While, In General, Uncertainty Quantification (UQ) Is NP-Hard, Many Practical UQ Problems Can Be Made Feasible*

1st Ander Gray and Scott Ferson
Institute for Risk and Uncertainty
University of Liverpool
Liverpool, UK
{Ander.Gray,Scott.Ferson}@liverpool.ac.uk

2nd Olga Kosheleva
Department of Teacher Education
University of Texas at El Paso
El Paso, Texas, USA
olgak@utep.edu

3rd Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, Texas, USA
vladik@utep.edu

***Abstract*—In general, many general mathematical formulations of uncertainty quantification problems are NP-hard, meaning that (unless it turned out that P = NP) no feasible algorithm is possible that would always solve these problems. In this paper, we argue that if we restrict ourselves to practical problems, then the correspondingly restricted problems become feasible – namely, they can be solved by using linear programming techniques.**

***Index Terms*—uncertainty quantification, p-box, NP-hard, decision making, linear programming**

I. Formulation of the Problem

A. Uncertainty Quantification (UQ) is important

Most of our knowledge about the world comes ultimately from measurements. We also use theoretical models, but these models – even those that are not directly inspired by measurements results – have to be confirmed by measurements, and are only as reliable as the measurements that confirm them.

No matter how accurate are the measurements, they are never absolutely accurate. In general, the result $\widetilde{x}$ of measuring a quantity x is different from the actual (unknown) value x of this quantity. In other words, we have a non-zero measurement error $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$; see, e.g., [13].

Because of this uncertainty, the estimates that we make based on these measurements are also uncertain. To make proper decisions, we need to understand how accurate are these estimates. For example, if we are prospecting for oil in a certain area and the current estimate is that this area contains 100 million tons of oil, then whether it is very good news or just maybe good news depends on how accurate is this estimate:

- if it is 100 ± 20, then we should start exploiting this oil field;
- however, if it is 100 ± 200, then maybe there is no oil there at all, so it is advisable to do some more measurements before investing money into this area.

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes), and by the AT&T Fellowship in Information Technology. It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

B. How uncertainty is usually described

Let us denote all the quantities that we measure by $x_1, \ldots, x_m$. In these terms, we want to know which tuples $X = (x_1, \ldots, x_m)$ are possible – i.e., what is the set $\mathcal{X}$ of possible tuples, and what are the probabilities of different tuples X from this set.

Ideally, we should therefore find the probability distribution on the set of all possible tuples. But where can we get the corresponding probabilities? There is only one source of all the information about the world – and, in particular, of the information about the probabilities – measurements.

During any period of time, we can only perform finitely many measurements. A general probability distribution requires infinitely many parameters to describe – and, based on a finite set of measurements, we cannot uniquely determine the probability distribution. So, in practice, we never know the exact probability distribution – we only have partial information about the actual (unknown) probability distribution.

We may get different types of information about this distribution:

- For example, we can have bounds on the distribution's moments.
- We can have bounds $\underline{F}_i(v_i) \le F_i(v_i) \le \overline{F}_i(v_i)$ on the values of the cumulative distribution functions

$$F_i(v_i) \stackrel{\text{def}}{=} \text{Prob}(x_i \le v_i);$$

  such bounds are known as *probability boxes*, or *p-boxes*, for short.
- We may have many other different types of information about uncertainty.

C. In general, UQ problems are NP-hard

In general, uncertainty quantification problems are NP-hard; see, e.g., [1], [9], [12] and references therein. This means, crudely speaking, that, unless P = NP (which most computer

scientists believe to be false), no feasible algorithm is possible that would solve all possible UQ problems.

D. What we do in this paper

In this paper, we argue that if we go down to the level of the original measurement results, then most (if not all) *practical* UQ problems become feasible.

To make this argument, we need to formulate a general UQ problem, namely, we need to specify:

- what information can we have – this we analyze in Section 2, and
- what do we want to estimate – this we analyze in Section 3.

Based on this general description, we need to explain how we can feasibly estimate what we want based on the information that we have – this we describe in Section 4.

II. What Information Can We Have

A. The main claim of this section

The main claim of this section is that many types of partial information about the probability distribution consist of linear inequalities, i.e., inequalities of the type

$$\int a(X) \cdot f(X)\, dX \le b \tag{1}$$

for some function $a(X)$, where $f(X)$ is the actual (unknown) probability density function.

To support this claim, we will first describe typical types of partial information (many of them are mentioned in [3]), and show that they indeed have this form (1) – or at least can be equivalently reformulated in this form.

After that, we provide general arguments that *any* reasonable partial information has this form.

B. Upper bounds vs. lower bounds

The formula (1) provides an upper bound b on the statistical characteristic

$$\int a(X) \cdot f(X)\, dX.$$

In some practical situations, we have lower bounds:

$$b \le \int a(X) \cdot f(X)\, dX. \tag{2}$$

From the physical viewpoint, lower bounds are different. However, from the computational viewpoint, each lower bound can be easily reformulated in terms of equivalent upper bound. Namely, if we multiply both sides of the inequality (2) by -1, we get the inequality

$$\int a'(X) \cdot f(X) \le b', \tag{3}$$

where we denoted $a'(X) \stackrel{\text{def}}{=} -a(X)$ and $b' \stackrel{\text{def}}{=} -b$.

Because of this possibility, in the following text, we will mostly consider upper bounds.

C. What if we know the exact value

If we know the exact value b of the corresponding statistical characteristics, i.e., if we know that

$$\int a(X) \cdot f(X)\, dX = b,$$

then this knowledge can be described as two bounds:

$$b \le \int a(X) \cdot f(X)\, dX$$

and

$$\int a(X) \cdot f(X)\, dX \le b,$$

i.e., equivalently, as two upper bounds: (1) and

$$\int (-a(X)) \cdot f(X)\, dX \le (-b).$$

D. Continuous vs. discrete

From the purely mathematical viewpoint, most physical quantities x can take any real-number values – and there are infinitely many possible real numbers. However, from the practical viewpoint, we need to take into account that values which are too close to each other are practically indistinguishable.

For example, theoretically, we can determine a person's weight with microgram accuracy, but then the weight will change when a person breathes in or breathes out, sweats, or drinks a cup of water. Measuring this weight with accuracy better than ± 200 grams makes no practical sense. Similarly, when we measure outside temperature, making too accurate measurements makes no sense: the temperature at different parts of the street may differ by half a degree or even more. The same is true for all possible quantities.

Also, for each measuring instrument, there are natural bounds within which it can provide meaningful measurements. For example:

- a ruler cannot be used to measure distances larger than a certain amount,
- all scales have their limitations after which the scale will be simply crushed by the weight,
- thermometers melt if the temperature is too high and freeze when it is too low, etc.

Because of this, in reality, each quantity x_i has only finitely many practically distinguishable values:

- the smallest detectable value $x_{i,0}$,
- the next value $x_{i,1} = x_{i,0} + h$, where h_i is the smallest difference that makes practical sense,
- the value $x_{i,2} = x_{i,0} + 2h_i$, etc., all the way to
- the largest possible value $x_{i,n_i} = x_{i,0} + N_i \cdot h_i$ for an appropriate N_i.

As a result, there are finitely many possible values of the tuple $X = (x_1, \ldots, x_m)$: namely, only the tuples

$$X_{n_1,\ldots,n_m} \stackrel{\text{def}}{=} (x_{1,n_1}, \ldots, x_{m,n_m}).$$

In these terms, to describe the probability distribution, it is sufficient to describe the probability

$$p_{n_1,\ldots,n_m} \stackrel{\text{def}}{=} \text{Prob}(X = X_{n_1,\ldots,n_m})$$

of each possible tuple. In terms of these probabilities, the inequality (1) takes the form

$$\sum_{n_1,\ldots,n_m} a(X_{n_1,\ldots,n_m}) \cdot p_{n_1,\ldots,n_m} \le b. \tag{4}$$

The left-hand side of the formula (4) is nothing else but an integral sum of the integral from the formula (1). When h is small – and usually, it is small – the integral sum is very close to the actual integral: it is sufficient to recall that the usual way of estimating an integral is to compute the corresponding integral sum.

Thus, from the practical viewpoint, we will consider formulas (1) and (4) as interchangeable, and use both.

E. Moments and bounds on moments

Let us start providing examples of partial information about probabilities that can be described in the equivalent form (1)–(4). Our first example is *moments*, i.e., expressions of the type

$$M_{k_1,\ldots,k_m} \stackrel{\text{def}}{=} \int x_1^{k_1} \cdot \ldots \cdot x_m^{k_m} \cdot f(x_1,\ldots,x_m)\, dx_1\, \ldots\, dx_m,$$

or, in discrete form,

$$M_{k_1,\ldots,k_m} = \sum_{n_1=1}^{N_1} \ldots \sum_{n_m=1}^{N_m} x_{1,n_1}^{k_1} \cdot \ldots \cdot x_{m,n_m}^{k_m} \cdot p_{n_1,\ldots,n_m}.$$

Both expressions are clearly linear in terms of the corresponding probabilities – i.e., in terms of $f(X)$ or of the probabilities $p_{n_1,\ldots}$. Thus, any bound on the moments is a linear inequality – i.e., has the desired form (1)–(4).

F. Cumulative distribution functions (cdf) and p-boxes

Another possible information is information about (i.e., bounds on) the cumulative distribution function $F(v)$ of:

- either the quantities x_i themselves,
- or, more generally, of a quantity

$$y = s(X) = s(x_1,\ldots,x_n)$$

depending on these quantities.

The value $F(v) = \text{Prob}(y \le v)$ can be described as

$$F(v) = \int_{X:\ s(X)\le v} f(X)\, dX,$$

i.e., equivalently, as

$$F(v) = \int_X a(X) \cdot f(X)\, dX,$$

where:

- $a(X) = 1$ if $s(X) \le v$, and
- $a(X) = 0$ otherwise.

Similarly, in the discrete case, the value $F(v)$ can be described as

$$\sum_{X_{n_1,\ldots,n_m}} a(X_{n_1,\ldots,n_m}) \cdot p_{n_1,\ldots,n_m}.$$

In both continuous and discrete cases, constraints

$$\underline{F}(v) \le F(v) \le \overline{F}(v)$$

become inequalities which are linear with respect to $f(X)$ or $p_{n_1,\ldots,n_m}$.

G. Information about the probability density function

We can also have information about (i.e., bounds on) the values $f(X)$ (or, in the discrete case, $p_{n_1,\ldots,n_m}$) of the probability density function $f(X)$ itself.

In this case, of course, the bounds $f(X) \le b$ or $p_{n_1,\ldots,n_m} \le b$ are already inequalities which are linear in terms of $f(X)$ or $p_{n_1,\ldots,n_m}$.

H. Symmetry information

In addition to bounds on the values of different statistical characteristics, we may also have information about the invariance of these characteristics with respect to some transformations.

For example, we may know that a 1-D distribution $f(x_1)$ is symmetric with respect to the transformation $x_1 \mapsto -x_1$. This invariance means that $f(-x_1) = f(x_1)$, i.e., that $f(-x_1) - f(x_1) = 0$. This is also a linear equality in terms of the function $f(X)$ – and can, thus, be described as two linear inequalities.

I. What about the general case

In general, how do we estimate a general statistical characteristic? Like every other knowledge about the words, we estimates these characteristics based on the measurement results.

Specifically, we have several tuples $X^{(1)},\ldots,X^{(K)}$ corresponding to different independent measurements, and we want to estimate the desired characteristics based on this K-element sample.

As we have mentioned, in practice, at any given moment of time, we can have only finitely many measurement results. Based on these measurement results, we can only determine the values of finitely many parameters describing the corresponding probability distribution. Thus, no matter what estimation method we use, we restrict ourselves – explicitly or implicitly – to a finite-parametric family of distributions

$$f(X, c_1,\ldots,c_t),$$

in which each distribution is obtained by specifying the values of the parameters c_j.

For example, in engineering applications, we often explicitly restrict ourselves to Gaussian distributions, for which the known formula describes the pdf in terms of means and the covariance matrix. On the other hand, a widely used way to determine a probability distribution based on partial

information is the Maximum Entropy approach (see, e.g., [7]), in which, among all probability distributions $f(X)$ which are consistent with observations, we select the distribution for which the entropy

$$-\int f(X)\cdot\ln(f(X))\,dX$$

attains the largest possible value.

Similarly, if we use machine learning [2] – e.g., deep learning [6] – to determine the desired distribution, we still get a finite-parametric family of the distributions – but this time, this family is implicitly described, there is no simple analytical expression for distributions from this family.

In terms of the family $f(X,c_1,\ldots,c_t)$, identifying a distribution means estimating the values of all the parameters c_j. How can we estimate these parameters based on the sample? For each combination of values c_j, the probability of observing each tuple $X^{(k)}$ is equal to $f\left(X^{(k)},c_1,\ldots,c_t\right)$. Since these measurements are independent, the probability that we have observed all tuples is equal to the product of these probabilities, i.e., to the value

$$f\left(X^{(1)},c_1,\ldots,c_t\right)\cdot\ldots\cdot f\left(X^{(K)},c_1,\ldots,c_t\right). \qquad (5)$$

This value represents, in effect, the probability that the values c_j are the good fit for the observed tuples. Since we need to select a single combination of the parameters c_j, a reasonable idea is to select the most probable combination

$$c=(c_1,\ldots,c_t),$$

i.e., the combination for which the product (5) attains the largest possible value. This approach – known as the *Maximum Likelihood* approach – is indeed one of the most widely used techniques for estimating the values of different statistical characteristics; see, e.g., [15].

How does this lead to linear inequalities? Maximizing the product (5) is equivalent to maximizing its logarithm, i.e., the sum

$$\sum_{k=1}^{K}\ln\left(f\left(X^{(k)},c_1,\ldots,c_t\right)\right). \qquad (6)$$

According to calculus, the maximum of a function is attained when all its partial derivatives are equal to 0, i.e., when for each $j=1,\ldots,t$, we have

$$\sum_{k=1}^{K}\frac{\partial}{\partial c_j}\left(\ln\left(f\left(X^{(k)},c_1,\ldots,c_t\right)\right)\right)=0. \qquad (7)$$

So, we have

$$\sum_{k=1}^{K}a_j\left(X^{(k)}\right)=0. \qquad (8)$$

where we denoted

$$a_j(X)\stackrel{\text{def}}{=}\frac{\partial}{\partial c_j}\left(\ln\left(f\left(X,c_1,\ldots,c_t\right)\right)\right). \qquad (9)$$

For any function $a(X)$, a natural estimate of its mean value $\int a(X)\cdot f(X)\,dX$ based on sample is the sample average:

$$\frac{a\left(X^{(1)}\right)+\ldots+a\left(X^{(K)}\right)}{K}.$$

The formula (8) implies that the sample average of the values $a_j(X)$ is equal to 0:

$$\frac{a_j\left(X^{(1)}\right)+\ldots+a_j\left(X^{(K)}\right)}{K}=0. \qquad (10)$$

Thus, the corresponding mean value $\int a_j(X)\cdot f(X)\,dX$ is also close to 0 – with accuracy ε with which we can estimate this mean by a sample mean. In other words, the estimates c_j are equivalent to the following linear inequalities

$$-\varepsilon\le\int a_j(X)\cdot f(X)\,dX\le\varepsilon$$

for the functions $a_j(X)$ defined by the formula (9).

In other words, in the general case, we can also formulate any partial knowledge about a probability distribution in terms of linear inequalities.

III. What Do We Want to Estimate

A. The main claim of this section

The main claim of this section is that the decisions of a rational decision makers are equivalent to maximizing some expression

$$c=\int c(X)\cdot f(X)\,dX \qquad (11)$$

which is linear in terms of the probabilities $f(X)$.

In the discrete case, this maximized expression takes the form

$$c=\sum_{n_1=1}^{N_1}\ldots\sum_{n_m=1}^{N_m}c_{n_1,\ldots,n_m}\cdot p_{n_1,\ldots,n_m}. \qquad (12)$$

To justify this claim, we will use general ideas of decision theory; see, e.g., [5], [8], [10]–[12], [14].

B. How to describe preferences in numerical terms

Computers have been invented to deal with numbers. Numbers are still what computers process most efficiently. So, to enable computers to help us make decisions – and making decisions is one of the main objectives of science and engineering – it is desirable to describe all available information into numbers. In particular, it is desirable to transform information about our preferences into numbers.

To perform this transformation, we need to have a numerical scale for preferences. This scale can be constructed as follows. First, let us select two alternatives:

- an alternative A_- which is worse than anything that we can potentially encounter; we will call this alternative *very bad*; and
- an alternative A_+ which is better than anything that we can potentially encounter; we will call this alternative *very good*.

Then, for all numbers p from the interval $[0,1]$, we can form a lottery in which:

- we get a very good alternative A_+ with probability p, and
- we get a very bad alternative A_- with the remaining probability $1-p$.

We will denote this lottery by $L(p)$.

Now, let us consider any of the actual alternatives A. Then:

- when p is close to 0, then the lottery $L(p)$ is close to the very bad alternative A_- and is, thus, worse than A; we will denote it by $L(p) < A$;
- on the other hand, when p is close to 1, then the lottery $L(p)$ is close to the very good alternative A_+ and is, thus, better than A: $A < L(p)$.

As the probability p of getting the very good alternative increases, the lottery $L(p)$ becomes better and better. At some point, we will switch from $L(p) < A$ to $A < L(p)$. This threshold value $u(A)$, for which:

- $L(p) < A$ for $p < u(A)$, and
- $A < L(p)$ for $p > u(A)$

is called the *utility* of the alternative A.

By definition of the utility, for every $\varepsilon > 0$, we have

$$L(u(A)-\varepsilon) < A < L(u(A)+\varepsilon).$$

As we have mentioned earlier, when the value ε is sufficiently small, there is no way to practically distinguish probabilities $u(A)$ and $u(A) \pm \varepsilon$. Thus, from the practical viewpoint, the alternative A is equivalent to the lottery $L(u(A))$; we will denote this practical equivalence by $A \equiv L(u(A))$.

For lotteries $L(p)$, the larger the probability, the better:

$$p < q \Leftrightarrow L(p) < L(q).$$

Thus, in general, $A < B$ if and only $u(A) < u(B)$. So, utilities indeed provide a numerical representation of preferences.

C. How to describe preferences under uncertainty

In practice, we often cannot predict with 100% certainty what will be the consequence of each action. Suppose that for some action a, possible results are $A_1, \ldots, A_r$ with probabilities $p_1, \ldots, p_r$. So, this action is equivalent to a lottery in which we get each alternative A_i with probability p_i.

Each alternative A_i, in its turn, is equivalent to a lottery in which we get A_+ with probability $u(A_i)$ and A_- with the remaining probability $1-u(A_i)$. Thus, the action a is equivalent to a two-stage lottery, in which:

- first, we select one of the alternatives A_i with probability p_i, and
- then, depending on which alternative A_i we selected on the first stage, we select A_+ with probability $u(A_i)$ or A_- with the remaining probability $1-u(A_i)$.

As a result of this two-stage lottery, we get either A_+ or A_-. One can see that the probability of selecting A_+ in this two-stage lottery is equal to the sum

$$\sum_{i=1}^{r} p_i \cdot u(A_i). \tag{13}$$

By definition of utility, this means that the formula (13) describes the utility of the action a. So, this formula describes the quality of each action – as a linear combination of the corresponding probabilities. Thus, to decide which action is better, we need to estimate this expression (13) for different actions.

In particular, in our case, when alternatives correspond to different tuples $X_{n_1,\ldots,n_m}$ with probabilities $p_{n_1,\ldots,n_m}$, we get the expression

$$\sum_{n_1=1}^{N_1} \cdots \sum_{n_m=1}^{N_m} c_{n_1,\ldots,n_m} \cdot p_{n_1,\ldots,n_m}, \tag{14}$$

where

$$c_{n_1,\ldots,n_m} \stackrel{\text{def}}{=} u(X_{n_1,\ldots,n_m})$$

is the utility of the corresponding tuple.

In the integral form, this expression takes the form

$$\int c(X) \cdot f(X)\, dX. \tag{15}$$

Thus, to make decisions, we need to be able to estimate, based on the available knowledge, the value of the expression (14)–(15).

IV. How Can We Feasibly Estimate the Desired Quantities: General Idea and an Example

A. General case

According to the above analysis, our knowledge of the actual (unknown) probabilities) $p_{n_1,\ldots,n_m}$ can be described in terms of linear inequalities

$$\begin{aligned} \sum_{n_1=1}^{N_1} \cdots \sum_{n_m=1}^{N_m} a^{(1)}_{n_1,\ldots,n_m} \cdot p_{n_1,\ldots,n_m} &\le b^{(1)}, \\ &\cdots \\ \sum_{n_1=1}^{N_1} \cdots \sum_{n_m=1}^{N_m} a^{(L)}_{n_1,\ldots,n_m} \cdot p_{n_1,\ldots,n_m} &\le b^{(L)}. \end{aligned} \tag{16}$$

Based on this information, we want to estimate the value of the objective function

$$c = \sum_{n_1=1}^{N_1} \cdots \sum_{n_m=1}^{N_m} c_{n_1,\ldots,n_m} \cdot p_{n_1,\ldots,n_m}. \tag{17}$$

In general, due to uncertainty, the value of the objective function is not uniquely determined by the available data, we can have the whole range $[\underline{c}, \overline{c}]$ of possible values. Here:

- The lower endpoint $\underline{c}$ of this range can be found if we minimize the expression (17) under constraints (16).
- The upper endpoint $\overline{c}$ of this range can be found if we maximize the expression (17) under constraints (16).

In both cases, we optimize a linear expression under linear inequalities. Such optimization problems are known as *linear programming* problems. Good news is that there exists feasible algorithms for solving linear programming problems; see, e.g., [16].

Thus, indeed, we can conclude that many practical uncertainty optimization problems are feasible.

B. Example

Suppose that we know:

- the joint probability of x_1 and x_2, and
- the joint probability of x_2 and x_3.

What can we then conclude about the joint probability of x_1 and x_3?

In other words, for each n_1, n_2, and n_3, we know the values

$$p^{(1,2)}_{n_1,n_2} = \mathrm{Prob}(x_1 = x_{1,n_1} \,\&\, x_2 = x_{2,n_2})$$

and

$$p^{(2,3)}_{n_2,n_3} = \mathrm{Prob}(x_2 = x_{1,n_2} \,\&\, x_3 = x_{3,n_3}).$$

Based on this information, we need to estimate the values

$$p^{(1,3)}_{n_1,n_3} = \mathrm{Prob}(x_1 = x_{1,n_1} \,\&\, x_3 = x_{3,n_3}),$$

i.e., to find the ranges

$$\left[\underline{p}^{(1,3)}_{n_1,n_3}, \overline{p}^{(1,3)}_{n_1,n_3}\right].$$

In terms of the unknowns p_{n_1,n_2,n_3}, each desired value $p^{(1,3)}_{n_1,n_3}$ has the form

$$\sum_{n_2=1}^{N_2} p_{n_1,n_2,n_3}, \tag{18}$$

while the available information has the form

$$\sum_{n_3=1}^{N_3} p_{n_1,n_2,n_3} = p^{(1,2)}_{n_1,n_2} \tag{19}$$

and

$$\sum_{n_1=1}^{N_1} p_{n_1,n_2,n_3} = p^{(2,3)}_{n_2,n_3}. \tag{20}$$

Thus, to solve our problem, we must solve, for each n_1 and n_3, the following two linear programming problems:

- to find the lower endpoint of the corresponding range, we minimize the expression (18) under the conditions (19) and (20); and
- to find the upper endpoint of the corresponding range, we maximize the expression (18) under the conditions (19) and (20).

C. Can we get non-trivial bounds this way?

Sometimes, we may get trivial bounds $\underline{p}^{(1,3)}_{n_1,n_3} = 0$ and $\overline{p}^{(1,3)}_{n_1,n_3} = 1$.

However, there are cases when we get non-trivial bounds. For example, suppose that for all three variables, we have the same sequence of values $x_{1,i} = x_{2,i} = x_{3,i}$ for all i, and that $p^{(1,2)}_{i,j} = 0$ for $i \neq j$ and $p^{(2,3)}_{j,k} = 0$ when $j \neq k$. This means that:

- with probability 1, $x_1 = x_2$, and
- with probability 1, $x_2 = x_3$.

In this case, we conclude that $x_1 = x_3$, i.e., that $p^{(1,3)}_{i,k} = 0$ for all $i \neq k$.

One can see that in this situation, the above linear programming problems lead to $\underline{p}^{(1,3)}_{i,k} = \overline{p}^{(1,3)}_{i,k} = 0$ for all $i \neq k$.

References

[1] D. J. Berleant, O. Kosheleva, V. Kreinovich, and H. T. Nguyen, "Unimodality, independence lead to NP-hardness of interval probability problems", Reliable Computing, vol. 13, no. 3, pp. 261–282, 2007.

[2] C. M. Bishop, Pattern Recognition and Machine Learning, New York: Springer, 2006.

[3] M. Daub, S. Marelli, and B. Sudret, "On constrained distribution-free p-boxes and their propagation", Proceedings of the 9th International Workshop on Reliable Engineering Computing REC'2021, Taormina, Italy, May 16–20, 2021, pp. 365–379.

[4] S. Ferson and A. Gray, "Distribution-free uncertainty propagation", Proceedings of the 9th International Workshop on Reliable Engineering Computing REC'2021, Taormina, Italy, May 16–20, 2021, pp. 395–407.

[5] P. C. Fishburn, Utility Theory for Decision Making, New York, NY: John Wiley & Sons Inc., 1969.

[6] I. Goodfellow, Y. Bengio, and A. Courville, Deep Leaning, Cambridge, Massachusetts: MIT Press, 2016.

[7] E. T. Jaynes and G. L. Bretthorst, Probability Theory: The Logic of Science, Cambridge, UK: Cambridge University Press, 2003.

[8] V. Kreinovich, "Decision making under interval uncertainty (and beyond)", In: P. Guo and W. Pedrycz (Eds), Human-Centric Decision-Making Models for Social Sciences, Cham, Switzerland: Springer Verlag, 2014, pp. 163–193.

[9] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational Complexity and Feasibility of Data Processing And Interval Computations*, Kluwer, Dordrecht, 1998.

[10] R. D. Luce and R. Raiffa, Games and Decisions: Introduction and Critical Survey, New York, NY: Dover, 1989.

[11] H. T. Nguyen, O. Kosheleva, and V. Kreinovich, "Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction", International Journal of Intelligent Systems, vol. 24, no. 1, pp. 27–47, 2009.

[12] H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Berlin, Heidelberg: Springer Verlag, 2012.

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

[14] H. Raiffa, Decision Analysis, Columbus, Ohio: McGraw-Hill, 1997.

[15] D. J. Sheskin, Handbook of Parametric and Non-Parametric Statistical Procedures, London, UK: Chapman & Hall/CRC, 2011.

[16] R. J. Vanderbei, Linear Programming: Foundations and Extensions, Cham, Switzerland: Springer, 2020.