Exact Upper Bound on the Mean of the Product of Many Random Variables With Known Expectations

Vladik Kreinovich[1], Scott Ferson[2], and Lev Ginzburg[2]

[1]Computer Science Department, University of Texas at El Paso
El Paso, TX 79968, USA, vladik@cs.utep.edu
[2]Applied Biomathematics, 100 North Country Road,
Setauket, NY 11733, USA, {scott,lev}@ramas.com

Abstract

In practice, in addition to the intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ of possible values of inputs $x_1, \ldots, x_n$, we sometimes also know their means E_i. For such cases, we provide an explicit exact (= best possible) upper bound for the mean of the product $x_1 \cdot \ldots \cdot x_n$ of positive values x_i.

1 Formulation of the Problem

Case study: practical problem from ecology. In many ecological applications (see, e.g., [2] and references therein), we have some information about the (positive) parameters $x_1, \ldots, x_n$, and we are interested in the product $y = x_1 \cdot \ldots \cdot x_n$. For example, pollutant often comes from the industrial source to, say, a lake, via a chain of transitions, so the resulting concentration can be estimated as $x_1 \cdot x_2 \cdot \ldots \cdot x_n$, where x_1 is the original pollutant amount and the parameters x_i ($i \geq 2$) describe what portion of the pollutant goes from one link to the next one. For example, x_2 may describe the portion of the pollutant that seeps into the soil, x_3 the portion of the soil pollutant that goes from the soil into the creeks, and x_4 describes the portion of the creek's pollutant that stays in the lake. For each of these parameters, we usually know the interval $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ of possible values.

In addition to the intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ of possible values of x_i, we often know the mean values E_i. Our goal is then to find the interval of possible values of the product y, and the bounds on the *mean* of this product.

Since in ecological problems, we are mainly interested in the worst-case estimates, so we mainly interested in the upper bound $\overline{y}$ for the interval $\mathbf{y}$ and in

the upper bound $\overline{E}$ for the mean E.

Particular case when we only know intervals. If for each variable x_i, the only information we have is an interval $[\underline{x}_i, \overline{x}_i]$ of possible values, then the only thing that we can conclude about the product y is that it belongs to the interval $[\underline{y}, \overline{y}]$, where $\underline{y} = \underline{x}_1 \cdot \ldots \cdot \underline{x}_n$ and $\overline{y} = \overline{x}_1 \cdot \ldots \cdot \overline{x}_n$.

This problem is a simple particular case of interval computations [4, 5]; more precisely, it is a particular application of interval computations to *indirect measurement*, when we are interested in the value of some quantity y that is difficult (or even impossible) to measure directly). To estimate y, we therefore measure the values of several easier-to-measure quantities $x_1, \ldots, x_n$, and then use the known relation $y = f(x_1, \ldots, x_n)$ between x_i and y to reconstruct the value y as $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$, where $\widetilde{x}_i$ is the result of measuring x_i.

In many real-life situations, the only information that we have about the measurement error $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$ is that this error cannot exceed a known bound Δ_i, i.e., that $|\Delta x_i| \leq \Delta_i$. In such situations, after measuring x_i, the only information that we get about the actual (unknown) value of x_i is that this value belongs to the interval $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i] = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$. In this case, we are interested in the interval $\mathbf{y}$ of possible value of y, i.e., in the range of the function $f(x_1, \ldots, x_n)$ over the corresponding box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$.

Interval computations provide the exact range for the case when $f(x_1, \ldots, x_n)$ is a simple arithmetic operation, and provide an enclosure for the general case.

What if we also know expectations E_i of variables x_i: what is known. In some practical situations, in addition to the upper bound on the measurement error Δx_i, we have partial information about the probabilities of different values within this interval. A very typical case is when we know the mean value of this error. Thus, in addition to knowing the interval of possible values $\mathbf{x}_i$ for x_i, we know the mathematical expectation E_i for x_i [6]. In such situations, in addition to the interval of possible values of $y = f(x_1, \ldots, x_n)$, we want to know the range of possible values of the mathematical expectation E of y.

In [3], we have shown how to compute the exact range of E for the case when $f(x_1, \ldots, x_n)$ is a simple arithmetic operation – i.e., when $n = 2$ and $f(x_1, x_2)$ is equal either to the sum $x_1 + x_2$, or to the difference $x_1 - x_2$, or to the product $x_1 \cdot x_2$, etc. – and provide an enclosure for the general case.

How can we apply the known results to our ecological problem. In the above ecological problem, in addition to the intervals $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ of possible values of x_i, we also often know the mean value E_i. Our goal is then to find not only the upper bound on the interval of possible values of the product y, but also the upper bound $\overline{E}$ on the *mean* E of this product.

As we have mentioned, computing $\overline{y}$ is easy: since all the values x_i are positive, we have $\overline{y} = \overline{x}_1 \cdot \ldots \cdot \overline{x}_n$. When x_i are independent, computing $\overline{E}$ is also easy: in this case, $\overline{E} = E = E_1 \cdot \ldots \cdot E_n$.

The situation becomes less trivial in the general case when we cannot assume independence, and we therefore have to consider all possible distributions on the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$. For this case, in principle, we can use algorithms presented in [2] for a product of two variables, and, by applying this algorithm $n-1$ times, get estimates for $x_1 \cdot x_2$, $(x_1 \cdot x_2) \cdot x_3$, ..., and finally, for $y = x_1 \cdot \ldots \cdot x_n$.

For our practical problem, it is desirable to have an explicit expression for $\overline{E}$. The problem with the approach that we have just described is that the resulting algorithmic estimate cannot be easily described in an explicit form and therefore, it is difficult to analyze – and the analysis of possible changes is one of the main objectives of ecological research. It is therefore desirable to produce an explicit easy-to-analyze expression for $\overline{E}$. Such an expression is provided in this paper.

2 Main Result

In formal terms, in this paper, we solve the following problem:

GIVEN: positive values $\underline{x}_1$, $\overline{x}_1$, ..., $\underline{x}_n$, $\overline{x}_n$, E_1, ..., E_n,

FIND: the value

$$\overline{E} \stackrel{\text{def}}{=} \max\{E(x_1 \cdot \ldots \cdot x_n) \mid \text{ all distributions of } (x_1, \ldots, x_n) \text{ for which}$$

$$x_1 \in [\underline{x}_1, \overline{x}_1], \ldots, x_n \in [\underline{x}_n, \overline{x}_n], E[x_1] = E_1, \ldots, E[x_n] = E_n\}.$$

To describe the value $\overline{E}$, we first compute the values $p_i \stackrel{\text{def}}{=} (E_i - \underline{x}_i)/(\overline{x}_i - \underline{x}_i)$ and then order the variables in the decreasing order of p_i. Without losing generality, we can assume that the variables $x_1, \ldots, x_n$ are already ordered in this way, i.e., that $p_1 \geq p_2 \geq \ldots \geq p_n$. Then:

$$\begin{aligned}
\overline{E} = {} & (1 - p_1) \cdot \underline{x}_1 \cdot \underline{x}_2 \ldots \cdot \underline{x}_n + \\
& (p_1 - p_2) \cdot \overline{x}_1 \cdot \underline{x}_2 \cdot \ldots \cdot \underline{x}_n + \\
& \ldots + \\
& (p_i - p_{i+1}) \cdot \overline{x}_1 \cdot \ldots \cdot \overline{x}_i \cdot \underline{x}_{i+1} \cdot \ldots \cdot \underline{x}_n + \\
& \ldots + \\
& p_n \cdot \overline{x}_1 \cdot \ldots \cdot \overline{x}_n.
\end{aligned} \tag{1}$$

The proof of this result is given in the Proofs section. Before we present this proof, we will analyze the computational complexity of this algorithm, describe the intuitive meaning of the above formula, and present a (toy) numerical example (that will be easy to trace by hand).

3 Algorithm for Computing $\overline{E}$: Description and Computational Complexity

At first glance, the formula (1) provides a straightforward algorithm for computing $\overline{E}$. Indeed, according to this formula, $\overline{E}$ is the sum of $n+1$ products; so, to compute $\overline{E}$, we can simply compute all these products, and then add them up. Of course, first, we compute the values p_i (this requires $O(n)$ steps), and then sort the variables in the decreasing order of p_i. It is well known that we can sort a list of n elements in $O(n \cdot \log(n))$ steps; see, e.g., [1]. Once the sorting is done, to compute each product, we need one subtraction and n multiplications – i.e., $n+1$ arithmetic operations. Thus, totally, we need $(n+1)^2$ operations to compute all the products – and n additions to add them up, to the total of $O(n^2)$ operations.

This is reasonable when n is small, but when we have many factors, the quadratic time algorithm may become too long. It turns out to be possible to compute $\overline{E}$ much faster if we represent the formula (1) in the following equivalent form:

$$\overline{E} = (1-p_1) \cdot \Pi_0 + (p_1 - p_2) \cdot \Pi_1 + \ldots + (p_i - p_{i+1}) \cdot \Pi_i + \ldots + p_n \cdot \Pi_n, \quad (2)$$

where we denoted

$$\Pi_i \stackrel{\text{def}}{=} \overline{x}_1 \cdot \ldots \cdot \overline{x}_i \cdot \underline{x}_{i+1} \cdot \ldots \cdot \underline{x}_n. \quad (3)$$

The advantage of this representation is that we do not need to use all n multiplications to compute each product Π_i: once we know Π_i, we can compute Π_{i-1} as

$$\Pi_{i-1} = \Pi_i \cdot \frac{\underline{x}_i}{\overline{x}_i}. \quad (4)$$

Thus, we can compute $\overline{E}$ by using the following algorithm:

- First, we sort all the variables x_i in the decreasing order of x_i; this requires $O(n \cdot \log(n))$ steps. After the sorting, we have $p_1 \geq p_2 \ldots \geq p_n$
- Second, we compute $\Pi_n \stackrel{\text{def}}{=} \overline{x}_1 \cdot \ldots \cdot \overline{x}_n$; this computation requires $n-1 = O(n)$ steps.
- After that, we consequently compute Π_{n-1}, Π_{n-2}, $\ldots$, Π_0 by using the formula (4). Computing each new value of Π_i requires 2 arithmetic operations, so we have a total of $2n = O(n)$ operations.
- Finally, we compute $\overline{E}$ by using the formula (2): this computation requires n subtractions, $n+1$ multiplications, and n additions, i.e., totally, $O(n)$ operations.

Overall, this algorithm requires $O(n \cdot \log(n)) + O(n) = O(n \cdot \log(n))$ operations – which, for large n, is much smaller than n^2.

4 Intuitive Meaning of the Above Formula

The probability p_i can be interpreted as follows: if we only allow values $\underline{x}_i$ and $\overline{x}_i$, then there is only one probability distribution on x_i for which the average is exactly E_i. In this probability distribution, the probability $p[\overline{x}_i]$ of $\overline{x}_i$ is equal to p_i, and the probability $p[\underline{x}_i]$ of $\underline{x}_i$ is equal to $1 - p_i$.

In general, when we have two events A and B with known probabilities $p(A)$ and $p(B)$, then the probability of $A \,\&\, B$ can take any value from the interval $[p(A) \mathbin{\underline{\&}} p(B), p(A) \mathbin{\overline{\&}} p(B)]$, where $a \mathbin{\underline{\&}} b \stackrel{\text{def}}{=} \max(a+b-1, 0)$ and $a \mathbin{\overline{\&}} b \stackrel{\text{def}}{=} \min(a, b)$ (see, e.g., [7]).

Let us explain where these formulas come from. Let us first show that $a \mathbin{\overline{\&}} b = \min(a, b)$ is indeed the largest possible value of $p(A \,\&\, B)$ for all possible pairs of random events A and B for which $p(A) = a$ and $p(B) = b$. Indeed:

- Since $p(A \,\&\, B) \le p(A) = a$ and $p(A \,\&\, B) \le p(B) = b$, we can conclude that $p(A \,\&\, B) \le \min(a, b)$, i.e., that the probability $p(A \,\&\, B)$ cannot exceed $\min(a, b)$.
- So, to complete the proof, it is sufficient to show that there exist events A and B for which $p(A) = a$, $p(B) = b$, and $p(A \,\&\, B) = \min(a, b)$. To produce such events, let us consider a random variable ξ that is uniformly distributed on the interval $[0, 1]$; for this random variable, we can define A and B as follows:
  - A is true if $\xi \in [0, a]$, and
  - B to be true if $\xi \in [0, b]$.

  In this case, $A \,\&\, B$ means that ξ belongs to both sets, i.e., that $\xi \in [0, a] \cap [0, b] = [0, \min(a, b)]$. By definition of a uniform distribution, here, $p(A) = a$, $p(B) = b$, and $p(A \,\&\, B) = \min(a, b)$ – so $\min(a, b)$ is indeed possible.

Similarly, one can show that $a \mathbin{\underline{\&}} b = \max(a + b - 1, 0)$ is the smallest possible value of $p(A \,\&\, B)$ for all possible pairs of random events A and B for which $p(A) = a$ and $p(B) = b$. Indeed:

- It is known that for every two events A and B, we have $p(A \,\&\, B) = p(A) + p(B) - p(A \vee B)$. Since $p(A) = a$, $p(B) = b$, and $p(A \vee B) \le 1$, we can conclude that $p(A \,\&\, B) \ge a + b - 1$. Since $p(A \,\&\, B) \ge 0$, we can therefore conclude that $p(A \,\&\, B) \ge \max(a + b - 1, 0)$, i.e., that the probability $p(A \,\&\, B)$ cannot be smaller than $\max(a + b - 1, 0)$.
- To complete the proof, it is sufficient to show that there exist events A and B for which $p(A) = a$, $p(B) = b$, and $p(A \,\&\, B) = \max(a+b-1, 0)$. To produce such events, let us consider the same random variable ξ uniformly distributed on the interval $[0, 1]$ as we considered for $\overline{\&}$; we can define A and B as follows:

- A is the same as before: A is true if $\xi \in [0, a]$;
- b is defined differently: B is true if $\xi \in [1-b, 1]$.

In this case, $p(A) = a$, $p(B) = b$, and $A \,\&\, B$ means that ξ belongs to both sets, i.e., that $\xi \in [0, a] \cap [1-b, 1]$. The probability $p(A \,\&\, B)$ depends on whether these two intervals intersect:

- When $a < 1-b$, e.g., when $a+b<1$, the intersection is empty hence $p(A \,\&\, B) = 0$.
- When $a \geq 1-b$, the intersection is equal to the interval $[1-b, a]$ of width $a-(1-b) = a+b-1$, so the probability $p(A \,\&\, B)$ is equal to $a+b-1$.

In both cases, $p(A \,\&\, B) = \max(a+b-1, 0)$ – so the value $\max(a+b-1, 0)$ is indeed a possible value of $p(A \,\&\, B)$.

Thus, we can introduce a natural notation $\neg p \stackrel{\text{def}}{=} 1-p$ and rewrite the above formula as follows:

$$\overline{E} = \sum_{I \subseteq N} E_I,$$

where, for $I = \{i_1, \ldots, i_k\}$ and $N - I = \{j_1, \ldots, j_l\}$, we denoted:

$$E_I \stackrel{\text{def}}{=} (p_{i_1} \mathbin{\overline{\&}} \ldots \mathbin{\overline{\&}} p_{i_k}) \mathbin{\underline{\&}} (\neg p_{j_1} \mathbin{\overline{\&}} \ldots \mathbin{\overline{\&}} \neg p_{j_l}) \cdot \overline{x}_{i_1} \cdot \ldots \cdot \overline{x}_{i_k} \cdot \underline{x}_{j_1} \cdot \ldots \cdot \underline{x}_{j_l}.$$

Indeed, we have

$$p_{i_1} \mathbin{\overline{\&}} \ldots \mathbin{\overline{\&}} p_{i_k} = \min(p_{i_1}, \ldots, p_{i_k}),$$

$$\neg p_{j_1} \mathbin{\overline{\&}} \ldots \mathbin{\overline{\&}} \neg p_{j_l} = \min(1-p_{j_1}, \ldots, 1-p_{j_l}) = 1 - \max(p_{j_1}, \ldots, p_{j_l}),$$

and therefore, a p_i-dependent factor in E_I can be rewritten as

$$\max(\min(p_{i_1}, \ldots, p_{i_k}) - \max(p_{j_1}, \ldots, p_{j_l}), 0).$$

The only possibility for the corresponding difference to be ≥ 0 is when each value p_{i_m} is larger than each value p_{j_q} – in other words, when all the values $p_{i_1}, \ldots, p_{i_k}$ precede all the values $p_{j_1}, \ldots, p_{j_l}$ in the decreasing order of p_i.

5 Numerical (Toy) Example

To illustrate our algorithm, let us consider the following simple example. Suppose that we have 3 variables x_1, x_2, and x_3. We know:

- that $x_1 \in [1, 3]$, with the mean $E_1 = 2$;
- that $x_2 \in [1, 6]$, with the mean $E_2 = 3$; and

- that $x_3 \in [10, 2]$, with the mean $E_3 = 17$.

In this case, the range of possible values of the product $y = x_1 \cdot x_2 \cdot x_3$ is

$$[1, 3] \cdot [1, 6] \cdot [10, 20] = [1 \cdot 1 \cdot 10, 3 \cdot 6 \cdot 20] = [10, 360].$$

What is the largest possible value $\overline{E}$ of the mean of y? According to our algorithm, first, we compute the values p_i. Here,

$$p_1 = \frac{E_1 - \underline{x}_1}{\overline{x}_1 - \underline{x}_1} = \frac{2 - 1}{3 - 1} = 0.5.$$

Similarly, we compute $p_2 = (3-1)/(6-1) = 0.4$ and $p_3 = (17-10)/(20-10) = 0.7$.

Then, we sort the variables x_i in the decreasing order of p_i. Here, $0.7 \geq 0.5 \geq 0.4$, so the original variable x_3 is now the first variable, the original x_1 is now the second one, and the original x_2 is now the third one. If we list the variables in this new order, then these variables are:

- x_1 with range $[10, 20]$, with $E_1 = 17$ and $p_1 = 0.7$;
- x_2 with range $[1, 3]$, with $E_2 = 2$ and $p_1 = 0.5$;
- x_3 with range $[1, 6]$, with $E_3 = 4$ and $p_1 = 0.4$.

For these variables, we compute

$$\Pi_n = \Pi_3 = \overline{x}_1 \cdot \overline{x}_2 \cdot \overline{x}_3 = 20 \cdot 3 \cdot 6 = 360.$$

After that, we consequently compute

$$\Pi_2 = \Pi_3 \cdot \frac{\underline{x}_3}{\overline{x}_3} = 360 \cdot \frac{1}{6} = 60;$$

$$\Pi_1 = \Pi_2 \cdot \frac{\underline{x}_2}{\overline{x}_2} = 60 \cdot \frac{1}{3} = 20;$$

$$\Pi_0 = \Pi_1 \cdot \frac{\underline{x}_1}{\overline{x}_1} = 20 \cdot \frac{10}{20} = 10.$$

Finally, we compute

$$\overline{E} = (1 - p_1) \cdot \Pi_0 + (p_1 - p_2) \cdot \Pi_1 + p_2 \cdot \Pi_2 =$$

$$(1 - 0.7) \cdot 10 + (0.7 - 0.5) \cdot 20 + (0.5 - 0.4) \cdot 60 + 0.4 \cdot 360 = 3 + 4 + 6 + 144 = 157.$$

As we will see from the proof, this value is attained for the following joint probability distribution on the set of all possible vectors (x_1, x_2, x_3):

- with probability 0.3, we have $x_1 = 10$, $x_2 = 1$, and $x_3 = 1$;

- with probability 0.2, we have $x_1 = 20$, $x_2 = 1$, and $x_3 = 1$;
- with probability 0.1, we have $x_1 = 20$, $x_2 = 3$, and $x_3 = 1$;
- with probability 0.4, we have $x_1 = 20$, $x_2 = 3$, and $x_3 = 6$.

One can easily check that for this distribution, $E[x_1] = 17$, $E[x_2] = 2$, $E[x_3] = 3$, and $E[x_1 \cdot x_2 \cdot x_3] = 157$.

6 Proof of the Main Result

1°. To get the desired bound for $\overline{E}$, we must consider the values $E[x_1 \cdot \ldots \cdot x_n]$ for all possible probability distributions on the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$ for which $E[x_1] = E_1$, ..., $E[x_n] = E_n$. To describe a general probability distribution, we must use infinitely many parameters, and hence, this problem is difficult to solve directly.

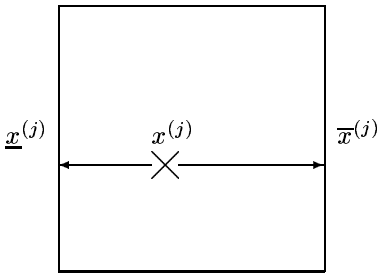
To make the problem simpler, we will show that a general distribution with $E[x_i] = E_i$ can be simplified without changing the values $E[x_i]$ and $E[x_1 \cdot \ldots \cdot x_n]$. Thus, to describe possible values of $E[x_1 \cdot \ldots \cdot x_n]$, we do not need to consider all possible distributions, it is sufficient to consider only the simplified ones.

We will describe the simplification for discrete distributions that concentrate on finitely many points $x^{(j)} = (x_1^{(j)}, \ldots, x_n^{(j)})$, $1 \le j \le N$. An arbitrary probability distribution can be approximated by such distributions, so we do not lose anything by this restriction.

So, we have a probability distribution in which the point $x^{(1)}$ appears with the probability $p^{(1)}$, the point $x^{(2)}$ appears with the probability $p^{(2)}$, etc. Let us modify this distribution as follows: pick a point $x^{(j)} = (x_1^{(j)}, x_2^{(j)}, \ldots)$ that occurs with probability $p^{(j)}$, and replace it with two points: $\overline{x}^{(j)} = (\overline{x}_1, x_2^{(j)}, \ldots)$ with probability $p^{(j)} \cdot \overline{p}^{(j)}$ and $\underline{x}^{(j)} = (\underline{x}_1, x_2^{(j)}, \ldots)$ with probability $p^{(j)} \cdot \underline{p}^{(j)}$, where

$$\overline{p}^{(j)} \stackrel{\text{def}}{=} \frac{x_1^{(j)} - \underline{x}_1}{\overline{x}_1 - \underline{x}_1}$$

and $\underline{p}^{(j)} \stackrel{\text{def}}{=} 1 - \overline{p}^{(j)}$:



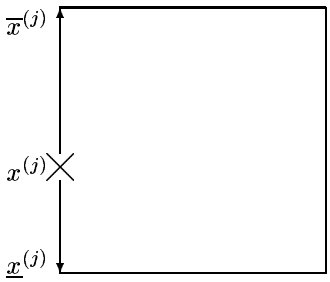
Here, the values $\overline{p}^{(j)}$ and $\underline{p}^{(j)} = 1 - \overline{p}^{(j)}$ are chosen in such a way that $\overline{p}^{(j)} \cdot \overline{x}_1 + \underline{p}^{(j)} \cdot \underline{x}_1 = x_1^{(j)}$. Due to this choice,

$$p^{(j)} \cdot \overline{p}^{(j)} \cdot \overline{x}_1 + p^{(j)} \cdot \underline{p}^{(j)} \cdot \underline{x}_1 = p^{(j)} \cdot x_1^{(j)},$$

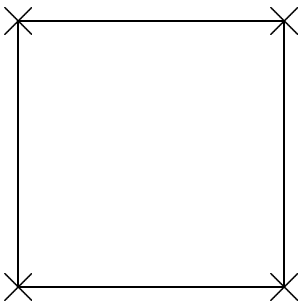
hence for the new distribution, the mathematical expectation $E[x_1]$ is the same as for the old one. Similarly, we can prove that the values $E[x_2]$, ..., $E[x_n]$, and $E[x_1 \cdot \ldots \cdot x_n]$ do not change.

We started with a general discrete distribution with N points for each of which $x_1^{(j)}$ could be inside the interval $\mathbf{x}_1$, and we have a new distribution for which $\leq N - 1$ points have the value x_1 inside this interval. We can perform a similar replacement for all N points and get a distribution with the same values of $E[x_1]$, ..., $E[x_n]$, and $E[x_1 \cdot \ldots \cdot x_n]$ as the original one but for which, for every point, x_1 is equal either to $\underline{x}_1$, or to $\overline{x}_1$.

For the new distribution, we can perform a similar transformation relative to x_1 and end up – without changing the values x_1 – with the distribution for which always either $x_2 = \underline{x}_1$ or $x_2 = \overline{x}_2$:



Similarly, we can perform such a transformation for x_3, etc. Thus, instead of considering all possible distributions, it is sufficient to consider only distributions for which $x_1 \in \{\underline{x}_1, \overline{x}_1\}$, ..., $x_n \in \{\underline{x}_n, \overline{x}_n\}$. In other words, it is sufficient to consider only distributions which are located in 2^n corner points of the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$:



2°. Let us now show that, if we are looking for the maximum $\overline{E}$ of E, it is sufficient to consider only distributions with the following property: for every two points $x^{(i)}$ and $x^{(j)}$ with non-zero probability, if $x_k^{(i)} < x_k^{(j)}$ for some coordinate k, then $x_l^{(i)} \le x_l^{(j)}$ for all other coordinates l.

We will prove this statement as follows. Let us assume that the above property is not satisfied. This means that for some k and l, we have $x_k^{(i)} < x_k^{(j)}$ and $x_l^{(i)} > x_l^{(j)}$.

Let $p^{(i)} > 0$ and $p^{(j)} > 0$ be the probabilities of these two points. We will show that, if with probability $p \stackrel{\text{def}}{=} \min(p^{(i)}, p^{(j)})$, we "swap" the coordinates of the points $x^{(i)}$ and $x^{(j)}$, we thus increase (or keep unchanged) the value E. Therefore, when we are looking for the maximum of E, it is sufficient to consider only distributions for which the above property holds.

Specifically, let $k_1, \ldots, k_q$ be coordinates for which $x_{k_m}^{(i)} \le x_{k_m}^{(j)}$, and let $l_1, \ldots, l_s$ be coordinates for which $x_{l_t}^{(i)} > x_{l_t}^{(j)}$. With probability p, we replace the points $x^{(i)}$ and $x^{(j)}$ with two new points $x_{\text{new}}^{(i)}$ and $x_{\text{new}}^{(j)}$ for which coordinates k_m remain the same while the coordinates l_t are swapped: $x_{\text{new},k_m}^{(i)} = x_{k_m}^{(i)}$, $x_{\text{new},k_m}^{(j)} = x_{k_m}^{(j)}$, $x_{\text{new},l_t}^{(i)} = x_{l_t}^{(j)}$, and $x_{\text{new},l_t}^{(j)} = x_{l_t}^{(i)}$. It is easy to see that this swap does not change the averages $E[x_i]$. How does it affect the mathematical expectation of the product $E[x_1 \cdot \ldots \cdot x_n]$? The only two terms that changed are terms corresponding to $x^{(i)}$ and $x^{(j)}$ with probability p:

- For the original points, the sum of these two terms is equal to

$$p \cdot \left(\prod_{z=1}^{n} x_z^{(i)} + \prod_{z=1}^{n} x_z^{(j)} \right) = p \cdot (\Pi_k^{(i)} \cdot \Pi_l^{(i)} + \Pi_k^{(j)} \cdot \Pi_l^{(j)}),$$

  where we denoted:

$$\Pi_k^{(i)} \stackrel{\text{def}}{=} \prod_{m=1}^{q} x_{k_m}^{(i)}, \quad \Pi_l^{(i)} \stackrel{\text{def}}{=} \prod_{t=1}^{s} x_{l_t}^{(i)},$$

$$\Pi_k^{(j)} \stackrel{\text{def}}{=} \prod_{m=1}^{q} x_{k_m}^{(j)}, \quad \Pi_l^{(j)} \stackrel{\text{def}}{=} \prod_{t=1}^{s} x_{l_t}^{(j)}.$$

- For the new points, the corresponding sum is equal to

$$p \cdot (\Pi_k^{(i)} \cdot \Pi_l^{(j)} + \Pi_k^{(j)} \cdot \Pi_l^{(i)}).$$

- Therefore, the difference between the new and the old values of $E[x_1 \cdot \ldots \cdot x_n]$ is equal to:

$$p \cdot (\Pi_k^{(i)} \cdot \Pi_l^{(j)} + \Pi_k^{(j)} \cdot \Pi_l^{(i)} - \Pi_k^{(i)} \cdot \Pi_l^{(i)} - \Pi_k^{(j)} \cdot \Pi_l^{(j)}).$$

One can easily see that this difference is equal to

$$p \cdot (\Pi_k^{(i)} - \Pi_k^{(j)}) \cdot (\Pi_l^{(j)} - \Pi_l^{(i)}).$$

By definition of k_m, we have $x_{k_m}^{(i)} \le x_{k_m}^{(j)}$; multiplying these inequalities between positive numbers, we conclude that $\Pi_k^{(i)} \le \Pi_k^{(j)}$. Similarly, from $x_{l_t}^{(i)} > x_{l_t}^{(j)}$, we conclude that $\Pi_l^{(i)} > \Pi_l^{(j)}$. Thus, the difference between the new and the old values is indeed non-negative.

The statement is proven.

3°. Due to Part 2° of the proof, for every two different points $x^{(i)} \ne x^{(j)}$:

- either $x_k^{(i)} \le x_k^{(j)}$ for all k and $x_k^{(i)} < x_k^{(j)}$ some all k; we will denote this by $x^{(i)} \prec x^{(j)}$;
- or $x_k^{(j)} \le x_k^{(i)}$ for all k and $x_k^{(j)} < x_k^{(i)}$ some all k – i.e., $x^{(j)} \prec x^{(i)}$.

So, the relation $\prec$ defines a linear (total) order on the set of all the points $x^{(i)}$. Without losing generality, let us assume that the points $x^{(i)}$ are ordered according to this order, i.e., that

$$x^{(1)} \prec x^{(2)} \prec \ldots \prec x^{(N)}.$$

By definition of $\prec$, we can conclude that for each coordinate k, we have:

$$x_k^{(1)} \le x_k^{(2)} \le \ldots \le x_k^{(N)}.$$

In Part 1° of the proof, we have already shown that for every point $x^{(i)}$, each coordinate $x_k^{(i)}$ is equal either to smallest possible value $\underline{x}_k$ or to the largest possible value $\overline{x}_k$. Due to the above inequality, once $x_k^{(i)}$ is equal to its largest possible value, i.e., once $x_k^{(i)} = \overline{x}_k$, all the following values of x_k must also be equal to the same largest possible value, i.e., $x_k^{(i+1)} = \ldots = x_k^{(N)} = \overline{x}_k$.

Therefore, when we move from $x^{(i)}$ to $x^{(i+1)}$, the overall number of coordinates equal to $\overline{x}_k$ cannot decrease; it cannot also stay the same because otherwise, we would have $x^{(i)} = x^{(i+1)}$. Thus, this number can only increase. This overall number can take values from 0 to n, and this overall number increases once we go from $x^{(i)}$ to $x^{(i+1)}$; thus, we cannot have more than n such increases, and so, we can have no more than $n + 1$ different points $x^{(i)}$.

Based on the order between the points $x^{(i)}$, we can defined the order between the coordinates x_k: namely, we say that x_k precedes x_l if in the sequence $x^{(i)}$, the first appearance of $\overline{x}_k$ precedes the first appearance of $\overline{x}_l$. One can easily see that this relation is an order. This is, in general, partial order; let us arbitrarily extend it to a linear order on the set of n coordinates $x_1, \ldots, x_n$.

For simplicity, let us assume that the variables $x_1, \ldots, x_n$ are already ordered according to this order, i.e., that $\overline{x}_1$ first appears in the sequence $x^{(i)}$ before (or

at the same time as) $\overline{x}_2$, etc. Due to this order, if for some point $x^{(i)}$, we have a "small" value of some coordinate $x_k^{(i)} = \underline{x}_k$, then all the following coordinates are also "small": $x_{k+1}^{(i)} = \underline{x}_{k+1}$, ..., $x_n^{(i)} = \underline{x}_n$. In other words, each vector $x^{(i)}$ can take one of the following values:

$$(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n), (\overline{x}_1, \underline{x}_2, \ldots, \underline{x}_n), \ldots, (\overline{x}_1, \ldots, \overline{x}_i, \underline{x}_{i+1}, \ldots, \underline{x}_n), \ldots, (\overline{x}_1, \ldots, \overline{x}_n).$$

These are exactly the vectors corresponding to the expression for $\overline{E}$ that we are proving. To complete the proof, we must therefore show that these expressions occur with probabilities, correspondingly, $1 - p_1$, $p_1 - p_2$, etc.

Indeed:

- let $p^{(1)}$ be the probability of $(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n)$;
- let $p^{(2)}$ be the probability of $(\overline{x}_1, \underline{x}_2, \ldots, \underline{x}_n)$;
- ...
- let $p^{(i+1)}$ be the probability of $(\overline{x}_1, \ldots, \overline{x}_i, \underline{x}_{i+1}, \ldots, \underline{x}_n)$;
- ...
- let $p^{(n+1)}$ be the probability of $(\overline{x}_1, \ldots, \overline{x}_n)$.

The sum of all these probabilities should be equal to 1:

$$p^{(1)} + p^{(2)} + \ldots + p^{(n+1)} = 1.$$

For each i, the mean value of x_i (that should be equal to E_i) is equal to

$$\underline{x}_i \cdot (p^{(1)} + \ldots + p^{(i)}) + \overline{x}_i \cdot (p^{(i+1)} + \ldots + p^{(n+1)}).$$

By definition, p_i is the probability with which we must take $\overline{x}_i$ so that if we take $\underline{x}_i$ with probability $1 - p_i$, we get the desired mean p_i. Thus, for every i, we have:

$$p_i = p^{(i+1)} + \ldots + p^{(n+1)}.$$

In particular, for $i < n$, we have

$$p_{i+1} = p^{(i+2)} + \ldots + p^{(n+1)};$$

thus,

$$p_i - p_{i+1} = (p^{(i+1)} + p^{(i+2)} + \ldots + p^{(n+1)}) - (p^{(i+2)} + \ldots + p^{(n+1)}) = p^{(i+1)}.$$

For $i = n$, we have $p_n = p^{(n+1)}$. Finally, the probability $p^{(1)}$ can be determined as

$$p^{(1)} = 1 - (p^{(2)} + \ldots + p^{(n+1)}) =$$

$$1 - ((p_1 - p_2) + (p_2 - p_3) + \ldots + (p_{n-1} - p_n) + p_n) = 1 - p_1.$$

For the values $x^{(i)}$ with these probabilities, the mathematical expectation of the product $x_1 \cdot \ldots \cdot x_n$ is exactly equal to the expression from our Main Result. The theorem is proven.

7 Open Problem and Future Work

In this paper, we started with a practical problem in which all the values x_i are non-negative. For this practical problem, we came up with an exact solution.

It is desirable to generalize our result to the case when some variables x_i may take negative values. Our result does not directly apply to this case, because in our proof, we used the fact that for $x_i \geq 0$, the function $f(x_1, \ldots, x_n) = x_1 \cdot \ldots \cdot x_n$ is non-decreasing in each of the variables; this property is not true for negative x_i. We hope, however, that although our result is not directly generalizable, our techniques will be. What would probably help is an example of a practical problem with $x_i < 0$; an intuition behind such an example may help.

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209 and grant NCC2-1232, by NSF grants CDA-9522207, EAR-0112968, EAR-0225670, and 9710940 Mexico/Conacyt, by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research (AFOSR), Air Force Materiel Command, USAF, under grants numbers F49620-95-1-0518 and F49620-00-1-0365, by the IEEE/ACM SC2001 and 2002 Minority Serving Institutions Participation Grants, and by Small Business Innovation Research grant 9R44CA81741 to Applied Biomathematics from the National Cancer Institute (NCI), a component of the National Institutes of Health (NIH). The opinions expressed herein are those of the author(s) and not necessarily those of NASA, NSF, AFOSR, NCI, or the NIH. This work was partly done when one of the authors (V.K.) was a Visiting Researcher at the Euler International Mathematical Institute, St. Petersburg, Russia.

The authors are thankful to all the participants of the Logical Seminar of the Mathematical Institute of the Russian Academy of Science, especially to Gennady Davydov and Yuri Matiyasevich, for useful discussions, to Daniel Berleant for his encouragement, and to anonymous referees for helpful suggestions and discussions.

References

[1] Th. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*, MIT Press, Cambridge, MA, and Mc-Graw Hill Co., N.Y., 2001.

[2] S. Ferson, *RAMAS Risk Calc 4.0: Risk Assessment with Uncertain Numbers*, CRC Press, Boca Raton, Florida, 2002.

[3] S. Ferson, L. Ginzburg, V. Kreinovich, and J. Lopez, "Absolute Bounds on the Mean of Sum, Product, etc.: A Probabilistic Extension of Interval Arithmetic", *Extended Abstracts of the 2002 SIAM Workshop on Validated Computing*, Toronto, Canada, May 23–25, 2002, pp. 70–72 (detailed paper can be found at http://www.cs.utep.edu/vladik/2002/tr02-12b.ps.gz or http://www.cs.utep.edu/vladik/2002/tr02-12b.pdf).

[4] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, London, 2001.

[5] R. B. Kearfott and V. Kreinovich (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 1996.

[6] S. Rabinovich, *Measurement Errors: Theory and Practice*, American Institute of Physics, New York, 1993.

[7] P. Walley, *Statistical reasoning with imprecise probabilities*, Chapman and Hall, N.Y., 1991.