Absolute Bounds on the Mean of Sum, Product, Max, and Min: A Probabilistic Extension of Interval Arithmetic

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

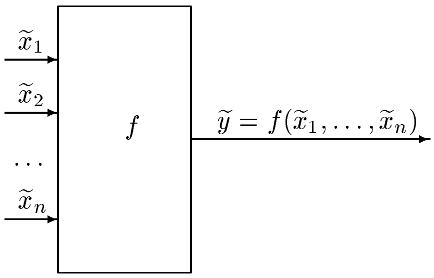
[1]Applied Biomathematics, 100 North Country Road,
Setauket, NY 11733, USA, {scott,lev}@ramas.com
[2]Computer Science Department, University of Texas at El Paso
El Paso, TX 79968, USA, vladik@cs.utep.edu

Abstract

We extend the main formulas of interval arithmetic $x_1 \circledast x_2$ for different arithmetic operations $\circledast$ to the case when, for each input x_i, in addition to the interval $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ of possible values, we also know its mean E_i (or an interval $\mathbf{E}_i$ of possible values of the mean), and we want to find the corresponding bounds for $x_1 \circledast x_2$ and its mean.

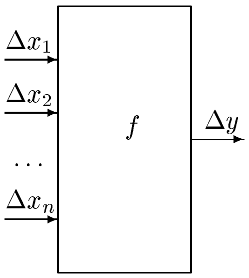
1 Error Estimation for Indirect Measurements: An Important Practical Problem

A practically important class of statistical problems is related to data processing (indirect measurements). Some physical quantities y – such as the distance to a star or the amount of oil in a given well – are impossible or difficult to measure directly. To estimate these quantities, we use *indirect* measurements, i.e., we measure some easier-to-measure quantities $x_1, \ldots, x_n$ which are related to y by a known relation $y = f(x_1, \ldots, x_n)$, and then use the measurement results $\widetilde{x}_i$ $(1 \le i \le n)$ to compute an estimate $\widetilde{y}$ for y as $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$:



For example, to find the resistance R, we measure current I and voltage V, and then use the known relation $R = V/I$ to estimate resistance as $\widetilde{R} = \widetilde{V}/\widetilde{I}$.

Measurement are never 100% accurate, so in reality, the actual value x_i of i-th measured quantity can differ from the measurement result $\widetilde{x}_i$. In probabilistic terms, x_i is a random variable; its probability distribution describes the probabilities of different possible value of measurement error $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$. It is desirable to describe the error $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$ of the result of data processing:



2 Robust Statistics and Interval Computations

Often, we know (or assume) that the measurement error Δx_i of each direct measurement is normally distributed with a known standard deviation σ_i, and that measurement errors corresponding to different measurements are independent. These assumptions – justified by the central limit theorem, according to which sums of independent identically distributed random variables with finite moments tend quickly toward the Gaussian distribution – underly the traditional engineering approach to estimating measurement errors.

In some situations, the error distributions are not Gaussian, but we know their exact shape (e.g., lognormal). In many practical measurement situations, however, we only have *partial* information about the probability distributions [13, 14].

Traditional statistical techniques deal (see, e.g., [20]) with the situations when we know the exact shape of the probability distributions. To deal with practical situations in which we only have a partial information about the distributions, special techniques have to be invented. Such techniques are called methods of *robust statistics*. They are called robust because they are usually designed to provide guaranteed estimates, i.e., estimates which are valid for all possible distributions from a given class [9, 20].

An important case of partial information about a random variable x is when we know (with probability 1) that x is within a given interval $\mathbf{x} = [\underline{x}, \overline{x}]$, but we have no information about the probability distribution within this interval. In other words, x may be uniformly distributed on this interval, it may be deterministic (i.e., distributed in a single value with probability 1), distributed according to a truncated Gaussian, bimodal distribution – we do not know.

This situation occurs often in measurement practice [16] when, for a measuring instrument, we do not know the probability distribution of the measurement error Δx, and instead, we only know the upper bound Δ on the measurement error Δx: $|\Delta x| \le \Delta$. This upper bound is usually provided by the manufacturer of the measuring instrument. (If the manufacturer does not guarantee any upper bound on the measurement error, then this "measuring instrument" is practically useless.) In this case, after we performed a measurement and got a measurement result $\widetilde{x}$, the only information that we have about the actual value x of the measured quantity is that it belongs to the interval $\mathbf{x} = [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$.

In some case, instead of providing us with a single bound on $|\Delta x|$, the manufacturer provides us with two different bounds: the lower bound $\underline{\Delta}$ and the upper bound $\overline{\Delta}$ on Δx. In this case, after we performed a measurement and got a measurement result $\widetilde{x}$, the only information that we have about the actual value x of the measured quantity is that it belongs to the interval $\mathbf{x} = [\widetilde{x} + \underline{\Delta}, \widetilde{x} + \overline{\Delta}]$.

In principle, for every measuring instrument, we can *calibrate* it by comparing the results of measuring with this instrument with the results of measuring the same quantity by a standard (much more accurate) measuring instrument. The difference between these two measurement results is practically equal to the sensor's measurement error, and thus, the empirical distribution of this difference is close to the desired probability distribution for measurement error. There are two cases, however, when this calibration is not done:

- First is the case of cutting-edge measurements, e.g., measurements in fundamental science. When a Hubble telescope detects the light from a distant galaxy, there is no "standard" (much more accurate) telescope floating nearby that we can use to calibrate the Hubble: the Hubble telescope is the best we have.

- The second case is the case of measurements on the shop floor. In this case, in principle, every sensor can be thoroughly calibrated, but sensor calibration is so costly – usually costing ten times more than the sensor itself – that manufacturers rarely do it.

In both cases, we have interval uncertainty.

When the uncertainty of each data points is of this interval type, we arrive at the following problem: for each of n random variables $x_1, \ldots, x_n$, we know that it is located (with probability 1) within a given interval $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$. We do not know the distributions within the intervals, and we do not know whether the random variables x_i are independent or not. What can we then conclude about the probability distribution of $y = f(x_1, \ldots, x_n)$?

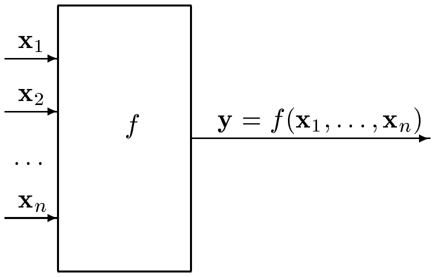
Since the only information we have about each variable x_i consists of its lower bound $\underline{x}_i$ and upper bound $\overline{x}_i$, it is natural to ask for similar bounds $\mathbf{y} = [\underline{y}, \overline{y}]$ for y. As a result, we arrive at the following problem:

GIVEN: an algorithm computing a function $f(x_1, \ldots, x_n)$ from R^n to R and n intervals $\mathbf{x}_1, \ldots, \mathbf{x}_n$,

TAKE: all possible joint probability distributions on R^n for which, for each i, $x_i \in \mathbf{x}_i$ with probability 1;

FIND: the set $\mathbf{Y}$ of all possible values of a random variable $y = f(x_1, \ldots, x_n)$ for all such distributions.

One can easily prove that $\mathbf{Y}$ is equal to the range $f(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ of the given function f on given intervals, i.e., to $\{f(x_1, \ldots, x_n) \,|\, x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}$:



Indeed:

- If $x_i \in \mathbf{x}_i$ with probability 1, then y belongs to the range with probability 1.

- Vice versa, any point y from the range can be represented as $f(x_1, \ldots, x_n)$ for some $x_i \in \mathbf{x}_i$, so to show that $y \in \mathbf{Y}$, we can take degenerate distributions located at x_i with probability 1.

Q.E.D.

The problem of computing the range of a given function $f(x_1, \ldots, x_n)$ on given intervals $\mathbf{x}_i$ is exactly the problem solved by interval computations. The main interval computations approach to solving this problem is to take into consideration that inside the computer, every algorithm consists of elementary operations (arithmetic operations, min, max, etc.). For each elementary operation $f(x, y)$, if we know the intervals $\mathbf{x}$ and $\mathbf{y}$ for x and y, we can compute the exact range $f(\mathbf{x}, \mathbf{y})$:

$$[\underline{x}_1, \overline{x}_1] + [\underline{x}_2, \overline{x}_1] = [\underline{x}_1 + \underline{x}_2, \overline{x}_1 + \overline{x}_2];$$

$$[\underline{x}_1, \overline{x}_1] - [\underline{x}_2, \overline{x}_1] = [\underline{x}_1 - \overline{x}_2, \overline{x}_1 - \underline{x}_2];$$

etc.; the corresponding formulas form the so-called *interval arithmetic*; see, e.g., [10, 11, 12]. We can therefore repeat the computations forming the program f step-by-step, replacing each operation with real numbers by the corresponding operation of interval arithmetic. It is known that, as a result, we get an enclosure for the desired range.

In the above proof, we considered the case when we have no information about the correlation between the random variables. One can easily see that the same proof shows that if we assume independence, we still get the same range.

For functions of two variables, we can consider two additional cases:

- when x_1 and x_2 are highly positively correlated, i.e., when there exists a random variable t such that both x_1 and x_2 are non-decreasing functions of t (crudely speaking, x_1 is (non-strictly) increasing in x_2), and
- when x_i is highly negatively correlated, i.e., when there exists a random variable t such that both x_1 and x_2 are non-decreasing functions of t (crudely speaking, when x_1 is decreasing in x_2).

The above simple proof shows that in both cases, we get the same range $\mathbf{Y}$ as in the above case of no information about the correlation.

In some practical situations, in addition to the lower and upper bounds on each random variable x_i, we know the bounds $\mathbf{E}_i = [\underline{E}_i, \overline{E}_i]$ on its mean E_i. Let us give examples of such situations.

The first example is from indirect measurements. In measurement practice (see, e.g., [16]), the overall measurement error Δx is usually represented as a sum of two components:

- a *systematic* error component $\Delta_s x$ which is defined as the expected value $E[\Delta x]$, and
- a *random* error component $\Delta_r x$ which is defined as the difference between the overall measurement error and the systematic error component: $\Delta_r x \stackrel{\text{def}}{=} \Delta x - \Delta_s x$.

In addition to the bound Δ on the overall measurement error, the manufacturers of the measuring instrument often provide an upper bound Δ_s on the systematic error component: $|\Delta_s x| \le \Delta_s$.

This additional information is provided because, with this additional information, we not only get a bound on the accuracy of a single measurement, but we also get an idea of what accuracy we can attain if we use repeated measurements to increase the measurement accuracy. Indeed, the very idea that repeated measurements can improve the measurement accuracy is natural: we measure the same quantity by using the same measurement instrument several (N) times, and then take, e.g., an arithmetic average

$$\bar{x} = \frac{\widetilde{x}^{(1)} + \ldots + \widetilde{x}^{(N)}}{N}$$

of the corresponding measurement results $\widetilde{x}^{(1)} = x + \Delta x^{(1)}, \ldots, \widetilde{x}^{(N)} = x + \Delta x^{(N)}$.

- If systematic error is the only error component, then all the measurements lead to exactly the same value $\widetilde{x}^{(1)} = \ldots = \widetilde{x}^{(N)}$, and averaging does not change the value – hence does not improve the accuracy.

- On the other hand, if we know that the systematic error component is 0, i.e., $E[\Delta x] = 0$ and $E[\widetilde{x}] = x$, then, as $N \to \infty$, the arithmetic average tends to the actual value x. In this case, by repeating the measurements sufficiently many times, we can determine the actual value of x with an arbitrary given accuracy.

In general, by repeating measurements sufficiently many times, we can arvitrarily decrease the random error component and thus attain accuracy as close to Δ_s as we want.

When this additional information is given, then, after we performed a measurement and got a measurement result $\widetilde{x}$, then not only we get the information that the actual value x of the measured quantity belongs to the interval $\mathbf{x} = [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$, but we can also conclude that the expected value of $x = \widetilde{x} - \Delta x$ (which is equal to $E[x] = \widetilde{x} - E[\Delta x] = \widetilde{x} - \Delta_s x$) belongs to the interval $\mathbf{E} = [\widetilde{x} - \Delta_s, \widetilde{x} + \Delta_s]$.

Just like sometimes, manufacturers provide us with two different bounds $\underline{\Delta}$ and $\overline{\Delta}$ for the overall measurement error, they also sometimes supply us with two different bounds for the systematic error component: $\underline{\Delta}_s$ and $\overline{\Delta}_s$. In this case, after we performed a measurement and got a measurement result $\widetilde{x}$, then we conclude that the actual (unknown) value x of the measured quantity belongs to the interval $\mathbf{x} = [\widetilde{x} + \underline{\Delta}, \widetilde{x} + \overline{\Delta}]$, and that its expected value $E[x]$ belongs to the interval $\mathbf{E} = [\widetilde{x} + \underline{\Delta}_s, \widetilde{x} + \overline{\Delta}_s]$.

The width of the interval $\mathbf{x}$ describes how accurate the measurement is. The width of the interval $\mathbf{E}$ describes how biased is this measurement, i.e., how accurate results we can get if we repeat this same measurement several times.

In indirect measurements, we directly measure auxiliary quantities $x_1, \ldots, x_n$, and then use the measurement results $\widetilde{x}_1, \ldots, \widetilde{x}_n$ and the known dependence $y = f(x_1, \ldots, x_n)$ between the desired quantity y and the quantities x_i to estimate the value of y as $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$. When for each direct measurement, we have not only the interval bound $\mathbf{x}_i$ on the value x_i, but also the bounds $\mathbf{E}_i$ on the corresponding expectation $E[x_i]$ – the bounds that characterize the potential accuracy of repeated measurement – then it is natural to try to find the similar information for the resulting quantity y. In other words, we would like to know, in addition to the interval $\mathbf{y}$ of possible value of y, also the interval of possible values of $E[y]$. This additional interval will hopefully provide us with the information on how repeated measurements can improve the accuracy of this indirect measurement.

Similar situations also occur in *risk analysis*. In this discipline, predictions are made about the magnitudes or probabilities of structural failures or other adverse extreme events such as patients receiving toxic doses of therapeutic drugs or endangered species going extinct. These forecasts are often computed from limited empirical information. In traditional "worst case" analyses, the elementary methods of interval analysis are applied to risk formulations estimating, for instance, the difference between a structure's strength and some stress acting on it, or the delivered dose of a drug, or the population size of the endangered species, etc.

In probabilistic risk assessments, analysts usually assume statistical distributions for each of the variables in the risk expression. Of course, such assumptions are often difficult to justify when empirical data are sparse. For instance, it may be implausible that the distribution family is known empirically. This can be a serious problem for analysts because the choice of distribution family may strongly influence the results of the analysis [4].

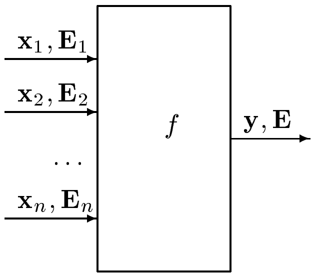
To avoid committing to untenable assumptions about the input variables, many authors have suggested propagating the means and variances of the variables through the risk expressions as a crude form of risk analysis (e.g., [17, 19, 25]). This approach is sometimes called first-order error analysis, and is a widely used approach for making risk estimates. In tradtional probability theory, this approach is called moment propagation and is considered a fundamental part of mathematical statistics (see, for example, [23]).

In such situations, when in addition to the lower and upper bounds on each random variable x_i, we know the bounds $\mathbf{E}_i = [\underline{E}_i, \overline{E}_i]$ on its mean E_i, we arrive at the following problem:

GIVEN: an algorithm computing a function $f(x_1, \ldots, x_n)$ from R^n to R; n intervals $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and n intervals $\mathbf{E}_1, \ldots, \mathbf{E}_n$,

TAKE: all possible joint probability distributions on R^n for which, for each i, $x_i \in \mathbf{x}_i$ with probability 1 and the mean E_i belongs to $\mathbf{E}_i$;

FIND: the set $\mathbf{Y}$ of all possible values of a random variable $y = f(x_1, \ldots, x_n)$ and the set $\mathbf{E}$ of all possible values of $E \stackrel{\text{def}}{=} E[y]$ for all such distributions:



One can easily show that the interval part $[\underline{y}, \overline{y}]$ of the result is the same as for interval arithmetic, so what we really need to compute is the range $\mathbf{E}$ for E.

A similar problem can be formulated for the case when x_i are known to be independent, and for the cases when $n = 2$ and the values x_i are highly positively or highly negatively correlated.

In more precise terms, we want to solve the following problem:

GIVEN: an algorithm computing a function $f(x_1, \ldots, x_n)$ from R^n to R; and values $\underline{x}_1$, $\overline{x}_1$, $\ldots$, $\underline{x}_n$, $\overline{x}_n$, $\underline{E}_1$, $\overline{E}_1$, $\ldots$, $\underline{E}_n$, $\overline{E}_n$.

FIND: the values

$$\underline{E} \stackrel{\text{def}}{=} \min\{E(f(x_1, \ldots, x_n) \mid \text{ all distributions of } (x_1, \ldots, x_n) \text{ for which}$$

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

and

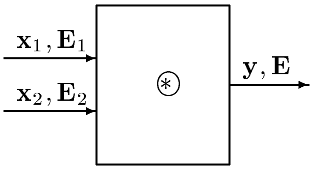
$$\overline{E} \stackrel{\text{def}}{=} \max\{E(f(x_1, \ldots, x_n) \mid \text{ all distributions of } (x_1, \ldots, x_n) \text{ for which}$$

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

(plus restrictions on the correlation).

These problems are in line with the problems traditionally solved by robust statistics (see, e.g., [9]): many known algorithms in the area of robust statistics also return a guaranteed robust estimate for the mean, which holds for a collection of distributions. One would expect that these problems have already been solved in robust statistics. To our surprise, it turned out that while robust statistics does have a lot of useful and interesting results about the guaranteed bounds on the mean for many classes of distributions, the problem of how to compute guaranteed bounds on the mean of a simple arithmetic expression, has not yet been solved satisfactorily.

Moreover, such bounds were not even known for elementary arithmetic operations, i.e., for the cases when $n = 2$ and $f = \circledast$ is one of the basic arithmetic operations ($+$, $-$, $\cdot$, $1/x$, min, max). In these cases, the problem takes the following form: if we know two tuples $(\underline{x}_i, \underline{E}_i, \overline{E}_i, \overline{x}_i)$, $(i = 1, 2)$, what tuple describes possible values of $y = x_1 \circledast x_2$?



In more precise terms, the following class of problems was open:

GIVEN: values $\underline{x}_1$, $\overline{x}_1$, $\underline{x}_2$, $\overline{x}_2$, $\underline{E}_1$, $\overline{E}_1$, $\underline{E}_2$, $\overline{E}_2$, and an operation $\circledast$,

FIND: the values

$$\underline{E} \stackrel{\text{def}}{=} \min\{E(x_1 \circledast x_2) \mid \text{ all distributions of } (x_1, x_2) \text{ for which}$$

$$x_1 \in [\underline{x}_1, \overline{x}_1], x_2 \in [\underline{x}_2, \overline{x}_2], E[x_1] \in [\underline{E}_1, \overline{E}_1], E[x_2] \in [\underline{E}_2, \overline{E}_2]\}$$

and

$$\overline{E} \stackrel{\text{def}}{=} \max\{E(x_1 \circledast x_2) \mid \text{ all distributions of } (x_1, x_2) \text{ for which}$$

$$x_1 \in [\underline{x}_1, \overline{x}_1], x_2 \in [\underline{x}_2, \overline{x}_2], E[x_1] \in [\underline{E}_1, \overline{E}_1], E[x_2] \in [\underline{E}_2, \overline{E}_2]\}$$

(plus restrictions on the correlation).

The main objective of this paper is to provide formulas for $\underline{E}$ and $\overline{E}$ for the arithmetic operations.

Once such formulas are found, we can handle the case of a general algorithm f in the same manner as in interval computations. Indeed, in general, the information about i-th input is represented by a pair of nested intervals $\langle \mathbf{x}_i, \mathbf{E}_i \rangle$, where $\mathbf{E}_i \subseteq \mathbf{x}_i$. Formulas for arithmetic operations allows us to transform the information $\langle \mathbf{x}_1, \mathbf{E}_1 \rangle$, $\langle \mathbf{x}_2, \mathbf{E}_2 \rangle$ corresponding to the inputs x_1 and x_2 into a similar information $\langle \mathbf{x}, \mathbf{E} \rangle$ about the result $x \stackrel{\text{def}}{=} x_1 \circledast x_2$ of the corresponding arithmetic operation. Suppose now that, instead of a simple arithmetic operation $x_1 \circledast x_2$, we have a more complex algorithm $f(x_1, \ldots, x_n)$, and we have an information $\langle \mathbf{x}_i, \mathbf{E}_i \rangle$ $(1 \le n)$ corresponding to n inputs. Then, in order to get a similar information about $y = f(x_1, \ldots, x_n)$, we repeat the computations forming the program f step-by-step, replacing each operation with real numbers by the corresponding operation on pairs of intervals $\langle \mathbf{x}_i, \mathbf{E}_i \rangle$. Similar to interval arithmetic, we can prove that, as a result, we get enclosures for the desired intervals $\mathbf{y}$ and $\mathbf{E}_y$ corresponding to y.

Viewed in this way, our paper in very much in line with the traditional interval computations. However, the main technical difficulty of our results is not in this interval-like idea, but in deriving the formulas for $\underline{E}$ and $\overline{E}$ for the main arithmetic operations. In deriving these formulas, we mainly used – as one can see from our proofs – statistical techniques: more precisely, we used techniques developed and used in robust statistics.

Overall, in this paper, we suggest that one can combine the methods of moment propagation with elementary interval analysis to obtain results that are better than can be obtained from either analysis separately. Rowe [18] considered the problem of computing moments of certain kinds of transformations such as exp, log, sqrt, etc. from sparse structural information such as first moments and ranges of the operands. We extend this approach to the context of convolutions between poorly characterized random variables. In particular, in risk analysis, Rowe's methods, together with the present extension, creates what may be characterized as a *distribution-free* risk analysis that lets analysts compute bounds on risk expressions without making assumptions about the precise distributions of the underlying variables.

3 Main Results for the Case When We Know the Exact Values of E_1 and E_2

In this section, we describe the formulas for the interval $\mathbf{E}$ of possible values of $E = E[y]$ for the case when the intervals $\mathbf{E}_1$ and $\mathbf{E}_2$ are degenerate: $\mathbf{E}_i = [E_i, E_i]$ – i.e., when we know the exact values of E_1 and E_2. In the following section, we show how to extend these formulas to the general case of non-degenerate intervals $\mathbf{E}_i$.

For *addition*, the answer is simple. Since $E[x_1 + x_2] = E[x_1] + E[x_2]$, if $y = x_1 + x_2$, there is only one possible value for $E = E[y]$: the value $E = E_1 + E_2$. This value does not depend on whether we have correlation or nor, and whether we have any information about the correlation.

Similarly, the answer is simple for *subtraction:* if $y = x_1 - x_2$, there is only one possible value for $E = E[y]$: the value $E = E_1 - E_2$.

For *multiplication*, if the variables x_1 and x_2 are independent, then $E[x_1 \cdot x_2] = E[x_1] \cdot E[x_2]$. Hence, if $y = x_1 \cdot x_2$ and x_1 and x_2 are independent, there is only one possible value for $E = E[y]$: the value $E = E_1 \cdot E_2$.

The first non-trivial case is the case of multiplication in the presence of possible correlation:

Theorem 1. *For multiplication $y = x_1 \cdot x_2$, when we have no information about the correlation,*

$$\underline{E} = \max(p_1 + p_2 - 1, 0) \cdot \overline{x}_1 \cdot \overline{x}_2 + \min(p_1, 1 - p_2) \cdot \overline{x}_1 \cdot \underline{x}_2 + \min(1 - p_1, p_2) \cdot \underline{x}_1 \cdot \overline{x}_2 +$$
$$\max(1 - p_1 - p_2, 0) \cdot \underline{x}_1 \cdot \underline{x}_2;$$

and

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

where $p_i \stackrel{\text{def}}{=} (E_i - \underline{x}_i)/(\overline{x}_i - \underline{x}_i)$.

Before we move to other results, let us make three comments:

- about the proofs;
- about the use of this result in computations, and
- about the intuitive meaning of this result.

The first comment is about the *proofs:* for the convenience of the readers who are more interested in the resulting formulas than in the details of the proofs, we have placed the proofs of all our results in the special Appendix.

The second comment is about the use of Theorem 1 in actual *computations.* As we have mentioned, the main objective of this paper is to produce computational tools. As of now, Theorem 1 simply provides a mathematical formula for computing $\underline{E}$ and $\overline{E}$. How can we actually use this formula in computations?

The answer to this question depends on which real numbers are represented in the actual computational device. If our computer system enables us to represent arbitrarily large integers, and also enables us to represent arbitrary rational numbers (fractions) and exact operations on rational numbers, then the formulas from Theorem 1 provide us with the exact bounds on $\underline{E}$ and $\overline{E}$.

In more traditional computer systems, only certain rational numbers are represented exactly. Our objective is to produce guaranteed bounds for E; therefore, we must provide a guaranteed lower bound for $\underline{E}$ and a guaranteed upper bound for $\overline{E}$. A natural way to do that is to use *interval computations* [10, 11, 12] to perform operations from the formulas from Theorem 1. As a result, due to round-off errors, for each of the values $\underline{E}$ and $\overline{E}$, instead of the exact rational value, we will have an interval of possible values: an interval $[\underline{E}^-, \underline{E}^+]$ for $\underline{E}$ and the interval $[\overline{E}^-, \overline{E}^+]$ for $\overline{E}$. Since we are interested in a lower bound for $\underline{E}$ and an upper bound for $\overline{E}$, we should therefore return $\underline{E}^-$ as the desired lower bound for $\underline{E}$ and $\overline{E}^+$ as the desired upper bound for $\overline{E}$.

Since round-off errors are the only reason why the resulting bounds $\underline{E}^-$ and $\overline{E}^+$ are different from the actual values $\underline{E}$ and $\overline{E}$, the more digits we allow in representing real numbers, the more accurate our bounds.

Is this the best way to compute the desired bounds for $\underline{E}$ and $\overline{E}$? Probably not. By using interval computations, we compute *four* bounds $\underline{E}^-$, $\underline{E}^+$, $\overline{E}^-$, and $\overline{E}^+$, while we are only interested in two of them. Probably, by appropriately specifying a rounding mode for each elementary operation involved in computing these formulas, we may be able to get the desired two bounds without computing the two unnecessary ones.

A positive answer to this question may not lead to a huge practical computation speedup – after all, the formulas from Theorem 1 are pretty simple – but it is an interesting and challenging question.

This comment applies not only to Theorem 1, but to all the results from this paper. As we will see, in all our theorems, we get simple algebraic expressions for $\underline{E}$ and $\overline{E}$. These expressions can be computed exactly if we use a computer system in which fractions are exactly represented, and they require interval computations (or maybe an appropriate selection of rounding) in more traditional computer systems, where some rational numbers can only be represented with a finite accuracy.

Our third comment is about *understanding* the result of Theorem 1, i.e., about the intuitive meaning of the corresponding formulas. At first glance, these formulas for multiplication may sound very technical, but in reality, as we will show, they make intuitive sense:

- 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 $[\underline{p}(A \,\&\, B), \overline{p}(A \,\&\, B)]$, where $\underline{p}(A \,\&\, B) \stackrel{\rm def}{=} \max(p(A) + p(B) - 1, 0)$ and $\overline{p}(A \,\&\, B) \stackrel{\rm def}{=} \min(p(A), p(B))$ (see, e.g., [21]). Indeed:
  - the largest possible intersection is the smallest of the two sets, and

- the smallest possible intersection is when they are as far apart as possible:
  * if $p(A)+p(B)\le 1$, they can be completely disjoint hence $\underline{p}(A\,\&\,B)=0$,
  * else we spread them as much as possible, so that $p(A\vee B)=1$ hence $p(A\,\&\,B)=p(A)+p(B)-p(A\,\&\,B)=p(A)+p(B)-1$.

From this viewpoint, since $p_1=p[\overline{x}_1]$ and $p_2=p[\overline{x}_2]$, we can interpret $\min(p_1,p_2)$ as $\overline{p}[\overline{x}_1\,\&\,\overline{x}_2]$. Similarly, we can interpret all other terms in the above formulas, so we can rewrite the formulas for $\underline{E}$ and $\overline{E}$ as follows:

$$\underline{E}=\underline{p}[\overline{x}_1\,\&\,\overline{x}_2]\cdot\overline{x}_1\cdot\overline{x}_2+\overline{p}[\overline{x}_1\,\&\,\underline{x}_2]\cdot\overline{x}_1\cdot\underline{x}_2+\overline{p}[\underline{x}_1\,\&\,\overline{x}_2]\cdot\underline{x}_1\cdot\overline{x}_2+$$
$$\underline{p}[\underline{x}_1\,\&\,\underline{x}_2]\cdot\underline{x}_1\cdot\underline{x}_2;$$
$$\overline{E}=\overline{p}[\overline{x}_1\,\&\,\overline{x}_2]\cdot\overline{x}_1\cdot\overline{x}_2+\underline{p}[\overline{x}_1\,\&\,\underline{x}_2]\cdot\overline{x}_1\cdot\underline{x}_2+\underline{p}[\underline{x}_1\,\&\,\overline{x}_2]\cdot\underline{x}_1\cdot\overline{x}_2+$$
$$\overline{p}[\underline{x}_1\,\&\,\underline{x}_2]\cdot\underline{x}_1\cdot\underline{x}_2.$$

Theorem 2. *For multiplication $y=x_1\cdot x_2$, when x_1 and x_2 are highly positively correlated, we have:*

$$\underline{E}=E_1\cdot E_2$$

and

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

Theorem 3. *For multiplication $y=x_1\cdot x_2$, when x_1 and x_2 are highly negatively correlated, we have:*

$$\underline{E}=\max(p_1+p_2-1,0)\cdot\overline{x}_1\cdot\overline{x}_2+\min(p_1,1-p_2)\cdot\overline{x}_1\cdot\underline{x}_2+\min(1-p_1,p_2)\cdot\underline{x}_1\cdot\overline{x}_2+$$
$$\max(1-p_1-p_2,0)\cdot\underline{x}_1\cdot\underline{x}_2;$$

and

$$\overline{E}=E_1\cdot E_2.$$

Comment. One can easily see that the only difference between these formulas (corresponding to high correlation) and the formulas from Theorem 1 (corresponding to all possible values of correlation) is that one of the bounds is replaced by the value $E_1\cdot E_2$. From the common sense viewpoint, it is very natural that the value $E_1 \cdot E_2$ corresponding to the absence of correlation separates cases with positive correlation (for which $E[x_1 \cdot x_2]$ is higher) and cases with negative correlation (for which $E[x_1 \cdot x_2]$ is lower).

For the *inverse* $y = 1/x_1$, the finite range is possible only when $0 \notin \mathbf{x}_1$. Without losing generality, we can consider the case when $0 < \underline{x}_1$. In this case, methods presented in [18] lead to the following bound:

Theorem 4. *For the inverse $y = 1/x_1$, the range of possible values of E is*

$$\mathbf{E} = \left[\frac{1}{E_1}, \frac{p_1}{\overline{x}_1} + \frac{1-p_1}{\underline{x}_1}\right].$$

(Here p_1 denotes the same value as in Theorems 1–3).

For $x_1 \circledast x_2 = \min(x_1, x_2)$ and $x_1 \circledast x_2 = \max(x_1, x_2)$, there is an easy case: if all the points from one of the intervals are not larger than all the points from another interval:

- If $\overline{x}_1 \le \underline{x}_2$, then $x_1 \in \mathbf{x}_1$ and $x_2 \in \mathbf{x}_2$ imply $x_1 \le x_2$. In this case, $\min(x_1, x_2) = x_1$ and $\max(x_1, x_2) = x_2$, hence $E[\min(x_1, x_2)] = E[x_1] = E_1$ and $E[\max(x_1, x_2)] = E[x_2] = E_2$.
- Similarly, if $\overline{x}_2 \le \underline{x}_1$, then $x_1 \in \mathbf{x}_1$ and $x_2 \in \mathbf{x}_2$ imply $x_2 \le x_1$. In this case, $\min(x_1, x_2) = x_2$ and $\max(x_1, x_2) = x_1$, hence $E[\min(x_1, x_2)] = E[x_2] = E_2$ and $E[\max(x_1, x_2)] = E[x_1] = E_1$.

What happens in all the other possible cases, i.e., when the intervals $\mathbf{x}_1$ and $\mathbf{x}_2$ have a non-degenerate intersection?

For min and max in case of independence, the results are as follows:

Theorem 5. *For minimum $y = \min(x_1, x_2)$, when x_1 and x_2 are independent, we have $\overline{E} = \min(E_1, E_2)$ and*

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

One can check that when $\overline{x}_1 \le \underline{x}_2$ (corr., when $\overline{x}_2 \le \underline{x}_1$), these formulas return the correct value E_1 (corr., E_2). The same is true for all the following theorems about min and max.

Theorem 6. *For maximum $y = \max(x_1, x_2)$, when x_1 and x_2 are independent, we have $\underline{E} = \max(E_1, E_2)$ and*

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

$$(1 - p_1) \cdot (1 - p_2) \cdot \max(\underline{x}_1, \underline{x}_2).$$

Comment. Both formulas have a natural probabilistic interpretation similar to the formulas from Theorem 1: indeed, e.g., $p_1 \cdot p_2$ is the probability $p[\overline{x}_1 \,\&\, \overline{x}_2]$ under the condition that x_1 and x_2 are independent random variables.

For the case when we have no information about the correlation between x_1 and x_2, we have the following results:

Theorem 7. *For minimum $y = \min(x_1, x_2)$, when we have no information about the correlation between x_1 and x_2, we have $\overline{E} = \min(E_1, E_2)$ and*

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

$$\min(1 - p_1, p_2) \cdot \min(\underline{x}_1, \overline{x}_2) + \max(1 - p_1 - p_2, 0) \cdot \min(\underline{x}_1, \underline{x}_2).$$

Theorem 8. *For maximum $y = \max(x_1, x_2)$, when we have no information about the correlation between x_1 and x_2, we have $\underline{E} = \max(E_1, E_2)$ and*

$$\overline{E} = \min(p_1, p_2) \cdot \max(\overline{x}_1, \overline{x}_2) + \max(p_1 - p_2, 0) \cdot \max(\overline{x}_1, \underline{x}_2) +$$

$$\max(p_2 - p_1, 0) \cdot \max(\underline{x}_1, \overline{x}_2) + \min(1 - p_1, 1 - p_2) \cdot \max(\underline{x}_1, \underline{x}_2).$$

What if we have high correlation? Let us first describe two easier cases:

Theorem 9. *For minimum $y = \min(x_1, x_2)$, when x_1 and x_2 are highly negatively correlated, the bounds $\underline{E}$ and $\overline{E}$ are the same as in Theorem 7.*

Theorem 10. *For maximum $y = \min(x_1, x_2)$, when x_1 and x_2 are highly positively correlated, the bounds $\underline{E}$ and $\overline{E}$ are the same as in Theorem 8.*

The other two cases are somewhat more complex:

Theorem 11. *For minimum $y = \min(x_1, x_2)$, when x_1 and x_2 are highly positively correlated, $\overline{E} = \min(E_1, E_2)$, and $\underline{E}$ is the smallest of the following values:*

- *the value $\min(E_1, E_2)$ that corresponds to a 1-point distribution;*
- *the values $p_1 \cdot \min(\overline{x}_1, E_2) + (1 - p_1) \cdot \min(\underline{x}_1, E_2)$ and $p_2 \cdot \min(E_1, \overline{x}_2) + (1 - p_2) \cdot \min(E_1, \underline{x}_2)$ that correspond to 2-point distributions;*
- *the solutions to the problems*

$$p^{(1)} \cdot \underline{x}_1 + p^{(2)} \cdot x_2^{(1)} + p^{(3)} \cdot x_1^{(2)} \to \min$$

under the conditions

$$\underline{x}_1 < x_2^{(1)} < x_1^{(2)} < \overline{x}_2; \quad x_1^{(2)} \le \overline{x}_1; \quad \underline{x}_2 \le x_2^{(1)};$$

$$p^{(1)} + p^{(2)} + p^{(3)} = 1;$$

$$p^{(1)} \cdot \underline{x}_1 + (p^{(2)} + p^{(3)}) \cdot x_1^{(2)} = E_1;$$

$$(p^{(1)} + p^{(2)}) \cdot x_2^{(1)} + p^{(3)} \cdot \overline{x}_2 = E_2;$$

and

$$p^{(1)} \cdot \underline{x}_2 + p^{(2)} \cdot x_1^{(1)} + p^{(3)} \cdot x_2^{(2)} \to \min$$

under the conditions

$$\underline{x}_2 < x_1^{(1)} < x_2^{(2)} < \overline{x}_1; \quad x_2^{(2)} \le \overline{x}_2; \quad \underline{x}_1 \le x_1^{(1)};$$

$$p^{(1)} + p^{(2)} + p^{(3)} = 1;$$

$$(p^{(1)} + p^{(2)}) \cdot x_2^{(1)} + p^{(3)} \cdot \overline{x}_2 = E_1;$$

$$p^{(1)} \cdot \underline{x}_2 + (p^{(2)} + p^{(3)}) \cdot x_2^{(2)} = E_2.$$

Comment. The corresponding optimization problems are not difficult to solve. Indeed, e.g., in the first problem:

$$p^{(1)} \cdot \underline{x}_1 + p^{(2)} \cdot x_2^{(1)} + p^{(3)} \cdot x_1^{(2)} \to \min$$

under the conditions

$$\underline{x}_1 < x_2^{(1)} < x_1^{(2)} < \overline{x}_2; \quad x_1^{(2)} \le \overline{x}_1; \quad \underline{x}_2 \le x_2^{(1)};$$

$$p^{(1)} + p^{(2)} + p^{(3)} = 1;$$

$$p^{(1)} \cdot \underline{x}_1 + (p^{(2)} + p^{(3)}) \cdot x_1^{(2)} = E_1;$$

$$(p^{(1)} + p^{(2)}) \cdot x_2^{(1)} + p^{(3)} \cdot \overline{x}_2 = E_2;$$

once we fix $p^{(2)}$ and $p^{(3)}$, we can describe $p^{(1)}$a s $1-p^{(2)}-p^{(3)}$, and then explicitly describe $x_2^{(1)}$ and $x_1^{(2)}$ from the equations containing these values:

$$x_1^{(2)} = \frac{E_1 - p^{(1)}\underline{x}_1}{p^{(2)} + p^{(3)}}; \quad \cdot x_2^{(1)} = \frac{E_2 - p^{(3)} \cdot \overline{x}_2}{1 - p^{(3)}}.$$

Substituting these expressions into the optimized function, we get an explicit expression for it in terms of $p^{(2)}$ and $p^{(3)}$:

$$(1 - p^{(2)} - p^{(3)}) \cdot \underline{x}_1 + p^{(2)} \cdot \frac{E_2 - p^{(3)} \cdot \overline{x}_2}{1 - p^{(3)}} + p^{(3)} \cdot \frac{E_1 - p^{(1)}\underline{x}_1}{p^{(2)} + p^{(3)}} \to \min$$

The minimum is attained either at the endpoints of the corresponding intervals, or at a stationary point, where both partial derivatives are 0. Differentiating with respect to $p^{(2)}$ and $p^{(3)}$, we get simple equations relating $p^{(2)}$ and $p^{(3)}$.

Theorem 12. *For maximum $y = \max(x_1, x_2)$, when x_1 and x_2 are highly positively correlated, $\underline{E} = \max(E_1, E_2)$, and $\overline{E}$ is the largest of the following values:*

- *the value $\max(E_1, E_2)$ that corresponds to a 1-point distribution;*
- *the values $p_1 \cdot \max(\overline{x}_1, E_2) + (1 - p_1) \cdot \max(\underline{x}_1, E_2)$ and $p_2 \cdot \max(E_1, \overline{x}_2) + (1 - p_2) \cdot \max(E_1, \underline{x}_2)$ that correspond to 2-point distributions;*
- *the solutions to the problems*

$$p^{(1)} \cdot \overline{x}_1 + p^{(2)} \cdot x_2^{(1)} + p^{(3)} \cdot x_1^{(2)} \to \max$$

under the conditions

$$\overline{x}_1 > x_2^{(1)} > x_1^{(2)} > \underline{x}_2; \quad x_1^{(2)} \geq \underline{x}_1; \quad \overline{x}_2 \geq x_2^{(1)};$$

$$p^{(1)} + p^{(2)} + p^{(3)} = 1;$$

$$p^{(1)} \cdot \overline{x}_1 + (p^{(2)} + p^{(3)}) \cdot x_1^{(2)} = E_1;$$

$$(p^{(1)} + p^{(2)}) \cdot x_2^{(1)} + p^{(3)} \cdot \underline{x}_2 = E_2;$$

and

$$p^{(1)} \cdot \overline{x}_2 + p^{(2)} \cdot x_1^{(1)} + p^{(3)} \cdot x_2^{(2)} \to \max$$

under the conditions

$$\overline{x}_2 > x_1^{(1)} > x_2^{(2)} > \underline{x}_1; \quad x_2^{(2)} \geq \underline{x}_2; \quad \overline{x}_1 \geq x_1^{(1)};$$

$$p^{(1)} + p^{(2)} + p^{(3)} = 1;$$

$$(p^{(1)} + p^{(2)}) \cdot x_2^{(1)} + p^{(3)} \cdot \underline{x}_1 = E_1;$$

$$p^{(1)} \cdot \overline{x}_2 + (p^{(2)} + p^{(3)}) \cdot x_2^{(2)} = E_2.$$

4 General Case: When Intervals $\mathbf{E}_i$ Are Non-Degenerate

In the previous section, we showed the bounds $\underline{E}$ and $\overline{E}$ for the moment $E = E[x_1 \circledast x_2]$ for the case when we know the moments $E_i = E[x_i]$. We described these bounds for each of four correlation situations:

i : when the variables x_1 and x_2 are *i*ndependent;

u : when the correlation between x_1 and x_2 is *u*nknown;

p : when we know that x_1 and x_2 are highly *p*ositively correlated;

n : when we know that x_1 and x_2 are highly *n*egatively correlated.

Namely, for each arithmetic operation $\circledast$ and for each correlation situation c, we described these bounds as explicit functions of E_1 and E_2:

$$\underline{E} = \underline{f}^c_{\circledast}(E_1, E_2); \quad \overline{E} = \overline{f}^c_{\circledast}(E_1, E_2).$$

If we only know the intervals $\mathbf{E}_1 = [\underline{E}_1, \overline{E}_1]$ and $\mathbf{E}_2 = [\underline{E}_2, \overline{E}_2]$ of possible values of E_1 and E_2, then the set of possible values for E is a union of the sets of possible bounds for all $E_1 \in \mathbf{E}_1$ and for all $E_2 \in \mathbf{E}_2$. Thus, the resulting bounds $\underline{E}$ and $\overline{E}$ can be described by the following formulas:

$$\underline{E} = \inf_{E_1 \in \mathbf{E}_1, E_2 \in \mathbf{E}_2} \underline{f}^c_{\circledast}(E_1, E_2); \quad \overline{E} = \sup_{E_1 \in \mathbf{E}_1, E_2 \in \mathbf{E}_2} \overline{f}^c_{\circledast}(E_1, E_2).$$

Let us show that for the elementary arithmetic operations, these formulas can be simplified into explicit analytical expressions.

For *addition* $\circledast = +$, $\underline{f}_+(E_1, E_2) = \overline{f}_+(E_1, E_2) = E_1 + E_2$, therefore, the resulting bounds $\underline{E}$ and $\overline{E}$ can be obtained by simply applying interval arithmetic: $\underline{E} = \underline{E}_1 + \underline{E}_2$ and $\overline{E} = \overline{E}_1 + \overline{E}_2$.

For *subtraction* $\circledast = -$, $\underline{f}_-(E_1, E_2) = \overline{f}_-(E_1, E_2) = E_1 - E_2$, so we can also use interval arithmetic: $\underline{E} = \underline{E}_1 - \overline{E}_2$ and $\overline{E} = \overline{E}_1 - \underline{E}_2$.

For *multiplication* $\circledast = \times$, for the case of independent variables ($c = i$),

$$\underline{f}^i_\times(E_1, E_2) = \overline{f}^i_\times(E_1, E_2) = E_1 \cdot E_2,$$

so we can use interval arithmetic as well: $\mathbf{E} = \mathbf{E}_1 \cdot \mathbf{E}_2$, i.e.,

$$\underline{E} = \min(\underline{E}_1 \cdot \underline{E}_2, \underline{E}_1 \cdot \overline{E}_2, \overline{E}_1 \cdot \underline{E}_2, \overline{E}_1 \cdot \overline{E}_2);$$

$$\overline{E} = \max(\underline{E}_1 \cdot \underline{E}_2, \underline{E}_1 \cdot \overline{E}_2, \overline{E}_1 \cdot \underline{E}_2, \overline{E}_1 \cdot \overline{E}_2).$$

For multiplication under no information about dependence ($c = u$), the formulas for $\underline{f}^u_\times(E_1, E_2)$ and $\underline{f}^u_\times(E_1, E_2)$ are given (by Theorem 1) in terms of the related probabilities $p_i = (E_i - \underline{x}_1)/(\overline{x}_i - \underline{x}_i)$, so it is reasonable to replace the intervals $\mathbf{E}_i$ by the corresponding intervals for probabilities:

$$\mathbf{p}_i = \frac{\mathbf{E}_i - \underline{x}_i}{\overline{E}_i - \underline{x}_i},$$

i.e., $\mathbf{p}_i = [\underline{p}_i, \overline{p}_i]$, where:

$$\underline{p}_i = \frac{\underline{E}_i - \underline{x}_i}{\overline{E}_i - \underline{x}_i}; \quad \overline{p}_i = \frac{\overline{E}_i - \underline{x}_i}{\overline{E}_i - \underline{x}_i}.$$

In terms of probability intervals, we get the following results:

Theorem 13. *For multiplication under no information about dependence, to find $\underline{E}$, it is sufficient to consider the following combinations of p_1 and p_2:*

- $p_1 = \underline{p}_1$ *and* $p_2 = \underline{p}_2$; $p_1 = \underline{p}_1$ *and* $p_2 = \overline{p}_2$; $p_1 = \overline{p}_1$ *and* $p_2 = \underline{p}_2$; $p_1 = \overline{p}_1$ *and* $p_2 = \overline{p}_2$;
- $p_1 = \max(\underline{p}_1, 1 - \overline{p}_2)$ *and* $p_2 = 1 - p_1$ (*if* $1 \in \mathbf{p}_1 + \mathbf{p}_2$); *and*
- $p_1 = \min(\overline{p}_1, 1 - \underline{p}_2)$ *and* $p_2 = 1 - p_1$ (*if* $1 \in \mathbf{p}_1 + \mathbf{p}_2$).

The smallest value of $\underline{f}^u_\times(p_1, p_2)$ for all these cases is the desired lower bound $\underline{E}$.

Theorem 14. *For multiplication under no information about dependence, to find $\overline{E}$, it is sufficient to consider the following combinations of p_1 and p_2:*

- $p_1 = \underline{p}_1$ *and* $p_2 = \underline{p}_2$; $p_1 = \underline{p}_1$ *and* $p_2 = \overline{p}_2$; $p_1 = \overline{p}_1$ *and* $p_2 = \underline{p}_2$; $p_1 = \overline{p}_1$ *and* $p_2 = \overline{p}_2$;
- $p_1 = p_2 = \max(\underline{p}_1, \underline{p}_2)$ (*if* $\mathbf{p}_1 \cap \mathbf{p}_2 \neq \emptyset$); *and*
- $p_1 = p_2 = \min(\overline{p}_1, \overline{p}_2)$ (*if* $\mathbf{p}_1 \cap \mathbf{p}_2 \neq \emptyset$).

The largest value of $\overline{f}^u_\times(p_1, p_2)$ for all these cases is the desired upper bound $\overline{E}$.

Comment. The fact that we need to consider several cases, and then take the maximum for find $\overline{E}$ and the minimum to find $\underline{E}$, is not surprising: a similar procedure is used in interval arithmetic to compute the range for the product of two intervals.

For multiplication in the case of high positive correlation, we have

$$\underline{f}^p_\times(E_1, E_2) = E_1 \cdot E_2,$$

hence the smallest possible value $\underline{E}$ of E when $E_1 \in \mathbf{E}_1$ and $E_2 \in \mathbf{E}_2$ can be computed as

$$\underline{E} = \min(\underline{E}_1 \cdot \underline{E}_2, \underline{E}_1 \cdot \overline{E}_2, \overline{E}_1 \cdot \underline{E}_2, \overline{E}_1 \cdot \overline{E}_2).$$

The upper bound $\overline{f}^p_\times(E_1, E_2)$ is the same as for the case $c = u$, so to compute $\overline{E}$, we can use the algorithm presented in Theorem 14.

For multiplication in the case of high negative correlation, we have

$$\overline{f}^p_\times(E_1, E_2) = E_1 \cdot E_2,$$

hence the largest possible value $\overline{E}$ of E when $E_1 \in \mathbf{E}_1$ and $E_2 \in \mathbf{E}_2$ can be computed as

$$\overline{E} = \max(\underline{E}_1 \cdot \underline{E}_2, \underline{E}_1 \cdot \overline{E}_2, \overline{E}_1 \cdot \underline{E}_2, \overline{E}_1 \cdot \overline{E}_2).$$

The lower bound $\underline{f}^p_\times(E_1, E_2)$ is the same as for the case $c = u$, so to compute $\underline{E}$, we can use the algorithm presented in Theorem 13.

Theorem 15. *When $0 < \underline{x}_1$, for the inverse $y = 1/x_1$, the range of possible values of E is*

$$\mathbf{E} = \left[\frac{1}{\overline{E}_1}, \frac{\underline{p}_1}{\overline{x}_1} + \frac{1 - \underline{p}_1}{\underline{x}_1}\right].$$

Theorem 16. *For minimum $y = \min(x_1, x_2)$, when x_1 and x_2 are independent, we have:*

$$\underline{E} = \underline{p}_1 \cdot \underline{p}_2 \cdot \min(\overline{x}_1, \overline{x}_2) + \underline{p}_1 \cdot (1 - \underline{p}_2) \cdot \min(\overline{x}_1, \underline{x}_2) + (1 - \underline{p}_1) \cdot \underline{p}_2 \cdot \min(\underline{x}_1, \overline{x}_2) +$$
$$(1 - \underline{p}_1) \cdot (1 - \underline{p}_2) \cdot \min(\underline{x}_1, \underline{x}_2);$$

and

$$\overline{E} = \min(\overline{E}_1, \overline{E}_2).$$

Theorem 17. *For maximum $y = \max(x_1, x_2)$, when x_1 and x_2 are independent, we have:*

$$\underline{E} = \max(\underline{E}_1, \underline{E}_2)$$

and

$$\overline{E} = \overline{p}_1 \cdot \overline{p}_2 \cdot \max(\overline{x}_1, \overline{x}_2) + \overline{p}_1 \cdot (1 - \overline{p}_2) \cdot \max(\overline{x}_1, \underline{x}_2) + (1 - \overline{p}_1) \cdot \overline{p}_2 \cdot \max(\underline{x}_1, \overline{x}_2) +$$
$$(1 - \overline{p}_1) \cdot (1 - \overline{p}_2) \cdot \max(\underline{x}_1, \underline{x}_2).$$

Theorem 18. *For minimum under no information about dependence,* $\overline{E} = \min(\overline{E}_1, \overline{E}_2)$*; to find* $\underline{E}$*, we must consider all combinations of* p_1 *and* p_2 *from Theorem* 13 *and take the smallest possible value of* $\underline{f}^u_{\min}(p_1, p_2)$ *for all these combinations.*

Theorem 19. *For maximum under no information about dependence,* $\underline{E} = \max(\underline{E}_1, \underline{E}_2)$*; to find* $\overline{E}$*, we must consider all combinations of* p_1 *and* p_2 *from Theorem* 14 *and take the largest possible value of* $\overline{f}^u_{\max}(p_1, p_2)$ *for all these combinations.*

Theorem 20. *For minimum* $y = \min(x_1, x_2)$*, when* x_1 *and* x_2 *are highly negatively correlated, the bounds* $\underline{E}$ *and* $\overline{E}$ *are the same as in Theorem* 18*.*

Theorem 21. *For maximum* $y = \min(x_1, x_2)$*, when* x_1 *and* x_2 *are highly positively correlated, the bounds* $\underline{E}$ *and* $\overline{E}$ *are the same as in Theorem* 19*.*

For the other two cases, we only have explicit formulas for one of the bounds:

Theorem 22. *For minimum under highly positive correlation,* $\overline{E} = \min(\overline{E}_1, \overline{E}_2)$.

Theorem 23. *For maximum under highly negative correlation,* $\underline{E} = \max(\underline{E}_1, \underline{E}_2)$.

5 From Elementary Arithmetic Operations to General Algorithms

So far, we have discussed how to find the intervals of possible values for $E[y]$ for the case when $y = f(x_1, \ldots, x_n)$ is an elementary arithmetic operation. In practice, of course, we are interested largely in the situations when f is a complex algorithm. How can we find the set of possible values of E for this more complex case?

One possibility, as we have mentioned, is to follow the example of straightforward interval computations: represent the algorithm f as a "code list" (a sequence of elementary arithmetic operation), and replace each operation by the corresponding operation with the pairs $\langle \mathbf{x}_i, \mathbf{E}_i \rangle$. Similarly to the case of interval computations, we can prove that the resulting interval $\widetilde{\mathbf{E}}$ is an enclosure for the desired interval $\mathbf{E}$.

Sometimes we get the exact interval, but often we get a proper superset of the desired interval $\mathbf{E}$. How can we find the actual range of $E = E[y]$? At first glance, the exact formulation of this problem requires that we use infinitely many variables, because we must describe all possible probability distributions on the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$ (or, in the independent case, all possible tuples consisting of distributions on all n intervals $\mathbf{x}_1, \ldots, \mathbf{x}_n$). It turns out, however, that we can reformulate these problems in equivalent forms that require only finitely many variables:

Theorem 24. *Let n be an integer, $\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{E}_1\subseteq\mathbf{x}_1,\ldots,\mathbf{E}_n\subseteq\mathbf{x}_n$ be intervals, and let $f(x_1,\ldots,x_n)$ be a continuous function of n real variables. Then, the range $\mathbf{E}=[\underline{E},\overline{E}]$ of possible values of $E[y]$, where $y=f(x_1,\ldots,x_n)$, over all possible distributions on the box $\mathbf{x}_1\times\ldots\times\mathbf{x}_n$ for which $x_i\in\mathbf{x}_i$ and $E[x_i]\in\mathbf{E}_i$, can be determined as follows:*

- *$\underline{E}$ is a solution to the following constraint optimization problem:*

$$\sum_{j=0}^{n} p^{(j)}\cdot f(x_1^{(j)},\ldots,x_n^{(j)})\to\min$$

under the conditions

$$\sum_{j=0}^{n} p^{(j)}=1;$$

$$p^{(j)}\ge 0,\quad \text{for all } j;$$

$$\underline{x}_i\le x_i^{(j)}\le\overline{x}_i\quad\text{for all } i,j;$$

$$\underline{E}_i\le\sum_{j=0}^{n} p^{(j)}\cdot x_i^{(j)}\le\overline{E}_i\quad\text{for all } i.$$

- *$\overline{E}$ is a solution to the constraint optimization problem:*

$$\sum_{j=0}^{n} p^{(j)}\cdot f(x_1^{(j)},\ldots,x_n^{(j)})\to\max$$

under the same constraints.

Theorem 25. *Let n be an integer, $\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{E}_1\subseteq\mathbf{x}_1,\ldots,\mathbf{E}_n\subseteq\mathbf{x}_n$ be intervals, and let $f(x_1,\ldots,x_n)$ be a continuous function of n real variables. Then, the range $\mathbf{E}=[\underline{E},\overline{E}]$ of possible values of $E[y]$, where $y=f(x_1,\ldots,x_n)$, over all possible independent distributions $\mathbf{x}_1,\ldots,\mathbf{x}_n$ for which $x_i\in\mathbf{x}_i$ and $E[x_i]\in\mathbf{E}_i$, can be determined as follows:*

- *$\underline{E}$ is a solution to the following constraint optimization problem:*

$$\sum_{\varepsilon_1\in\{-,+\}}\cdots\sum_{\varepsilon_n\in\{-,+\}} p_1^{\varepsilon_1}\cdot\ldots\cdot p_n^{\varepsilon_n}\cdot f(x_1^{\varepsilon_1},\ldots,x_n^{\varepsilon_n})\to\min$$

under the conditions

$$p_i^-+p_i^+=1\quad\text{for all } i;$$

$$p_i^-\ge 0 \text{ and } p_i^+\ge 0,\quad\text{for all } i;$$

$$\underline{x}_i\le x_i^-\le x_i^+\le\overline{x}_i\quad\text{for all } i;$$

$$\underline{E}_i\le p^-\cdot x_i^-+p_i^+\cdot x_i^+\le\overline{E}_i\quad\text{for all } i.$$

- $\overline{E}$ *is a solution to the constraint optimization problem:*

$$\sum_{\varepsilon_1 \in \{-,+\}} \ldots \sum_{\varepsilon_n \in \{-,+\}} p_1^{\varepsilon_1} \cdot \ldots \cdot p_n^{\varepsilon_n} \cdot f(x_1^{\varepsilon_1}, \ldots, x_n^{\varepsilon_n}) \to \max$$

  under the same constraints.

For convex and concave functions, these results can be further simplified:

Theorem 26. *Let n be an integer, $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{E}_1 \subseteq \mathbf{x}_1, \ldots, \mathbf{E}_n \subseteq \mathbf{x}_n$ be intervals, and let $f(x_1, \ldots, x_n)$ be a convex function of n real variables. Then, the range $\mathbf{E} = [\underline{E}, \overline{E}]$ of possible values of $E[y]$, where $y = f(x_1, \ldots, x_n)$, over all possible distributions on the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$ for which $x_i \in \mathbf{x}_i$ and $E[x_i] \in \mathbf{E}_i$, is as follows:*

$$\underline{E} = f(E_1, \ldots, E_n),$$

and $\overline{E}$ is a solution to the following constraint optimization problem:

$$\sum_{j=0}^{n} p^{(j)} \cdot f(x_1^{(j)}, \ldots, x_n^{(j)}) \to \max$$

under the conditions

$$\sum_{j=0}^{n} p^{(j)} = 1;$$

$$p^{(j)} \geq 0, \quad \text{for all } j;$$

$$x_i^{(j)} = \underline{x}_i \text{ or } x_i^{(j)} = \overline{x}_i \quad \text{for all } i, j;$$

$$\underline{E}_i \leq \sum_{j=0}^{n} p^{(j)} \cdot x_i^{(j)} \leq \overline{E}_i \quad \text{for all } i.$$

Theorem 27. *Let n be an integer, $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{E}_1 \subseteq \mathbf{x}_1, \ldots, \mathbf{E}_n \subseteq \mathbf{x}_n$ be intervals, and let $f(x_1, \ldots, x_n)$ be a concave function of n real variables. Then, the range $\mathbf{E} = [\underline{E}, \overline{E}]$ of possible values of $E[y]$, where $y = f(x_1, \ldots, x_n)$, over all possible distributions on the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$ for which $x_i \in \mathbf{x}_i$ and $E[x_i] \in \mathbf{E}_i$, is as follows:*

$$\overline{E} = f(E_1, \ldots, E_n),$$

and $\underline{E}$ is a solution to the following constraint optimization problem:

$$\sum_{j=0}^{n} p^{(j)} \cdot f(x_1^{(j)}, \ldots, x_n^{(j)}) \to \min$$

under the conditions

$$\sum_{j=0}^{n} p^{(j)} = 1;$$

$$p^{(j)} \geq 0, \quad \text{for all } j;$$

$$x_i^{(j)} = \underline{x}_i \text{ or } x_i^{(j)} = \overline{x}_i \quad \text{for all } i, j;$$

$$\underline{E}_i \leq \sum_{j=0}^{n} p^{(j)} \cdot x_i^{(j)} \leq \overline{E}_i \quad \text{for all } i.$$

Theorem 28. *Let n be an integer, $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{E}_1 \subseteq \mathbf{x}_1, \ldots, \mathbf{E}_n \subseteq \mathbf{x}_n$ be intervals, and let $f(x_1, \ldots, x_n)$ be a convex function of n real variables. Then, the range $\mathbf{E} = [\underline{E}, \overline{E}]$ of possible values of $E[y]$, where $y = f(x_1, \ldots, x_n)$, over all possible independent distributions $\mathbf{x}_1, \ldots, \mathbf{x}_n$ for which $x_i \in \mathbf{x}_i$ and $E[x_i] \in \mathbf{E}_i$, is as follows:*

$$\underline{E} = f(E_1, \ldots, E_n);$$

$$\overline{E} = \sum_{\varepsilon_1 \in \{-,+\}} \cdots \sum_{\varepsilon_n \in \{-,+\}} p_1^{\varepsilon_1} \cdot \ldots \cdot p_n^{\varepsilon_n} \cdot f(x_1^{\varepsilon_1}, \ldots, x_n^{\varepsilon_n}),$$

where $p_i^+ \stackrel{\text{def}}{=} p_i$, $p_i^- \stackrel{\text{def}}{=} 1 - p_i$, $x_i^- \stackrel{\text{def}}{=} \underline{x}_i$ and $x_i^+ \stackrel{\text{def}}{=} \overline{x}_i$.

Theorem 29. *Let n be an integer, $\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{E}_1 \subseteq \mathbf{x}_1, \ldots, \mathbf{E}_n \subseteq \mathbf{x}_n$ be intervals, and let $f(x_1, \ldots, x_n)$ be a concave function of n real variables. Then, the range $\mathbf{E} = [\underline{E}, \overline{E}]$ of possible values of $E[y]$, where $y = f(x_1, \ldots, x_n)$, over all possible independent distributions $\mathbf{x}_1, \ldots, \mathbf{x}_n$ for which $x_i \in \mathbf{x}_i$ and $E[x_i] \in \mathbf{E}_i$, is as follows:*

$$\underline{E} = \sum_{\varepsilon_1 \in \{-,+\}} \cdots \sum_{\varepsilon_n \in \{-,+\}} p_1^{\varepsilon_1} \cdot \ldots \cdot p_n^{\varepsilon_n} \cdot f(x_1^{\varepsilon_1}, \ldots, x_n^{\varepsilon_n});$$

$$\overline{E} = f(E_1, \ldots, E_n),$$

where $p_i^+ \stackrel{\text{def}}{=} p_i$, $p_i^- \stackrel{\text{def}}{=} 1 - p_i$, $x_i^- \stackrel{\text{def}}{=} \underline{x}_i$ and $x_i^+ \stackrel{\text{def}}{=} \overline{x}_i$.

Similar results can be proven for the case when $n = 2$ and x_1 and x_2 are highly correlated:

Theorem 30. *Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{E}_1 \subseteq \mathbf{x}_1, \mathbf{E}_2 \subseteq \mathbf{x}_2$ be intervals, and let $f(x_1, x_2)$ be a continuous function of two real variables. Then, the range $\mathbf{E} = [\underline{E}, \overline{E}]$ of possible values of $E[y]$, where $y = f(x_1, x_2)$, over all possible distributions highly positively correlated distributions on the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$ for which $x_i \in \mathbf{x}_i$ and $E[x_i] \in \mathbf{E}_i$, can be determined as follows:*

- *$\underline{E}$ is a solution to the following constraint optimization problem:*

$$\sum_{j=0}^{2} p^{(j)} \cdot f(x_1^{(j)}, x_n^{(j)}) \to \min$$

under the conditions

$$\sum_{j=0}^{2} p^{(j)} = 1;$$

$$p^{(j)} \ge 0, \quad \text{for all } j;$$

$$\underline{x}_i \le x_i^{(j)} \le \overline{x}_i \quad \text{for all } i, j;$$

$$\underline{E}_i \le \sum_{j=0}^{n} p^{(j)} \cdot x_i^{(j)} \le \overline{E}_i \quad \text{for all } i;$$

$$x_i^{(j)} \le x_i^{(j+1)} \quad \text{for all } i, j.$$

- $\overline{E}$ *is a solution to the constraint optimization problem:*

$$\sum_{j=0}^{2} p^{(j)} \cdot f(x_1^{(j)}, x_2^{(j)}) \to \max$$

under the same constraints.

Theorem 31. *Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{E}_1 \subseteq \mathbf{x}_1, \mathbf{E}_2 \subseteq \mathbf{x}_2$ be intervals, and let $f(x_1, x_2)$ be a continuous function of two real variables. Then, the range $\mathbf{E} = [\underline{E}, \overline{E}]$ of possible values of $E[y]$, where $y = f(x_1, x_2)$, over all possible distributions highly negatively correlated distributions on the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$ for which $x_i \in \mathbf{x}_i$ and $E[x_i] \in \mathbf{E}_i$, can be determined as follows:*

- $\underline{E}$ *is a solution to the following constraint optimization problem:*

$$\sum_{j=0}^{2} p^{(j)} \cdot f(x_1^{(j)}, x_n^{(j)}) \to \min$$

under the conditions

$$\sum_{j=0}^{2} p^{(j)} = 1;$$

$$p^{(j)} \ge 0, \quad \text{for all } j;$$

$$\underline{x}_i \le x_i^{(j)} \le \overline{x}_i \quad \text{for all } i, j;$$

$$\underline{E}_i \le \sum_{j=0}^{n} p^{(j)} \cdot x_i^{(j)} \le \overline{E}_i \quad \text{for all } i;$$

$$x_1^{(j)} \le x_1^{(j+1)} \text{ and } x_2^{(j)} \ge x_2^{(j+1)} \quad \text{for all } j.$$

- $\overline{E}$ *is a solution to the constraint optimization problem:*

$$\sum_{j=0}^{2} p^{(j)} \cdot f(x_1^{(j)}, x_2^{(j)}) \to \max$$

under the same constraints.

6 From Intervals and First-Order Moments to Higher-Order Moments

So far, we have provided explicit formulas for the elementary arithmetic operations $f(x_1, \ldots, x_n)$ for the case when we know the first order moments. What if, in addition to that, we have some information about second order (and/or higher order) moments of x_i? What will we be then able to conclude about the moments of y? Partial answers to this question are given in [6, 18]; it is desirable to find a general answer.

Similarly to the case of first moments, we can reduce the corresponding problems to the constraint optimization problems with finitely many variables (the proof is similar). For example, when, in addition to intervals $\mathbf{E}_i$ that contain the first moments $E[x_i]$, we know the intervals $\mathbf{E}_{ik}$ that contain the second moments $E[x_i \cdot x_k]$, then the corresponding bounds $\underline{E}$ and $\overline{E}$ on $E[y]$ can be computed by solving the following problems:

$$\sum_{j=0}^{N} p^{(j)} \cdot f(x_1^{(j)}, \ldots, x_n^{(j)}) \to \min(\max)$$

under the conditions

$$\sum_{j=0}^{N} p^{(j)} = 1;$$

$$p^{(j)} \geq 0, \quad \text{for all } j;$$

$$\underline{x}_i \leq x_i^{(j)} \leq \overline{x}_i \quad \text{for all } i, j;$$

$$\underline{E}_i \leq \sum_{j=0}^{n} p^{(j)} \cdot x_i^{(j)} \leq \overline{E}_i \quad \text{for all } i;$$

$$\underline{E}_{ik} \leq \sum_{j=0}^{n} p^{(j)} \cdot x_i^{(j)} \cdot x_k^{(j)} \leq \overline{E}_{ik} \quad \text{for all } i, k,$$

where $N = n(n+1)/2$.

Bounds on $E[y^2]$ can be obtained in a similar way if instead of the original function $f(x_1, \ldots, x_n)$ that describes y in terms of x_i, we consider a new function $\widetilde{f}(x_1, \ldots, x_n) \stackrel{\text{def}}{=} f^2(x_1, \ldots, x_n)$ that describes y^2 in terms of x_i.

If we know that the variables x_i are *highly correlated*, then (similarly to Theorems 18 and 19) we can add the constraints

$$x_i^{(j)} \leq x_i^{(j+1)} \quad \text{for all } i, j$$

(for positive correlation) or

$$x_1^{(j)} \leq x_1^{(j+1)} \text{ and } x_2^{(j)} \geq x_2^{(j+1)} \quad \text{for all } j$$

(for negative correlation).

If we know that the variables x_i are *independent*, and we know the bounds $\mathbf{E}_i$ on $E[x_i]$ and $\mathbf{M}_i$ on $E[x_i^2]$, then the corresponding bounds $\underline{E}$ and $\overline{E}$ on $E[y]$ can be computed by solving the following problems:

$$\sum_{j_1=1}^{3} \ldots \sum_{j_n=1}^{3} p_1^{(j_1)} \cdot \ldots \cdot p_n^{(j_n)} \cdot f(x_1^{(j_1)}, \ldots, x_n^{(j_n)}) \to \min(\max)$$

under the conditions

$$\sum_{j=1}^{3} p_i^{(j)} = 1 \quad \text{for all } i;$$

$$p_i^{(j)} \geq 0, \quad \text{for all } i, j;$$

$$\underline{x}_i \leq x_i^{(j)} \leq \overline{x}_i \quad \text{for all } i, j;$$

$$\underline{E}_i \leq \sum_{j=1}^{3} p_i^{(j)} \cdot x_i^{(j)} \leq \overline{E}_i \quad \text{for all } i;$$

$$\underline{M}_i \leq \sum_{j=1}^{3} p_i^{(j)} \cdot \left(x_i^{(j)}\right)^2 \leq \overline{M}_i \quad \text{for all } i.$$

Conclusions

In addition to intervals, we sometimes know the means E_i or intervals $[\underline{E}_i, \overline{E}_i]$ that contain the corresponding means. In this paper, we have described how to generalize interval arithmetic to this new case.

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 SC2002 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 Faculty Member at the Fields Institute for Research in Mathematical Sciences (Toronto, Canada).

The authors are thankful to all the participants of the 2002 SIAM Workshop on Validated Computing and of the 2002 Informal Working Group on Validated Methods for Optimization, especially to Dan Berleant, George Corliss, R. Baker Kearfott, Weldon Lodwick, and Bill Walster, for valuable discussions.

We are specially thankful to the anonymous referee for valuable suggestions.

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] J. Bukowski, L. Korn, and D. Wartenberg, "Correlated inputs in quantitative risk assessments: the effects of distributional shape", *Risk Analysis*, 1995, Vol. 15, pp. 215–219.

[5] S. Ferson, W. Root and R. Kuhn, *RAMAS Risk Calc: Risk Assessment with Uncertain Numbers*, Applied Biomathematics, Setauket, NY, 1999.

[6] S. Ferson, D. Myers, and D. Berleant, *Distribution-free risk analysis: I. Range, mean, and variance*, Applied Biomath. Technical Report, 2001.

[7] M. J. Frank, R. B. Nelson, and B. Schweizer, "Best-possible bounds for the distribution of a sum – a problem of Kolmogorov", *Probability Theory and Related Fields*, 1987, Vol. 74, pp. 199–211.

[8] L. A. Goodman, "On the exact variance of products", *Journal of American Statistical Association*, 1960, Vol. 55, pp. 718–713.

[9] P. J. Huber, *Robust statistics,* Wiley, New York, 1981.

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

[11] R. B. Kearfott, *Rigorous Global Search: Continuous Problems*, Kluwer, Dordrecht, 1996.

[12] R. E. Moore, *Methods and Applications of Interval Analysis*, SIAM, Philadelphia, 1979.

[13] P. V. Novitskii and I. A. Zograph, *Estimating the Measurement Errors*, Leningrad: Energoatomizdat, 1991 (in Russian).

[14] A. I. Orlov, "How often are the observations normal?", *Industrial Laboratory*, 1991, Vol. 57, No. 7, pp. 770–772.

[15] J. Pesonen and E. Hyvönen, "Interval Approach Challenges Monte Carlo Simulation", *Reliable Computing*, 1996, Vol. 2, No. 2, pp. 155–160.

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

[17] K. H. Reckow and S. C. Chapra, *Engineering Approaches for Lake Management. Volume 1: Data Analysis and Empirical Modeling*, Butterworth Publishers, Boston, 1983.

[18] 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.

[19] W. Slob, "Uncertainty analysis in multiplicative models", *Risk Analysis*, 1994, Vol. 14, pp. 571–576.

[20] H. M. Wadsworth, Jr. (eds.), *Handbook of statistical methods for engineers and scientists*, McGraw-Hill Publishing Co., New York, 1990.

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

[22] G. W. Walster, "Philosophy and practicalities of interval arithmetic", In: R. E. Moore (ed.), *Reliability in Computing*, 1988, pp. 307–323.

[23] S. S. Wilks, *Mathematical Statistics*, John Wiley & Sons, New York, 1962.

[24] 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.

[25] P. Wiwatandate and H. G. Claycamp, "Exact propagation of uncertainties in multiplicative models", *Human and Ecological Risk Assessment*, 2000, Vol. 6, pp. 355–368.

Appendix: Proofs

Proof of Theorem 1. To get the desired bounds $\underline{E}$ and $\overline{E}$, we must consider the values $E[x_1 \cdot x_2]$ for all possible probability distributions on the box $\mathbf{x}_1 \times \mathbf{x}_2$ for which $E[x_1] = E_1$ and $E[x_2] = x_2$. To describe a general probability distribution, we must use infinitely many parameters, and hence, this problem is difficult to solve directly.

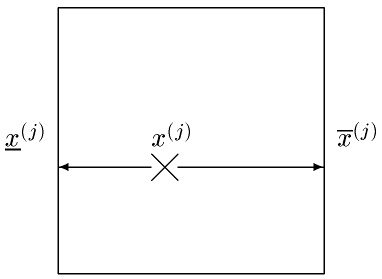
To make the problem simpler, we will show that a general distribution with $E[x_i] = E_i$ can be simplified without changing the values $E[x_i]$ and $E[x_1 \cdot x_2]$. Thus, to describe possible values of $E[x_1 \cdot x_2]$, we do not need to consider all possible distributions, it is sufficient to consider only the simplified ones.

We will describe the simplification for discrete distributions that concentrate on finitely many points $x^{(j)} = (x_1^{(j)}, x_2^{(j)})$, $1 \le j \le N$. 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 $x^{(j)} = (x_1^{(j)}, x_2^{(j)})$ that occurs with probability $p^{(j)}$, and replace it with two points: $\overline{x}^{(j)} = (\overline{x}_1, x_2^{(j)})$ with probability $p^{(j)} \cdot \overline{p}^{(j)}$ and $\underline{x}^{(j)} = (\underline{x}_1, x_2^{(j)})$ 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 $\underline{p}^{(j)} \stackrel{\text{def}}{=} 1 - \overline{p}^{(j)}$:



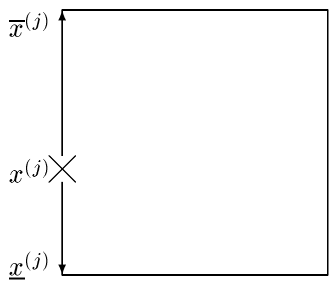
Here, the values $\overline{p}^{(j)}$ and $\underline{p}^{(j)} = 1 - \overline{p}^{(j)}$ are chosen in such a way that $\overline{p}^{(j)} \cdot \overline{x}_1 + \underline{p}^{(j)} \cdot \underline{x}_1 = x_1^{(j)}$. Due to this choice,

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

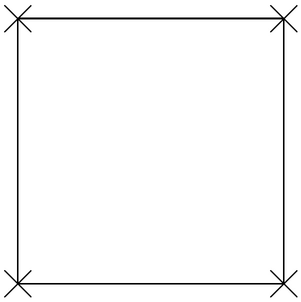
hence for the new distribution, the mathematical expectation $E[x_1]$ is the same as for the old one. Similarly, we can prove that the values $E[x_2]$ and $E[x_1 \cdot x_2]$ do not change.

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 $\le N-1$ 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 $E[x_1]$, $E[x_2]$, and $E[x_1 \cdot x_2]$ 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 $x_2 = \overline{x}_2$:



Thus, instead of considering all possible distributions, it is sufficient to consider only distributions for which $x_1 \in \{\underline{x}_1, \overline{x}_1\}$ and $x_2 \in \{\underline{x}_2, \overline{x}_2\}$. In other words, it is sufficient to consider only distributions which are located in the four corner points $(\underline{x}_1, \underline{x}_2)$, $(\underline{x}_1, \overline{x}_2)$, $(\overline{x}_1, \underline{x}_2)$, and $(\overline{x}_1, \overline{x}_2)$ of the box $\mathbf{x}_1 \times \mathbf{x}_2$:



Such distribution can be characterized by the probabilities of these four points; we will denote these probabilities by $p(\underline{x}_1 \,\&\, \underline{x}_2)$, $p(\underline{x}_1 \,\&\, \overline{x}_2)$, $p(\overline{x}_1 \,\&\, \underline{x}_2)$, and $p(\overline{x}_1 \,\&\, \overline{x}_2)$. So, we have a finite-parametric family of distributions.

These four probabilities cannot be given arbitrarily, they must satisfy three equations: their sum must be equal to 1:

$$p(\underline{x}_1 \,\&\, \underline{x}_2) + p(\underline{x}_1 \,\&\, \overline{x}_2) + p(\overline{x}_1 \,\&\, \underline{x}_2) + p(\overline{x}_1 \,\&\, \overline{x}_2) = 1;$$

and we must have the given values of $E[x_1]$ and $E[x_2]$.

Since x_1 takes only two possible values $\underline{x}_1$ and $\overline{x}_1$, the condition $E[x_1] = E_1$ can be described as

$$p(\underline{x}_1) \cdot \underline{x}_1 + p(\overline{x}_1) \cdot \overline{x}_1 = E_1,$$

where

$$p(\overline{x}_1) \stackrel{\text{def}}{=} p(\overline{x}_1 \,\&\, \underline{x}_2) + p(\overline{x}_1 \,\&\, \overline{x}_2)$$

is the total probability of $\overline{x}_1$, and $p(\underline{x}_1) = 1 - p(\overline{x}_1)$ is the total probability of $\underline{x}_1$. From the condition that

$$(1 - p(\overline{x}_1)) \cdot \underline{x}_1 + p(\overline{x}_1) \cdot \overline{x}_1 = E_1,$$

we conclude that

$$p(\overline{x}_1) = \frac{E_i - \underline{x}_i}{\overline{x}_i - \underline{x}_i},$$

i.e., that the probability $p(\overline{x}_1)$ coincides with the quantity p_1 defined in Theorem 1. Thus, we must have $p(\overline{x}_1 \,\&\, \underline{x}_2) + p(\overline{x}_1 \,\&\, \overline{x}_2) = p_1$.

Similarly, we can conclude that the probability $p(\overline{x}_2)$ (that $x_2 = \overline{x}_2$) coincides with the quantity p_2 defined in Theorem 1, and thus, that $p(\underline{x}_1 \,\&\, \overline{x}_2) + p(\overline{x}_1 \,\&\, \overline{x}_2) = p_2$.

We have three conditions on four probabilities. Thus, we have, in effect, a 1-dimensional family of distributions, for which it is much easier to find the smallest and the largest possible values of $E[x_1 \cdot x_2]$.

Reduction to a 4-corner distribution is a general fact, true for both desired bounds $\underline{E}$ and $\overline{E}$. Now, depending on which bound we want to estimate, we will perform different transformations. Since we are interested not only in the general case, but also in the cases of "co-monotonicity" (highly positive and highly negative correlation), it makes sense to ask a natural question: when does "co-monotonicity" increase $E[x_1 \cdot x_2]$? when does it decrease it?

To answer this question, we will also consider discrete distributions, this time – distributions in which N points $x^{(j)} = (x_1^{(j)}, x_2^{(j)})$ all have equal probabilities. It is also well known that every probability distribution on the box can be approximated by such distributions (in such a way that moments are approximated as well). For such distributions:

- highly positive correlation means that we can order the points $x^{(j)}$ in such a way that when $i < j$, we have $x_1^{(i)} \le x_1^{(j)}$ and $x_2^{(i)} \le x_2^{(j)}$;
- highly negative correlation means that we can order the points $x^{(j)}$ in such a way that when $i < j$, we have $x_1^{(i)} \le x_1^{(j)}$ and $x_2^{(i)} \ge x_2^{(j)}$.

Let us first consider the case of highly positive correlation. Reformulating the above property, we can see that a distribution is not highly positively correlated if and only if there exist points i and j for which $x_1^{(i)} < x_1^{(j)}$ and $x_2^{(i)} > x_2^{(j)}$. We can "correct" this obstacle to highly positive correlation by "swapping" the second components of these points, i.e., by replacing $x^{(i)}$ and $x^{(j)}$ with two new points $x^{(i)}_{\rm new} = (x_1^{(i)}, x_2^{(j)})$ and $x^{(j)}_{\rm new} = (x_1^{(j)}, x_2^{(i)})$. It is easy to see that this swap does not change $E[x_1] = (1/N) \cdot \sum\limits_{k=1}^{N} x_1^{(k)}$ and $E[x_2] = (1/N) \cdot \sum\limits_{k=1}^{N} x_2^{(k)}$. How does it affect $E[x_1 \cdot x_2] = (1/N) \cdot \sum\limits_{k=1}^{N} x_1^{(k)} \cdot x_2^{(k)}$? The only two terms that are changed are terms corresponding to $k = i$ and $k = j$:

- For the original points, the sum of these two terms is equal to $x_1^{(i)} \cdot x_2^{(i)} + x_1^{(j)} \cdot x_2^{(j)}$.
- For the new points, the corresponding sum is equal to $x_1^{(i)} \cdot x_2^{(j)} + x_1^{(j)} \cdot x_2^{(i)}$.
- Therefore, the difference between the new and the old values of $E[x_1 \cdot x_2]$ is equal to:

$$\frac{1}{N} \cdot (x_1^{(i)} \cdot x_2^{(j)} + x_1^{(j)} \cdot x_2^{(i)} - (x_1^{(i)} \cdot x_2^{(i)} + x_1^{(j)} \cdot x_2^{(j)})).$$

One can easily see that this difference is equal to

$$\frac{1}{N} \cdot (x_1^{(i)} - x_1^{(j)}) \cdot (x_2^{(j)} - x_2^{(i)}).$$

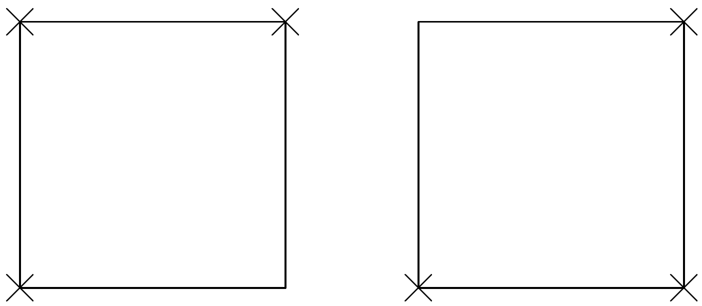
Thus, when we have a distribution which is not highly positively correlated, i.e., for which $x_1^{(i)} < x_1^{(j)}$ and $x_2^{(i)} > x_2^{(j)}$ for some i and j, the above swap not only deletes this violation of highly positive correlation, it also increases the value $E[x_1 \cdot x_2]$.

So, if we are interested in the maximum of $E[x_1 \cdot x_2]$, we can perform these swaps until no further swaps are possible (since there are only finitely many rearrangement of the original second components, and each swap increases the value of $E[x_1 \cdot x_2]$, this process has to stop). So, we arrive at the following conclusion: when we are looking for the maximum of $E[x_1 \cdot x_2]$, we only need to consider highly positively correlated distributions.

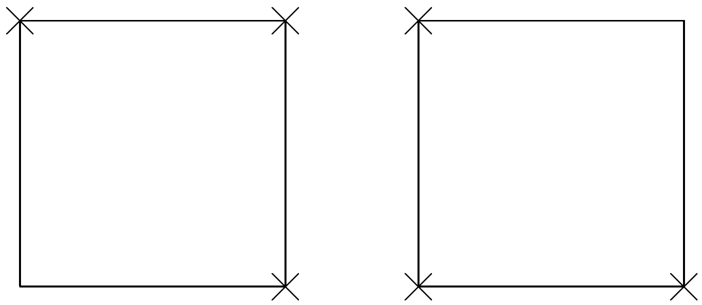
Similarly, a distribution is not highly negatively correlated if and only if there exist points i and j for which $x_1^{(i)} < x_1^{(j)}$ and $x_2^{(i)} < x_2^{(j)}$. In this case, a similar swap leaves $E[x_1]$ and $E[x_2]$ unchanged but decreases the value $E[x_1 \cdot x_2]$. Thus, when we are looking for the minimum of $E[x_1 \cdot x_2]$, we only need to consider highly negatively correlated distributions.

We have already proven that it is sufficient to consider only distributions located at the four corners of the box $\mathbf{x}_1 \times \mathbf{x}_2$. Now, we know something extra:

- If we look for the maximum of $E[x_1 \cdot x_2]$, then it is sufficient to consider the case of highly positive correlation. For a 4-corner distribution this means that we cannot have both points $(\underline{x}_1, \overline{x}_2)$ and $(\overline{x}_1, \underline{x}_2)$ – for one of these points, the probability should be equal to 0:



- If we look for the minimum of $E[x_1 \cdot x_2]$, then it is sufficient to consider the case of highly negative correlation. For a 4-corner distribution this means that we cannot have both points $(\underline{x}_1, \underline{x}_2)$ and $(\overline{x}_1, \overline{x}_2)$ – for one of these points, the probability should be equal to 0:



Let us first consider the case of $\overline{E}$.

- If $p(\overline{x}_1 \,\&\, \underline{x}_2) = 0$, then $p(\overline{x}_1 \,\&\, \overline{x}_2) = p_1$. So, from $p(\underline{x}_1 \,\&\, \overline{x}_2) + p(\overline{x}_1 \,\&\, \overline{x}_2) = p_2$, we can conclude that $p_2 \geq p_1$, and $p(\underline{x}_1 \,\&\, \overline{x}_2) = p_2 - p_1$. The remaining probability $p(\underline{x}_1 \,\&\, \underline{x}_2)$ can be determined from the condition that the sum of all three probabilities is 1, and is, therefore, equal to $1 - p_2$.
- If $p(\underline{x}_1 \,\&\, \overline{x}_2) = 0$, then $p(\overline{x}_1 \,\&\, \overline{x}_2) = p_2$. So, from $p(\overline{x}_1 \,\&\, \underline{x}_2) + p(\overline{x}_1 \,\&\, \overline{x}_2) = p_1$, we can conclude that $p_1 \geq p_2$, and $p(\overline{x}_1 \,\&\, \underline{x}_2) = p_1 - p_2$. The remaining probability $p(\underline{x}_1 \,\&\, \underline{x}_2)$ can be determined from the condition that the sum of all three probabilities is 1, and is, therefore, equal to $1 - p_1$.

One can see that in both cases,

$$p(\overline{x}_1 \,\&\, \overline{x}_2) = \min(p_1, p_2), p(\overline{x}_1 \,\&\, \underline{x}_2) = \max(p_1 - p_2, 0),$$

$$p(\underline{x}_1 \,\&\, \overline{x}_2) = \max(p_2 - p_1, 0), p(\underline{x}_1 \,\&\, \underline{x}_2) = \min(1 - p_1, 1 - p_2).$$

Therefore, the mathematical expectation $E[x_1 \cdot x_2]$ of $x_1 \cdot x_2$ is equal exactly to the expression from Theorem 1.

For $\underline{E}$, the situation is similar:

- If $p(\underline{x}_1 \,\&\, \underline{x}_2) = 0$, then $p(\underline{x}_1 \,\&\, \overline{x}_2) = p(\underline{x}_1) = 1 - p_1$ and $p(\overline{x}_1 \,\&\, \underline{x}_2) = p(\underline{x}_1) = 1 - p_2$. The remaining probability $p(\overline{x}_1 \,\&\, \overline{x}_2)$ can be determined from the condition that the sum of all three probabilities is 1, and is, therefore, equal to $1 - (1 - p_1) - (1 - p_2) = p_1 + p_2 - 1$. Since probabilities are non-negative, this is only possible when $p_1 + p_2 \geq 1$.

- If $p(\overline{x}_1 \,\&\, \overline{x}_2) = 0$, then $p(\overline{x}_1 \,\&\, \underline{x}_2) = p_1$ and $p(\overline{x}_1 \,\&\, \underline{x}_2) = p_2$. The remaining probability $p(\underline{x}_1 \,\&\, \underline{x}_2)$ can be determined from the condition that the sum of all three probabilities is 1, and is, therefore, equal to $1 - p_1 - p_2$. Since probabilities are non-negative, this is only possible when $p_1 + p_2 \leq 1$.

One can see that in both cases, $p(\overline{x}_1 \,\&\, \overline{x}_2) = \max(p_1 + p_2 - 1, 0)$, $p(\overline{x}_1 \,\&\, \underline{x}_2) = \min(p_1, 1 - p_2)$, $p(\underline{x}_1 \,\&\, \overline{x}_2) = \min(1 - p_1, p_2)$, and $p(\underline{x}_1 \,\&\, \underline{x}_2) = \max(1 - p_1 - p_2, 0)$. Therefore, the mathematical expectation $E[x_1 \cdot x_2]$ of $x_1 \cdot x_2$ is equal exactly to the expression from Theorem 1.

The theorem is proven.

Proof of Theorem 2. In the proof of Theorem 1, we have shown that the maximum of $E[x_1 \cdot x_2]$ is attained at a highly positively correlated distribution; therefore, the maximum $\overline{E}$ of $E[x_1 \cdot x_2]$ over all possible distributions should be equal to the maximum of $E[x_1 \cdot x_2]$ over all highly positively correlated distributions.

To complete the proof of the theorem, it is therefore sufficient to prove that the lower bound $\underline{E}$ is equal to $E_1 \cdot E_2$. The value $E_1 \cdot E_2$ can be attained by a degenerate distribution concentrated on the point (E_1, E_2) with probability 1. So, to complete the proof of the theorem, we will show that by replacing the original distribution by the degenerate distribution concentrated on its average (E_1, E_2), we decrease $E[x_1 \cdot x_2]$.

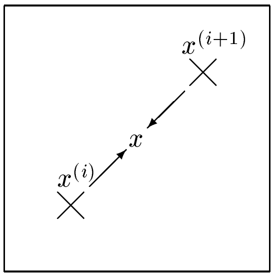
Similar to Theorem 1, we can, without loss of generality, consider discrete distributions that concentrate on finitely many points $x^{(j)} = (x_1^{(j)}, x_2^{(j)})$, $1 \leq j \leq N$. So, we have a probability distribution in which the point $x^{(1)}$ appears with the probability $p^{(1)}$, the point $x^{(2)}$ appears with the probability $p^{(2)}$, etc. As we have mentioned in the proof of Theorem 1, the fact that the distribution is highly positively correlated means that we can order the points $x^{(j)}$ in such a way that when $i < j$, we have $x_1^{(i)} \leq x_1^{(j)}$ and $x_2^{(i)} \leq x_2^{(j)}$. Let us therefore assume that the points are ordered this way.

We will show that when we replace, in our distribution, the two neighboring points $x^{(i)}$ and $x^{(i+1)}$ by their average $x = \alpha \cdot x^{(i+1)} + (1 - \alpha) \cdot x^{(i)}$ (where $\alpha = p^{(i+1)}/(p^{(i)} + p^{(i+1)})$) with probability $p^{(i)} + p^{(i+1)}$, then:

(1) we preserve $E[x_1]$ and $E[x_2]$;

(2) we preserve the highly positive correlation property, and

(3) the value $E[x_1 \cdot x_2]$ either decreases or stays the same.



Once this is proven, we will have a new highly positively correlated distribution with $N-1$ points and smaller (or same) value of $E[x_1 \cdot x_2]$. We can apply the same reduction to the new distribution, etc., until we have only a single point $x = (x_1, x_2)$ left. Since our transformation preserves the values $E[x_1]$ and $E[x_2]$, we have $x_1 = E_1$ and $x_2 = E_2$ and hence, $E[x_1 \cdot x_2] = E_1 \cdot E_2$. Since the value $E[x_1 \cdot x_2]$ cannot increase under our transformations, we can therefore conclude that the original value of $E[x_1 \cdot x_2]$ was larger or equal to $E_1 \cdot E_2$ – i.e., that $E_1 \cdot E_2$ is indeed the lower bound.

Let us proceed with proof of the three properties of the above transformation.

(1) Let us first prove that the above transformation preserves $E[x_1]$ (for $E[x_2]$ the proof is the same). Indeed, after the transformation, the only change in the original expression $E[x_1] = \sum\limits_{k=1}^{N} p^{(k)} \cdot x_1^{(k)}$ is that we replace the sum $p^{(i)} \cdot x_1^{(i)} + p^{(i+1)} \cdot x_1^{(i+1)}$ of i-th and $(i+1)$-th terms by the new term $(p^{(i)} + p^{(i+1)}) \cdot x_1$, i.e., by definition of x, by the term

$$(p^{(i)} + p^{(i+1)}) \cdot (\alpha \cdot x_1^{(i+1)} + (1-\alpha) \cdot x_1^{(i)}).$$

By definition of α, this expression is exactly equal to the original sum (this is why we chose the above α). So, the values $E[x_i]$ are indeed unchanged.

(2) It is also easy to show that the distribution continues to be strictly positively correlated. To prove it, we must show two things:

(2a) that if $k < i$, then $x_1^{(k)} \le x_1$ and $x_2^{(k)} \le x_2$; and

(2b) that if $i < k$, then $x_i \le x_1^{(k)}$ and $x_2 \le x_2^{(k)}$.

Let us prove the first property (2a) (the second property (2b) is proven in the same way). Since the original distribution was highly positively correlated, we have $x_1^{(k)} \le x_1^{(k)}$ and $x_1^{(k)} \le x_1^{(i+1)}$. Multiplying the first inequality by $\alpha \ge 0$ and the second one by $1-\alpha \ge 0$, we conclude that $x_k^{(k)} \le x_1$ (the proof is the same for the second component).

(3) Finally, let us prove that the above transformation decreases $E[x_1 \cdot x_2]$. Indeed, originally, we had $E[x_1 \cdot x_2] = \sum_{k=1}^{N} p^{(k)} \cdot x_1^{(k)} \cdot x_2^{(k)}$. After the transformation, the only change is that we replace the sum $p^{(i)} \cdot x_1^{(i)} \cdot x_2^{(i)} + p^{(i+1)} \cdot x_1^{(i+1)} \cdot x_2^{(i+1)}$ of i-th and $(i+1)$-th terms by the new term $(p^{(i)} + p^{(i+1)}) \cdot x_1 \cdot x_2$. By definition of α, we have $p^{(i+1)} = (p^{(i+1)} + p^{(i)}) \cdot \alpha$ and $p^{(i+1)} = (p^{(i+1)} + p^{(i)}) \cdot (1-\alpha)$. Thus, the first sum can be represented as $(p^{(i)} + p^{(i+1)}) \cdot ((1-\alpha) \cdot x_1^{(i)} \cdot x_2^{(i)} + \alpha \cdot x_1^{(i+1)} \cdot x_2^{(i+1)})$. Thus, to prove that the change cannot increase $E[x_1 \cdot x_2]$, it is sufficient to prove that

$$(1-\alpha) \cdot x_1^{(i)} \cdot x_2^{(i)} + \alpha \cdot x_1^{(i+1)} \cdot x_2^{(i+1)} - x_1 \cdot x_2 \ge 0.$$

By definition of x, this is equivalent to:

$$(1-\alpha) \cdot x_1^{(i)} \cdot x_2^{(i)} + \alpha \cdot x_1^{(i+1)} \cdot x_2^{(i+1)} -$$

$$(\alpha \cdot x_1^{(i+1)} + (1-\alpha) \cdot x_1^{(i)}) \cdot (\alpha \cdot x_2^{(i+1)} + (1-\alpha) \cdot x_2^{(i)}) \ge 0.$$

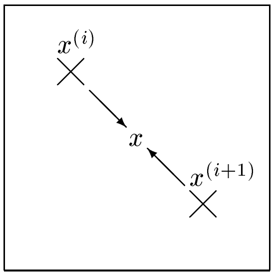
If we actually multiply the terms in the left-hand side and then combine together similar terms, we will conclude that the left-hand side is equal to:

$$\alpha \cdot (1-\alpha) \cdot (x_1^{(i+1)} - x_1^{(i)}) \cdot (x_2^{(i+1)} - x_2^{(i)}).$$

Since the original distribution is highly positively correlated and $i < i+1$, we have $x_1^{(i+1)} - x_1^{(i)} \ge 0$ and $x_2^{(i+1)} - x_2^{(i)} \ge 0$, so the left-hand side is indeed non-negative.

The theorem is proven.

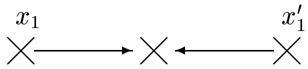
Proof of Theorem 3 is similar to the proof of Theorem 2, with the same transformation:



Proof of Theorem 4. For $x_1 > 0$, the function $f(x_1) \stackrel{\text{def}}{=} 1/x_1$ is convex: for every x_1, x'_1, and $\alpha \in [0,1]$, we have

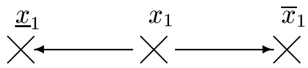
$$f(\alpha \cdot x_1 + (1-\alpha)\cdot x'_1) \le \alpha \cdot f(x_1) + (1-\alpha)\cdot f(x'_1).$$

Hence, if we are looking for a minimum of $E[1/x_1]$, we can replace every two points from the probability distribution with their average, and the resulting value of $E[1/x_1]$ will only decrease:



So, the minimum is attained when the probability distribution is concentrated on a single value – which has to be E_1. Thus, the smallest possible value of $E[1/x_1]$ is $1/E_1$.

Due to the same convexity, if we want maximum of $E[1/x_1]$, we should replace every value $x_1 \in [\underline{x}_1, \overline{x}_1]$ by a probabilistic combination of the values $\underline{x}_1, \overline{x}_1$:



So, the maximum is attained when the probability distribution is concentrated on these two endpoints $\underline{x}_1$ and $\overline{x}_1$. Since the average of x_1 should be equal to E_1, we can, similarly to the proof of Theorem 1, conclude that in this distribution, $\overline{x}_1$ occurs with probability p_1, and $\underline{x}_1$ occurs with probability $1-p_1$. For this distribution, the value $E[1/x_1]$ is exactly the upper bound from the formulation of Theorem 4. The theorem is proven.

Proof of Theorem 5. Since $\min(x_1, x_2) \le x_1$, we have $E[\min(x_1, x_2)] \le E[x_1] = E_1$. Similarly, $E[\min(x_1, x_2)] \le E_2$, hence, $E[\min(x_1, x_2)] \le \min(E_1, E_2)$. The value $\min(E_1, E_2)$ is possible when $x_1 = E_1$ with probability 1 and $x_2 = E_2$ with probability 1. Thus, $\min(E_1, E_2)$ is the exact upper bound for $E[\min(x_1, x_2)]$.

For each x_2, the function $x_1 \to \min(x_1, x_2)$ is concave; therefore, similarly to the proof of Theorem 4, if we replace each point $x^{(j)} = (x_1^{(j)}, x_2^{(j)})$ by the corresponding probabilistic combination of the points $(\underline{x}_1, x_2^{(j)})$ and $(\overline{x}_1, x_2^{(j)})$ (as in the proof of Theorem 1), we preserve $E[x_1]$ and $E[x_2]$ and decrease the value $E[\min(x_1, x_2)]$. Thus, when we are looking for the smallest possible value of $E[\min(x_1, x_2)]$, it is sufficient to consider only the distributions for which x_1 is located at one of the endpoints $\underline{x}_1$ or $\overline{x}_1$. Similarly to the proof of Theorem 1, the probability of $\overline{x}_1$ is equal to p_1.

Similarly, we can conclude that to find the largest possible value of $E[\min(x_1, x_2)]$, it is sufficient to consider only distributions in which x_2 can take only two values: $\underline{x}_2$ and $\overline{x}_2$. To get the desired value of E_2, we must have $\overline{x}_2$ with probability p_1 and $\underline{x}_2$ with probability $1-p_2$.

Since we consider the case when x_1 and x_2 are independent, and each of them takes two possible values, we can conclude that $x = (x_1, x_2)$ can take four possible values $(\underline{x}_1, \underline{x}_2)$, $(\underline{x}_1, \overline{x}_2)$, $(\overline{x}_1, \underline{x}_2)$, and $(\overline{x}_1, \overline{x}_2)$, and the probability of each of these values is equal to the product of the probabilities corresponding to x_1 and x_2. For this distribution, $E[\min(x_1, x_2)]$ is exactly the expression from the formulation of Theorem 4. Q.E.D.

Proof of Theorem 6 is similar to the proof of Theorem 5, with $x_1 \le \max(x_1, x_2)$ to prove the lower bound and convexity of $\max(x_1, x_2)$ to prove the upper bound.

Proof of Theorem 7. Similarly to the proof of Theorem 5, we can conclude that $\min(E_1, E_2)$ is the attainable upper bound for $E[\min(x_1, x_2)]$, and that to find the lower bound for $E[\min(x_1, x_2)]$, it is sufficient to consider distributions located at the four corners of the box $\mathbf{x}_1 \times \mathbf{x}_2$.

Similarly to the proof of Theorem 1, we will show that to find the desired minimum of $E[\min(x_1, x_2)]$, it is sufficient to consider distributions with highly negative correlation. The proof of this statement is based on the same idea as in the proof of Theorem 1: that if we have two points $x^{(i)}$ and $x^{(j)}$ that contradict the assumption of highly negative correlation, i.e., $x_1^{(i)} < x_1^{(j)}$ and $x_2^{(i)} < x_2^{(j)}$, then by swapping the second components, we can decrease $E[\min(x_1, x_2)]$.

In other words, we claim that if $x_1^{(i)} < x_1^{(j)}$ and $x_2^{(i)} < x_2^{(j)}$, then

$$\min(x_1^{(i)}, x_2^{(i)}) + \min(x_1^{(j)}, x_2^{(j)}) \ge \min(x_1^{(i)}, x_2^{(j)}) + \min(x_1^{(j)}, x_2^{(i)}).$$

This claim can be easily proven case-by-case by considering all possible orders between the four numbers $x_1^{(i)}$, $x_1^{(j)}$, $x_2^{(i)}$, and $x_2^{(j)}$.

Thus, as in the proof of Theorem 1, we conclude that the smallest possible value for $E[\min(x_1, x_2)]$ is attained when we have a highly negatively correlated distribution on the corner points. In Theorem 1, we have already found the probabilities corresponding to this distribution and thus, we can compute the corresponding mathematical expectation $E[\min(x_1, x_2)]$. Q.E.D.

Proof of Theorem 8 is similar to the proof of Theorem 7; the only difference is that for max, instead of a highly negative correlation, we need a distribution with a highly positive correlation.

Proof of Theorem 9. In general, since not every distribution is highly negatively correlated, the interval of possible values $\mathbf{E}^n$ corresponding to highly negative distributions is a subset of the interval $\mathbf{E}^u = [\underline{E}^u, \overline{E}^u]$ corresponding to all possible distributions. In Theorem 7, however, both the distribution for which $\underline{E}^u$ is attained and the distribution for which $\overline{E}^u$ is attained are highly negatively correlated. Thus, both $\underline{E}^u$ and $\overline{E}^u$ belong to the desired interval $\mathbf{E}^n$, so the intervals $\mathbf{E}^n$ and $\mathbf{E}^u$ coincide. Q.E.D.

Proof of Theorem 10 is similar to the proof of Theorem 9.

Proof of Theorem 11. Similarly to the proof of Theorem 1, let us consider a discrete N-point highly positively correlated distribution in which each point $x^{(j)} = (x_1^{(j)}, x_2^{(j)})$ occurs with probability $p^{(j)}$, and the points are sorted: $x_1^{(1)} \le x_1^{(2)} \le \ldots \le x_1^{(N)}$ and $x_2^{(1)} \le x_2^{(2)} \le \ldots \le x_2^{(N)}$. Let us temporarily fix the points $x^{(j)}$ and allow the probabilities $p^{(j)}$ to change; the probabilities $p^{(j)} \ge 0$ must satisfy three conditions:

$$p^{(1)} + \ldots + p^{(N)} = 1;$$

$$p^{(1)} \cdot x_1^{(1)} + \ldots + p^{(N)} \cdot x_1^{(N)} = E_1;$$

$$p^{(1)} \cdot x_2^{(1)} + \ldots + p^{(N)} \cdot x_2^{(N)} = E_2.$$

Under these conditions, we want to minimize the mean $E[\min(x_1, x_2)]$, i.e., the expression

$$p^{(1)} \cdot \min(x_1^{(1)}, x_2^{(1)}) + \ldots + p^{(N)} \cdot \min(x_1^{(N)}, x_2^{(N)}).$$

With respect to the values $p^{(j)}$, 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 constrains are equalities. Thus, in our case, all but three probabilities $p^{(j)}$ must be equal to 0.

So, to find the smallest possible value of $E[x_1 \cdot x_2]$, it is sufficient to consider probability distributions that are located on $N \le 3$ points $x^{(j)}$. We will prove that these points and the corresponding probabilities $p^{(j)}$ are the ones that lead to the formulas from the formulation of the theorem.

$N = 1$. If we have a probability distribution that is located on a single point x, then, since we know $E[x_1] = E_1$ and $E[x_2] = E_2$, this point has to be $x = (E_1, E_2)$. For this point, $\min(x_1, x_2) = \min(E_1, E_2)$.

$N > 1$. Let us now consider the case when the probability distribution is located on at least two points.

We start with the first point $x^{(1)}$. For this first point, the minimum $\min(x_1^{(1)}, x_2^{(1)})$ is equal either to $x_1^{(1)}$ or to $x_2^{(1)}$. We will consider the case when the minimum is equal to $x_1^{(1)}$, i.e., when $x_1^{(1)} \le x_2^{(1)}$ (the proof for the second case is similar).

Let us now consider the second point. If for the second point $x^{(2)}$, we also have $x_1^{(2)} \le x_2^{(2)}$, then we can replace both points $x^{(1)}$ and $x^{(2)}$ with a single point

$$x \stackrel{\text{def}}{=} \frac{p^{(1)}}{p^{(1)} + p^{(2)}} \cdot x^{(1)} + \frac{p^{(2)}}{p^{(1)} + p^{(2)}} \cdot x^{(2)}.$$

One can easily check that after this replacement, we have the same value of $E[x_1]$ and the same value of $E[x_2]$.

Also, from $x_1^{(1)} \le x_2^{(1)}$ and $x_1^{(2)} \le x_2^{(2)}$, we can conclude that $x_1 \le x_2$, so $\min(x_1, x_2) = x_1$. Therefore, after this replacement, we have the same value of $E[\min(x_1, x_2)]$. In this case, we get a probability distribution with fewer points. If necessary, we can apply this replacement again and again until we arrive at the situation when this replacement is no longer possible, i.e., when either we are left with only one point $x^{(1)}$ (which will then be equal to (E_1, E_2)), or we will have $x_1^{(1)} < x_2^{(1)}$ and $x_1^{(2)} > x_2^{(2)}$. Thus, to find the smallest possible value $\underline{E}$, it is sufficient to consider cases for which the order between x_1 and x_2 alternates: $x_1^{(1)} < x_2^{(1)}$ and $x_1^{(2)} > x_2^{(2)}$. Similarly, if we have the third point, it is sufficient to consider only cases when $x_1^{(3)} < x_2^{(3)}$.

These inequalities allow us to simplify the situation even further. Indeed, we know (since the distribution is highly positively correlated) that $x_2^{(1)} \le x_2^{(2)}$. Let us show that strict inequality is impossible. Indeed, if $x_2^{(1)} < x_2^{(2)}$, we can replace both values $x_2^{(1)} \le x_2^{(2)}$ by their average

$$x_2 \stackrel{\text{def}}{=} \frac{p^{(1)}}{p^{(1)} + p^{(2)}} \cdot x_2^{(1)} + \frac{p^{(2)}}{p^{(1)} + p^{(2)}} \cdot x_2^{(2)}.$$

In this replacement, we increase $x_2^{(1)}$ and decrease $x_2^{(2)}$.

After this replacement, the value $E[x_2]$ remains the same. The value $E[x_1]$ does not change – since we did not change the first components at all. The value $E[\min(x_1, x_2)]$ actually decreases:

- since we increased $x_2^{(1)}$, and we had $x_1^{(1)} < x_2^{(1)}$, the new value of $\min(x_1^{(1)}, x_1^{(2)})$ remains the same – equal to $x_1^{(1)}$;
- since we decreased $x_2^{(2)}$, and we had $x_1^{(2)} > x_2^{(2)}$, the new value of $\min(x_1^{(2)}, x_2^{(2)})$ is equal to the new value of $x_2^{(2)}$ – i.e., smaller than before;
- finally, the value of $\min(x_1^{(3)}, x_2^{(3)})$ does not change, because we did not change $x_1^{(3)}$ or $x_2^{(3)}$.

So, two terms in the sum $\sum p^{(j)} \cdot \min(x_1^{(j)}, x_2^{(j)})$ remain the same, one decreases – hence the entire sum decreases too.

Thus, if we are looking for the smallest possible value of E, it is sufficient to consider only cases when $x_2^{(1)} = x_2^{(2)}$. Let us consider separately the two cases: when the distribution is located on two points (i.e., when $N = 2$), and when the distribution is located on three points (i.e., $N = 3$).

$N = 2$. If we have only two points, the equality $x_2^{(1)} = x_2^{(2)}$ means the value x_2 is the same for both points. Therefore, the mathematical expectation $E[x_2] = E_2$

coincides with this value, hence, this value must be equal to E_1. Hence, for this case, $E[\min(x_1, x_2)] = E[\min(x_1, E_2)]$. Now, from the fact that minimum is a concave function, we can conclude that the smallest possible value of this expectation is when x_1 is located at the endpoints of the interval $\mathbf{x}_1$. Thus, we get an expression presented in the formulation of Theorem 11.

$N = 3$. Let us now consider the case when we have three points. In this case, not only we have $x_2^{(1)} = x_2^{(2)}$, but, similarly, we can prove that it is sufficient to consider only cases when $x_1^{(2)} = x_1^{(3)}$.

In this case, the alternating inequalities take the form $x_1^{(1)} < x_2^{(1)} < x_1^{(2)} < x_2^{(3)}$, and the problem becomes as follows:

$$p^{(1)} \cdot x_1^{(1)} + p^{(2)} \cdot x_2^{(1)} + p^{(3)} \cdot x_1^{(2)} \to \min$$

under the conditions

$$p^{(1)} + p^{(2)} + p^{(3)} = 1;$$

$$p^{(1)} \cdot x_1^{(1)} + p^{(2)} \cdot x_1^{(2)} + p^{(3)} \cdot x_1^{(2)} = E_1;$$

$$p^{(1)} \cdot x_2^{(1)} + p^{(2)} \cdot x_2^{(1)} + p^{(3)} \cdot x_2^{(3)} = E_2.$$

The last two conditions can be reformulated as follows:

$$p^{(1)} \cdot x_1^{(1)} + (p^{(2)} + p^{(3)}) \cdot x_1^{(2)} = E_1;$$

$$(p^{(1)} + p^{(2)}) \cdot x_2^{(1)} + p^{(3)} \cdot x_2^{(3)} = E_2.$$

The minimized expression differs from the expression for E_1 in only one term, so this expression can be represented as $E_1 - p^{(2)} \cdot (x_1^{(2)} - x_2^{(1)})$. Therefore, the values $p^{(j)}$ and $x^{(j)}$ minimize the desired expression if and only if they maximize the expression

$$p^{(2)} \cdot (x_1^{(2)} - x_2^{(1)}) \to \max.$$

Let us use this representation to prove that $x_1^{(1)} = \underline{x}_1$. Specifically, we will show that if $x_1^{(1)} > \underline{x}_1$, then we can further decrease $E[\min(x_1, x_2)]$. Indeed, if $x_1^{(1)} > \underline{x}_1$, this means that $x_1^{(1)}$ is strictly inside the interval $\mathbf{x}_1$, and thus, when a real number Δx is sufficiently small, the value $x_1^{(1)} + \Delta x$ is still within this interval. Let us show that by appropriately changing $p^{(1)}$ and $p^{(2)}$ and leaving all other variables intact, we can preserve $E[x_1]$ and $E[x_2]$ and decrease $E[\min(x_1, x_2)]$.

Indeed, if we change $p^{(1)}$ to a new value $p^{(1)} + \Delta p$, then, to preserve the sum of the probabilities, we must change $p^{(2)}$ to $p^{(2)} - \Delta p$. In this case, the sum $p^{(1)} + p^{(2)}$ remains the same, hence the equality containing E_2 remains valid. For the equality containing E_1 to remain valid, we must choose Δp

in such a way that the difference between the new and the old combinations $p^{(1)} \cdot x_1^{(1)} + (p^{(2)} + p^{(3)}) \cdot x_1^{(2)}$ is 0, i.e., that

$$p^{(1)} \cdot \Delta x + \Delta p \cdot x_1^{(1)} - \Delta p \cdot x_1^{(2)} + o(\Delta x) = 0.$$

In other words, we must have

$$\Delta p = \Delta x \cdot \frac{p^{(1)}}{x_1^{(2)} - x_1^{(1)}} + o(\Delta x).$$

For this value, the change in $p^{(2)} \cdot (x_1^{(2)} - x_2^{(1)})$ is equal to $-\Delta p \cdot (x_1^{(2)} - x_2^{(1)})$, i.e., to:

$$-\Delta x \cdot \frac{p^{(1)}}{x_1^{(2)} - x_2^{(1)}} \cdot (x_1^{(2)} - x_2^{(1)}) + o(\Delta x).$$

For small $\Delta x < 0$, this value is positive, thus we further increase $p^{(2)} \cdot (x_1^{(2)} - x_2^{(1)})$, hence decrease $E[\min(x_1, x_2)]$. This proves that the minimum of $E[\min(x_1, x_2)]$ is attained when $x_1^{(1)} = \underline{x}_1$.

Similarly, the minimum is attained when $x_2^{(3)} = \overline{x}_2$. So, the problem becomes as follows:

$$p^{(1)} \cdot \underline{x}_1 + p^{(2)} \cdot x_2^{(1)} + p^{(3)} \cdot x_1^{(2)} \to \min$$

under the conditions

$$\underline{x}_1 < x_2^{(1)} < x_1^{(2)} < \overline{x}_2,$$

$$p^{(1)} + p^{(2)} + p^{(3)} = 1;$$

$$p^{(1)} \cdot \underline{x}_1 + (p^{(2)} + p^{(3)}) \cdot x_1^{(2)} = E_1;$$

$$(p^{(1)} + p^{(2)}) \cdot x_2^{(1)} + p^{(3)} \cdot \overline{x}_2 = E_2.$$

This is exactly the problem described in the formulation of Theorem 11 (the second problem from the formulation of Theorem 11 corresponds to the case when $\min(x_1^{(1)}, x_2^{(1)}) = x_2^{(1)}$). The theorem is proven.

Proof of Theorem 12 is similar to the proof of Theorem 11.

Proof of Theorems 13 and 14. Let us first prove the result (Theorem 14) for the upper bound $\overline{E}$. The formula for $\overline{E}$ given in Theorem 1 can be simplified if we consider two cases: $p_1 \le p_2$ and $p_1 \ge p_2$. Indeed:

- in the first case, when $p_1 \le p_2$, we have:

$$\overline{f}_\times^u = p_1 \cdot \overline{x}_1 \cdot \overline{x}_2 + (p_2 - p_1) \cdot \underline{x}_1 \cdot \overline{x}_2 + (1 - p_2) \cdot \underline{x}_1 \cdot \underline{x}_2;$$

- in the second case, when $p_1 \ge p_2$, we have:

$$\overline{f}_\times^u = p_2 \cdot \overline{x}_1 \cdot \overline{x}_2 + (p_1 - p_2) \cdot \overline{x}_1 \cdot \underline{x}_2 + (1 - p_1) \cdot \underline{x}_1 \cdot \underline{x}_2.$$

To find the largest possible value $\overline{E}$ of E, it is sufficient to consider the largest possible values for each of these cases, and then take the largest of the resulting two numbers.

In each case, for a fixed p_2, the formula is linear in p_1. To find the maximum of a linear function on an interval, it is sufficient to consider this interval's endpoints. Thus, the maximum in p_1 is attained when either p_1 attains its smallest possible value $\underline{p}_1$, or when p_1 attains the largest possible value within this case; depending on p_2, this value is either $p_1 = \overline{p}_1$ or $p_1 = p_2$.

Thus, to find the maximum for each cases, it is sufficient to consider only the following cases: $p_1 = \underline{p}_1$, $p_1 = \overline{p}_1$, and $p_1 = p_2$. Similarly, it is sufficient to consider only the following cases for p_2: $p_2 = \underline{p}_2$, $p_2 = \overline{p}_2$, and $p_1 = p_2$.

When $p_1 \neq p_2$, we therefore have one of the first four cases described in Theorem 14. The case $p_1 = p_2$ is possible only when the intervals $\mathbf{p}_1$ and $\mathbf{p}_2$ of possible values of p_1 and p_2 have a common point: $\mathbf{p}_1 \cap \mathbf{p}_2 \neq \emptyset$ (i.e., when $\max(\underline{p}_1, \underline{p}_2) \leq \min(\overline{p}_1, \overline{p}_2)$). In this case, the probability $p_1 = p_2$ can take all possible values from the intersection

$$\mathbf{p}_1 \cap \mathbf{p}_2 = [\max(\underline{p}_1, \underline{p}_2), \min(\overline{p}_1, \overline{p}_2)]$$

of the intervals $\mathbf{p}_1$ and $\mathbf{p}_2$. In case $p_1 = p_2$, the formula for $\overline{f}^u_\times$ can be further simplified:

$$\overline{f}^u_\times = p_1 \cdot \overline{x}_1 \cdot \overline{x}_2 + (1 - p_1) \cdot \underline{x}_1 \cdot \underline{x}_2.$$

This formula is linear in p_1, so to find its maximum, it is sufficient to consider the endpoints of the interval $\mathbf{p}_1 \cap \mathbf{p}_2$, i.e., the values $p_1 = p_2 = \max(\underline{p}_1, \underline{p}_2)$ and $p_1 = p_2 = \min(\overline{p}_1, \overline{p}_2)$ – the remaining cases from Theorem 13. For $\overline{E}$, the statement is proven.

Let us now prove Theorem 13 – for the lower bound $\underline{E}$. The formula for $\underline{E}$ given in Theorem 1 can be simplified if we consider two cases: $p_1 + p_2 \leq 1$ and $p_1 + p_2 \geq 1$:

- in the first case, when $p_1 + p_2 \leq 1$, we have:

$$\underline{f}^u_\times = p_1 \cdot \overline{x}_1 \cdot \underline{x}_2 + p_2 \cdot \underline{x}_1 \cdot \overline{x}_2 + (1 - p_1 - p_2) \cdot \underline{x}_1 \cdot \underline{x}_2;$$

- in the second case, when $p_1 + p_2 \geq 1$, we have:

$$\underline{f}^u_\times = (p_1 + p_2 - 1) \cdot \overline{x}_1 \cdot \overline{x}_2 + (1 - p_2) \cdot \overline{x}_1 \cdot \underline{x}_2 + (1 - p_1) \cdot \underline{x}_1 \cdot \overline{x}_2.$$

To find the smallest possible value $\underline{E}$ of E, it is sufficient to consider the smallest possible values for each of these cases, and then take the smallest of the resulting two numbers.

In each case, for a fixed p_2, the formula is linear in p_1. To find the maximum of a linear function on an interval, it is sufficient to consider this interval's endpoints. Thus, the maximum in p_1 is attained when either p_1 attains its smallest possible value $\underline{p}_1$, or when p_1 attains the largest possible value within this case; depending on p_2, this value is either $p_1 = \overline{p}_1$ or $p_1 = 1 - p_2$.

Thus, to find the minimum for each cases, it is sufficient to consider only the following cases: $p_1 = \underline{p}_1$, $p_1 = \overline{p}_1$, and $p_1 = 1 - p_2$. Similarly, it is sufficient to consider only the following cases for p_2: $p_2 = \underline{p}_2$, $p_2 = \overline{p}_2$, and $p_2 = 1 - p_1$ (the last case is equivalent to $p_1 = 1 - p_2$).

When $p_1 \neq 1 - p_2$, we therefore have one of the first four cases described in Theorem 13. The case $p_1 = 1 - p_2$ (i.e., $p_1 + p_2 = 1$) is possible only when the number 1 belongs to the sum $\mathbf{p}_1 + \mathbf{p}_2$ of the intervals $\mathbf{p}_1$ and $\mathbf{p}_2$, i.e., when $\underline{p}_1 + \underline{p}_2 \le 1 \le \overline{p}_1 + \overline{p}_2$. In this case, the probability $p_1 = 1 - p_2$ can take all possible values from the intersection

$$\mathbf{p}_1 \cap (1 - \mathbf{p}_2) = [\max(\underline{p}_1, 1 - \overline{p}_2), \min(\overline{p}_1, 1 - \underline{p}_2)]$$

of the intervals $\mathbf{p}_1$ and $1 - \mathbf{p}_2$. For $p_1 = 1 - p_2$, the formula for $\underline{f}^u_\times$ is linear in p_1, so to find its minimum, it is sufficient to consider the endpoints of the interval $\mathbf{p}_1 \cap (1 - \mathbf{p}_2)$, i.e., the values $p_1 = 1 - p_2 = \max(\underline{p}_1, 1 - \overline{p}_2)$ and $p_1 = 1 - p_2 = \min(\overline{p}_1, 1 - \underline{p}_2)$ – the remaining two cases from Theorem 13. Theorem 13 is proven as well. Q.E.D.

Proof of Theorem 15. For each value $E_1 \in [\underline{E}_1, \overline{E}_1]$, possible values of $E[1/x_1]$ are described by Theorem 4.

In particular, for each E_1, the lower bound is equal to $1/E_1$. To find the lower bound $\underline{E}$ among all $E_1 \in \mathbf{E}_1$, we must therefore find the smallest of the lower bounds $1/E_1$ when $E_1 \in \mathbf{E}_1$. This smallest value is clearly attained when E_1 is the largest possible $E_1 = \overline{E}_1$, so the desired lower bound is equal to $1/\overline{E}_1$.

Similarly, for each E_1, the upper bound is equal to

$$\frac{p_1}{\overline{x}_1} + \frac{1 - p_1}{\underline{x}_1} = p_1 \cdot \left(\frac{1}{\overline{x}_1} - \frac{1}{\underline{x}_1} \right) + \frac{1}{\underline{x}_1}.$$

To find the upper bound $\overline{E}$ among all $E_1 \in \mathbf{E}_1$, we must therefore find the largest of the upper bounds when $E_1 \in \mathbf{E}_1$. Since the coefficient at p_1 is negative, this largest value is clearly attained when p_1 is the smallest possible $p_1 = \underline{p}_1$, so the desired upper bound is exactly as in the formulation of Theorem 15. Q.E.D.

Proof of Theorem 16. According to Theorem 5, for each $E_1 \in \mathbf{E}_1$ and $E_2 \in \mathbf{E}_2$, we have $\overline{f}^i_{\min}(E_1, E_2) = \min(E_1, E_2)$. This function is non-decreasing in each of the variables hence its largest possible value is attained when $E_1 = \overline{E}_1$ and $E_2 = \overline{E}_2$. The resulting value $\min(\overline{E}_1, \overline{E}_2)$ is exactly the bound described in the formulation of Theorem 16.

For each $E_i \in \mathbf{E}_i$, the corresponding lower bound is described by Theorem 5. To find the smallest possible value of this bound for all $p_i \in \mathbf{p}_i$, let us simplify the above expression by gathering together terms proportional to p_1. As a result, we get the following expression:

$$p_1 \cdot p_2 \cdot [\min(\overline{x}_1, \overline{x}_2) - \min(\underline{x}_1, \overline{x}_2)]+$$
$$p_1 \cdot (1-p_2) \cdot [\min(\overline{x}_1, \underline{x}_2) - \min(\underline{x}_1, \underline{x}_2)]+$$
$$p_2 \cdot \min(\underline{x}_1, \overline{x}_2) + (1-p_2) \cdot \min(\underline{x}_1, \underline{x}_2).$$

From $\overline{x}_1 \geq \underline{x}_1$, we conclude that $\min(\overline{x}_1, \overline{x}_2) \geq \min(\underline{x}_1, \overline{x}_2)$ and therefore, the coefficient at $p_1 \cdot p_2$ is non-negative. Similarly, the coefficient at $p_1 \cdot (1-p_2)$ is non-negative. Thus, when we fix p_2, the above expression becomes a non-decreasing linear function of p_1. Since this expression is non-decreasing, its minimum is attained when p_1 takes the smallest possible value $p_1 = \underline{p}_1$. Similarly, we can prove that the minimum is attained when p_2 takes the smallest possible value $p_2 = \underline{p}_2$. Thus, the minimum is attained when $p_1 = \underline{p}_1$ and $p_2 = \underline{p}_2$. For these p_1 and p_2, the resulting bound is exactly what we formulated in Theorem 16. Q.E.D.

Proof of Theorem 17 is similar to the proof of Theorem 16.

Proof of Theorems 18 and 19. In both cases, one bound is the simple minimum or maximum of E_i, and other bound is described by a complex expression.

- For the easy min (max) bound, the proof is similar to the proof of Theorems 16 and 17.
- For the more complex bound, the dependence on p_1 and p_2 is similar to the dependence that we used in the proof of Theorems 13 and 14. so our result that it is sufficient to consider six possible pairs (p_1, p_2) to find the minimum (maximum) is true here as well.

Proof of Theorems 20–23 is similar to the proof of Theorem 9.

Proof of Theorem 24. For fixed $E_i \in \mathbf{E}_i$, the fact that it is sufficient to consider only distributions concentrated on $\leq (n+1)$ points can be proven similarly to the proof of Theorem 11. The range of E corresponding to non-degenerate $\mathbf{E}_i$ is the union of the ranges corresponding to different $E_i \in \mathbf{E}_i$, so both the upper and the lower endpoints for this range correspond to some $E_i \in \mathbf{E}_i$ and thus, to some distributions concentrated on $\leq (n+1)$ points.

Proof of Theorem 25 is similar to the proof of Theorem 24:

- The optimized function is linear with respect to each of n probability distributions.
- For each of these distributions, there are only two constraints: that the sum of probabilities is 1, and that the mathematical expectation is E_i.
- So, similarly to the proof of Theorem 11, it is sufficient to consider only distributions concentrated on no more than 2 points.

Let us denote the smallest of these two points by x_i^- and the largest by x_i^+. The corresponding probabilities will be denoted by p_i^- and $p_i^+ = 1 - p_i^-$. Then, we get the desired formulas.

Proof of Theorems 26–29 is similar to the proofs of Theorems 4 and 5.

Proof of Theorems 30 and 31 is similar to the proof of Theorem 24.