Towards Adding Probabilities and Correlations to Interval Computations ⋆

Daniel Berleant

Department of Information Science, University of Arkansas at Little Rock
Little Rock, Arkansas 72204, USA

Martine Ceberio, Gang Xiang, and Vladik Kreinovich *

Department of Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA

Abstract

The paper is a continuation of our previous work towards the use of probability information in interval computations. While in the previous work, bounds on the first order moments are taken into account, the contribution of this article is to deal with correlations. Specifically, in this paper, we develop a new method that takes into account both correlation among measured parameters and bounds on their expected values when doing interval computation.

Key words: interval computations, imprecise probabilities, correlation

1 Formulation of the Problem

Why data processing? In many real-life situations, we are interested in the value of a physical quantity y that is difficult or impossible to measure directly. Examples of such quantities are the distance to a star and the amount of oil

⋆ This work was supported in part by NSF grants EAR-0225670 and DMS-0532645 and by Texas Department of Transportation grant No. 0-5453. The authors are thankful to the anonymous referees for valuable suggestions.
* Corresponding author
Email addresses: jdberleant@ualr.edu (Daniel Berleant), mceberio@cs.utep.edu (Martine Ceberio), gxiang@utep.edu (Gang Xiang), vladik@utep.edu (Vladik Kreinovich).

in a given well. Since we cannot measure y directly, a natural strategy is to measure y *indirectly*. Specifically, we find some easier-to-measure quantities $x_1, \ldots, x_n$ which are related to y by a known relation $y = f(x_1, \ldots, x_n)$. To estimate y, we

first measure the quantities $x_1, \ldots, x_n$, and then use the relation $y = f(x_1, \ldots, x_n)$ to compute an estimate for y.

Why interval computations? Measurement are never 100% accurate, so after the measurement, we only know the values x_i with some uncertainty [13]. It is desirable to describe the resulting uncertainty in $y = f(x_1, \ldots, x_n)$.

In chemistry and environmental sciences, there are many measuring techniques where we only get the interval of possible values of the desired quantity. For example, if we did not detect any pollution, the pollution value v can be anywhere between 0 and the sensor's detection limit DL. In other words, the only information that we have about v is that v belongs to the interval $[0, DL]$; we have no information about the probability of different values from this interval.

Another example: to study the effect of a pollutant on the fish, we check on the fish daily; if a fish was alive on Day 5 but dead on Day 6, then the only information about the lifetime of this fish is that it is somewhere within the interval $[5, 6]$; we have no information about the distribution of different values in this interval.

In such cases, after performing a measurement, the only information that we have about the actual value x_i of the measured quantity is that it belongs to the interval $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$.[1] In this situation, the only information that we have about the (unknown) actual value of $y = f(x_1, \ldots, x_n)$ is that y belongs to the range $\mathbf{y} = [\underline{y}, \overline{y}]$ of the function f over the box $\mathbf{x}_1 \times \ldots \times \mathbf{x}_n$:

$$\mathbf{y} = [\underline{y}, \overline{y}] = \{f(x_1, \ldots, x_n) \,:\, x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}.$$

The process of computing this interval range based on the input intervals $\mathbf{x}_i$ is part of *interval computations*; see, e.g., [7].

Interval computations techniques: brief reminder. Historically what is often called the "straightforward" method was the first for estimating the desired range of a function. This method is based on the fact that inside the computer, every algorithm for processing real numbers is implemented as a sequence of elementary operations $a + b$, $a - b$, $a \cdot b$, and a/b; usually, a/b is computed as $a \cdot (1/b)$, making $a + b$, $a - b$, $a \cdot b$, and $1/a$ sufficient. For each

[1] We use the convention of bold, non-italic symbols for naming intervals.

of these elementary operations $f(a, b)$, if we know the intervals $\mathbf{a}$ and $\mathbf{b}$ for a and b, we can compute the exact range $f(\mathbf{a}, \mathbf{b})$. The corresponding formulas form the so-called *interval arithmetic*:

$$[\underline{a}, \overline{a}] + [\underline{b}, \overline{b}] = [\underline{a} + \underline{b}, \overline{a} + \overline{b}]; \quad [\underline{a}, \overline{a}] - [\underline{b}, \overline{b}] = [\underline{a} - \overline{b}, \overline{a} - \underline{b}];$$

$$[\underline{a}, \overline{a}] \cdot [\underline{b}, \overline{b}] = [\min(\underline{a} \cdot \underline{b}, \underline{a} \cdot \overline{b}, \overline{a} \cdot \underline{b}, \overline{a} \cdot \overline{b}), \max(\underline{a} \cdot \underline{b}, \underline{a} \cdot \overline{b}, \overline{a} \cdot \underline{b}, \overline{a} \cdot \overline{b})];$$
$$1/[\underline{a}, \overline{a}] = [1/\overline{a}, 1/\underline{a}] \text{ if } 0 \notin [\underline{a}, \overline{a}].$$

In straightforward interval computations, we replace each floating point operation in the program f by the corresponding interval operation; as a result, after all the operations, we get an interval $\mathbf{Y}$. It is known that this resulting interval $\mathbf{Y}$ is an enclosure of the desired range $\mathbf{y}$, i.e., that $\mathbf{Y} \supseteq \mathbf{y}$.

In some cases, $\mathbf{Y} = \mathbf{y}$. In more complex cases, the enclosure has excess width ($\mathbf{Y} \supset \mathbf{y}$). There exist more sophisticated techniques for producing narrower enclosures, e.g., centered form methods [7]. However, for each of these techniques, there are cases when we still get excess width. Reason: it is known (see, e.g., [10]), that the problem of computing the exact range is NP-hard even for polynomial functions $f(x_1, \ldots, x_n)$ (indeed, even for quadratic functions f).

Motivating practical problem. In some practical situations, in addition to lower and upper bounds on each random variable x_i, we know bounds $\mathbf{E}_i = [\underline{E}_i, \overline{E}_i]$ on its mean E_i; see, e.g., [13].

If we have this information for every x_i, then, in addition to the interval $\mathbf{y}$ of possible values of y, we can also try to estimate the interval of possible values of $E[y]$. Thus, we arrive at the following problem.

New problem in precise terms. Given an algorithm computing a function $f(x_1, \ldots, x_n)$ from $\mathbb{R}^n$ to $\mathbb{R}$, and values $\underline{x}_1$, $\overline{x}_1$, …, $\underline{x}_n$, $\overline{x}_n$, $\underline{E}_1$, $\overline{E}_1$, …, $\underline{E}_n$, $\overline{E}_n$, we want to find

$$\underline{E} \stackrel{\text{def}}{=} \min\{E[f(x_1, \ldots, x_n)] : \text{ all distributions of}$$

$$(x_1, \ldots, x_n) \text{ for which } x_1 \in [\underline{x}_1, \overline{x}_1], \ldots, x_n \in [\underline{x}_n, \overline{x}_n],$$
$$E[x_1] \in [\underline{E}_1, \overline{E}_1], \ldots E[x_n] \in [\underline{E}_n, \overline{E}_n]\};$$

and $\overline{E}$ which is the maximum of $E[f(x_1, \ldots, x_n)]$ for all such distributions.

In addition to considering all possible distributions, we can also consider the case when all the variables x_i are independent, or, more generally, when we know the correlations among the x_i.

Comment. This problem is a particular case of *imprecise probability* problems, when we have a partial information about the probability distribution; see, e.g., the monographs [11,15]. These monographs also describe techniques for solving such problems; most of these techniques are based on the fact that in many of these problems, the optimized function is a linear function(al) of the values of (unknown) probability density function $d(x_1, \ldots, x_n)$, and the constraints on $d(x_1, \ldots, x_n)$ are linear inequalities in terms of these unknown values. In other words, many such problems are (infinite-dimensional) linear programming (LP) problems.

It is known that there exist efficient algorithms for solving (finite-dimensional) linear programming problems. We can approximate the LP problem with infinitely many unknowns $d(x_1, \ldots, x_n)$ by a problem with finitely many unknown if we consider, as new unknowns, the probabilities within certain n-dimensional boxes. The more boxes we consider and the narrower these boxes, the more accurate the corresponding approximation. Thus, we can get more and more accurate approximations to the desired values $\underline{E}$ and $\overline{E}$; see, e.g., [1–5].

However, the more accurate the computations, the more boxes we need to take and thus, the longer the running time of the corresponding algorithms. To speed up these computations, it is desirable to find (whenever possible) explicit *analytical* expressions for $\underline{E}$ and $\overline{E}$. Once such expressions are known, we can compute the *exact* values of $\underline{E}$ and $\overline{E}$ – and thus, the number of elementary computational operations needed for these computations no longer increases with accuracy (as for the LP-based methods).

Let us describe situations when such analytical expressions are possible.

2 What Is Known

Extending interval arithmetic to handle expectations. The main idea behind standard interval computations can be applied here as well. First we find out how to solve the problem when $n = 2$ and $f(x_1, x_2)$ is one of the standard arithmetic operations. Then, once we have an arbitrary algorithm $f(x_1, \ldots, x_n)$, we parse it and replace each elementary operation on real numbers with the corresponding operation on quadruples $(\underline{x}, \underline{E}, \overline{E}, \overline{x})$.

To implement this idea, we must therefore know how to solve the above problem for elementary operations.

For *addition*, the answer is straightforward: $E[x_1 + x_2] = E[x_1] + E[x_2]$. So, if we know the values $E_1 \stackrel{\text{def}}{=} E[x_1]$ and $E_2 \stackrel{\text{def}}{=} E[x_2]$, then for $y = x_1 + x_2$, the only possible value of $E \stackrel{\text{def}}{=} E[y]$ is $E = E_1 + E_2$. This value does not depend on whether we have correlation or whether we have any information about the correlation. Thus, if we only know the ranges $\mathbf{E}_1$ and $\mathbf{E}_2$ of possible values of E_1 and E_2, then the range of possible values of E is $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$.

Similarly, the answer is straightforward for *subtraction:* if $y = x_1 - x_2$, there is only one possible value for $E = E[y]$: the value $E = E_1 - E_2$. Thus, $\mathbf{E} = \mathbf{E}_1 - \mathbf{E}_2$.

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 $E = E_1 \cdot E_2$; hence $\mathbf{E} = \mathbf{E}_1 \cdot \mathbf{E}_2$.

The only non-trivial case is the case of multiplication in the presence of possible correlation. When we know the exact values of E_1 and E_2, the solution to the above problem is known [8]:

Theorem 1 *If $y = x_1 \cdot x_2$, and we have no information about the correlation, then the range $[\underline{E}, \overline{E}]$ of $E[x_1 \cdot x_2]$ is $[E_{\min}, E_{\max}]$, where $p_i \stackrel{\text{def}}{=} (E_i - \underline{x}_i)/(\overline{x}_i - \underline{x}_i)$, and:*

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

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

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

$$\max(p_2 - p_1, 0) \cdot \underline{x}_1 \cdot \overline{x}_2 + \min(1 - p_1, 1 - p_2) \cdot \underline{x}_1 \cdot \underline{x}_2. \tag{2}$$

Comment. In this case, $\mathbf{E} = [E_{\min}, E_{\max}]$. In the following text, we will use the expressions (1) and (2) to describe the ranges of E for other cases, when the expression for the range $\mathbf{E} = [\underline{E}, \overline{E}]$ is different from the above expression $[E_{\min}, E_{\max}]$.

For the *inverse* $y = 1/x_1$, a finite range is possible only when $0 \notin \mathbf{x}_1$. Without loss of generality, we can consider the case when $0 < \underline{x}_1$. In this case, we have the following bound [8]:

Theorem 2 *For the inverse $y = 1/x_1$, the range of possible values of E is $\mathbf{E} = [1/E_1, p_1/\overline{x}_1 + (1 - p_1)/\underline{x}_1]$.*

(Here p_1 denotes the same value as in Theorem 1.)

Taking correlation into account. As we have seen, for elementary arithmetic operations other than multiplication, the range of the result's expectation is uniquely determined by the ranges of the input expectations. For multiplication, the range of $E[x_1 \cdot x_2]$ depends on both the ranges of $E[x_i]$ and the correlation between the x_i.

For multiplication, we know the bounds on $E[x_1 \cdot x_2]$ for two cases: when x_1 and x_2 are independent, and when we have no information about their correlation. In reality, we may have partial information about the correlation. For example, we may know the exact value ρ of the correlation

$$\rho(x_1, x_2) \stackrel{\text{def}}{=} \frac{E[x_1 \cdot x_2] - E_1 \cdot E_2}{\sigma_1 \cdot \sigma_2} \tag{3}$$

(where σ_i is the standard deviation of x_i). Or more generally we might have an interval $[\underline{\rho}, \overline{\rho}]$ of possible values of ρ.

Analytical expressions are desirable. In [5], a linear programming-based numerical method is described for computing the ranges of binary functions under constraints on the correlation of its arguments. For example, this method can be applied to the problem of estimating the range of $E[x_1 \cdot x_2]$ under known correlation.

In the cases of independence and unknown correlation, there are explicit analytical expressions for the range of $E[x_1 \cdot x_2]$. In general, as we have mentioned earlier, analytical expressions are much faster to compute than numerical methods. In this paper, we provide analytical expressions for the correlation case as well.

3 Main Results

Preliminaries. Our objective is, given the intervals $[\underline{x}_1, \overline{x}_1]$, $[\underline{x}_2, \overline{x}_2]$, the values $E_1 = E[x_1]$, $E_2 = E[x_2]$, and $\rho = \rho(x_1, x_2)$, to find the range $[\underline{E}, \overline{E}]$ of possible values of $E[x_1 \cdot x_2]$.

Before we derive an expression for the general situation, let us identify the quantitative values for Pearson correlation coefficient ρ corresponding to the known cases – independence and unknown correlation. For the independence case, $\rho = 0$.

For the case of unknown correlation, according to [8] both the smallest value $E_{\min}$ of $E[x_1 \cdot x_2]$ and the largest value $E_{\max}$ of $E[x_1 \cdot x_2]$ are attained when each of the variables x_i has a 2-point (2-impulse) marginal distribution: $p(x_i = \overline{x}_i) = p_i$ and $p(x_i = \underline{x}_i) = 1 - p_i$. (Probability p_i is uniquely determined by expected value $E[x_i]$.) For this marginal distribution,

$$\sigma^2[x_i] = E[(x_i - E_i)^2] = p_i \cdot (\overline{x}_i - E_i)^2 + (1 - p_i) \cdot (E_i - \underline{x}_i)^2.$$

Since $p_i = (E_i - \underline{x}_i)/(\overline{x}_i - \underline{x}_i)$, algebraic manipulation yields

$$\sigma^2[x_i] = (\overline{x}_i - E_i) \cdot (E_i - \underline{x}_i).$$

Thus, using eq. (3), the correlation coefficients $\rho_{\rm min}$ and $\rho_{\rm max}$ corresponding to these extreme distributions are equal to $\rho_{\rm min} = \dfrac{E_{\rm min} - E_1 \cdot E_2}{\sigma}$ and $\rho_{\rm max} = \dfrac{E_{\rm max} - E_1 \cdot E_2}{\sigma}$, where

$$\sigma \stackrel{\rm def}{=} \sigma_1 \cdot \sigma_2 = \sigma[x_1] \cdot \sigma[x_2] =$$

$$\sqrt{(\overline{x}_1 - E_1) \cdot (E_1 - \underline{x}_1)} \cdot \sqrt{(\overline{x}_2 - E_2) \cdot (E_2 - \underline{x}_2)}.$$

The case of unknown correlation includes the case of independence $\rho = 0$ as a particular case. For $\rho = 0$, we have $E[x_1, \cdot x_2] = E_1 \cdot E_2$; thus, the interval $[E_{\rm min}, E_{\rm max}]$ of possible values of $E[x_1 \cdot x_2]$ contains the value $E_1 \cdot E_2$: $E_{\rm min} \le E_1 \cdot E_2 \le E_{\rm max}$. Hence, we get $\rho_{\rm min} \le 0$ and $\rho_{\rm max} \ge 0$.

Case of exactly known non-zero correlation. The non-positive value $\rho_{\rm min}$ corresponds to the smallest possible value $E_{\rm min}$ of $E[x_1 \cdot x_2]$, and the non-negative value $\rho_{\rm max}$ corresponds to the largest possible value $E_{\rm max}$.

It is desirable to extend these results to intermediate values of $\rho \in [\rho_{\rm min}, \rho_{\rm max}]$.

Theorem 3 *Let $[\underline{x}_1, \overline{x}_1]$ and $[\underline{x}_2, \overline{x}_2]$ be given intervals, $E_1 \in [\underline{x}_1, \overline{x}_1]$ and $E_2 \in [\underline{x}_2, \overline{x}_2]$ be given numbers, and ρ be a number from the interval $[\rho_{\rm min}, \rho_{\rm max}]$. Then the closure $[\underline{E}, \overline{E}]$ of the range of possible values $E[x_1, x_2]$ for all possible distributions for which:*

- *x_1 is located in $[\underline{x}_1, \overline{x}_1]$, and x_2 is located in $[\underline{x}_2, \overline{x}_2]$;*
- *$E[x_1] = E_1$, and $E[x_2] = E_2$; and*
- *$\rho[x_1, x_2] = \rho$,*

is

- *for $\rho \ge 0$: $[E_1 \cdot E_2, E_1 \cdot E_2 + \rho \cdot \sigma]$;*
- *for $\rho \le 0$: $[E_1 \cdot E_2 + \rho \cdot \sigma, E_1 \cdot E_2]$.*

Comment. It should be mentioned that Theorem 3 does not claim that the range R of possible values of $E[x_1, x_2]$ *coincides* with the corresponding interval $[E_1 \cdot E_2, E_1 \cdot E_2 + \rho \cdot \sigma]$ or $[E_1 \cdot E_2 + \rho \cdot \sigma, E_1 \cdot E_2]$; this theorem only states that this interval coincides with the *closure* of the range R (i.e., with the set of all limits of all the sequences from the range R).

The difference between the range R and the corresponding (closed) interval comes from the fact that ρ is only defined when $\sigma_i > 0$. Thus, e.g., for $\rho > 0$, eq. (3) implies $E[x_1 \cdot x_2] > E[x_1] \cdot E[x_2]$. So, under the standard definition of (Pearson) correlation, the lower endpoint $E_1 \cdot E_2$ of the interval $[E_1 \cdot E_2, E_1 \cdot E_2 + \rho \cdot \sigma]$ might be unattainable.

If we instead define a distribution with correlation ρ as a distribution for which

$$E[x_1 \cdot x_2] = E[x_1] \cdot E[x_2] + \rho \cdot \sigma[x_1] \cdot \sigma[x_2],$$

then the degenerate distribution $x_1 \equiv E_1$, $x_2 \equiv E_2$, with $\sigma[x_1] = \sigma[x_2] = 0$, is a distribution with a given ρ for which $E[x_1 \cdot x_2] = E_1 \cdot E_2$. Under this alternative definition, the range R coincides with the corresponding interval – and there is no need to make the formulation more complex by referring to the closure.

Proof. When $\rho = 0$, then, by definition of the correlation, $E[x_1 \cdot x_2] = E_1 \cdot E_2$. So, it is sufficient to consider values of $\rho \neq 0$. In this proof, we will only consider the case $\rho > 0$; the case $\rho < 0$ is similar.

We first prove that the value $E[x_1 \cdot x_2]$ always belongs to the interval $[E_1 \cdot E_2, E_1 \cdot E_2 + \rho \cdot \sigma]$. $E_1 \cdot E_2$ is the lower bound because, since $\rho > 0$, we have $E[x_1 \cdot x_2] = E_1 \cdot E_2 + \rho \cdot \sigma[x_1] \cdot \sigma[x_2] > E_1 \cdot E_2$.

To prove the upper bound, we show that for each x_i, $\sigma^2[x_i] \le (E_i - \underline{x}_i) \cdot (\overline{x}_i - E_i)$. Let us first consider discrete distributions that take values $x_i^{(j)} \in [\underline{x}_i, \overline{x}_i]$ $(1 \le j \le N)$ with probabilities $p^{(j)} \ge 0$ such that $\sum\limits_{j=1}^{N} p^{(j)} = 1$. For such distributions, the constraint $E[x_i] = E_i$ takes the form $\sum\limits_{j=1}^{N} p^{(j)} \cdot x_i^{(j)} = E_i$. Under these constraints, let us find the largest possible value of

$$\sigma^2[x_i] = E[x_i^2] - E_i^2 = \sum_{j=1}^{N} p^{(j)} \cdot \left(x_i^{(j)}\right)^2 - E_i^2.$$

In terms of the unknown probabilities $p_i^{(j)}$, 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 is the idea 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 two probabilities $p_i^{(j)}$ must be equal to 0, i.e., the distribution must be located at two points x_i^- and x_i^+. Since the mean is E_i, these values must be on different sides of E_i. Without losing generality, we can thus assume that $x_i^- \le E_i \le x_i^+$.

We have already mentioned that for 2-point distributions, once the points x_i^- and x_i^+ are fixed, the condition that the mean equals E_i uniquely determines the probabilities, and the resulting variance is $(x_i^+ - E_i) \cdot (E_i - x_i^-)$. When $x_i^+ \le \overline{x}_i$ and $x_i^- \ge \underline{x}_i$, the largest value of this product is attained when x_i^+ attains its largest possible value $\overline{x}_i$, and x_i^- attains its smallest possible value $\underline{x}_i$. Thus, for discrete distributions, $\sigma^2[x_i] \le (\overline{x}_i - E_i) \cdot (E_i - \underline{x}_i)$.

An arbitrary distribution can be approximated by discrete ones to arbitrary accuracy (in weak topology), so this inequality is true for all distributions. Thus, $\sigma[x_1] \cdot \sigma[x_2] \le \sigma$, and the equality $E[x_1 \cdot x_2] = E_1 \cdot E_2 + \rho \cdot \sigma[x_2] \cdot \sigma[x_2]$ implies that $E[x_1 \cdot x_2] \le E_1 \cdot E_2 + \rho \cdot \sigma$.

We now prove that both endpoints are exact. For every $\varepsilon > 0$, if we take a distribution in which each x_i is located in the ε-vicinity of E_i, then $x_1 \cdot x_2$ (and hence $E[x_1 \cdot x_2]$) is located in the close vicinity of $E_1 \cdot E_2$. When $\varepsilon \to 0$, we conclude that $E[x_1 \cdot x_2]$ can be arbitrarily close to $E_1 \cdot E_2$, so the lower endpoint is indeed exact.

To complete the proof, we next show that the upper endpoint $E_1 \cdot E_2 + \rho \cdot \sigma$ is attainable, and thus also exact. Indeed, as we have mentioned, the largest possible value $E_{\max}$ is attained for a joint distribution in which both marginal distributions are 2-point ones, located on the endpoints of the corresponding interval $[\underline{x}_i, \overline{x}_i]$, and that for such distributions, $\sigma^2[x_i] = (\overline{x}_i - E_i) \cdot (E_i - \underline{x}_i)$. In general, distributions with such marginals are located at 4 vertices of the rectangle $[\underline{x}_1, \overline{x}_1] \times [\underline{x}_2, \overline{x}_2]$. The set of such distributions is determined by linear constraints and is, thus, connected. Along this set, the correlation ranges from 0 to the value $\rho_{\max}$. Since $\rho \in [0, \rho_{\max}]$ and correlation continuously depends on the probabilities, these exists an intermediate value of these probabilities where the correlation exactly equals the given value ρ.

The theorem is proven.

Case of correlation known with interval uncertainty. We can handle the case of an interval $[\underline{\rho}, \overline{\rho}]$ of possible values for ρ instead of an exact value of ρ by simply combining the intervals from Theorem 3 and using the fact that the corresponding formulas monotonically depend on ρ.

Theorem 4 *Let $[\underline{x}_1, \overline{x}_1]$ and $[\underline{x}_2, \overline{x}_2]$ be given intervals, $E_1 \in [\underline{x}_1, \overline{x}_1]$ and $E_2 \in [\underline{x}_2, \overline{x}_2]$ be given numbers, and $[\underline{\rho}, \overline{\rho}]$ be a subinterval of the interval $[\rho_{\min}, \rho_{\max}]$. Then the closure $[\underline{E}, \overline{E}]$ of the range of possible values $E[x_1, x_2]$ for all possible distributions for which:*

- *x_1 is located in $[\underline{x}_1, \overline{x}_1]$, and x_2 is located in $[\underline{x}_2, \overline{x}_2]$;*
- *$E[x_1] = E_1$, and $E[x_2] = E_2$; and*
- *$\rho[x_1, x_2] \in [\underline{\rho}, \overline{\rho}]$*

equals

- *for* $0 \le \underline{\rho}$*:* $[E_1 \cdot E_2, E_1 \cdot E_2 + \overline{\rho} \cdot \sigma]$*;*
- *for* $\overline{\rho} \le 0$*:* $[E_1 \cdot E_2 + \underline{\rho} \cdot \sigma, E_1 \cdot E_2]$*;*
- *for* $\underline{\rho} \le 0 \le \overline{\rho}$*:* $[E_1 \cdot E_2 + \underline{\rho} \cdot \sigma, E_1 \cdot E_2 + \overline{\rho} \cdot \sigma]$*.*

4 Auxiliary Results

Computationally efficient expressions for $E_{\min}$ and $E_{\max}$.

Proposition 1

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

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

Proof. Let us first simplify the expression for $E_{\max}$ from Theorem 1. When $p_1 \le p_2$, we get

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

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

Substituting the definitions of p_i, we conclude that

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

Opening parentheses, we get

$$E_{\max} = E^{(1)} \stackrel{\text{def}}{=} E_1 \cdot \overline{x}_2 - \underline{x}_1 \cdot \overline{x}_2 + E_2 \cdot \underline{x}_1.$$

By using the symmetry between x_1 and x_2, we can now conclude that when $p_1 \ge p_2$,

$$E_{\max} = E^{(2)} \stackrel{\text{def}}{=} E_2 \cdot \overline{x}_1 - \overline{x}_1 \cdot \underline{x}_2 + E_1 \cdot \underline{x}_2.$$

The condition $p_1 \le p_2$ is equivalent to

$$(E_1 - \underline{x}_1) \cdot (\overline{x}_2 - \underline{x}_2) \le (E_2 - \underline{x}_2) \cdot (\overline{x}_1 - \underline{x}_1),$$

i.e.,

$$E_1 \cdot \overline{x}_2 - E_1 \cdot \underline{x}_2 - \underline{x}_1 \cdot \overline{x}_2 + \underline{x}_1 \cdot \underline{x}_2 \le E_2 \cdot \overline{x}_1 - E_2 \cdot \underline{x}_1 - \overline{x}_1 \cdot \underline{x}_2 + \underline{x}_1 \cdot \underline{x}_2.$$

Subtracting the common term $\underline{x}_1 \cdot \underline{x}_2$ from both sides and moving terms to other sides, we get an equivalent form of this inequality:

$$E_1 \cdot \overline{x}_2 - \underline{x}_1 \cdot \overline{x}_2 + E_2 \cdot \underline{x}_1 \le E_2 \cdot \overline{x}_1 - \overline{x}_1 \cdot \underline{x}_2 + E_1 \cdot \underline{x}_2,$$

i.e., $E^{(1)} \le E^{(2)}$. So, if $p_1 \le p_2$, i.e., if $E^{(1)} \le E^{(2)}$, we get $E_{\max} = E^{(1)}$; otherwise, we get $E_{\max} = E^{(2)}$. These two cases can be combined into a single formula $E_{\max} = \min(E^{(1)}, E^{(2)})$, i.e.,

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

By adding $-E_1 \cdot E_2$ to both expressions $E^{(1)}$ and $E^{(2)}$, we get the desired expression for $E_{\max}$.

Since $E[x_1 \cdot x_2] = -E[(-x_1) \cdot x_2]$, where $-x_1 \in [-\overline{x}_1, \underline{x}_1]$ with $E[-x_1] = -E_1$, we have

$$E_{\min} \stackrel{\text{def}}{=} \min E[x_1 \cdot x_2] = -\max E[(-x_1) \cdot x_2].$$

Hence, the new expression for $E_{\max}$ leads to the desired expression for $E_{\min}$. The proposition is proven.

Can we propagate correlations through computations? In straightforward interval computations, we propagate intervals through computations; can we similarly propagate correlations? The following result shows that it is not easy even for addition:

Proposition 2 *If we know that $\rho[x_1, x_2] = \rho$, then the only possible conclusion about $\rho' = \rho[x_1, x_1 + x_2]$ is that $\rho' \in [\rho, 1]$.*

Proof. If we take $x_1 \ll x_2$, we get $\rho' \approx \rho$, and if we take $x_2 \ll x_1$, we get $\rho' \approx 1$. The smaller the corresponding ratio x_1/x_2 or x_2/x_1, the closer we are, correspondingly, to ρ and to 1.

Let us prove that ρ' cannot be smaller than ρ. Since correlation can be defined in terms of the differences $x_i - E[x_i]$, we can shift both variables to $E[x_i] = 0$ without changing the correlations $\rho[x_1, x_2]$ and $\rho[x_1, x_1 + x_2]$; thus, is it sufficient to prove the desired inequality $\rho' \ge \rho$ for the case when $E[x_i] = 0$. In this case, if we denote $\sigma_i \stackrel{\text{def}}{=} \sigma[x_i]$, we get

$$\rho' = \frac{E[x_1 \cdot (x_1 + x_2)]}{\sigma_1 \cdot \sigma[x_1 + x_2]} = \frac{\sigma_1^2 + E[x_1 \cdot x_2]}{\sigma_1 \cdot \sigma[x_1 + x_2]}.$$

Here, since $E_i = 0$, we have $E[x_1 \cdot x_2] = \rho \cdot \sigma_1 \cdot \sigma_2$. Similarly,

$$\sigma^2[x_1 + x_2] = E[(x_1 + x_2)^2] = E[x_1^2] + E[x_2^2] + 2 \cdot E[x_1 \cdot x_2] =$$

$$\sigma_1^2 + \sigma_2^2 + 2\rho \cdot \sigma_1 \cdot \sigma_2,$$

so the above expression for ρ' takes the form: $\rho' = \dfrac{\sigma_1 + \rho \cdot \sigma_1 \cdot \sigma_2}{\sigma_1 \cdot \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho \cdot \sigma_1 \cdot \sigma_2}}$,

and the desired inequality $\rho' \ge \rho$ takes the form $\dfrac{\sigma_1^2 + \rho \cdot \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho \cdot \sigma_1 \cdot \sigma_2}} \ge \rho$.

Multiplying both sides by the denominator, we get the equivalent inequality

$$\sigma_1 + \rho \cdot \sigma_2 \geq \rho \cdot \sqrt{\sigma_1^2 + \sigma_2^2 + 2\rho \cdot \sigma_1 \cdot \sigma_2}. \tag{4}$$

If $\rho \geq 0$, then we can square both sides and get an equivalent inequality

$$\sigma_1^2 + 2\rho \cdot \sigma_1 \cdot \sigma_2 + \rho^2 \cdot \sigma_2^2 \geq \rho^2 \cdot (\sigma_1^2 + \sigma_2^2 + 2\rho \cdot \sigma_1 \cdot \sigma_2).$$

Subtracting $\rho^2 \cdot \sigma_2^2$ from both sides, and moving all the terms to the right-hand side, we get an equivalent inequality

$$\sigma_1^2 \cdot (1 - \rho^2) + 2\rho \cdot \sigma_1 \cdot \sigma_2 \cdot (1 - \rho^2) \geq 0,$$

which is always true for $\rho \geq 0$ (since $\rho \leq 1$).

If $\rho < 0$, the right-hand side of (4) is negative, so we consider two possible cases. The first case is when

$$\sigma_1 + \rho \cdot \sigma_2 \geq 0.$$

Then inequality (4) is automatically true.

The second case is when $\sigma_1 + \rho \cdot \sigma_2 < 0$. In this case, (4) is equivalent to

$$0 < -\sigma_1 + |\rho| \cdot \sigma_2 \leq |\rho| \cdot \sqrt{\sigma_1^2 + \sigma_2^2 - 2|\rho| \cdot \sigma_1 \cdot \sigma_2}.$$

By squaring both sides, we get an equivalent inequality

$$\sigma_1^2 - 2|\rho| \cdot \sigma_1 \cdot \sigma_2 + \rho^2 \cdot \sigma_2^2 \leq \rho^2 \cdot (\sigma_1^2 + \sigma_2^2 - 2|\rho| \cdot \sigma_1 \cdot \sigma_2).$$

Subtracting $\rho^2 \cdot \sigma_2^2$ from both sides, and moving all the terms to the right-hand side, we get an equivalent inequality

$$\sigma_1^2 \cdot (1 - \rho^2) - 2|\rho| \cdot \sigma_1 \cdot \sigma_2 \cdot (1 - \rho^2) \leq 0.$$

Dividing both sides by $\sigma_1 \cdot (1 - \rho^2) > 0$, we get an equivalent inequality $\sigma_1 - 2|\rho| \cdot \sigma_2 \leq 0$. We consider the case when $\sigma_1 - |\rho| \cdot \sigma_2 < 0$, hence $\sigma_1 - 2|\rho| \cdot \sigma_2 \leq \sigma_1 - |\rho| \cdot \sigma_2 < 0$. The inequality is proven.

Since $x_1 - x_2 = x_1 + (-x_2)$, and $\rho[x_1, -x_2] = -\rho[x_1, x_2]$, we have the following corollary:

Proposition 3 *If we know that $\rho[x_1, x_2] = \rho$, then:*

- *the best possible conclusion about $\rho' = \rho[x_1, x_1 - x_2]$ is that $\rho' \in [-\rho, 1]$;*
- *the best possible conclusion about $\rho'' = \rho[x_2, x_1 - x_2]$ is that $\rho'' \in [-1, \rho]$.*

For multiplication $x_1 \cdot x_2$, we get an even wider range of value for the correlation:

Proposition 4 *Let $\rho \in [-1,1]$ be a given number. Then, the smallest interval $[\underline{\rho}', \overline{\rho}']$ that contains all possible values of $\rho' = \rho[x_1, x_1 \cdot x_2]$ for all pairs of random variables x_1 and x_2 for which $\rho[x_1, x_2] = \rho$ is the entire interval $[\underline{\rho}', \overline{\rho}'] = [-1, 1]$.*

Proof. Let x_1 and x_2 be an arbitrary pair with $\rho[x_1, x_2] = \rho$. Once can easily check that adding an arbitrary number a to x_2 does not change the value of the correlation, i.e., that for $x_2' \stackrel{\text{def}}{=} x_2 + a$, we still have $\rho[x_1, x_2'] = \rho$. For this new pair (x_1, x_2'), the correlation $\rho' = \rho[x_1, x_1 \cdot x_2']$ takes the form

$$\rho' = \frac{E[x_1^2 \cdot x_2'] - E[x_1] \cdot E[x_1 \cdot x_2']}{\sigma[x_1] \cdot \sigma[x_1 \cdot x_2']} =$$

$$\frac{E[x_1^2 \cdot (x_2 + a)] - E[x_1] \cdot E[x_1 \cdot (x_2 + a)]}{\sigma[x_1] \cdot \sigma[x_1 \cdot x_2 + x_1 \cdot a]} =$$

$$\frac{E[x_1^2 \cdot x_2] + a \cdot E[x_1^2] - E[x_1] \cdot E[x_1 \cdot x_2] - a \cdot (E[x_1])^2}{\sigma[x_1] \cdot \sigma[x_1 \cdot x_2 + x_1 \cdot a]}.$$

When $a \to \infty$, in the numerator, the prevailing term is $a \cdot (E[x_1^2] - (E[x_1])^2) = a \cdot \sigma^2[x_1]$. In the denominator, $x_1 \cdot a$ prevails over $x_1 \cdot x_2$, and thus, the denominator is asymptotically equal to $\sigma[x_1] \cdot \sigma[a \cdot x_1] = \sigma[x_1] \cdot |a| \cdot \sigma[x_1] = |a| \cdot \sigma^2[x_1]$. Therefore, when $a \to \infty$, we get

$$\rho' \sim \frac{a \cdot \sigma^2[x_1]}{|a| \cdot \sigma^2[x_1]} = \frac{a}{|a|} = \operatorname{sgn}(a).$$

In other words, when $a \to +\infty$, we get values of the correlation ρ' arbitrarily close to 1, and when $a \to -\infty$, we get values of the correlation ρ' arbitrarily close to -1. The proposition is proven.

For a *unary* linear function $f(x_1) = a \cdot x_1 + b$, we get $\rho[x_1, f(x_1)] = 1$ for $a > 0$ and $\rho[x_1, f(x_1)] = -1$ for $a < 0$.

For non-linear unary functions $f(x_1)$, instead of a single value $\rho[x_1, f(x_1)]$, we can get the interval of possible values.

As an example, let us take a simple non-linear function $f(x_1) = x_1^2$. For an arbitrary real number a and for $\varepsilon > 0$, we can consider a 2-point distribution located at $a - \varepsilon$ and $a + \varepsilon$ with probability 1/2. For this distribution, we get $E[x_1] = \dfrac{1}{2} \cdot ((a - \varepsilon) + (a + \varepsilon)) = a$. For $x_2 = f(x_1) = x_1^2$, we have $E[x_2] = \dfrac{1}{2} \cdot ((a - \varepsilon)^2 + (a + \varepsilon)^2) = a^2 + \varepsilon^2$, and $E[x_1 \cdot x_2] = \dfrac{1}{2} \cdot ((a - \varepsilon)^3 + (a + \varepsilon)^3) = a^3 + 3a \cdot \varepsilon^2$.

Here, the absolute value $|x_1 - E[x_1]|$ of the difference $x_1 - E[x_1]$ is equal to ε with probability 1, so $\sigma[x_1] = \varepsilon$; similarly, $\sigma[x_2] = 2|a| \cdot \varepsilon$. Therefore, we have

$$\rho[x_1, f(x_1)] = \frac{E[x_1 \cdot x_2] - E[x_1] \cdot E[x_2]}{\sigma[x_1] \cdot \sigma[x_2]} = \frac{a^3 + 3a \cdot \varepsilon^2 - a \cdot (a^2 + \varepsilon^2)}{\varepsilon \cdot 2|a| \cdot \varepsilon} =$$

$$\frac{a^3 + 3a \cdot \varepsilon^2 - a^3 - a \cdot \varepsilon^2}{2|a| \cdot \varepsilon^2} = \frac{2a \cdot \varepsilon^2}{2|a| \cdot \varepsilon^2} = \mathrm{sgn}(a).$$

Hence, for $a > 0$, we get $\rho = 1$, and for $a < 0$, we get $\rho = -1$.

In this case, the smallest interval containing possible values of the correlation $\rho[x_1, f(x_1)]$ is the entire interval $[-1, 1]$. A similar conclusion can be made for an arbitrary non-monotonic function $f(x_1)$: if we pick a on the increasing part of $f(x_1)$, we get $\rho \approx 1$, and if we pick a on the decreasing side, we get $\rho \approx -1$.

5 Conclusion and Open Problems

Conclusion. In many practical situations, in addition to intervals $\mathbf{x}_i$ of possible values of directly measured quantities $x_1, \ldots, x_n$, we also have partial information about the probabilities of different values within these intervals. For example, we may know the bounds on the first order moments and/or the bounds on the correlations.

The paper is a continuation of our previous work towards the use of probability information in interval computations. In the previous work, we explain how to take into account bounds on the first order moments. In this paper, we develop a new method that takes into account both correlation among measured parameters and bounds on their expected values when doing interval computation.

Open problems. Several open problems remain. What if we have a multiple product? For the case of unknown correlation, analytical formulas were obtained in [9].

What if we use different correlation characteristics [14], e.g., the Spearman and Kendall correlations, or copulas [6,12]?

What about the ranges for $E[\min(x_1, x_2)]$ and $E[\max(x_1, x_2)]$ under a given correlation (for the case of unknown correlation, such ranges were described in [8]).

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, editors, *Applications of Interval Computations*, Kluwer, Dordrecht, 1996.

[3] D. Berleant, M.-P. Cheong, C. Chu, Y. Guan, A. Kamal, G. Sheblé, S. Ferson, and J. F. Peters, "Dependable handling of uncertainty", *Reliable Computing*, 2003, Vol. 9, No. 6, pp. 407–418.

[4] 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.

[5] D. Berleant and J. Zhang, "Using Pearson correlation to improve envelopes around the distributions of functions," *Reliable Computing*, 2004, Vol. 10, No. 2, pp. 139–161.

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

[7] L. Jaulin, M. Keiffer, O. Didrit, E. Walter, *Applied Interval Analysis*, Springer-Verlag, London, 2001.

[8] V. Kreinovich, "Probabilities, Intervals, What Next? Optimization Problems Related to Extension of Interval Computations to Situations with Partial Information about Probabilities", *Journal of Global Optimization*, 2004, Vol. 29, No. 3, pp. 265–280.

[9] V. Kreinovich, S. Ferson, and L. Ginzburg, "Exact Upper Bound on the Mean of the Product of Many Random Variables With Known Expectations", Reliable Computing, 2003, Vol. 9, No. 6, pp. 441–463.

[10] V. Kreinovich, A. Lakeyev, J. Rohn, P. Kahl, *Computational complexity and feasibility of data processing and interval computations*, Kluwer, Dordrecht, 1997.

[11] V. P. Kuznetsov, *Interval statistical models*, Moscow, Radio i Svyaz Publ., 1991 (in Russian).

[12] R. B. Nelsen, *Introduction to Copulas*, Springer Verlag, New York, 1999.

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

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

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