12-2006

Estimating Variance under Interval and Fuzzy Uncertainty: Case of Hierarchical Estimation

Gang Xiang

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

Comments:
Published in: Patricia Melin, Oscar Castillo, Luis T. Aguilar, Janusz Kacprzyk, and Witold Pedrycz (eds.), Foundations of Fuzzy Logic and Soft Computing, Proceedings of the World Congress of the International Fuzzy Systems Association IFSA'2007, Cancun, Mexico, June 18-21, 2007, Springer Lecture Notes on Artificial Intelligence, 2007, Vol. 4529, pp. 3-12.

Estimating Variance under Interval and Fuzzy Uncertainty: Case of Hierarchical Estimation

Gang Xiang and Vladik Kreinovich

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

1 Estimating Variance under Interval and Fuzzy Uncertainty: Motivations and Known Results

Computing statistics is important. Traditional data processing in science and engineering starts with computing the basic statistical characteristics such as the population mean and population variance:

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

Additional problem. Traditional engineering statistical formulas assume that we know the *exact* values x_i of the corresponding quantity. In practice, these values come either from measurements or from expert estimates. In both case, we get only *approximations* $\widetilde{x}_i$ to the actual (unknown) values x_i.

When we use these approximate values $\widetilde{x}_i \neq x_i$ to compute the desired statistical characteristics such as E and V, we only get approximate valued $\widetilde{E}$ and $\widetilde{V}$ for these characteristics. It is desirable to estimate the accuracy of these approximations.

Case of measurement uncertainty. Measurements are never 100% accurate. As a result, the result $\widetilde{x}$ of the measurement is, in general, different from the (unknown) actual value x of the desired quantity. The difference $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$ between the measured and the actual values is usually called a *measurement error.*

The manufacturers of a measuring device usually provide us with an upper bound Δ for the (absolute value of) possible errors, i.e., with a bound Δ for which we guarantee that $|\Delta x| \leq \Delta$. The need for such a bound comes from the very nature of a measurement process: if no such bound is provided, this means that the difference between the (unknown) actual value x and the observed value $\widetilde{x}$ can be as large as possible.

Since the (absolute value of the) measurement error $\Delta x = \tilde{x} - x$ is bounded by the given bound Δ, we can therefore guarantee that the actual (unknown) value of the desired quantity belongs to the interval $[\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.

Traditional probabilistic approach to describing measurement uncertainty. In many practical situations, we not only know the interval $[-\Delta, \Delta]$ of possible values of the measurement error; we also know the probability of different values Δx within this interval [7].

In practice, we can determine the desired probabilities of different values of Δx by comparing the results of measuring with this instrument with the results of measuring the same quantity by a standard (much more accurate) measuring instrument. Since the standard measuring instrument is much more accurate than the one use, the difference between these two measurement results is practically equal to the measurement error; thus, the empirical distribution of this difference is close to the desired probability distribution for measurement error.

Interval approach to measurement uncertainty. As we have mentioned, in many practical situations, we do know the probabilities of different values of the measurement error. There are two cases, however, when this determination is not done:

- First is the case of cutting-edge measurements, e.g., measurements in fundamental science. When a Hubble telescope detects the light from a distant galaxy, there is no "standard" (much more accurate) telescope floating nearby that we can use to calibrate the Hubble: the Hubble telescope is the best we have.
- The second case is the case of measurements on the shop floor. In this case, in principle, every sensor can be thoroughly calibrated, but sensor calibration is so costly – usually costing ten times more than the sensor itself – that manufacturers rarely do it.

In both cases, we have no information about the probabilities of Δx; the only information we have is the upper bound on the measurement error.

In this case, after performing a measurement and getting a measurement result $\widetilde{x}$, the only information that we have about the actual value x of the measured quantity is that it belongs to the interval $\mathbf{x} = [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$. In this situation, for each i, we know the interval $\mathbf{x}_i$ of possible values of x_i, and we need to find the ranges $\mathbf{E}$ and $\mathbf{V}$ of the characteristics E and V over all possible tuples $x_i \in \mathbf{x}_i$.

Case of expert uncertainty. An expert usually describes his/her uncertainty by using words from the natural language, like "most probably, the value of the quantity is between 6 and 7, but it is somewhat possible to have values between 5 and 8". To formalize this knowledge, it is natural to use *fuzzy set theory*, a formalism specifically designed for describing this type of informal ("fuzzy") knowledge [3, 6].

As a result, for every value x_i, we have a fuzzy set $\mu_i(x_i)$ which describes the expert's prior knowledge about x_i: the number $\mu_i(x_i)$ describes the expert's degree of certainty that x_i is a possible value of the i-th quantity.

An alternative user-friendly way to represent a fuzzy set is by using its α-cuts $\{x_i \mid \mu_i(x_i) > \alpha\}$ (or $\{x_i \mid \mu_i(x_i) \ge \alpha\}$). For example, the α-cut corresponding to $\alpha = 0$ is the set of all the values which are possible at all, the α-cut corresponding to $\alpha = 0.1$ is the set of all the values which are possible with degree of certainty at least 0.1, etc. In these terms, a fuzzy set can be viewed as a nested family of intervals $[\underline{x}_i(\alpha), \overline{x}_i(\alpha)]$ corresponding to different level α.

Estimating statistics under fuzzy uncertainty: precise formulation of the problem. In general, we have fuzzy knowledge $\mu_i(x_i)$ about each value x_i; we want to find the fuzzy set corresponding to a given characteristic $y = C(x_1, \ldots, x_n)$. Intuitively, the value y is a reasonable value of the characteristic if $y = f(x_1, \ldots, x_n)$ for some reasonable values x_i, i.e., if for some values $x_1, \ldots, x_n$, x_1 is reasonable, and x_2 is reasonable, …, and $f = f(x_1 \ldots, x_n)$. If we interpret "and" as min and "for some" ("or") as max, then we conclude that the corresponding degree of certainty $\mu(y)$ in y is equal to $\mu(y) = \max\{\min(\mu_1(x_1), \ldots, \mu_n(x_n)) | C(x_1, \ldots, x_n) = y\}$.

Reduction to the case of interval uncertainty. It is known that the above formula (called *extension principle*) can be reformulated as follows: for each α, the α-cut $\mathbf{y}(\alpha)$ of y is equal to the range of possible values of $C(x_1, \ldots, x_n)$ when $x_i \in \mathbf{x}_i(\alpha)$ for all i. Thus, from the computational viewpoint, the problem of computing the statistical characteristic under fuzzy uncertainty can be reduced to the problem of computing this characteristic under interval uncertainty; see, e.g., [2]

In view of this reduction, in the following text, we will consider the case of interval uncertainty.

Estimating statistics under interval uncertainty: a 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

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

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

Estimating mean under interval uncertainty. The 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 $[E(\underline{x}_1, \ldots, x_n), E(\overline{x}_1, \ldots, \overline{x}_n)]$. In other words, $\underline{E} = \frac{1}{n} \cdot (\underline{x}_1 + \ldots + \underline{x}_n)$ and $\overline{E} = \frac{1}{n} \cdot (\overline{x}_1 + \ldots + \overline{x}_n)$.

Estimating variance under interval uncertainty. It is known that the problem of computing the exact range $\mathbf{V} = [\underline{V}, \overline{V}]$ for the variance V over interval data $x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ is, in general, NP-hard; see, e.g., [4, 5]. Specifically, there is a polynomial-time algorithm for computing $\underline{V}$, but computing $\overline{V}$ is, in general, NP-hard.

In many practical situations, there are efficient algorithms for computing $\overline{V}$: e.g., an $O(n \cdot \log(n))$ time algorithm exists when no two narrowed intervals $[x_i^-, x_i^+]$, where $x_i^- \stackrel{\text{def}}{=} \widetilde{x}_i - \frac{\Delta_i}{n}$ and $x_i^+ \stackrel{\text{def}}{=} \widetilde{x}_i + \frac{\Delta_i}{n}$, are proper subsets of one another, i.e., when $[x_i^-, x_i^+] \not\subseteq (x_j^-, x_j^+)$ for all i and j [1].

2 Hierarchical Case: Formulation of the Problem

Hierarchical case: description. In some practical situations, we do not know the individual values of the observations x_i, we only have average values corresponding to several ($m < n$) groups $I_1, \ldots, I_m$ of observations. Typically, for each group j, we know the frequency p_j of this group (i.e., the probability that a randomly selected observation belongs to this group), the average E_j over this group, and the standard deviation σ_j within j-th group.

Hierarchical case: analysis. In this case, the overall average E can be described as

$$E = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{n} \cdot \sum_{j=1}^{m} \sum_{i \in I_j} x_i.$$

By definition of the group average E_j, we have $E_j = \frac{1}{n_j} \cdot \sum\limits_{i \in I_j} x_i$, where $n_j = p_j \cdot n$ denotes the overall number of elements in the j-th group. Thus, $\sum\limits_{i \in I_j} x_i = n_j \cdot E_j = p_j \cdot n \cdot E_j$, hence

$$E = \sum_{j=1}^{m} p_j \cdot E_j. \tag{1}$$

Similarly, the overall variance $V = \sigma^2$ can be described as

$$V = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 - E^2 = \frac{1}{n} \cdot \sum_{j=1}^{m} \sum_{i \in I_j} x_i^2 - E^2.$$

For each j and for each $i \in I_j$, we have $x_i = (x_i - E_j) + E_j$, hence $x_i^2 = (x_i - E_j)^2 + E_j^2 + 2(x_i - E_j) \cdot E_j$. Therefore,

$$\sum_{i \in I_j} x_i^2 = \sum_{i \in I_j} (x_i - E_j)^2 + n_j \cdot E_j^2 + 2E_j \cdot \sum_{i \in I_j} (x_i - E_j).$$

The first sum, by definition of population variance σ_j, is equal to $n_j \cdot \sigma_j^2$; the third sum, by definition of the population mean, is equal to 0. Thus, $\sum\limits_{i \in I_j} x_i^2 = n_j \cdot (\sigma_j^2 + E_j^2)$, where $n_j = p_j \cdot n$, and thus,

$$V = V_E + V_\sigma, \tag{2}$$

where

$$V_E \stackrel{\text{def}}{=} M_E - E^2, \tag{3}$$

$$M_E \stackrel{\text{def}}{=} \sum_{j=1}^{m} p_j \cdot E_j^2, \tag{4}$$

$$V_\sigma \stackrel{\text{def}}{=} \sum_{j=1}^{m} p_j \cdot \sigma_j^2. \tag{5}$$

Hierarchical case: situation with interval uncertainty. It is reasonable to consider the situations when we only know the values E_j and σ_j with interval uncertainty, i.e., when we only know the intervals $\mathbf{E}_j = [\underline{E}_j, \overline{E}_j]$ and $[\underline{\sigma}_j, \overline{\sigma}_j]$ that contain the actual (unknown) values of E_j and σ_j. In such situations, we must find the ranges of the possible values for the population mean E (as described by the formula (1)) and for the population variance V (as described by the formula (2)).

Analysis of the interval problem. The formula (1) that describes the dependence of E on E_j is monotonic in E_j. Thus, we get an explicit formula for the range $[\underline{E}, \overline{E}]$ of the population mean E:

$$\underline{E} = \sum_{j-1}^{m} p_j \cdot \underline{E}_j; \quad \overline{E} = \sum_{j-1}^{m} p_j \cdot \overline{E}_j.$$

Since the terms V_E and V_σ in the expression for V depend on different variables, the range $[\underline{V}, \overline{V}]$ of the population variance V is equal to the sum of the ranges $[\underline{V}_E, \overline{V}_E]$ and $[\underline{V}_\sigma, \overline{V}_\sigma]$ of the corresponding terms:

$$\underline{V} = \underline{V}_E + \underline{V}_\sigma; \quad \overline{V} = \overline{V}_E + \overline{V}_\sigma.$$

Due to similar monotonicity, we can find an explicit expression for the range $[\underline{V}_\sigma, \overline{V}_\sigma]$ for V_σ:

$$\underline{V}_\sigma = \sum_{j=1}^{m} p_j \cdot (\underline{\sigma}_j)^2; \quad \overline{V}_\sigma = \sum_{j=1}^{m} p_j \cdot (\overline{\sigma}_j)^2.$$

Thus, to find the range of the population variance V, it is sufficient to find the range of the term V_E. So, we arrive at the following problem:

Formulation of the problem in precise terms.

GIVEN:

- an integer $m \ge 1$;
- m numbers $p_j > 0$ for which $\sum\limits_{j=1}^m p_j = 1$; and
- m intervals $\mathbf{E}_j = [\underline{E}_j, \overline{E}_j]$.

COMPUTE the range $\mathbf{V}_E = \{V_E(E_1, \ldots, E_m) \,|\, E_1 \in \mathbf{E}_1, \ldots, E_m \in \mathbf{E}_m\}$, where

$$V_E \stackrel{\text{def}}{=} \sum_{j=1}^m p_j \cdot E_j^2 - E^2; \quad E \stackrel{\text{def}}{=} \sum_{j=1}^m p_j \cdot E_j.$$

3 Main Result

Since the function V_E is convex, we can compute its minimum $\underline{V}_E$ on the box $\mathbf{E}_1 \times \ldots \times \mathbf{E}_m$ by using known polynomial-time algorithms for minimizing convex functions over interval domains; see, e.g., [8].

For computing maximum $\overline{V}_E$, even the particular case when all the values p_j are equal $p_1 = \ldots = p_m = 1/m$, is known to be NP-hard. Thus, the more general problem of computing $\overline{V}_E$ is also NP-hard. Let us show that in a reasonable class of cases, there exists a feasible algorithm for computing $\overline{V}_E$.

For each interval $\mathbf{E}_j$, let us denote its midpoint by $\widetilde{E}_j \stackrel{\text{def}}{=} \dfrac{\underline{E}_j + \overline{E}_j}{2}$, and its half-width by $\Delta_j \stackrel{\text{def}}{=} \dfrac{\overline{E}_j - \underline{E}_j}{2}$. In these terms, the j-th interval $\mathbf{E}_j$ takes the form $[\widetilde{E}_j - \Delta_j, \widetilde{E}_j + \Delta_j]$.

In this paper, we consider narrowed intervals $[E_j^-, E_j^+]$, where

$$E_j^- \stackrel{\text{def}}{=} \widetilde{E}_j - p_j \cdot \Delta_j, \quad E_j^+ \stackrel{\text{def}}{=} \widetilde{E}_j + p_j \cdot \Delta_j.$$

We show that there exists an efficient $O(m \cdot \log(m))$ algorithm for computing $\overline{V}_E$ for the case when no two narrowed intervals are proper subsets of each other, i.e., when $[E_j^-, E_j^+] \not\subseteq (E_k^-, E_k^+)$ for all j and k.

Algorithm.

- First, we sort the midpoints $\widetilde{E}_1, \ldots, \widetilde{E}_m$ into an increasing sequence. Without losing generality, we can assume that
$$\widetilde{E}_1 \le \widetilde{E}_2 \le \ldots \le \widetilde{E}_m.$$
- Then, for every k from 0 to m, we compute the value $V_E^{(k)} = M^{(k)} - (E^{(k)})^2$ of the quantity V_E for the vector $E^{(k)} = (\underline{E}_1, \ldots, \underline{E}_k, \overline{E}_{k+1}, \ldots, \overline{E}_m)$.
- Finally, we compute $\overline{V}_E$ as the largest of $m+1$ values $V_E^{(0)}, \ldots, V_E^{(m)}$.

To compute the values $V_E^{(k)}$, first, we explicitly compute $M^{(0)}$, $E^{(0)}$, and $V_E^{(0)} = M^{(0)} - E^{(0)}$. Once we computed the values $M^{(k)}$ and $E^{(k)}$, we can compute

$$M^{(k+1)} = M^{(k)} + p_{k+1} \cdot (\underline{E}_{k+1})^2 - p_{k+1} \cdot (\overline{E}_{k+1})^2$$

and

$$E^{(k+1)} = E^{(k)} + p_{k+1} \cdot \underline{E}_{k+1} - p_{k+1} \cdot \overline{E}_{k+1}.$$

4 Proof

Number of computation steps.

- It is well known that sorting requires $O(m \cdot \log(m))$ steps.
- Computing the initial values $M^{(0)}$, $E^{(0)}$, and $V_E^{(0)}$ requires linear time $O(m)$.
- For each k from 0 to $m-1$, we need a constant number $O(1)$ of steps to compute the next values $M^{(k+1)}$, $E^{(k+1)}$, and $V_E^{(k+1)}$.
- Finally, finding the largest of $m+1$ values $V_E^{(k)}$ also requires $O(m)$ steps.

Thus, overall, we need

$$O(m \cdot \log(m)) + O(m) + m \cdot O(1) + O(m) = O(m \cdot \log(m)) \text{ steps.}$$

Proof of correctness. The function V_E is convex. Thus, its maximum $\overline{V}_E$ on the box $\mathbf{E}_1 \times \ldots \times \mathbf{E}_m$ is attained at one of the vertices of this box, i.e., at a vector $(E_1, \ldots, E_m)$ in which each value E_j is equal to either $\underline{E}_j$ or to $\overline{E}_j$.

To justify our algorithm, we need to prove that this maximum is attained at one of the vectors $E^{(k)}$ in which all the lower bounds $\underline{E}_j$ precede all the upper bounds $\overline{E}_j$. We will prove this by reduction to a contradiction. Indeed, let us assume that the maximum is attained at a vector 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 followed by the lower bound; then, in the optimal vector $(E_1, \ldots, E_m)$, we have $E_i = \overline{E}_i$ and $E_{i+1} = \underline{E}_{i+1}$.

Since the maximum is attained for $E_i = \overline{E}_i$, replacing it with $\underline{E}_i = \overline{E}_i - 2\Delta_i$ will either decrease the value of V_E or keep it unchanged. Let us describe how V_E changes under this replacement. Since V_E is defined in terms of M and E, let us first describe how E and M change under this replacement. In the sum for M, we place $(\overline{E}_i)^2$ with

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

Thus, the value M changes into $M + \Delta_i M$, where

$$\Delta_i M = -4 \cdot p_i \cdot \Delta_i \cdot \overline{E}_i + 4 \cdot p_i \cdot \Delta_i^2. \tag{6}$$

The population mean E changes into $E + \Delta_i E$, where

$$\varDelta_i E = -2 \cdot p_i \cdot \varDelta_i. \tag{7}$$

Thus, the value E^2 changes into $(E + \varDelta_i E)^2 = E^2 + \varDelta_i(E^2)$, where

$$\varDelta_i(E^2) = 2 \cdot E \cdot \varDelta_i E + (\varDelta_i E)^2 = -4 \cdot p_i \cdot E \cdot \varDelta_i + 4 \cdot p_i^2 \cdot \varDelta_i^2. \tag{8}$$

So, the variance V changes into $V + \varDelta_i V$, where

$$\varDelta_i V = \varDelta_i M - \varDelta_i(E^2) = -4 \cdot p_i \cdot \varDelta_i \cdot \overline{E}_i + 4 \cdot p_i \cdot \varDelta_i^2 + 4 \cdot p_i \cdot E \cdot \varDelta_i - 4 \cdot p_i^2 \cdot \varDelta_i^2 =$$

$$4 \cdot p_i \cdot \varDelta_i \cdot (-\overline{E}_i + \varDelta_i + E - p_i \cdot \varDelta_i).$$

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

$$\varDelta_i V = 4 \cdot p_i \cdot \varDelta_i \cdot (-\widetilde{E}_i + E - p_i \cdot \varDelta_i). \tag{9}$$

So, the fact that $\varDelta_i V \le 0$ means that

$$E \le \widetilde{E}_i + p_i \cdot \varDelta_i = E_i^+. \tag{10}$$

Similarly, since the maximum of V_E is attained for $E_{i+1} = \underline{E}_{i+1}$, replacing it with $\overline{E}_{i+1} = \underline{E}_{i+1} + 2\varDelta_{i+1}$ will either decrease the value of V_E or keep it unchanged. In the sum for M, we replace $(\underline{E}_{i+1})^2$ with

$$(\overline{E}_{i+1})^2 = (\underline{E}_{i+1} + 2\varDelta_{i+1})^2 = (\underline{E}_{i+1})^2 + 4 \cdot \varDelta_{i+1} \cdot \underline{E}_{i+1} + 4 \cdot \varDelta_{i+1}^2.$$

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

$$\varDelta_{i+1} M = 4 \cdot p_{i+1} \cdot \varDelta_{i+1} \cdot \underline{E}_{i+1} + 4 \cdot p_{i+1} \cdot \varDelta_{i+1}^2. \tag{11}$$

The population mean E changes into $E + \varDelta_{i+1} E$, where

$$\varDelta_{i+1} E = 2 \cdot p_{i+1} \cdot \varDelta_{i+1}. \tag{12}$$

Thus, the value E^2 changes into $E^2 + \varDelta_{i+1}(E^2)$, where

$$\varDelta_{i+1}(E^2) = 2 \cdot E \cdot \varDelta_{i+1} E + (\varDelta_{i+1} E)^2 = 4 \cdot p_{i+1} \cdot E \cdot \varDelta_{i+1} + 4 \cdot p_{i+1}^2 \cdot \varDelta_{i+1}^2. \tag{13}$$

So, the term V_E changes into $V_E + \varDelta_{i+1} V$, where

$$\varDelta_{i+1} V = \varDelta_{i+1} M - \varDelta_{i+1}(E^2) =$$

$$4 \cdot p_{i+1} \cdot \varDelta_{i+1} \cdot \underline{E}_{i+1} + 4 \cdot p_{i+1} \cdot \varDelta_{i+1}^2 - 4 \cdot p_{i+1} \cdot E \cdot \varDelta_{i+1} - 4 \cdot p_{i+1}^2 \cdot \varDelta_{i+1}^2 =$$

$$4 \cdot p_{i+1} \cdot \varDelta_{i+1} \cdot (\underline{E}_{i+1} + \varDelta_{i+1} - E - p_{i+1} \cdot \varDelta_{i+1}).$$

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

$$\varDelta_{i+1} V = 4 \cdot p_{i+1} \cdot (\widetilde{E}_{i+1} - E - p_{i+1} \cdot \varDelta_{i+1}). \tag{14}$$

Since V_E attains maximum at $(E_1, \ldots, E_i, E_{i+1}, \ldots, E_m)$, we have $\Delta_{i+1} V \leq 0$, hence

$$E \geq \widetilde{E}_{i+1} - p_{i+1} \cdot \Delta_{i+1} = E^-_{i+1}. \tag{15}$$

We can also change both E_i and E_{i+1} at the same time. In this case, from the fact that V_E attains maximum, we conclude that $\Delta V_E \leq 0$.

Here, the change ΔM in M is simply the sum of the changes coming from E_i and E_{i+1}:

$$\Delta M = \Delta_i M + \Delta_{i+1} M, \tag{16}$$

and the change in E is also the sum of the corresponding changes:

$$\Delta E = \Delta_i E + \Delta_{i+1} E. \tag{17}$$

So, for

$$\Delta V = \Delta M - \Delta(E^2),$$

we get

$$\Delta V = \Delta_i M + \Delta_{i+1} M - 2 \cdot E \cdot \Delta_i E - 2 \cdot E \cdot \Delta_{i+1} E - (\Delta_i E)^2 - (\Delta_{i+1} E)^2 -$$

$$2 \cdot \Delta_i E \cdot \Delta_{i+1} E.$$

Hence,

$$\Delta V = (\Delta_i M - 2 \cdot E_i \cdot \Delta_i E - (\Delta_i E)^2) + (\Delta_{i+1} M - 2 \cdot E_{i+1} \cdot \Delta_{i+1} E - (\Delta_{i+1} E)^2) -$$

$$2 \cdot \Delta E_i \cdot \Delta E_{i+1},$$

i.e.,

$$\Delta V = \Delta_i V + \Delta_{i+1} V - 2 \cdot \Delta_i E \cdot \Delta_{i+1} E. \tag{18}$$

We already have expressions for $\Delta_i V$, $\Delta_{i+1} V$, $\Delta_i E$, and $\Delta_{i+1} E$, and we already know that $E^-_{i+1} \leq E \leq D^+_i$. Thus, we have $D(E) \leq 0$ for some $E \in [E^-_{i+1}, E^+_i]$, where

$$D(E) \stackrel{\text{def}}{=} 4 \cdot p_i \cdot \Delta_i \cdot (-E^+_i + E) + 4 \cdot p_{i+1} \cdot \Delta_{i+1} \cdot (E^- i + 1 - E) + 8 \cdot p_i \cdot \Delta_i \cdot p_{i+1} \cdot \Delta_{i+1}.$$

Since the narrowed intervals are not subsets of each other, we can sort them in lexicographic order; in which order, midpoints are sorted, left endpoints are sorted, and right endpoints are sorted, hence $E^-_i \leq E^-_{i+1}$ and $E^+_i \leq E^+_{i+1}$.

For $E = E^-_{i+1}$, we get

$$D(E^-_{i+1}) = 4 \cdot p_i \cdot \Delta_i \cdot (-E^+_i + E^-_{i+1}) + 8 \cdot p_i \cdot \Delta_i \cdot p_{i+1} \cdot \Delta_{i+1} =$$

$$4 \cdot p_i \cdot \Delta_i \cdot (-E^+_i + E^-_{i+1} + 2 \cdot p_{i+1} \cdot \Delta_{i+1}).$$

By definition, $E^-_{i+1} = E_{i+1} - p_{i+1} \cdot \Delta_{i+1}$, hence $E^-_{i+1} + 2 \cdot p_{i+1} \cdot \Delta_{i+1} = E^+_{i+1}$, and

$$D(E^-_{i+1}) = 4 \cdot p_i \cdot \Delta_i \cdot (E^+_{i+1} - E^+_i) \geq 0.$$

Similarly,

$$D(E_i^+) = 4 \cdot p_{i+1} \cdot \Delta_{i+1} \cdot (E_{i+1}^- - E_i^+) \geq 0.$$

The only possibility for both values to be 0 is when interval coincide; in this case, we can easily swap them. In all other cases, all intermediate values $D(E)$ are positive, which contradicts to our conclusion that $D(E) \leq 0$. The statement is proven.

Acknowledgments. This work was supported in part by NSF grants EAR-0225670 and DMS-0532645 and by the Texas Department of Transportation grant No. 0-5453.

References

1. E. Dantsin, V. Kreinovich, A. Wolpert, and G. Xiang, "Population Variance under Interval Uncertainty: A New Algorithm", *Reliable Computing*, 2006, Vol. 12, No. 4, pp. 273–280.
2. 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.
3. G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications.* Prentice Hall, Upper Saddle River, New Jersey, 1995.
4. 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*, 2007, Vol. 199, No. 2, pp. 418–423.
5. 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*, 2006, Vol. 12, No. 6, pp. 471–501.
6. H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic*, CRC Press, Boca Raton, Florida, 2005.
7. S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer-Verlag, New York, 2005.
8. S. A. Vavasis, *Nonlinear Optimization: Complexity Issues*, Oxford University Press, New York, 1991.