Departmental Technical Reports (CS)   Computer Science

7-2019

Beyond p-Boxes and Interval-Valued Moments: Natural Next Approximations to General Imprecise Probabilities

Olga Kosheleva
The University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Recommended Citation

Beyond p-Boxes and Interval-Valued Moments: Natural Next Approximations to General Imprecise Probabilities

Olga Kosheleva and Vladik Kreinovich

Abstract To make an adequate decision, we need to know the probabilities of different consequences of different actions. In practice, we only have partial information about these probabilities – this situation is known as *imprecise probabilities*. A general description of all possible imprecise probabilities requires using infinitely many parameters. In practice, the two most widely used few-parametric approximate descriptions are p-boxes (bounds on the values of the cumulative distribution function) and interval-valued moments (i.e., bounds on moments). In some situations, these approximations are not sufficiently accurate. So, we need more accurate more-parametric approximations. In this paper, we explain what are the natural next approximations.

1 How Decision Theory Describes Human Preferences: A Brief Reminder

One of the main objectives of decision making theory is to help people make decisions. To be able to provide this help, we need to be able to take into account people's preferences.

A natural way to elicit preferences is to ask people to compare different alternatives and select the one which is, in their opinion, the most preferable. As a result, we get a large number of pairs of alternatives (A, A') in which the person preferred the alternative A to the alternative A'. However, directly dealing with such sets of pairs is difficult. One of the main reasons for this difficulty is that computers have been originally designed to process numbers, not discrete objects like sets of pairs, and computers are still much more efficient in processing numbers than in processing

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso, El Paso, Texas 79968, USA
e-mail: olgak@utep.edu, vladik@utep.edu

other types of data. Thus, to enhance computer processing of people's preferences, it is desirable to describe these preferences by numbers.

Such a description is indeed possible, it is known as the *utility* approach; see, e.g., [5, 8, 9, 13, 15]. In this approach, we select two fixed alternatives:

- a very bad alternative A_- which is worse than anything that we will actually encounter, and
- a very good alternative A_+ which is better than anything that we will actually encounter.

Then, for each number p from the interval $[0, 1]$, we can form a lottery $L(p)$ in which:

- we get A_+ with probability p and
- we get A_- with the remaining probability $1 - p$.

When $p = 0$, the lottery $L(0)$ coincides with the very bad alternative A_- and is, thus, worse than any actual alternative A: $L(0) < A$. When $p = 1$, the lottery $L(1)$ coincides with the very good alternative A_+ and is thus, better than the actual alternative A: $A < A_+$. The larger the probability p of the very good alternative, the better the lottery $L(p)$. So:

- if $L(p) < A$ and $p' < p$, then also $L(p') < A$;
- similarly, if $A < L(p)$ and $p < p'$, then also $A < L(p')$.

So, when we compare the alternative A with lotteries $L(p)$ corresponding to increasing values p, at some point, we switch from $L(p) < A$ to $A < L(p)$. Thus, for each alternative A, there exists a threshold value p_0 such that:

- for $p < p_0$, we have $L(p) < A$, while
- for $p > p_0$, we have $A < L(p)$.

This threshold value can be formally defined as $\sup\{p : L(p) < A\}$ or, equivalently, as $\inf\{p : A < L(p)\}$. This threshold value is called the *utility* of the alternative A; it is usually denoted by $u(A)$.

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

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

Since the value ε can be arbitrarily small, and in practice, we do not notice very small differences, we can thus conclude that, from the practical purposes, the original alternative A is equivalent to the lottery $L(u(A))$. We will denote this practical equivalence by $A \equiv L(u(A))$.

So, if we have several alternatives, we should select the one whose utility is the largest.

Hence, to describe a person's preferences, we can use the utility function that assigns, to each alternative A, the utility $u(A)$ of this alternative to this person.

2 Which Action to Select: Discrete Case

How can use the utility-based description of preferences when making a decision? In decision making, we need to select between several possible actions – including, sometimes, an "action" of doing nothing. The difficulty is that usually, we cannot exactly predict the consequences of each action: usually, several different consequences are possible.

For example, when we invest money in a new company, we may gain (if the company prospers) or we may lose (if the company fails). When a medical doctor prescribes a strong medicine to a patient, this may lead to the patient's recovery – or it may lead to the appearance of very bad side effects, which are, unfortunately, possible for strong medicines.

Usually, we cannot predict the exact consequence of each action a, but, based on the previous experience, we can estimate the probabilities $p_1, \ldots, p_n$ of different consequences $A_1, \ldots, A_n$. Let $u(A_i)$ denote the utility of each of the possible consequences. By definition of utility, this means that each alternative A_i is equivalent to a lottery $L(u(A_i))$ in which:

- we get A_+ with probability $u(A_i)$ and
- we get A_- with the remaining probability $1 - u(A_i)$.

Thus, the action a is equivalent to a complex two-stage lottery, in which:

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

As a result of this two-stage lottery, we end up either with A_+ or with A_-. We can use the formula of complete probability to find the probability $u(a)$ of getting A_+:

$$u(a) = p_1 \cdot u(A_1) + \ldots + p_n \cdot u(A_n). \tag{1}$$

Thus, the action a is equivalent to lottery in which we get A_+ with probability $u(a)$ and A_- with the remaining probability. By definition of utility, this means that the utility of the action a is thus equal to $u(a)$.

When selecting an action, we need to select the action with the largest possible utility. The formula (1) for the action's utility is exactly the formula for the expected value of the utility. So, we can conclude that we need to select the action with the largest value of the expected utility.

3 Which Action to Select: General Case

In general, each action can have infinitely many possible consequences: e.g., if we give a strong medicine to a patient with fever, his resulting body temperature can

take any value from normal to – in unfortunate cases – an even higher fever. In such situations, we still need to compare expected values, but we now have an integral instead of the sum:

$$u(a) = \int \rho(x) \cdot u(x) \, dx, \tag{2}$$

where $\rho(x)$ is the probability density and $u(x)$ is the utility of an alternative corresponding to the value x of the corresponding parameter (or parameters).

4 We Usually Know Probabilities With Some Uncertainty

Based on the finite sample of past experiences, we can only determine the corresponding distribution $\rho(x)$ with some uncertainty. In other words, instead of a single probability distribution with probability density $\rho(x)$, we have a whole class P of probability distributions which are consistent with our knowledge.

If two different distributions $\rho_1(x)$ and $\rho_2(x)$ are possible, this means that we can also have a case in which:

- we have the first distribution $\rho_1(x)$ with some probability α and
- we have the second distribution $\rho_2(x)$ with the remaining probability $1 - \alpha$.

This case corresponds to the probability density function $\alpha \cdot \rho_1(x) + (1 - \alpha) \cdot \rho_2(x)$, a *convex combination* of the two original distributions. Thus, with every two distributions, the class P contains their convex combinations – i.e., the class P is *convex*.

For each utility function $u(x)$, for different probability distributions from the class P, we have, in general, different values of the utility $u(a)$. In other words, instead a single value $u(a)$, we have a set of possible values. For a convex combination of two distributions $\rho_1, \rho_2 \in P$, the corresponding value $u(a)$ is also a convex combination of values u_1 and u_2 corresponding to the combined distributions. Thus, the set of possible value is convex – and is, therefore, an interval $[\underline{u}(a), \overline{u}(a)]$. (It should be mentioned that this interval can be infinite or semi-infinite.)

These intervals form a known alternative *prevision-based* representation of imprecise probabilities, when to each function $u(x)$, we assign the interval of possible values of $u(a)$:

$$[\underline{u}(a), \overline{u}(a)] = \left\{ \int \rho(x) \cdot u(x) \, dx : \rho \in P \right\}; \tag{3}$$

see, e.g., [2, 17].

5 How Can We Represent Imprecise Probabilities in Practice: Usual Approaches of Interval-Valued Moments And P-Boxes

Ideally, we should store intervals (3) corresponding to all possible utility functions $u(x)$. However, there are very many different functions, and it is not possible to store

all the corresponding intervals in a computer. We should therefore limit ourselves to some few-parametric family of functions $u(x)$.

Which functions should we choose? We should select functions $u(x)$ corresponding to reasonable utilities. One class of such functions comes from the fact that usually, small changes in a parameter x leads to small changes in the alternative – and thus, to small changes in the corresponding utility. In mathematical terms, it is reasonable to describe this idea by saying that $u(x)$ smoothly depends on x. In many such cases, we can expand this function in Taylor series and keep only a few first terms in this expansion:

$$u(x) \approx u_0 + u_1 \cdot x + \ldots + u_k \cdot x^k.$$

For such functions, the expected value is simply a linear combination of the moments:

$$\int \rho(x) \cdot u(x) \, dx \approx u_0 + a_1 \cdot M_1 + \ldots + a_k \cdot M_k,$$

where $M_i \overset{\text{def}}{=} \int \rho(x) \cdot x^i \, dx$. Thus, in such situations, to represent the probability distribution, it is sufficient to represent its moments $M_1, \ldots, M_k$.

For imprecise probabilities, we do not know the exact moments, we only know moments with some uncertainty, i.e., for each i, we only know the intervals $[\underline{M_i}, \overline{M_i}]$. Such interval-valued moments are thus a natural way to present the corresponding uncertainty. For the case when we have several parameters $x = (x_1, \ldots, x_m)$, we similarly need to store intervals of possible values of the joint moments

$$M_{i_1 \ldots i_m} \overset{\text{def}}{=} \int \rho(x) \cdot x_1^{i_1} \cdot \ldots \cdot x_m^{i_m} \, dx.$$

Another important case is when we have an abrupt jump in utility for some value x_0 of the parameter x. For example, for a chemical plant:

- if the concentration of a potentially dangerous chemical in the air does not exceed the required very small threshold x_0, we are OK, but
- if $x > x_0$, the plant will suffer from fines and possible closure – which drastically decreases the corresponding utility.

In such cases, the utility function is approximately equal to a jump: $u(x) = u_-$ for $x < x_0$ and $u(x) = u_+$ for $x \geq x_0$. One can check that the resulting expected utility is equal to $u_+ + (u_- - u_+) \cdot F(x)$, where $F(x)$, as usual, denotes the corresponding value of the cumulative distribution function (cdf) – i.e., the probability that the corresponding random variable does not exceed x. In such cases, to determine the value of the expected utility, we need to know the values of the cdf $F(x)$.

In case of uncertainty, we do not know the exact values of $F(x)$. Instead, for each x, we only know the bounds $[\underline{F}(x), \overline{F}(x)]$ on $F(x)$. Such interval-valued cdf is known as the *probability box*, or a *p-box*, for short; see, e.g., [4].

In case of several variables, a natural idea is to describe similar bounds on the joint pdf

$$F(x_{01}, \ldots, x_{0m}) \overset{\text{def}}{=} \text{Prob}(x_1 \leq x_{01} \& \ldots \& x_m \leq x_{0m});$$

see, e.g., [14].

6 Data Processing under Interval-Valued Moments and p-Boxes

In both approaches, the information about the actual (unknown) probability density function $\rho(x)$ comes in the form of bounds on the corresponding integrals:

$$\underline{v}_i \leq \int \rho(x) \cdot f_i(x)\, dx \leq \overline{x}_i, \tag{4}$$

where:

- for moments, $f_i(x) = x^i$ and
- for p-boxes, $f_i(x) = 1$ for $x < x_i$ and $f_i(x) = 0$ for all other x.

Based on this information, we need to estimate the value of the expected utility:

$$\int \rho(x) \cdot u(x)\, dx. \tag{5}$$

In other words, we need to find the smallest and the largest values of the objective function (5) under constraints (4).

In some cases, there are explicit formulas for the corresponding smallest and largest values; see, e.g., [4]. In general, we need to optimize the objective function (5) (which is linear in terms of the unknowns $\rho(x)$) under constraints (4) which are linear inequalities. Such problems of optimizing linear objective functions under linear inequality constraints are known as *linear programming* problems. There exist efficient algorithms for solving such problems; see, e.g., [3, 10, 16].

7 Need to Go Beyond p-Boxes and Interval-Valued Moments

p-boxes and interval-valued moments provide a reasonable first approximation to the actual description of imprecise probabilities. In some cases, however, this approximation is not accurate enough. Let us explain why this can happen, on a simple p-box example.

Indeed, suppose that we know that the probability distribution has the from $F(x) = F_0(x, c)$ for some family of functions $F_0(x, c)$ and some (unknown) value of the parameter c. As the simplest possible example, we can consider the family $F_0(x, c) = \widetilde{F}(x) + c$, where $\widetilde{F}(x)$ is a given function and the parameter c can take any value from some reasonably narrow interval $[-\varepsilon, \varepsilon]$.

In this case, for each x, the smallest possible value of the corresponding cdf is $\underline{F}(x) = \widetilde{F}(x) - \varepsilon$ and the largest possible value is $\overline{F}(x) = \widetilde{F}(x) + \varepsilon$. Thus, in the p-box representation, the above uncertainty will be represented by the p-box

$$[\underline{F}(x), \overline{F}(x)] = [\widetilde{F}(x) - \varepsilon, \widetilde{F}(x) + \varepsilon].$$

Let us now consider the following natural task: selecting a symmetric confidence interval corresponding to a given confidence level $1 - \alpha$. In terms of the cumulative density function, the requirement that a symmetric interval $[-x_0, x_0]$ is a confidence interval with given confidence level means that

$$\text{Prob}(x \in [-x_0, x_0]) = F(x_0) - F(-x_0) \geq 1 - \alpha.$$

The narrowest interval with this property is the interval for which

$$F(x_0) - F(-x_0) = 1 - \alpha.$$

In our case, we know that the cdf has the form $F(x) = \widetilde{F}(x) + c$. Thus,

$$F(x_0) - F(-x_0) = \widetilde{F}(x_0) - \widetilde{F}(-x_0),$$

and the desired requirement takes the form

$$\widetilde{F}(x_0) - \widetilde{F}(-x_0) = 1 - \alpha. \tag{6}$$

But what will happen if we only use the information from the p-box? In this case, the only information that we have about $F(x_0)$ is that

$$F(x_0) \in \left[\widetilde{F}(x_0) - \varepsilon, \widetilde{F}(x_0) + \varepsilon\right]$$

and the only information that we have about $F(-x_0)$ is that

$$F(-x_0) \in \left[\widetilde{F}(-x_0) - \varepsilon, \widetilde{F}(-x_0) + \varepsilon\right].$$

We have no information about the dependence between the values $F(x_0)$ and $F(-x_0)$; thus, based on the p-box information, the only conclusion that we can make about the difference $F(x_0) - F(-x_0)$ is that this difference is somewhere in the interval

$$\left[\left(\widetilde{F}(x_0) - \varepsilon\right) - \left(\widetilde{F}(-x_0) + \varepsilon\right), \left(\widetilde{F}(x_0) + \varepsilon\right) - \left(\widetilde{F}(-x_0) - \varepsilon\right)\right] =$$

$$\left[\left(\widetilde{F}(x_0) - \widetilde{F}(x_0)\right) - 2\varepsilon, \left(\widetilde{F}(x_0) - \widetilde{F}(x_0)\right) + 2\varepsilon\right];$$

see, e.g., [6, 11, 12].

The only way to guarantee that $F(x_0) - F(-x_0) \geq 1 - \alpha$ is to require that the smallest possible value of this difference is larger than or equal to $1 - \alpha$, i.e., that

$$\left(\widetilde{F}(x_0) - \widetilde{F}(x_0)\right) - 2\varepsilon \geq 1 - \alpha$$

or, equivalently, that

$$\widetilde{F}(x_0) - \widetilde{F}(-x_0) \geq 1 - (\alpha - 2\varepsilon).$$

The narrowest possible symmetric interval for which we can provide such a guarantee is the interval for which

$$\widetilde{F}(x_0) - \widetilde{F}(-x_0) = 1 - (\alpha - 2\varepsilon). \tag{7}$$

By comparing the formula (7) that comes from the p-box approximation with the actual formula (6), we see that because of this approximation, we have to take an interval that is much wider than necessary – namely, an interval corresponding to a much larger confidence level $1 - (\alpha - 2\varepsilon)$.

Such examples show that we need to go beyond p-boxes and interval-valued moments. What are the natural next approximation?

8 Case of Interval-Valued Moments: What Are The Natural Next Approximations?

In the moments method, we approximate the actual utility function by a polynomial $a_0 + a_1 \cdot x + \ldots + a_k \cdot x^k$. Thus, to get an adequate description of decision making under the corresponding imprecise probability, we need to know, for each of these polynomials, the range of the corresponding integrals.

Strictly speaking, there is no need to try all possible polynomials. First, The term a_0 just adds a constant to the expected utility, so it is sufficient to consider the polynomials $a_1 \cdot x + \ldots + a_k \cdot x^k$.

Second, if we know the integral $\int u(x) \cdot \rho(x) \, dx$ corresponding to a function $u(x)$, then, for every real number c, we can determine the integral

$$\int u_c(x) \cdot \rho(x) \, dx$$

corresponding to the function $u_c(x) = c \cdot u(x)$ as $c \cdot \int u(x) \cdot \rho(x) \, dx$. Thus, if we know the range $[\underline{u}, \overline{u}(a)]$ of possible values of the integral $\int u(x) \cdot \rho(x) \, dx$, then:

- for each number $c > 0$, we can compute the range $[\underline{u}_c, \overline{u}_c]$ for the function $u_c(x)$ as $[\underline{u}_c, \overline{u}_c] = [c \cdot \underline{u}, c \cdot \overline{u}]$, and
- for each number $c < 0$, we can compute the range $[\underline{u}_c, \overline{u}_c]$ for the function $u_c(x)$ as $[\underline{u}_c, \overline{u}_c] = [c \cdot \overline{u}, c \cdot \underline{u}]$;

see, e.g., [6, 11, 12].

By multiplying each polynomial by an appropriate constant, we can always get $a_1 = 1$. Thus, it is sufficient to consider polynomials of the type

$$x + a_2 \cdot x^2 + \ldots + a_k \cdot x^k.$$

So, a natural next approximation is to select some value k and to consider, for all possible values of the parameters $a = (a_2, \ldots, a_k)$, the bounds $\underline{M}(a_2, \ldots, a_k)$ and

$\overline{M}(a_2, \ldots, a_k)$ on the integral

$$\int \rho(x) \cdot \left(x + a_2 \cdot x^2 + \ldots + a_k \cdot x^k \right) dx.$$

It is worth mentioning that for $k = 2$, we get linear combinations of the first and second moments; a similar idea – known as the Elastic Net (EN) method – is very successful in data processing; see, e.g., [1, 7, 18].

From the computational viewpoint, we are still Ok, since the corresponding constraints

$$\underline{M}(a_2, \ldots, a_k) \leq \int \rho(x) \cdot \left(x + a_2 \cdot x^2 + \ldots + a_k \cdot x^k \right) dx \leq \overline{M}(a_2, \ldots, a_k)$$

are still linear in terms of the unknowns $\rho(x)$ and thus, we can still use efficient linear programming algorithms to solve the corresponding computational problems.

9 Case of p-Boxes: What Are The Natural Next Approximations?

The use of p-boxes means we use bounds on the integrals $\rho(x) \cdot u(x) \, dx$ of the step-functions $H(a - x)$ corresponding to different thresholds a, where $H(x)$ is the *Heaviside function* which is:

- equal to 0 for negative x and
- 1 for non-negative x.

A natural idea is thus to consider bounds on the above integral for linear combinations of such step-functions, i.e., for functions of the type

$$u(x) = c_1 \cdot H(a_1 - x) + c_2 \cdot H(a_2 - x) + \ldots + c_k \cdot H(a_k - x).$$

Without losing generality, we can assume that $a_1 < a_2 < \ldots < a_k$. Such functions are piece-wise constant:

- for $x \leq a_1$, we have $u(x) = c_1 + c_2 + \ldots + c_k$;
- for $a_1 < x \leq a_2$, we have $u(x) = c_2 + \ldots + c_k$;
- …
- for $a_{k-1} < x \leq a_k$, we have $u(x) = c_k$; and
- for $x > a_k$, we have $u(x) = 0$.

Vice versa, every piece-wise constant function can be represented in this form.

In particular, for $k = 2$, $c_1 = -1$, and $c_2 = 1$, we get the characteristic function of the interval $[a_1, a_2]$ – thus, by considering such functions, we can find the exact confidence interval and hence, solve the problem with the p-boxes that we mentioned in the previous sections.

Every function can be thus approximated. The larger k we take, the more accurate the resulting description – and we can get any desired accuracy by selecting an appropriate values k.

Similarly to the previous section, we can multiply each such function by $1/c_1$ and thus, get the case when $c_1 = 1$. Thus, we need to consider bounds $\underline{F}(a_1, c_2, a_2, \ldots, c_k, a_k)$ and $\overline{F}(a_1, c_2, a_2, \ldots, c_k, a_k)$ on the integrals

$$\int \rho(x) \cdot (H(a_1 - x) + c_2 \cdot H(a_2 - x) + \ldots + c_k \cdot H(a_k - x)) \, dx =$$

$$(c_1 + c_2 + \ldots + c_k) \cdot \mathrm{Prob}(x \leq a_1) + (c_2 + \ldots + c_k) \cdot \mathrm{Prob}(a_1 < x \leq a_2) + \ldots +$$

$$c_k \cdot \mathrm{Prob}(a_{k-1} < x \leq a_k).$$

The resulting constraints

$$\underline{F}(a_1, c_2, a_2, \ldots, c_k, a_k) \leq$$

$$\int \rho(x) \cdot (H(a_1 - x) + c_2 \cdot H(a_2 - x) + \ldots + c_k \cdot H(a_k - x)) \, dx \leq$$

$$\overline{F}(a_1, c_2, a_2, \ldots, c_k, a_k)$$

are still linear in terms of the unknowns $\rho(x)$. So, when processing this information, we can still use efficient linear programming algorithms to solve the corresponding computational problems.

Acknowledgments

References

1. H. Alkhatib, I. Neumann, V. Kreinovich, and Chon Van Le, "Why LASSO, EN, and CLOT: invariance-based explanation", In: Nguyen Duc Trung, Nguyen Ngoc Thach, and V. Kreinovich, *Data Science for Financial Econometrics*, Springer, Cham, Switzerland, 2020, to appear.
2. T. Augustin, F. Coolen, G. de Cooman, and M. Troffaes (eds.), *Introduction to Imprecise Probabilities*, Wiley, New York, 2014.
3. Th. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, Massachusetts, 2009.
4. S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz, *Constructing Probability Boxes and Dempster-Shafer Structures*, Sandia National Laboratories, Report SAND2002-4015, January 2003.
5. P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.
6. L. Jaulin, M. Kiefer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics*, Springer, London, 2001.

7. B. Kargoll, M. Omidalizarandi, I. Loth, J.-A. Paffenholz, and H. Alkhatib, "An iteratively reweighted least-squares approach to adaptive robust adjustment of parameters in linear regression models with autoregressive and t-distributed deviations", *Journal of Geodesy*, 2018, Vol. 92, No. 3, pp. 271–297.

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*, Springer Verlag, 2014, pp. 163–193.

9. R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

10. D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, Springer, Cham, Switzerland, 2016.

11. G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017.

12. R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.

13. 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*, 2009, Vol. 24, No. 1, pp. 27–47.

14. R. Pelessoni, P. Vicig, I. Montes, and E. Miranda, *Bivariate p-boxes*, International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems, 2016, Vol. 24, No. 2, pp. 229–263.

15. H. Raiffa, *Decision Analysis*, McGraw-Hill, Columbus, Ohio, 1997.

16. R. J. Vanderbei, *Linear Programming: Foundations and Extensions*, Springer, New York, 2014.

17. P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London, UK, 1991.

18. H. Zou and T. Hastie, "Regularization and variable selection via the elastic net", *Journal of the Royal Statistical Society B*, 2005, Vol. 67, pp. 301–320.