided by the buying price of this instrument at the beginning of this period. It is usually assumed that we know the exact values $x_1, \ldots, x_n$ of the returns.

In practice, however, both the selling and the buying prices unpredictably fluctuate from day to day – and even within a single day. These minute-by-minute fluctuations are not always recorded; what we usually have recorded is the daily range of prices. As a result, we can only find the interval range $[\underline{x}_i, \overline{x}_i]$ of possible values of the return x_i.

3 Traditional approach to solving the problem of interval uncertainty

A traditional approach to solving this problem is to take the average

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

and to compute the characteristics based on these averages. In particular, as an estimate for the expected return, we use the value

$$\widetilde{E} = \frac{1}{n} \cdot \sum_{i=1}^{n} \widetilde{x}_i,$$

and as an estimate for the risk, we use the value

$$\widetilde{V} = \frac{1}{n} \cdot \sum_{i=1}^{n} (\widetilde{x}_i - \widetilde{E})^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (\widetilde{x}_i)^2 - \left(\frac{1}{n} \cdot \sum_{i=1}^{n} \widetilde{x}_i \right)^2.$$

4 Traditional approach: limitations

The traditional approach can lead to an underestimation of risk. For example, in the bull market, there may be dips leading to a small value of x_i, but overall, the values are increasing and therefore, the upper value $\overline{x}_i$ is a reasonable estimate for x_i. Thus, if we use the average, we underestimate the actual high price.

In contrast, for the bear market, spikes are accidental but lower values are more typical, so a more reasonable approximation for the actual x_i is the lower bound $\underline{x}_i$. Thus, if we use the average, we overestimate the actual low price.

As a result, when we compute the variance based on the values $\widetilde{x}_i$, we underestimate the low prices and underestimate the high prices – thus underestimating the variance (which is the measure of price variation).

5 Estimating statistics under interval uncertainty: a computational problem

In the case of interval uncertainty, instead of the true values $x_1, \ldots, x_n$, we only know the intervals $\mathbf{x}_1 = [\underline{x}_1, \overline{x}_1], \ldots, \mathbf{x}_n = [\underline{x}_n, \overline{x}_n]$ that contain the (unknown) true values of the measured quantities. For different values $x_i \in \mathbf{x}_i$, we get, in general, different values of the corresponding statistical characteristic $C(x_1, \ldots, x_n)$. Since all values $x_i \in \mathbf{x}_i$ are possible, we conclude that all the values $C(x_1, \ldots, x_n)$ corresponding to $x_i \in \mathbf{x}_i$ are possible estimates for the corresponding statistical characteristic. Therefore, for the interval data $\mathbf{x}_1, \ldots, \mathbf{x}_n$, a reasonable estimate for the corresponding statistical characteristic is the range

$$C(\mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{def}}{=} \{ C(x_1, \ldots, x_n) \,|\, x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n \}. \tag{3}$$

We must therefore modify the existing statistical algorithms so that they compute, or bound these ranges.

6 Estimating expected return under interval uncertainty

The expected return (arithmetic average) E is a monotonically increasing function of each of its n variables $x_1, \ldots, x_n$, so its smallest possible value $\underline{E}$ is attained when each value x_i is the smallest possible ($x_i = \underline{x}_i$) and its largest possible value is attained when $x_i = \overline{x}_i$ for all i. In other words, the range $\mathbf{E}$ of E is equal to $[E(\underline{x}_1, \ldots, \underline{x}_n), E(\overline{x}_1, \ldots, \overline{x}_n)]$. In other words, $\underline{E} = \dfrac{1}{n} \cdot (\underline{x}_1 + \ldots + \underline{x}_n)$ and $\overline{E} = \dfrac{1}{n} \cdot (\overline{x}_1 + \ldots + \overline{x}_n)$.

7 Linearized techniques

When the daily fluctuations are small, we can use the linearization techniques. Specifically, we represent the values x_i as $x_i = \widetilde{x}_i + \Delta x_i$, where the differences $\Delta x_i \stackrel{\text{def}}{=} x_i - \widetilde{x}_i$ are small, and we ignore quadratic terms in the formula for the variance.

The condition that $x_i \in [\underline{x}_i, \overline{x}_i]$ means that $\Delta x_i \in [-\Delta_i, \Delta_i]$, where $\Delta_i \stackrel{\text{def}}{=} \overline{x}_i - \widetilde{x}_i = \widetilde{x}_i - \underline{x}_i$ and hence,

$$\Delta_i = \frac{\overline{x}_i - \underline{x}_i}{2}.$$

In general, the variance has the form

$$V = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i \right)^2.$$

Linearization means that we replace the exact value

$$V(x_1, \ldots, x_n) = V(\widetilde{x}_1 + \Delta x_1, \ldots, \widetilde{x}_n + \Delta x_n)$$

with an approximate value

$$V(x_1, \ldots, x_n) = V(\widetilde{x}_1 + \Delta x_1, \ldots, \widetilde{x}_n + \Delta x_n) \approx$$

$$V(\widetilde{x}_1, \ldots, \widetilde{x}_n) + \sum_{i=1}^{n} \frac{\partial V}{\partial x_i}(\widetilde{x}_1, \ldots, \widetilde{x}_n) \cdot \Delta x_i.$$

Here, $\widetilde{V} = V(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ is the risk estimated based on the daily averages. The partial derivatives $\dfrac{\partial V}{\partial x_i}$ can be explicitly described, as

$$\frac{\partial V}{\partial x_i}(x_1, \ldots, x_n) = 2x_i - 2\left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i \right) = 2 \cdot (x_i - E),$$

so

$$\frac{\partial V}{\partial x_i}(\widetilde{x}_1, \ldots, \widetilde{x}_n) = 2(\widetilde{x}_i - \widetilde{E}).$$

Thus, the linearized expression for the variance takes the form

$$V = \widetilde{V} + 2\sum_{i=1}^{n}(\widetilde{x}_i - \widetilde{E}) \cdot \Delta x_i.$$

The expression for V is monotonic in each of the unknowns $\Delta x_i \in [-\Delta_i, \Delta_i]$:

- it is increasing when $\widetilde{x}_i \geq \widetilde{E}$ and
- it is decreasing when $\widetilde{x}_i \leq \widetilde{E}$.

Thus:

- When $\widetilde{x}_i \geq \widetilde{E}$, the maximum is attained when Δx_i attains its largest possible value Δ_i. For this value $\Delta x_i = \Delta_i$, the corresponding term in the expression for V takes the form $(\widetilde{x}_i - \widetilde{E}) \cdot \Delta_i$.

- When $\widetilde{x}_i \leq \widetilde{E}$, the maximum is attained when Δx_i attains its smallest possible value $-\Delta_i$. For this value $\Delta x_i = -\Delta_i$, the corresponding term in the expression for V takes the form $-(\widetilde{x}_i - \widetilde{E}) \cdot \Delta_i$.

Both cases can be described by the formula $|\widetilde{x}_i - \widetilde{E}| \cdot \Delta_i$.

Thus, the largest possible value for V takes the form $\overline{V} = \widetilde{V} + 2\Delta$, where

$$\Delta \stackrel{\text{def}}{=} \sum_{i=1}^{n} |\widetilde{x}_i - \widetilde{E}| \cdot \Delta_i.$$

One can similarly check that the smallest possible value for V takes the form $\underline{V} = \widetilde{V} - 2\Delta$. So, in the linear approximation, the range of possible values of the risk (variance) V has the form $[\widetilde{V} - 2\Delta, \widetilde{V} + 2\Delta]$.

8   Linearization approximation is not always adequate

In financial applications, the gain is usually obtained by having a small (often $<$ 1%) advantage over the competing financial instruments. From this viewpoint, it is desirable to have estimates which are as accurate as possible.

When the situation is stable, the daily fluctuations are low, and quadratic terms can be reasonable ignored. However, the whole purpose of estimating risk is to cover situations with high volatility. In such situations, the daily fluctuations $\overline{x}_i - \underline{x}_i = 2\Delta_i$ can also be sizeable, and thus, terms quadratic in Δ_i cannot be ignored if we want accurate estimates.

In such situations, we need the *exact* range of the variance (risk) V.

9 The exact estimation of risk under interval uncertainty is, in general, an NP-hard problem

It is known that the problem of computing the exact range $\mathbf{V} = [\underline{V}, \overline{V}]$ for the risk (variance) V over interval data $x_i \in [\underline{x}_i, \overline{x}_i]$ is, in general, computationally difficult (NP-hard); see, e.g., [11, 12].

Specifically, there is a $O(n \cdot \log(n))$ time algorithm for computing $\underline{V}$, but computing $\overline{V}$ is, in general, NP-hard.

10 Sequential algorithm for computing $\overline{V}$ in the no-proper-subset case

In many practical situations, there are efficient algorithms for computing $\overline{V}$. For example, there exists a $O(n \cdot \log(n))$ time algorithm which is applicable in all the cases when no two "narrowed" intervals, defined as

$$[x_i^-, x_i^+] \stackrel{\text{def}}{=} \left[\widetilde{x}_i - \frac{\Delta_i}{n}, \widetilde{x}_i + \frac{\Delta_i}{n} \right] \tag{4}$$

are proper subsets of one another, i.e., when $[x_i^-, x_i^+] \not\subset (x_j^-, x_j^+)$ for all i and j [4].

The algorithm from [4] is as follows:

1. First, we sort the values $\widetilde{x}_i$ into an increasing sequence. Without losing generality, we can assume that

$$\widetilde{x}_1 \leq \widetilde{x}_2 \leq \ldots \leq \widetilde{x}_n. \tag{5}$$

2. Then, for every k from 0 to n, we compute the value $V^{(k)} = M^{(k)} - (E^{(k)})^2$ of the sample variance V for the vector $x^{(k)} = (\underline{x}_1, \ldots, \underline{x}_k, \overline{x}_{k+1}, \ldots, \overline{x}_n)$. (For $k = 0$, $x^{(0)} = (\overline{x}_1, \ldots, \overline{x}_n)$.)

3. Finally, we compute $\overline{V}$ as the largest of $n + 1$ values $V^{(0)}, \ldots, V^{(n)}$.

To compute the values $V^{(k)}$ of Stage 2, first, we explicitly compute $M^{(0)} = \frac{1}{n} \cdot \sum_{i=1}^{n} (\overline{x}_i)^2$, $E^{(0)} = \frac{1}{n} \cdot \sum_{i=1}^{n} \overline{x}_i$, and

$$V^{(0)} = M^{(0)} - (E^{(0)})^2. \tag{6}$$

Once we know the values $M^{(k)}$ and $E^{(k)}$, we can compute

$$M^{(k+1)} = M^{(k)} + \frac{1}{n} \cdot (\underline{x}_{k+1})^2 - \frac{1}{n} \cdot (\overline{x}_{k+1})^2 \tag{7}$$

and

$$E^{(k+1)} = E^{(k)} + \frac{1}{n} \cdot \underline{x}_{k+1} - \frac{1}{n} \cdot \overline{x}_{k+1}. \tag{8}$$

Comment. Detailed justification for this algorithm is given in the Appendix.

11 Sequential algorithm: number of computation steps

Sorting requires $O(n \cdot \log(n))$ steps; see, e.g., [3]. Computing the initial values $M^{(0)}$, $E^{(0)}$, and $V^{(0)}$ requires linear time $O(n)$. For each k from 0 to $n-1$, we need a constant number of steps to compute the next values $M^{(k+1)}$, $E^{(k+1)}$, and $V^{(k+1)}$. Finally, finding the largest of $n+1$ values $V^{(k)}$ also requires $O(n)$ steps. Thus, overall, we need

$$O(n \cdot \log(n)) + O(n) + O(n) + O(n) = O(n \cdot \log(n)) \tag{9}$$

steps.

12 Comment about the possibility of linear-time algorithms

As we have seen, in the $O(n \cdot \log(n))$ algorithm, the main computation time is used on sorting. It is possible to avoid sorting when estimating variance under interval uncertainty (see, e.g., [6, 18]), and use instead the known fact that we can compute the median of a set of n elements in linear time (see, e.g., [3]). (This use of median is similar to the one from [2, 7].)

It is worth mentioning, however, that while asymptotically, the linear time algorithm for computing the median is faster than sorting, this median computing algorithm is still rather complex – so, for reasonable size n, sorting is faster than computing the median – and thus, sorting-based algorithms are actually faster than median-based ones.

13 Need for parallelization

Traditional algorithms for computing the variance V based on the exact values $x_1, \ldots, x_n$ require linear time $O(n)$. Algorithms for estimating variance under interval uncertainty require a larger amount of computation time – e.g., time $O(n \cdot \log(n))$.

In financial applications, it is often very important to produce the result as fast as possible. One way to speed up computations is to perform these algorithms in parallel on several processors, and thus, to speed up computations.

Let us we show how the algorithms for estimating variance under interval uncertainty can be parallelized.

14   Possibility of parallelization

For large n, we may want to further speed up computations if we have several processors working in parallel.

In the general case, all the stages of the above algorithm can be parallelized by known techniques. In particular, the computation of values $M^{(k)}$ and $E^{(k)}$ of Stage 2 is a particular case of a general *prefix-sum* problem, in which we must compute the values

$$a_1, \quad a_1 * a_2, \quad a_1 * a_2 * a_3, \dots, \tag{10}$$

for some associative operation $*$ (in our case, $* = +$).

15   Case of potentially unlimited number of processors

If we have a potentially unlimited number of processors, then we can do the following (see, e.g., [8], for the information on how to parallelize the corresponding stages):

- on Stage 1, we can sort the values $\widetilde{x}_i$ in time $O(\log(n))$;

- on Stage 2, we can compute the values $V^{(i)}$ (i.e., solve the prefix-sum problem) in time $O(\log(n))$;

- on Stage 3, we can compute the maximum of $V^{(i)}$ in time $O(\log(n))$.

As a result, we can check monotonicity in time

$$O(\log(n)) + O(\log(n)) + O(\log(n)) = O(\log(n)). \tag{11}$$

16   Case of a fixed number of processors

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

- on Stage 1, sort n values in time

$$O\left(\frac{n \cdot \log(n)}{p} + \log(n)\right); \tag{12}$$

see, e.g., [8];

- on Stage 2, compute the values $V^{(i)}$ in time

$$O\left(\frac{n}{p} + \log(p)\right); \tag{13}$$

see, e.g., [1];

- on Stage 3, compute the maximum of $V^{(i)}$ in time

$$O\left(\frac{n}{p} + \log(p)\right). \tag{14}$$

Overall, we thus need time

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

$$O\left(\frac{n \cdot \log(n)}{p} + \log(n) + \log(p)\right). \tag{15}$$

Acknowledgments

References

[1] G. E. Blelloch, "Prefix sums and their applications", In: J. H. Reif, *Synthesis of Parallel Algorithms*, Morgan Kaufmann, San Mateo, California, 1993, pp. 35–60.

[2] P. van der Broek and J. Noppen, "Fuzzy weighted average: alternative approach", *Proceedings of the 25th International Conference of the North American Fuzzy Information Processing Society NAFIPS'2006*, Montreal, Quebec, Canada, June 3–6, 2006.

[3] Th. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, MA, 2001.

[4] E. Dantsin, V. Kreinovich, A. Wolpert, and G. Xiang, "Population Variance under Interval Uncertainty: A New Algorithm", *Reliable Computing*, vol. 12, no. 4, pp. 273–280, 2006.

[5] D. Dubois, H. Fargier, and J. Fortin, "The empirical variance of a set of fuzzy intervals", *Proceedings of the 2005 IEEE International Conference on Fuzzy Systems FUZZ-IEEE'2005*, Reno, Nevada, May 22–25, 2005, pp. 885–890.

[6] S. Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, and L. Ginzburg, *Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty*, Sandia National Laboratories, Report SAND2007-0939, May 2007.

[7] P. Hansen, M. V. P. de Aragao, and C. C. Ribeiro, "Hyperbolic 0-1 programming and optimization in information retrieval", *Math. Programming*, vol. 52, pp. 255–263, 1991.

[8] J. Jájá, *An Introduction to Parallel Algorithms*, Addison-Wesley, Reading, MA, 1992.

[9] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: the logic of science*, Cambridge University Press, Cambridge, UK, 2003.

[10] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications.* Prentice Hall, Upper Saddle River, New Jersey, 1995.

[11] V. Kreinovich, L. Longpré, S. A. Starks, G. Xiang, J. Beck, R. Kandathi, A. Nayak, S. Ferson, and J. Hajagos, "Interval Versions of Statistical Techniques, with Applications to Environmental Analysis, Bioinformatics, and Privacy in Statistical Databases", *Journal of Computational and Applied Mathematics*, vol. 199, No. 2, pp. 418–423, 2007.

[12] V. Kreinovich, G. Xiang, S. A. Starks, L. Longpre, M. Ceberio, R. Araiza, J. Beck, R. Kandathi, A. Nayak, R. Torres, and J. Hajagos, "Towards combining probabilistic and interval uncertainty in engineering calculations: algorithms for computing statistics under interval uncertainty, and their computational complexity", *Reliable Computing*, vol. 12, no. 6, pp. 471–501, 2006.

[13] H. T. Nguyen, "A note on the extension principle for fuzzy sets", *J. Math. Anal. and Appl.*, vol. 64, pp. 369–380, 1978.

[14] H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic*, CRC Press, Boca Raton, Florida, 2005.

[15] S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer-Verlag, New York, 2005.

[16] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton, Florida, 2004.

[17] S. A. Vavasis, *Nonlinear Optimization: Complexity Issues*, Oxford University Press, New York, 1991.

[18] G. Xiang, M. Ceberio, and V. Kreinovich, "Computing Population Variance and Entropy under Interval Uncertainty: Linear-Time Algorithms", *Reliable Computing*, vol. 13, no. 6, , pp. 467–488, 2007.

A Justification for the $O(n \cdot \log(n))$ time algorithm

Let us first prove that the no-subset condition is equivalent to the condition that

$$|\widetilde{x}_i - \widetilde{x}_j| \geq \frac{|\Delta_i - \Delta_j|}{n}. \tag{16}$$

10

Indeed, the condition $|\widetilde{x}_i - \widetilde{x}_j| \geq \dfrac{|\Delta_i - \Delta_j|}{n}$ means that if $\widetilde{x}_i \geq \widetilde{x}_j$, then we have

$$\widetilde{x}_i - \widetilde{x}_j \geq \frac{\Delta_i - \Delta_j}{n}, \tag{17}$$

i.e.,

$$\widetilde{x}_i - \frac{\Delta_i}{n} \geq \widetilde{x}_j - \frac{\Delta_j}{n} \tag{18}$$

and also

$$\widetilde{x}_i - \widetilde{x}_j \geq \frac{\Delta_j - \Delta_i}{n}, \tag{19}$$

i.e.,

$$\widetilde{x}_i + \frac{\Delta_i}{n} \geq \widetilde{x}_j + \frac{\Delta_j}{n}. \tag{20}$$

This means that no narrowed interval (4) is a proper subinterval of the interior of another narrowed subinterval.

Vice versa, if one of the narrowed intervals is a proper subinterval of another one, then the condition (16) is not satisfied. Thus, the condition (16) indeed means that no *narrowed subintervals* are proper subintervals of each other.

Let us now prove that the above algorithm is indeed correct. With respect to each variable x_i, the population variance is a quadratic function which is non-negative for all x_i. It is well known that a maximum of such a function on each interval $[\underline{x}_i, \overline{x}_i]$ is attained at one of the endpoints of this interval. Thus, the maximum $\overline{V}$ of the population variance is attained at a vector $x = (x_1, \ldots, x_n)$ in which each value x_i is equal either to $\underline{x}_i$ or to $\overline{x}_i$.

We will first justify our algorithm for the case when $|\widetilde{x}_i - \widetilde{x}_j| > \dfrac{|\Delta_i - \Delta_j|}{n}$ for all $i \neq j$ and $\Delta_i > 0$ for all i.

To justify our algorithm, we need to prove that this maximum is attained at one of the vectors $x^{(k)}$ in which all the lower bounds $\underline{x}_i$ precede all the upper bounds $\overline{x}_i$. We will prove this by reduction to a contradiction. Indeed, let us assume that the maximum is attained at a vector x in which one of the lower bounds follows one of the upper bounds. In each such vector, let i be the largest upper bound index preceded by the lower bound; then, in the optimal vector x, we have $x_i = \overline{x}_i$ and $x_{i+1} = \underline{x}_{i+1}$.

Since the maximum is attained for $x_i = \overline{x}_i$, replacing it with $\underline{x}_i = \overline{x}_i - 2 \cdot \Delta_i$ will either decrease the value of the variance or keep it unchanged. Let us describe how variance changes under this replacement. In the sum for M, we replace $(\overline{x}_i)^2$ with

$$(\underline{x}_i)^2 = (\overline{x}_i - 2 \cdot \Delta_i)^2 = (\overline{x}_i)^2 - 4 \cdot \Delta_i \cdot \overline{x}_i + 4 \cdot \Delta_i^2. \tag{21}$$

Thus, the value M changes into $M + \Delta M_i$, where

$$\Delta M_i = -\frac{4}{n} \cdot \Delta_i \cdot \overline{x}_i + \frac{4}{n} \cdot \Delta_i^2. \tag{22}$$

11

The population mean E changes into $E + \Delta E_i$, where $\Delta E_i = -\dfrac{2 \cdot \Delta_i}{n}$. Thus, the value E^2 changes into $(E + \Delta E_i)^2 = E^2 + \Delta(E^2)_i$, where

$$\Delta(E^2)_i = 2 \cdot E \cdot \Delta E_i + \Delta E_i^2 = -\frac{4}{n} \cdot E \cdot \Delta_i + \frac{4}{n^2} \cdot \Delta_i^2. \qquad (23)$$

So, the variance V changes into $V + \Delta V_i$, where

$$\Delta V_i = \Delta M_i - \Delta(E^2)_i = -\frac{4}{n} \cdot \Delta_i \cdot \overline{x}_i + \frac{4}{n} \cdot \Delta_i^2 + \frac{4}{n} \cdot E \cdot \Delta_i - \frac{4}{n^2} \cdot \Delta_i^2 =$$

$$\frac{4}{n} \cdot \Delta_i \cdot \left(-\overline{x}_i + \Delta_i + E - \frac{\Delta_i}{n} \right). \qquad (24)$$

By definition, $\overline{x}_i = \widetilde{x}_i + \Delta_i$, hence $-\overline{x}_i + \Delta_i = -\widetilde{x}_i$. Thus, we conclude that

$$\Delta V_i = \frac{4}{n} \cdot \Delta_i \cdot \left(-\widetilde{x}_i + E - \frac{\Delta_i}{n} \right). \qquad (25)$$

Since V attains maximum at x, we have $\Delta V_i \leq 0$, hence

$$E \leq \widetilde{x}_i + \frac{\Delta_i}{n}. \qquad (26)$$

Similarly, since the maximum is attained for $x_{i+1} = \underline{x}_i$, replacing it with $\overline{x}_{i+1} = \underline{x}_{i+1} + 2 \cdot \Delta_{i+1}$ will either decrease the value of the variance or keep it unchanged. Let us describe how variance changes under this replacement. In the sum for M, we replace $(\underline{x}_{i+1})^2$ with

$$(\overline{x}_{i+1})^2 = (\underline{x}_{i+1} + 2 \cdot \Delta_{i+1})^2 = (\underline{x}_{i+1})^2 + 4 \cdot \Delta_{i+1} \cdot \underline{x}_{i+1} + 4 \cdot \Delta_{i+1}^2. \qquad (27)$$

Thus, the value M changes into $M + \Delta M_{i+1}$, where

$$\Delta M_{i+1} = \frac{4}{n} \cdot \Delta_{i+1} \cdot \underline{x}_{i+1} + \frac{4}{n} \cdot \Delta_{i+1}^2. \qquad (28)$$

The population mean E changes into $E + \Delta E_{i+1}$, where $\Delta E_{i+1} = \dfrac{2 \cdot \Delta_{i+1}}{n}$. Thus, the value E^2 changes into $(E + \Delta E_{i+1})^2 = E^2 + \Delta(E^2)_{i+1}$, where

$$\Delta(E^2)_{i+1} = 2 \cdot E \cdot \Delta E_{i+1} + \Delta E_{i+1}^2 = \frac{4}{n} \cdot E \cdot \Delta_{i+1} + \frac{4}{n^2} \cdot \Delta_{i+1}^2. \qquad (29)$$

So, the variance V changes into $V + \Delta V_{i+1}$, where

$$\Delta V_{i+1} = \Delta M_{i+1} - \Delta(E^2)_{i+1} =$$

$$\frac{4}{n} \cdot \Delta_{i+1} \cdot \underline{x}_{i+1} + \frac{4}{n} \cdot \Delta_{i+1}^2 - \frac{4}{n} \cdot E \cdot \Delta_{i+1} - \frac{4}{n^2} \cdot \Delta_{i+1}^2 =$$

$$\frac{4}{n} \cdot \Delta_{i+1} \cdot \left(\underline{x}_{i+1} + \Delta_{i+1} - E - \frac{\Delta_{i+1}}{n} \right). \qquad (30)$$

By definition, $\underline{x}_{i+1} = \widetilde{x}_{i+1} - \Delta_{i+1}$, hence $\underline{x}_{i+1} + \Delta_{i+1} = \widetilde{x}_{i+1}$. Thus, we conclude that

$$\Delta V_{i+1} = \frac{4}{n} \cdot \Delta_{i+1} \cdot \left(\widetilde{x}_{i+1} - E - \frac{\Delta_{i+1}}{n} \right). \tag{31}$$

Since V attains maximum at x, we have $\Delta V_{i+1} \leq 0$, hence

$$E \geq \widetilde{x}_{i+1} - \frac{\Delta_{i+1}}{n}. \tag{32}$$

We can also change *both* x_i and x_{i+1} at the same time. In this case, the change ΔM in M is simply the sum of the changes coming from x_i and x_{i+1}: $\Delta M = \Delta M_i + \Delta M_{i+1}$, and the change ΔE in E is also the sum of the corresponding changes: $\Delta E = \Delta E_i + \Delta E_{i+1}$. So, for

$$\Delta V = \Delta M - \Delta(E^2) = \Delta M - 2 \cdot E \cdot \Delta E - \Delta E^2, \tag{33}$$

we get

$$\Delta V = \Delta M_i + \Delta M_{i+1} - 2 \cdot E \cdot \Delta E_i - 2 \cdot E \cdot \Delta E_{i+1} - (\Delta E_i)^2 -$$
$$(\Delta E_{i+1})^2 - 2 \cdot \Delta E_i \cdot \Delta E_{i+1}. \tag{34}$$

Hence,

$$\Delta V = (\Delta M_i - 2 \cdot E \cdot \Delta E_i - (\Delta E_i)^2) +$$
$$(\Delta M_{i+1} - 2 \cdot E \cdot \Delta E_{i+1} - (\Delta E_{i+1})^2) - 2 \cdot \Delta E_i \cdot \Delta E_{i+1}, \tag{35}$$

i.e.,

$$\Delta V = \Delta V_i + \Delta V_{i+1} - 2 \cdot \Delta E_i \cdot \Delta E_{i+1}. \tag{36}$$

We already have the expressions for ΔV_i, ΔV_{i+1}, $\Delta E_i = -\dfrac{2 \cdot \Delta_i}{n}$, and $\Delta E_{i+1} = \dfrac{2 \cdot \Delta_{i+1}}{n}$, so we conclude that $\Delta V = \dfrac{4}{n} \cdot D(E)$, where x

$$D(E) \overset{\text{def}}{=} \Delta_i \cdot \left(-\widetilde{x}_i + E - \frac{\Delta_i}{n} \right) + \Delta_{i+1} \cdot \left(\widetilde{x}_{i+1} - E - \frac{\Delta_{i+1}}{n} \right) +$$
$$\frac{2}{n} \cdot \Delta_i \cdot \Delta_{i+1}. \tag{37}$$

Since the function V attains maximum at x, we have $\Delta V \leq 0$, hence $D(E) \leq 0$ (for the population mean E corresponding to the optimizing vector x).

The expression $D(E)$ is a linear function of E. From (26) and (32), we know that

$$\widetilde{x}_{i+1} - \frac{\Delta_{i+1}}{n} \leq E \leq \widetilde{x}_i + \frac{\Delta_i}{n}. \tag{38}$$

For $E = E^- \overset{\text{def}}{=} \widetilde{x}_{i+1} - \dfrac{\Delta_{i+1}}{n}$, we have

$$D(E^-) = \Delta_i \cdot \left(-\widetilde{x}_i + \widetilde{x}_{i+1} - \frac{\Delta_{i+1}}{n} - \frac{\Delta_i}{n} \right) + \frac{2}{n} \cdot \Delta_i \cdot \Delta_{i+1} =$$

$$\Delta_i \cdot \left(-\widetilde{x}_i + \widetilde{x}_{i+1} + \frac{\Delta_{i+1}}{n} - \frac{\Delta_i}{n} \right). \tag{39}$$

We consider the case when $|\widetilde{x}_{i+1} - x_i| > \dfrac{|\Delta_i - \Delta_{i+1}|}{n}$. Since the values $\widetilde{x}_i$ are sorted in increasing order, we have $\widetilde{x}_{i+1} \geq \widetilde{x}_i$, hence

$$\widetilde{x}_{i+1} - \widetilde{x}_i = |\widetilde{x}_{i+1} - \widetilde{x}_i| > \frac{|\Delta_i - \Delta_{i+1}|}{n} \geq \frac{\Delta_i}{n} - \frac{\Delta_{i+1}}{n}. \tag{40}$$

So, we conclude that $D(E^-) > 0$.

For $E = E^+ \stackrel{\text{def}}{=} \widetilde{x}_i + \dfrac{\Delta_i}{n}$, we have

$$D(E^+) = \Delta_{i+1} \cdot \left(\widetilde{x}_{i+1} - \widetilde{x}_i - \frac{\Delta_i}{n} - \frac{\Delta_{i+1}}{n} \right) + \frac{2}{n} \cdot \Delta_i \cdot \Delta_{i+1} =$$

$$\Delta_{i+1} \cdot \left(-\widetilde{x}_i + \widetilde{x}_{i+1} + \frac{\Delta_i}{n} - \frac{\Delta_{i+1}}{n} \right). \tag{41}$$

Here, from $|\widetilde{x}_{i+1} - \widetilde{x}_i| > \dfrac{|\Delta_i - \Delta_{i+1}|}{n}$, we also conclude that $D(E^+) > 0$.

Since the linear function $D(E)$ is positive on both endpoints of the interval $[E^-, E^+]$, it must be positive for every value E from this interval, which contradicts to our conclusion that $D(E) \geq 0$ for the actual population mean value $E \in [E^-, E^+]$. This contradiction shows that the maximum of the population variance V is indeed attained at one of the values $x^{(k)}$, hence the algorithm is justified.

The general case when $|\widetilde{x}_i - \widetilde{x}_j| \geq \dfrac{|\Delta_i - \Delta_j|}{n}$ and $\Delta_i \geq 0$ can be obtained as a limit of cases when we have strict inequalities. Since the function V is continuous, the value $\overline{V}$ continuously depends on the input bounds, so by tending to a limit, we can conclude that our algorithm works in the general case as well.