

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

Vladik Kreinovich¹, Scott Ferson², and Lev Ginzburg²

¹Computer Science Department, University of Texas at El Paso
El Paso, TX 79968, USA, vladik@cs.utep.edu

²Applied 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, \dots, 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 $x_1 \cdot \dots \cdot x_n$ of positive values x_i .

1 Formulation of the Problem

Case study: practical problem from ecology. In many ecological applications (see, e.g., [6] and references therein), we have some information about the (positive) parameters x_1, \dots, x_n , and we are interested in the product $y = x_1 \cdot \dots \cdot x_n$. 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 $x_1 \cdot x_2 \cdot \dots \cdot x_n$, 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 .

Comment: isn't this problem too specific? At first glance, it may look like this problem is too specific:

- it is *mathematically* specific: we only consider the simplest case of data combination – a product, and we only consider the case when we know the first moments;
- it is also *application specific*: we only consider a very specific class of applications – applications to ecology.

Should not we attempt to handle a more general mathematical problem? Or, if we cannot do that, should not we publish in a specialized ecological journal – as opposed to a more general journal such as *Reliable Computing*? Here are our answers to these doubts:

- First, as we will show, while we have an easy-to-compute (even explicit) solution for our problem, even simple natural generalizations lead to hard-to-solve (NP-hard) problems.
- Second, propagation of moments is an important practically useful statistical problem. We have discovered an explicit formula for solving this problem. Although we were motivated by a specific class of problems from ecology, we want to present this formula to the general community because we strongly believe that it will turn out to be useful in other practical problems as well.

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 $\underline{y} = \underline{x}_1 \cdot \dots \cdot \underline{x}_n$ and $\overline{y} = \overline{x}_1 \cdot \dots \cdot \overline{x}_n$.

This problem is a simple particular case of interval computations [10, 11]; 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, \dots, x_n , and then use the known relation $y = f(x_1, \dots, x_n)$ between x_i and y to reconstruct the value y as $\tilde{y} = f(\tilde{x}_1, \dots, \tilde{x}_n)$, where \tilde{x}_i is the result of measuring x_i .

In many real-life situations, the only information that we have about the measurement error $\Delta x_i \stackrel{\text{def}}{=} \tilde{x}_i - x_i$ 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 $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i] = [\tilde{x}_i - \Delta_i, \tilde{x}_i + \Delta_i]$. In this case, we are interested in the interval \mathbf{y} of possible value of y , i.e., in the range of the function $f(x_1, \dots, x_n)$ over the corresponding box $\mathbf{x}_1 \times \dots \times \mathbf{x}_n$.

Interval computations provide the exact range for the case when $f(x_1, \dots, x_n)$ 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 [15]. In such situations, in addition to the interval of possible values of $y = f(x_1, \dots, x_n)$, we want to know the range of possible values of the mathematical expectation E of y .

In [7], we have shown how to compute the exact range of E for the case when $f(x_1, \dots, x_n)$ 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 known results and techniques to our ecological problem: easy case. In the above ecological problem, in addition to the intervals $\mathbf{x}_i = [\underline{x}_i, \bar{x}_i]$ 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 \bar{E} on the *mean* E of this product.

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

How can we apply known results and techniques to our ecological problem: general case. 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 $\mathbf{x}_1 \times \dots \times \mathbf{x}_n$.

From the *practical* viewpoint, what we can do is try several different distributions with $E_i \in \mathbf{E}_i$, compute E for all such distributions (or estimate it by using Monte-Carlo simulations), and take the largest of the corresponding values E as an estimate for \bar{E} . In other words, as an estimate for the maximum \bar{E} of $E = E[x_1 \cdot \dots \cdot x_n]$ over all possible distributions, we take the largest of the values corresponding to finitely many distributions. The resulting estimate does not give us a guaranteed upper bound on the actual maximum \bar{E} , and in ecological problems, we need a guaranteed upper bound for \bar{E} .

Existing methods that provide a guaranteed upper bound for \bar{E} overestimate \bar{E} . There are several techniques that enable us to give a guaranteed upper bound for \bar{E} . We will show that none of these methods lead to the desired exact value of \bar{E} .

Copula-based method. One of these methods – see, e.g., [6, 8, 16] – is based on *copula techniques* pioneered in [19]. In these methods, we transform the

available information about the (unknown) probability distribution for each x_i into bounds on its CDF (so-called *p-box*). Copula-based formulas enable us to transform the p-boxes corresponding to individual variables x_i into a p-box for the product y . Based on this p-box, we can then produce guaranteed bounds on the mean $E[y]$ of y .

The problem with this method is that the resulting upper bound for E is often an overestimation for the desired value \overline{E} . This is true even in the simplest case when $n = 2$ and one of the variables – e.g., x_2 – is equal to 1 with probability 1 (i.e., it is actually a non-random deterministic real number). In this case, once we know the mean $E_1 = E[x_1]$, we can determine the p-box $[\underline{F}(x), \overline{F}(x)]$ containing all possible CDFs for which $E[x_1] = E_1$. Since based on the first moment, we cannot uniquely determine the CDF, this p-box is non-degenerate. Due to $x_2 = 1$, this p-box is identical with the p-box for $y = x_1 \cdot x_2$. Based on the p-box for y , we can determine the interval of possible values of $E[y]$ – it is easy to see that the endpoints of this interval are the mean values corresponding to the CDFs $\underline{F}(x)$ and $\overline{F}(x)$. Since the p-box is non-degenerate, these values are different.

Thus, by using the copula techniques, we get a non-degenerate interval of possible values for $E[y]$. However, by construction, we know the mean $E[y]$ to be precisely equal to E_1 . Thus, copula-based CDF bounds, while providing an enclosure for $E[y]$, lead to excess width. In other words, for our problem, the results of the copula techniques are, although rigorous, not best possible.

Histogram-based method. Another – *histogram-based* – method was proposed by Berleant and others in [1, 2, 3, 4]. In this method, a class of possible distributions is described by listing several intervals $[\underline{x}_i, \overline{x}_i]$ on the real line and describing intervals $[\underline{p}_i, \overline{p}_i]$ of possible values of the probability p_i that the random variable belongs to the corresponding interval $[\underline{x}_i, \overline{x}_i]$. For this method (similarly to the copula method), even when we know the exact value of $E[x_1]$ and we know that $x_2 = 1$ with probability 1, we still get a non-degenerate interval of possible values of $E[y]$ for $y = x_1 \cdot x_2$.

This fact can be independently confirmed, or it can be deduced from the general result [16] that in many problems, the histogram method gives the same answers as the copula-based method.

Interval Monte-Carlo method. A third method, described in Lodwick [13], uses an *interval* analogue of *Monte-Carlo* techniques – dividing the real line into subintervals and providing guaranteed bounds for each combination of subintervals corresponding to x_1, \dots, x_n . This method enables us to effectively compute lower and upper bounds on the distribution for $y = x_1 \cdot \dots \cdot x_n$, and, as a result, lower and upper bounds on $E[y]$. Again, even when we know the exact value of $E[x_1]$ and we know that $x_2 = 1$ with probability 1, we still get a guaranteed upper bound for $E[y]$ that is larger than the exact value $E[y]$.

Analytical method. A fourth – *analytical* – method is based on a pioneering work by Rowe [17]. For simple transformations of random variables (e.g., linear

transformations) and for simple combinations of random variables (e.g., sum or, more generally, linear combinations), textbooks on mathematical statistics contain explicit formulas for the expectation of such a combination. In his recent work [17], Rowe extended these formulas to more general transformations (such as $\log(x)$) and more general combinations.

In [7], we extended Rowe's formulas to the case when $y = x_1 \cdot x_2$. According to this result of ours, once we know the values E_1 and E_2 , we can compute the largest possible value \overline{E} of $E[y] = E[x_1 \cdot x_2]$ as follows: first, we compute the values $p_i \stackrel{\text{def}}{=} (E_i - \underline{x}_i)/(\overline{x}_i - \underline{x}_i)$, and then compute

$$\begin{aligned} \overline{E} = & \min(1-p_1, 1-p_2) \cdot \underline{x}_1 \cdot \underline{x}_2 + \max(p_1-p_2, 0) \cdot \overline{x}_1 \cdot \underline{x}_2 + \max(p_2-p_1, 0) \cdot \underline{x}_1 \cdot \overline{x}_2 + \\ & \min(p_1, p_2) \cdot \overline{x}_1 \cdot \overline{x}_2. \end{aligned}$$

This formula can be somewhat simplified if we order the variables in the decreasing order of p_i . Without losing generality, we can assume that the variables x_1 and x_2 are already ordered in this way, i.e., that $p_1 \geq p_2$. Then, the above formula takes a simplified form:

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

(It turns out that this formula is a particular case $n = 2$ of our general formula – which is the main result of this paper – so this $n = 2$ formula can be deduced from our more general result.)

Since we know how to estimate the mean of the product for two variables, we can, in principle, estimate the mean of the product of n variables as follows:

- first, we use the above formula to find the bounds on $E[x_1 \cdot x_2]$;
- then, we use the above formula once again to transform the known bounds on $E[x_1 \cdot x_2]$ and $E[x_3]$ into bounds on $E[(x_1 \cdot x_2) \cdot x_3]$;
- etc.

After applying the above formula $n - 1$ times, we get bounds on

$$E[y] = E[x_1 \cdot \dots \cdot x_n].$$

These bounds are valid enclosures – in particular, they lead to a correct upper bound for $E[y]$. The problem with this method is that this upper bound is sometimes an overestimation for \overline{E} .

Let us show this on a simple example of three variables. Let $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}_3 = [1, 2]$, $E_1 = 1$, $E_2 = 2$, and $E_3 = 1.5$. In this case, the variable x_1 is equal to 1 with probability 1, the variable x_2 is equal to 2 with probability 1, thus, $y = x_1 \cdot x_2 \cdot x_3 = 2x_3$ so $E[y] = 2E[x_3] = 2 \cdot 1.5 = 3$. Thus, the exact upper bound \overline{E} on $E[y]$ is equal to 3. Let us show that the above sequential estimate leads to an overestimation.

First, we combine x_1 and x_2 . Based on the fact that E_1 is equal to the lower endpoint of the corresponding interval \mathbf{x}_1 , we conclude that x_1 is (with probability 1) equal to 1. Similarly, x_2 is (with probability 1) equal to 2. Thus, $x_1 \cdot x_2$ is equal to $1 \cdot 2 = 2$. One can check that the above formula indeed leads to the correct value $\overline{E} = 2$ (and a similar formula for \underline{E} – also described in [7] – also leads to the correct value $\underline{E} = 2$). Thus, for $z_1 \stackrel{\text{def}}{=} x_1 \cdot x_2$, we know the exact value of $E[z_1]$: it is 2.

The interval of possible values for $z_1 = x_1 \cdot x_2$ can be computed by straightforward interval computations: it is $[1, 2] \cdot [1, 2] = [1, 4]$. According to the sequential algorithm, to find the upper bound for $E[(x_1 \cdot x_2) \cdot x_3]$, we can now apply the above formula for \overline{E} to the following two factors:

- $z_1 = x_1 \cdot x_2$ with interval of possible values $[1, 4]$ and mean 2, and
- $z_2 \stackrel{\text{def}}{=} x_3$ with interval of possible values $[1, 2]$ and mean 1.5.

Here, $p_1 = 1/3$, $p_2 = 1/2$. After ordering and renaming variables in decreasing order of p_i , we get:

- $p_1 = 1/2$, $\underline{z}_1 = 1$, $\overline{z}_1 = 2$,
- $p_1 = 1/3$, $\underline{z}_2 = 1$, $\overline{z}_2 = 4$,

and the above formula leads to

$$\left(\frac{1}{2}\right) \cdot 1 \cdot 1 + \left(\frac{1}{2} - \frac{1}{3}\right) \cdot 2 \cdot 1 + \frac{1}{3} \cdot 2 \cdot 4 = \frac{1}{2} + \frac{1}{3} + \frac{8}{3} = 3\frac{1}{2} > 3,$$

i.e., to an overestimation for \overline{E} .

For our practical problem, it is desirable to have an exact upper bound for \overline{E} . To make the best decision – and not to be over-cautious – we must use the most accurate (ideally, exact) estimates for \overline{E} . Since none of the existing method provides an exact upper bound \overline{E} , a new method is needed. Such a method is described in this paper.

2 Main Result

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

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

FIND: the value

$$\overline{E} \stackrel{\text{def}}{=} \max\{E[x_1 \cdot \dots \cdot x_n] \mid \text{all distributions of } (x_1, \dots, x_n) \text{ for which } x_1 \in [\underline{x}_1, \overline{x}_1], \dots, x_n \in [\underline{x}_n, \overline{x}_n], E[x_1] = E_1, \dots, E[x_n] = E_n\}.$$

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, \dots, x_n are already ordered in this way, i.e., that $p_1 \geq p_2 \geq \dots \geq p_n$. Then:

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

Comment. This formula can be represented in an alternative form. Namely, if we group together terms containing p_1 , p_2 , etc., in this formula, we end up with the following formula:

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

By definition of p_i , we have $p_i \cdot (\overline{x}_i - \underline{x}_i) = E_i - \underline{x}_i$, therefore (1a) can be rewritten as:

$$\begin{aligned}
\overline{E} = & \underline{x}_1 \cdot \underline{x}_2 \dots \underline{x}_n + \\
& (E_1 - \underline{x}_1) \cdot \underline{x}_2 \cdot \dots \underline{x}_n + \\
& \dots + \\
& \overline{x}_1 \cdot \dots \cdot \overline{x}_{i-1} \cdot (E_i - \underline{x}_i) \cdot \underline{x}_{i+1} \cdot \dots \underline{x}_n + \\
& \dots + \\
& \overline{x}_1 \cdot \dots \cdot \overline{x}_{n-1} \cdot (E_n - \underline{x}_n).
\end{aligned} \tag{1b}$$

The proof of the formula (1) 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., [5]. 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:

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

where we denoted

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

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 $p_1 \geq p_2 \dots \geq p_n$
- Second, we compute $\Pi_n \stackrel{\text{def}}{=} \overline{x}_1 \cdot \dots \cdot \overline{x}_n$; this computation requires $n - 1 = O(n)$ steps.
- After that, we consequently compute $\Pi_{n-1}, \Pi_{n-2}, \dots, \Pi_0$ 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 $[p(A) \underline{\&} p(B), p(A) \overline{\&} p(B)]$, where $a \underline{\&} b \stackrel{\text{def}}{=} \max(a + b - 1, 0)$ and $a \overline{\&} b \stackrel{\text{def}}{=} \min(a, b)$ (see, e.g., [18]).

Let us explain where these formulas come from. Let us first show that $a \overline{\&} b = \min(a, b)$ 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 $p(A \& B) \leq p(A) = a$ and $p(A \& B) \leq p(B) = b$, we can conclude that $p(A \& B) \leq \min(a, b)$, 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 $a \underline{\&} b = \max(a + b - 1, 0)$ 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 $p(A \vee B) \leq 1$, we can conclude that $p(A \& B) \geq a + b - 1$. Since $p(A \& B) \geq 0$, we can therefore conclude that $p(A \& B) \geq \max(a + b - 1, 0)$, 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 $I = \{i_1, \dots, i_k\}$ and $N - I = \{j_1, \dots, j_l\}$, we denoted:

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

Indeed, we have

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

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

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

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

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}, \dots, p_{i_k} precede all the values p_{j_1}, \dots, 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 $x_3 \in [10, 20]$, 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 $E[x_1 \cdot x_2 \cdot x_3] = 157$.

Let us try one more distribution to make sure that in this example, 157 is indeed the upper endpoint of the interval of possible values of $y = x_1 \cdot x_2 \cdot x_3$: the distribution in which all three variables x_i are statistically independent. For this distribution, we have $E = E[x_1 \cdot x_2 \cdot x_3] = E[x_1] \cdot E[x_2] \cdot E[x_3] = 2 \cdot 3 \cdot 17 = 102$, so indeed $E = 102 < 157 = \overline{E}$.

6 Proof of the Main Result

1°. To get the desired bound for \overline{E} , we must consider the values $E[x_1 \cdot \dots \cdot x_n]$ for all possible probability distributions on the box $\mathbf{x}_1 \times \dots \times \mathbf{x}_n$ for which $E[x_1] = E_1, \dots, E[x_n] = E_n$. The set of all such probability distributions is infinite-dimensional: indeed, in order to describe a probability distribution completely via moments, of which we have only one known, an infinite number of moments are required. Thus, the optimization problem that we have to solve the problem of maximizing an objective function E over an infinite-dimensional set. By definition of dimension, this means that we need infinitely many parameters to describe individual distributions from this set. It is known that in optimization problems, the larger the dimension of the set over which we optimize – i.e., the more parameters we need to describe the individual elements of this set – the more difficult the problem. Therefore, our problem – with infinite dimension – 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 $E[x_1 \cdot \dots \cdot x_n]$. Thus, to describe possible values of $E[x_1 \cdot \dots \cdot x_n]$, we do not need to consider all possible distributions, it is sufficient to consider only the simplified ones.

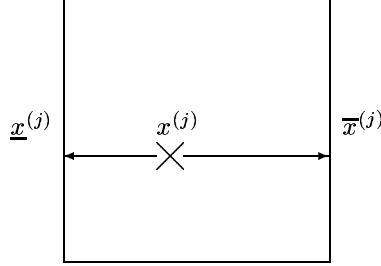
We will describe the simplification for discrete distributions that concentrate on finitely many points $x^{(j)} = (x_1^{(j)}, \dots, x_n^{(j)})$, $1 \leq j \leq N$. An arbitrary probability distribution can be approximated by such distributions, so we do not lose anything by this restriction.

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 $x^{(j)} = (x_1^{(j)}, x_2^{(j)}, \dots)$ that occurs with probability $p^{(j)}$, and replace it with two points: $\overline{x}^{(j)} = (\overline{x}_1, x_2^{(j)}, \dots)$ with probability $p^{(j)} \cdot \overline{p}^{(j)}$ and $\underline{x}^{(j)} = (\underline{x}_1, x_2^{(j)}, \dots)$ 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 $\underline{p}^{(j)} \stackrel{\text{def}}{=} 1 - \overline{p}^{(j)}$:



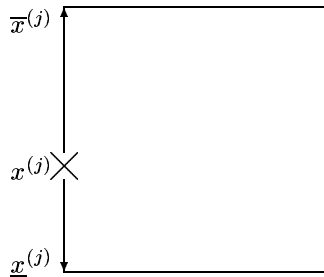
Here, the values $\overline{p}^{(j)}$ and $\underline{p}^{(j)} = 1 - \overline{p}^{(j)}$ are chosen in such a way that $\overline{p}^{(j)} \cdot \overline{x}_1 + \underline{p}^{(j)} \cdot \underline{x}_1 = x_1^{(j)}$. Due to this choice,

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

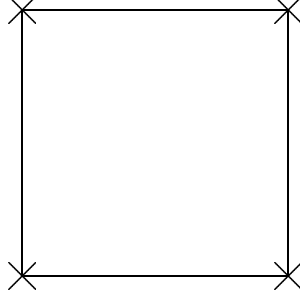
hence for the new distribution, the mathematical expectation $E[x_1]$ is the same as for the old one. Similarly, we can prove that the values $E[x_2], \dots, E[x_n]$, and $E[x_1 \cdot \dots \cdot x_n]$ do not change.

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 $E[x_1], \dots, E[x_n]$, and $E[x_1 \cdot \dots \cdot x_n]$ 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 $x_1 \in \{\underline{x}_1, \overline{x}_1\}$, \dots , $x_n \in \{\underline{x}_n, \overline{x}_n\}$. In other words, it is sufficient to consider only distributions which are located in 2^n corner points of the box $\mathbf{x}_1 \times \dots \times \mathbf{x}_n$:



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 $x_l^{(i)} \leq x_l^{(j)}$ 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, \dots, k_q be coordinates for which $x_{k_m}^{(i)} \leq x_{k_m}^{(j)}$, and let l_1, \dots, 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 $E[x_1 \cdot \dots \cdot x_n]$? 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:

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

- 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 $E[x_1 \cdot \dots \cdot x_n]$ 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 $x_{k_m}^{(i)} \leq x_{k_m}^{(j)}$; multiplying these inequalities between positive numbers, we conclude that $\Pi_k^{(i)} \leq \Pi_k^{(j)}$. 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 $x^{(i)} \neq x^{(j)}$:

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

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

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

$$x_k^{(1)} \leq x_k^{(2)} \leq \dots \leq x_k^{(N)}.$$

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 \bar{x}_k . Due to the above inequality, once $x_k^{(i)}$ is equal to its largest possible value, i.e., once $x_k^{(i)} = \bar{x}_k$, all the following values of x_k must also be equal to the same largest possible value, i.e., $x_k^{(i+1)} = \dots = x_k^{(N)} = \bar{x}_k$.

Therefore, when we move from $x^{(i)}$ to $x^{(i+1)}$, the overall number of coordinates equal to \bar{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 define 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 \bar{x}_k precedes the first appearance of \bar{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, \dots, x_n .

For simplicity, let us assume that the variables x_1, \dots, x_n are already ordered according to this order, i.e., that \bar{x}_1 first appears in the sequence $x^{(i)}$ before (or at the same time as) \bar{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”: $x_{k+1}^{(i)} = \underline{x}_{k+1}, \dots, x_n^{(i)} = \underline{x}_n$. In other words, each vector $x^{(i)}$ can take one of the following values:

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

These are exactly the vectors corresponding to the expression for \bar{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 $(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$;
- let $p^{(2)}$ be the probability of $(\bar{x}_1, \underline{x}_2, \dots, \underline{x}_n)$;
- ...
- let $p^{(i+1)}$ be the probability of $(\bar{x}_1, \dots, \bar{x}_i, \underline{x}_{i+1}, \dots, \underline{x}_n)$;
- ...
- let $p^{(n+1)}$ be the probability of $(\bar{x}_1, \dots, \bar{x}_n)$.

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

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

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

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

By definition, p_i is the probability with which we must take \bar{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:

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

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

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

thus,

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

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

$$\begin{aligned} p^{(1)} &= 1 - (p^{(2)} + \dots + p^{(n+1)}) = \\ &= 1 - ((p_1 - p_2) + (p_2 - p_3) + \dots + (p_{n-1} - p_n) + p_n) = 1 - p_1. \end{aligned}$$

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

7 Alternative Algorithm for Computing \bar{E} and Its Justification

In Section 1, we have shown that if we sequentially apply formula (1) with $n = 2$, we may end up with an overestimation for \bar{E} . It turns out, however, that if we apply the “ $n = 2$ ” formula in the order in which $p_1 \geq p_2 \geq \dots \geq p_n$, we get the exact upper bound \bar{E} . This alternative algorithm consists of $n - 1$ applications of the “ $n = 2$ ” formula and thus, after sorting (which takes $O(n \cdot \log(n))$ steps), requires linear computation time – just like the algorithm described above.

Let us outline the proof that this sequential algorithm is indeed correct. Indeed, first, we combine estimates for x_1 and x_2 into an estimate for $x_{12} = x_1 \cdot x_2$. For x_{12} , the upper bound E_{12} for the mean E is determined by the formula

$$E_{12} = (1 - p_1) \cdot \underline{x}_1 \cdot \underline{x}_2 + (p_1 - p_2) \cdot \bar{x}_1 \cdot \underline{x}_2 + p_2 \cdot \bar{x}_1 \cdot \bar{x}_2,$$

and the exact bounds on x_{12} come from interval multiplication: $\underline{x}_{12} = \underline{x}_1 \cdot \underline{x}_2$ and $\bar{x}_{12} = \bar{x}_1 \cdot \bar{x}_2$. Therefore,

$$p_{12} = \frac{E_{12} - \underline{x}_1 \cdot \underline{x}_2}{\bar{x}_1 \cdot \bar{x}_2 - \underline{x}_1 \cdot \underline{x}_2} = (p_1 - p_2) \cdot \frac{\bar{x}_1 \cdot \underline{x}_2 - \underline{x}_1 \cdot \underline{x}_2}{\bar{x}_1 \cdot \bar{x}_2 - \underline{x}_1 \cdot \underline{x}_2} + p_2 = p_1 \cdot r + p_2 \cdot (1 - r),$$

where

$$r \stackrel{\text{def}}{=} \frac{\bar{x}_1 \cdot \underline{x}_2 - \underline{x}_1 \cdot \underline{x}_2}{\bar{x}_1 \cdot \bar{x}_2 - \underline{x}_1 \cdot \underline{x}_2}.$$

Since $\underline{x}_1 \cdot \underline{x}_2 \leq \overline{x}_1 \cdot \underline{x}_2 \leq \overline{x}_1 \cdot \overline{x}_2$, we conclude that $0 \leq r \leq 1$, i.e., p_{12} is a convex combination of p_1 and p_2 . Since $p_1 \geq p_2 \geq p_3 \dots \geq p_n$, we thus conclude that $p_{12} \geq p_3 \geq \dots \geq p_n$. Thus, for the remaining variables x_{12}, x_3, \dots, x_n , we still have the right order of p_i .

That the resulting formula is the same can be seen if we use the form (1b) of the main formula. If we explicitly emphasize terms related to x_1 and x_2 in this formula, we get:

$$\begin{aligned} \overline{E} &= \underline{x}_1 \cdot \underline{x}_2 \cdot \underline{x}_3 \cdot \dots \cdot \underline{x}_n + \\ &\quad (E_1 - \underline{x}_1) \cdot \underline{x}_2 \cdot \underline{x}_3 \cdot \dots \cdot \underline{x}_n + \\ &\quad \overline{x}_1 \cdot (E_2 - \underline{x}_2) \cdot \underline{x}_3 \cdot \dots \cdot \underline{x}_n + \\ &\quad \overline{x}_1 \cdot \overline{x}_2 \cdot (E_3 - \underline{x}_3) \cdot \underline{x}_4 \cdot \dots \cdot \underline{x}_n + \\ &\quad \dots + \\ &\quad \overline{x}_1 \cdot \overline{x}_2 \cdot \overline{x}_3 \cdot \dots \cdot \overline{x}_{i-1} \cdot (E_i - \underline{x}_i) \cdot \underline{x}_{i+1} \cdot \dots \cdot \underline{x}_n + \\ &\quad \dots + \\ &\quad \overline{x}_1 \cdot \overline{x}_2 \cdot \overline{x}_3 \cdot \dots \cdot \overline{x}_{n-1} \cdot (E_n - \underline{x}_n). \end{aligned}$$

Since $\underline{x}_1 \cdot \underline{x}_2 = \underline{x}_{12}$ and $\overline{x}_1 \cdot \overline{x}_2 = \overline{x}_{12}$, we can reformulate this formula as:

$$\begin{aligned} \overline{E} &= \underline{x}_{12} \cdot \underline{x}_3 \cdot \dots \cdot \underline{x}_n + \\ &\quad (E_1 - \underline{x}_1) \cdot \underline{x}_2 \cdot \underline{x}_3 \cdot \dots \cdot \underline{x}_n + \\ &\quad \overline{x}_1 \cdot (E_2 - \underline{x}_2) \cdot \underline{x}_3 \cdot \dots \cdot \underline{x}_n + \\ &\quad \overline{x}_{12} \cdot (E_3 - \underline{x}_3) \cdot \underline{x}_4 \cdot \dots \cdot \underline{x}_n + \\ &\quad \dots + \\ &\quad \overline{x}_{12} \cdot \overline{x}_3 \cdot \dots \cdot \overline{x}_{i-1} \cdot (E_i - \underline{x}_i) \cdot \underline{x}_{i+1} \cdot \dots \cdot \underline{x}_n + \\ &\quad \dots + \\ &\quad \overline{x}_{12} \cdot \overline{x}_3 \cdot \dots \cdot \overline{x}_{n-1} \cdot (E_n - \underline{x}_n). \end{aligned}$$

A similar formula for $n = 2$ leads to

$$E_{12} = \underline{x}_1 \cdot \underline{x}_2 + (E_1 - \underline{x}_1) \cdot \underline{x}_2 + \overline{x}_1 \cdot (E_2 - \underline{x}_2),$$

thus

$$E_{12} - \underline{x}_1 \cdot \underline{x}_2 = (E_1 - \underline{x}_1) \cdot \underline{x}_2 + \overline{x}_1 \cdot (E_2 - \underline{x}_2).$$

Hence, the above formula for \overline{E} can be rewritten as:

$$\begin{aligned} \overline{E} &= \underline{x}_{12} \cdot \underline{x}_3 \cdot \dots \cdot \underline{x}_n + \\ &\quad (E_{12} - \underline{x}_{12}) \cdot \underline{x}_3 \cdot \dots \cdot \underline{x}_n + \end{aligned}$$

$$\begin{aligned}
& \bar{x}_{12} \cdot (E_3 - \underline{x}_3) \cdot \underline{x}_4 \cdot \dots \cdot \underline{x}_n + \\
& \quad \dots + \\
& \bar{x}_{12} \cdot \bar{x}_3 \cdot \dots \cdot \bar{x}_{i-1} \cdot (E_i - \underline{x}_i) \cdot \underline{x}_{i+1} \cdot \dots \cdot \underline{x}_n + \\
& \quad \dots + \\
& \bar{x}_{12} \cdot \bar{x}_3 \cdot \dots \cdot \bar{x}_{n-1} \cdot (E_n - \underline{x}_n),
\end{aligned}$$

i.e., exactly the formula (1b) for x_{12}, x_3, \dots, x_n . Thus, the sequential algorithm is indeed correct.

8 Can We Compute the Exact Lower Bound? First Auxiliary Result

In the previous sections, we have described an explicit analytical expression that enables us to compute the exact upper bound \bar{E} on the mean of the product of many positive random variables with known expectations. A natural question is: can we have a similar analytical expression for computing the exact *lower* bound \underline{E} for this mean?

In this section, we show that, unless $P=NP$, we cannot expect such an expression – in other words, that if such an expression exists, then we would have $P=NP$, which most computer scientists believe to be impossible (for definitions and detailed description of the $P=NP$ problem, see, e.g., [9, 12, 14]).

Specifically, we will show that if a feasible (computable in polynomial time) analytical expression for \underline{E} exists, then we would have a polynomial-time algorithm for solving a known NP-hard problem. Thus, we would conclude that the class NP of all “hard” problems would coincide with the class P of all problems that can be solved in polynomial time (e.g., in time bounded by a polynomial of a length of the input).

In our proof, as the known NP-hard problem \mathcal{P}_0 , we take a *subset* problem: given n positive integers s_1, \dots, s_n , to check whether there exist signs $\eta_i \in \{-1, +1\}$ for which the signed sum $\sum_{i=1}^n \eta_i \cdot s_i$ equals 0.

We will show that this problem can be reduced to the problem of computing \underline{E} , i.e., that to every instance (s_1, \dots, s_n) of the problem \mathcal{P}_0 , we can put into correspondence such an instance of the \underline{V} -computing problem that based on its solution, we can easily check whether the desired signs exist.

As this instance, we take the instance corresponding to the intervals $[\underline{x}_i, \bar{x}_i] = [2^{-s_i}, 2^{s_i}]$, with means $E_i = (1/2) \cdot (2^{-s_i} + 2^{s_i})$. We want to show that for the corresponding problem, always $\underline{E} \geq 1$, and $\underline{E} = 1$ if and only if there exist signs η_i for which $\sum \eta_i \cdot s_i = 0$.

Let us first show that in all cases, $\underline{E} \geq 1$. Similarly to the proof of the main result, we can show that the minimum of E is attained on distributions that are

located in points $x^{(1)}, \dots, x^{(N)}$ that are corner points of the box $\mathbf{x}_1 \times \dots \times \mathbf{x}_n$, i.e., in which, for every i and j , $x_i^{(j)} = \underline{x}_i$ or $x_i^{(j)} = \bar{x}_i$. Hence, $\underline{E} = \sum p^{(j)} \cdot v^{(j)}$, where $v^{(j)} \stackrel{\text{def}}{=} x_1^{(j)} \cdot \dots \cdot x_n^{(j)}$.

We can represent this expression as $\underline{E} = \sum p^{(j)} \cdot \exp(t^{(j)})$, where

$$t^{(j)} \stackrel{\text{def}}{=} \log(v^{(j)}) = \log(x_1^{(j)}) + \dots + \log(x_n^{(j)}).$$

Since $t^{(j)}$ is the sum, its expectation $t \stackrel{\text{def}}{=} \sum p^{(j)} \cdot t^{(j)}$ can be represented as the sum of n terms $t_i \stackrel{\text{def}}{=} \sum p^{(j)} \cdot \log(x_i^{(j)})$. In each of these n terms t_i , we add the values $\log(\underline{x}_i) = -s_i$ and $\log(\bar{x}_i) = s_i$. The total probability of $\log(\underline{x}_i) = -s_i$ and of $\log(\bar{x}_i) = s_i$ is equal to $1/2$ – because the mean E_i is exactly the average of 2^{-s_i} and 2^{s_i} . Thus, in each sum t_i , we add $\log(\underline{x}_i) = -s_i$ with a total probability $1/2$ and $\log(\bar{x}_i) = s_i$ with a total probability $1/2$. The resulting sum is 0, hence $t = t_1 + \dots + t_n = 0$, i.e., $\sum p^{(j)} \cdot t^{(j)} = 0$.

Since the function $\exp(x)$ is strictly convex, we have $\underline{E} = \sum p^{(j)} \cdot \exp(t^{(j)}) \geq \exp(\sum p^{(j)} \cdot t^{(j)}) = \exp(0) = 1$. Due to the same strict convexity, $\underline{E} = 1$ is and only if all the values $t^{(j)}$ are equal to 0 – and each $t^{(j)}$ is a sum of values $t_i^{(j)} = \pm s_i$. The reduction is proven.

9 Can We Compute the Exact Upper Bound if Some Values x_i Are Negative? Second Auxiliary Result

In the main part of 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. A natural question is: is it possible 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 $f(x_1, \dots, x_n) = x_1 \cdot \dots \cdot x_n$ is non-decreasing in each of the variables; this property is not true for negative x_i .

It turns out that we cannot easily compute the exact upper bound \overline{E} for the case when some of the values x_i may be negative. Indeed, if we could do that, then we could compute the exact lower bound for non-negative x_i as follows:

- for $x'_1 = -x_1, x'_2 = x_2, \dots, x'_n = x_n$, we have $y' = x'_1 \cdot \dots \cdot x'_n = -y$;
- hence, $E[y'] = E[-y] = -E[y]$;
- thus, the exact upper bound \overline{E}' for $E[y']$ and the exact lower bound \underline{E} for $E[y]$ are related by a simple formula: $\overline{E}' = -\underline{E}$;
- so, if we could easily compute \overline{E}' , we would thus be able to easily compute \underline{E} for non-negative x_i – and this, as we have shown, is not possible (unless $P=NP$).

So, no easy algorithm is possible for computing exact upper bounds in the general case of not necessarily non-negative x_i .

10 How Can We Actually Compute \underline{E} ? \overline{E} for Non-Negative x_i ?

In the previous sections, we have shown that it is provably difficult to compute the exact lower bound \underline{E} on the mean of the product of many random variables and to compute the exact upper bound for possibly negative x_i . In practical terms, this difficulty means that when the number n of variables increases, the amount of time required to compute the corresponding quantities may grow exponentially with n – so that when n becomes really large, in hundreds and thousands, the required computation time may exceed the lifetime of the Universe.

In many practical problems, n is not very large but moderate. How can we then compute, say, \underline{E} ? We have already shown that to find the minimum \underline{E} , it is sufficient to consider probability distributions located in corner points $x^{(1)}, \dots, x^{(N)}$ of the box $\mathbf{x}_1 \times \dots \times \mathbf{x}_n$. There are $N = 2^n$ such corner points, so, to find \underline{E} , it is sufficient to solve the following linear programming problem:

$$\sum p^{(j)} \cdot v^{(j)} \rightarrow \min,$$

where $v^{(j)} \stackrel{\text{def}}{=} x_1^{(j)} \cdot \dots \cdot x_n^{(j)}$, under the conditions

$$\begin{aligned} p^{(1)} + \dots + p^{(N)} &= 1; \\ p^{(1)} \cdot x_1^{(1)} + \dots + p^{(N)} \cdot x_1^{(N)} &= E_1; \\ &\dots \\ p^{(1)} \cdot x_n^{(1)} + \dots + p^{(N)} \cdot x_n^{(N)} &= E_n; \\ p^{(j)} &\geq 0. \end{aligned}$$

The unknown here are the values $p^{(j)}$. With respect to the these unknowns, we are minimizing a linear function under linear constraints (equalities and inequalities). Geometrically, the set of all points that satisfy several linear constraints is a polytope. It is well known that to find the minimum of a linear function on a polytope, it is sufficient to consider its vertices (this idea is behind linear programming). In algebraic terms, a vertex can be characterized by the fact that for N variables, N of the original constraints are equalities. In our case, we have $n + 1$ equalities and inequalities $p^{(j)} \geq 0$. Thus, in our case, all but $n + 1$ probabilities $p^{(j)}$ must be equal to 0. So, to find the smallest possible value \underline{E} of $E[x_1 \cdot \dots \cdot x_n]$, it is sufficient to consider probability distributions that are located on $n + 1$ points $x^{(j)}$.

In other words, to compute \underline{E} , we can do the following:

- For each combination $j_1 < \dots < j_{n+1} \leq N$ of $n + 1$ corners, we solve the system of $n + 1$ linear equations

$$\begin{aligned} p^{(j_1)} + \dots + p^{(j_{n+1})} &= 1; \\ p^{(j_1)} \cdot x_1^{(j_1)} + \dots + p^{(j_{n+1})} \cdot x_1^{(j_{n+1})} &= E_1; \\ &\dots \\ p^{(j_1)} \cdot x_n^{(j_1)} + \dots + p^{(j_{n+1})} \cdot x_n^{(j_{n+1})} &= E_n. \end{aligned}$$

- If the resulting values $p^{(j_1)}, \dots, p^{(j_{n+1})}$ are all non-negative, we compute

$$E = \sum_{k=1}^{n+1} p^{(j_k)} \cdot v^{(j_k)}.$$

- The smallest of such values E is the desired exact lower bound \underline{E} .

For moderate n , solving each system of $n+1$ linear equations with $n+1$ unknowns is fast. This algorithm requires solving

$$\binom{2^n}{n+1}$$

such systems. What does it mean for small n ?

- For $n = 2$, we need $\binom{4}{3} = 4$ systems (since for $n = 2$, we have exact formulas, we do not actually need to solve these systems).
- For $n = 3$, we need to solve $\binom{8}{4} = 70$ systems.
- For $n = 4$, we need to solve $\binom{16}{5} = 4,368$ systems.
- For $n = 5$, we need to solve $\binom{32}{6} = 906,192$ systems – it is still quite doable.
- For $n = 6$, we need to solve $\binom{64}{7} \approx 6.3 \cdot 10^9$ systems – potentially doable on a high performance computer.
- For $n = 7$, we need to solve $\binom{128}{8} \approx 1.4 \cdot 10^{12}$ systems – close to the limit of modern computers.
- For $n = 8$, we need to solve $\binom{256}{9} \approx 3.4 \cdot 10^{16}$ systems – way beyond the limit of modern computers.

A similar algorithm – but with taking the largest of the values $E = \sum p^{(j)} \cdot v^{(j)}$ instead of the smallest – can compute \overline{E} for the case when the values x_i can be negative.

11 What If, In Addition to the Means E_i , We Also Know Variances? Third Auxiliary Result

Often, in addition to the first moments (means) $E[x_i]$ of the variables x_i , we also have information about their variances $V[x_i]$. In this case, for an arbitrary function $f(x_1, \dots, x_n)$, it is natural, in addition to asking about the possible values of the mean $E[y]$ of y , to also ask about the possible values of the variance $V[y]$. It turns out that adding variances makes the problem very difficult to solve. Specifically, we will show that even for the simplest possible combination function $f(x_1, \dots, x_n) = x_1 + \dots + x_n$, the corresponding problem is NP-hard.

We will also show that it is NP-hard if we know second moments instead of variances.

Specifically, we show that the following problem is NP-hard:

GIVEN: values $\underline{x}_1, \bar{x}_1, \dots, \underline{x}_n, \bar{x}_n, E_1, \dots, E_n, V_1, \dots, V_n$;

FIND: the value

$$\underline{V} \stackrel{\text{def}}{=} \min\{V[x_1 + \dots + x_n] \mid \text{all distributions of } (x_1, \dots, x_n) \text{ for which} \\ x_1 \in [\underline{x}_1, \bar{x}_1], \dots, x_n \in [\underline{x}_n, \bar{x}_n], E[x_1] = E_1, \dots, E[x_n] = E_n, \\ V[x_1] = V_1, \dots, V[x_n] = V_n\}.$$

To prove this NP-hardness, we will reduce, to this problem, the same subset problem as in the previous section. Specifically, to every instance s_1, \dots, s_n of the subset problem, we put into correspondence the following particular case of the above problem: $\underline{x}_i = -s_i, \bar{x}_i = s_i, E_i = 0$, and $V_i = s_i^2$. We will show that $\underline{V} = 0$ if and only if the original subset problem has a solution.

Indeed, variance is always non-negative, so $\underline{V} \geq 0$. If the subset problem has a solution, i.e., if there exist signs $\eta_i \in \{-1, +1\}$ for which the signed sum $\sum_{i=1}^n \eta_i \cdot s_i$ equals 0, then we can take an auxiliary random variable α that takes values 1 and -1 with equal probability $1/2$, and then take $x_i = \eta_i \cdot s_i \cdot \alpha$. In this case, $y = \sum x_i = \alpha \cdot \sum (\eta_i \cdot s_i) = 0$, i.e., y is always equal to 0. Hence, the variance of y is 0.

Vice versa, if the variance of y is 0, this means that the sum $y = \sum x_i$ is equal to its average 0 with probability 1. For each variable x_i , the average is 0, so the variance is equal to the second moment $V[x_i] = E[x_i^2]$. The square x_i^2 takes possible values from the interval $[0, s_i^2]$. Since the expectation of x_i^2 is equal to its largest possible value, we can thus conclude that with probability 1, $x_i^2 = s_i^2$ – i.e., that either $x_i = s_i$ or $x_i = -s_i$. Since $x_1 + \dots + x_n = 0$, we can thus conclude that $\sum \pm x_i = 0$ – i.e., that the original particular case of a subset problem has a solution. For variances, NP-hardness is proven.

A variance $V[x_i]$ is naturally related to the second moment $M[x_i] \stackrel{\text{def}}{=} E[x_i^2]$: $V[x_i] = M[x_i] - (E[x_i])^2$ and $M[x_i] = V[x_i] + (E[x_i])^2$. So, knowing second moments is equivalent to knowing variances. Thus, for second moments, the corresponding problem is NP-hard as well.

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 Genady Davydov and Yuri Matiyasevich, for useful discussions, to Daniel Berleant for his encouragement, and to anonymous referees for helpful suggestions and discussions.

References

- [1] D. Berleant, “Automatically verified arithmetic with both intervals and probability density functions”, *Interval Computations*, 1993, No. 2, pp. 48–70.
- [2] D. Berleant, “Automatically verified arithmetic on probability distributions and intervals”, In: R. B. Kearfott and V. Kreinovich (eds.) *Applications of Interval Computations*, Kluwer, Dordrecht, 1996.
- [3] D. Berleant and C. Goodman-Strauss, “Bounding the results of arithmetic operations on random variables of unknown dependency using intervals”, *Reliable Computing*, 1998, Vol. 4, No. 2, pp. 147–165.
- [4] D. Berleant, L. Xie, and J. Zhang, “Statool: A Tool for Distribution Envelope Determination (DEnv), an Interval-Based Algorithm for Arithmetic on Random Variables”, *Reliable Computing*, 2003, Vol. 9, No. 2, pp. 91–108.

- [5] 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.
- [6] S. Ferson, *RAMAS Risk Calc 4.0: Risk Assessment with Uncertain Numbers*, CRC Press, Boca Raton, Florida, 2002.
- [7] 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>).
- [8] S. Ferson, D. Myers, and D. Berleant, *Distribution-free risk analysis: I. Range, mean, and variance*, Applied Biomathematics, Technical Report, 2001.
- [9] M. R. Garey and S. D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman and Company, New York, 1979.
- [10] 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.
- [11] R. B. Kearfott and V. Kreinovich (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 1996.
- [12] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1997.
- [13] W. A. Lodwick and K. D. Jamison, “Estimating and Validating the Cumulative Distribution of a Function of Random Variables: Toward the Development of Distribution Arithmetic”, *Reliable Computing*, 2003, Vol. 9, No. 2, pp. 127–141.
- [14] C. H. Papadimitriou, *Computational Complexity*, Addison Wesley, San Diego, 1994.
- [15] S. Rabinovich, *Measurement Errors: Theory and Practice*, American Institute of Physics, New York, 1993.
- [16] H. Regan, S. Ferson and D. Berleant, “Equivalence of five methods for bounding uncertainty”, *Journal of Approximate Reasoning* (to appear).
- [17] N. C. Rowe, “Absolute bounds on the mean and standard deviation of transformed data for constant-sign-derivative transformations”, *SIAM Journal of Scientific Statistical Computing*, 1988, Vol. 9, pp. 1098–1113.

- [18] P. Walley, *Statistical reasoning with imprecise probabilities*, Chapman and Hall, N.Y., 1991.
- [19] R. Williamson and T. Downs, “Probabilistic arithmetic I: numerical methods for calculating convolutions and dependency bounds”, *International Journal of Approximate Reasoning*, 1990, Vol. 4, pp. 89–158.