Towards Optimal Effort Distribution in Process Design under Uncertainty, with Application to Education

Olga Kosheleva[1] and Vladik Kreinovich[2]*

Departments of [1]Teacher Education and [2]Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
Contact e-mail: vladik@utep.edu
*Corresponding author

Abstract: In most application areas, we need to take care of several (reasonably independent) participants. For example, in controlling economics, we must make sure that all the economic regions prosper. In controlling environment, we want to guarantee that all the geographic regions have healthy environment. In education, we want to make sure that all the students learn all the needed knowledge and skills.

In real life, the amount of resources is limited, so we face the problem of optimally distributing these resources between different objects.

What is a reasonable way to formalize optimally? For each of the participants, preferences can be described by *utility functions*: namely, an action is better if its expected utility is larger. It is natural to require that the resulting group preference has the following property: if two actions have the same quality for all participants, then they are equivalent for the group as well. It turns out that under this requirement, the utility function of a group is a linear combination of individual utility functions.

To solve the resulting optimization problem, we need to know, for each participant i, the utility resulting from investing effort e in this participant. In practice, we only know this value with (interval) uncertainty. So, for each distribution of efforts, instead of a single value of the group utility, we only have an interval of possible values. To compare such intervals, we use Hurwicz optimism-pessimism criterion well justified in decision making.

In this paper, we propose a solution to the resulting optimization problems.

Keywords: Interval uncertainty; Distribution of effort; Optimization under uncertainty; Hurwicz optimism-pessimism criterion

Reference to this paper should be made as follows: Kosheleva, O., and Kreinovich, V. (xxxx) 'Towards Optimal Effort Distribution in Process Design under Uncertainty, with Application to Education', *Int. J. Reliability and Safety*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Olga Kosheleva is an Assistant Professor of Mathematics Education. She received her Ph.D. in Electrical and Computer Engineering from the University of Texas at El Paso in 2003. Her main research areas are Mathematics Education in Teacher Preparation and Curriculum Development with focus on Science, Technology, Engineering and Mathematics (STEM) integration and applied mathematics.

Vladik Kreinovich received his PhD in Mathematics and Physics in 1979, from the Institute of Mathematics of the Siberian Branch of the Academy of Sciences, Novosibirsk, Russia. He is a Professor of Computer Science at the University of Texas at El Paso. His main research interests include processing data under uncertainty, especially interval uncertainty. He has published 3 books, 6 edited books, and more than 700 papers.

1 Formulation of the Problem

Need for effort distribution. In most application areas, we need to take care of several (reasonably independent) participants.

- For example, when we control economics, we must make sure that all the economic regions prosper.
- When we control the environment, we want to guarantee that all the geographic regions have healthy environment.
- In education, we want to make sure that all the students learn all the needed knowledge and skills.
- In engineering, when we run a business that services numerous customers – e.g., that keeps their copy machines, or their computers, or their cars in good working conditions – we need to make sure that all the customers are satisfied.

In real life, the amount of resources is limited, so we face the problem of optimally distributing these resources between different objects.

Example: education. In a typical class, we have students at different levels of knowledge, student with different ability to learn the material. In the ideal world, we are able to devote unlimited individual attention to all the students and make sure that everyone learns all the material. In real life, our resources are finite. Based on this finite amount of resources, what is the best way to distribute efforts between different students?

How do we compare different effort distributions? What is a reasonable way to formalize optimally? According to decision theory, for each of the objects, preferences can be described by *utility functions* (see, e.g., Luce and Raiffa (1989)): namely, an action is better if the expected value of its utility is larger.

It is therefore necessary to combine these utility functions into a single objective function that characterizes the distribution of efforts. Such combination techniques will be discussed in this paper.

Uncertainty. In the ideal setting, we know, for each participant i, the utility resulting from investing effort e in this participant. In practice, we only know this value with (interval) uncertainty. So, for each distribution of efforts, instead of a single value of the group utility, we only have an interval of possible values.

We therefore need to be able to distribute efforts under such uncertainty. Methods of taking uncertainty into account when distributing effort will also be discussed in this paper.

Example: education. In education, we usually do not have exact information about the cognitive ability of each student, there is a large amount of uncertainty. We hope that our algorithms will be useful in comparing and designing teaching strategies.

2 Traditional Approach to Solving the Resource Distribution Problem and Its Limitations

How this problem is usually solved now: a brief description, on the example of the education problem. The success of each individual student i can be naturally gauged by this student's grade x_i. So, for two different effort distribution strategies T and T', we know the corresponding grades $x_1, \ldots, x_n$ and $x'_1, \ldots, x'_{n'}$. Which strategy is better?

In some cases, the answer to this question is straightforward. For example, when $n' = n$, $x_i \le x'_i$ for all i and $x_i < x'_i$ for some i, then clearly the strategy T' is better.

In practice, however, the comparison is rarely that straightforward. Often, when we change a strategy, some grades decrease while some other grades increase. In this case, how do we usually decide whether a new method is better or not?

In pedagogical research, the decision is usually made based on the comparison of the average grades

$$E \stackrel{\text{def}}{=} \frac{x_1 + \ldots + x_n}{n} \tag{1}$$

and

$$E' \stackrel{\text{def}}{=} \frac{x'_1 + \ldots + x'_{n'}}{n'}. \tag{2}$$

For example, we can use the t-test (see, e.g., Sheskin (2007)) and conclude that the method T' is better if the corresponding t-statistic

$$t \stackrel{\text{def}}{=} \frac{E' - E}{\sqrt{\dfrac{V}{n} + \dfrac{V'}{n'}}}, \tag{3}$$

where

$$V \stackrel{\text{def}}{=} \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - E)^2, \quad V' \stackrel{\text{def}}{=} \frac{1}{n'-1} \cdot \sum_{i=1}^{n'} (x'_i - E')^2, \tag{4}$$

exceeds the appropriate threshold t_α (depending on the level of confidence α with which we want to make this conclusion).

How this problem is usually solved now: limitations. The average grade is not always the most adequate way to gauging the success of a pedagogical strategy. Whether the average grade is a good criterion or not depends on our objective.

Let us illustrate this dependence on a simplified example. Suppose that after using the original teaching method T, we get the grades $x_1 = 60$ and $x_2 = 90$. The average value of these grades is

$$E = \frac{60 + 90}{2} = 75. \tag{5}$$

Suppose that the new teaching method T' leads to the grades $x_1' = x_2' = 70$. The average of the new grades is $E' = 70$.

Since the average grade decreases, the traditional conclusion would be that the new teaching method T' is not as efficient as the original method T. However, one possible objective may be to decrease the failing rate. Usually, 70 is the lowest numerical grade corresponding to a letter grade C, and any letter grade below C is considered failing. In this case,

- in the original teaching method, one of the two students failed, while
- in the new teaching method, both students passed the class.

Thus, with respect to this objective, the new teaching method is better.

3 Towards a More Adequate Solution to the Resource Distribution Problem: How to Combine Utilities of Different Participants

Need to combine utility values: reminder. We can describe the consequence of each possible action a for each participant i by his or her utility value $u_i(a)$. Thus, for n participants, each decision a is characterized by n utility values $u_1(a)$, $\ldots$, $u_n(a)$.

In order to compare different decisions a, we must combine these utility values $u_i(a)$ into a single value of an objective function

$$u(a) = f(u_1(a), \ldots, u_n(a)).$$

The meaning of this objective function is that the larger its value, the better the alternative a for the group as a whole. Thus, we should select the alternative for which this combined value $u(a)$ is the largest.

Need to select a combination function. There are many different ways to combine the utility values. In mathematical terms, this means that there are many different functions $f(u_1, \ldots, u_n)$ that can be used for such a combination. Which of these functions should we choose?

How to select a combination function: a reasonable requirement. Usually, utility theory describes preferences of individuals. However, in practice, utility theory (and related game theory) also describe interaction between groups. In such a description, we describe the preferences of each group participant (corporation, city, country, etc.) by using the same utility theory approach: that the preferences of a group can be described by a utility function in such a way that the preference of an action is determined by the corresponding expected utilities.

It is therefore desirable to select a combination function for which the combined function has the properties of the utility. In particular, it is natural to require that the resulting group preference has the following property:

- if two actions have the same quality for all participants,
- then these two actions are equivalent for the group as well.

Towards formulating the above requirement in precise mathematical terms. Actions can lead to different possible outcomes ω. We assume that for each possible outcome ω and for each participant i, we know the utility $u_i(\omega)$ that describes this participant's opinion of this outcome.

The traditional decision theory describes situations in which we know the probability of different possible outcomes. In other words, we assume that for every possible outcome ω and for every action a, we know the probability $p(\omega \mid a)$ of this outcome.

According to utility theory, the utility of an action a for a participant i is determined by the expected value of this participant's utility, i.e., by the value

$$u_i(a) = E[u_i \mid a] = \sum_{\omega} p(\omega \mid a) \cdot u_i(\omega).$$

For example, to the participant i, the action a is equivalent to the outcome ω_0 in which $u_i(\omega_0) = E[u_i]$.

For each outcome ω, the utility of the i-th participant is $u_i(\omega)$ and thus, the group utility is equal to

$$u(\omega) = f(u_1(\omega), \ldots, u_n(\omega)).$$

Thus, the utility u of the action a for the group should be equal to the expected value of the group utility:

$$u(a) = E[u \mid a] = \sum_{\omega} p(\omega \mid a) \cdot u(\omega),$$

i.e., to

$$u(a) = E[f(u_1, \ldots, u_n) \mid a] = \sum_{\omega} p(\omega \mid a) \cdot f(u_1(\omega), \ldots, u_n(\omega)).$$

On the other hand, for each participant i, the action a is simply equivalent to an alternative with utility $E[u_i]$. So, a situation in which each participant gets a reward of utility $E[u_i]$ (with probability 1) is, for each participant, equivalent to the action a. Our requirement states that this situation should be equivalent to the action a for the group as well.

For this equivalent situation, its group utility u_s can be found by applying the same combination function $f(u_1,\ldots,u_n)$ to the deterministic utility values $u_i = E[u_i]$:

$$u_s(a) = f(E[u_1],\ldots,E[u_n]).$$

Since the action a and the new situation are equivalent for the group, their group utilities must coincide. So, we must have

$$E[f(u_1,\ldots,u_n)] = f(E[u_1],\ldots,E[u_n]) \tag{6}$$

for every n random variables u_1, $\ldots$, u_n.

Resulting mathematical formulation. We are looking for functions $f(u_1,\ldots,u_n)$ for which, for all possible n random variables u_i, we have

$$E[f(u_1,\ldots,u_n)] = f(E[u_1],\ldots,E[u_n]).$$

Our result. We will show that the only functions which satisfy the above property are the linear functions

$$f(u_1,\ldots,u_n) = w_0 + \sum_{i=1}^{n} w_i \cdot u_i$$

for appropriate weights w_0, w_1, $\ldots$, w_n.

(For reader's convenience, the proofs of all the results are given in the Appendix.)

Comments. It is clear that linear functions satisfy the above property. What is less trivial – and what we prove – is that linear functions are the only ones that satisfy this property, i.e., that no matter what non-linear function we take, it will, in some cases, violate the above property.

This result is similar to the known results about the case when the participants are, in some reasonable sense, independent. This case has been actively analyzed in decision theory. In particular, it has been proven that the corresponding objective function can be represented as the sum of "marginal" objective functions representing different participants, i.e.,

$$f(x_1,\ldots,x_n) = f_1(x_1) + \ldots + f_n(x_n); \tag{7}$$

see, e.g., Fishburn (1969, 1988).

4 Towards the Optimal Effort Distribution: Constraint Optimization Problem

Formulation of the problem. Let $e_i(x_i)$ denote the amount of effort (time, etc.) that is needed for the i-th participant to achieve the value x_i of the corresponding quantity (grade, income, etc.). It is reasonable to assume that the better effect we want to achieve, the more effort we need, so each function $e_i(x_i)$ is strictly increasing.

Let $u_i(x_i)$ denote the utility of the i-th participant on achieving the value x_i, and let w_i denote the corresponding weight.

Let e denote the available amount of effort. In these terms, the problem of selecting the optimal teaching strategy means that we maximize the objective function under the constraint that the overall effort cannot exceed e:

$$\text{Maximize } f(x_1, \ldots, x_n) = w_0 + w_1 \cdot u_1(x_1) + \ldots w_n \cdot u_n(x_n) \tag{8}$$

under the constraint

$$e_1(x_1) + \ldots + e_n(x_n) \le e. \tag{9}$$

Comment. The objective function (8) depends only on the combinations $f_i(x) \stackrel{\text{def}}{=} w_i \cdot u_i(x)$ and not on the individual values of w_i and $u_i(x)$. Thus, to simplify the corresponding optimization problem, it makes sense to reformulate it in terms of the new "benefit" functions $f_i(x)$.

Similarly, the constant value w_0 is irrelevant for optimization, so we can safely ignore it. Thus, we arrive at the following simplified formulation:

$$\text{Maximize } f(x_1, \ldots, x_n) = f_1(x_1) + \ldots + f_n(x_n) \tag{10}$$

under the constraint (9).

Towards solution. First we note that, due to monotonicity, if the total effort is smaller than e, then we can spend more effort and get the better value of the objective function (10). In other words, the maximum is attained when all the effort is actually used, i.e., when we have the constraint

$$e_1(x_1) + \ldots + e_n(x_n) = e. \tag{11}$$

To maximize the objective function (10) under this constraint, we can use the Lagrange multiplier method. According to this method, the maximum of the function (10) under constraint (11) is attained when for some value λ, the auxiliary function

$$f_1(x_1) + \ldots + f_n(x_n) + \lambda \cdot (e_1(x_1) + \ldots + e_n(x_n)) \tag{12}$$

attains its (unconstrained) maximum. Differentiating this auxiliary function with respect to x_i and equating the derivatives to 0, we conclude that

$$f_i'(x_i) + \lambda \cdot e_i'(x_i) = 0, \tag{13}$$

where f_i' and e_i' denote the derivatives of the corresponding functions. From this formula, we can explicitly describe λ as

$$-\frac{f_i'(x_i)}{e_i'(x_i)} = \lambda. \tag{14}$$

So, once we know λ, we can find all the corresponding grades x_i – and the resulting efforts – by solving, for each i, a (non-linear) equation (14) with a single variable x_i.

The value λ can be found from the formula (11), i.e., from the condition that for the resulting values x_i, we get $\sum\limits_{i=1}^{n} e_i(x_i) = e$.

5 Need to Take Uncertainty Into Account

Assumptions: reminder. In the above text, we assumed that:

- we know *exactly* the benefits $f(x_1, \ldots, x_n)$ of participants achieving the levels $x_1, \ldots, x_n$; for example, we know the exact expressions for the benefit functions $f_i(x_i)$;
- we know *exactly* how much effort $e_i(x_i)$ is needed to bring each participant i to a given level x_i, and
- we know *exactly* the level of x_i of each participant – for example, in education, the level of knowledge x_i is exactly determined by the grade.

In practice, we have *uncertainty.*

Average benefit function. First, we rarely know the exact benefit function $f_i(x_i)$ characterizing each individual participant. At best, we know the *average* function $a(x)$ describing the average benefits of the level x to a participant.

Average effort function. Second, we rarely know the exact effort function $e_i(x_i)$ characterizing each individual participant. At best, we know the *average* function $e(x)$ describing the average effort needed to bring a participant to the level x.

Interval uncertainty. Finally, the known value $\widetilde{x}_i$ is only an approximate indication of the participant's level x_i. Once we know the estimate $\widetilde{x}_i$, we cannot conclude that the level x_i is exactly $\widetilde{x}_i$. At best, we know the accuracy ε_i of this estimate. In this case, the actual (unknown) level x_i can take any value from the interval $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i] \stackrel{\text{def}}{=} [\widetilde{x}_i - \varepsilon_i, \widetilde{x}_i + \varepsilon_i]$.

Under interval uncertainty, instead of a single value of the objective function $f(x_1, \ldots, x_n)$, we get an *interval* of possible values

$$\left[\underline{f}, \overline{f}\right] = f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{def}}{=} \{f(x_1, \ldots, x_n) \,|\, x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}. \tag{15}$$

Fuzzy uncertainty. In many practical situations, the estimates $\widetilde{x}_i$ come from experts. Experts often describe the inaccuracy of their estimates in terms of imprecise words from natural language, such as "approximately 0.1", etc. A natural way to formalize such words is to use special techniques developed for formalizing this type of estimates – specifically, the technique of fuzzy logic; see, e.g., Klir and Yuan (1995); Nguyen and Walker (2005).

In this technique, for each possible value of $x_i \in [\underline{x}_i, \overline{x}_i]$, we describe the degree $\mu_i(x_i)$ to which this value is possible. For each degree of certainty α, we can determine the set of values of x_i that are possible with at least this degree of certainty – the *α-cut* $\mathbf{x}_i(\alpha) = \{x \mid \mu_i(x) \ge \alpha\}$ of the original fuzzy set. Vice versa, if we know α-cuts for every α, then, for each object x, we can determine the degree of possibility that x belongs to the original fuzzy set; see, e.g., Dubois and Prade (1978); Klir and Yuan (1995); Moore and Lodwick (2003); Nguyen and Kreinovich (1996); Nguyen and Walker (2005). A fuzzy set can be thus viewed as a nested family of its (interval) α-cuts.

From the computational viewpoint, fuzzy uncertainty can be reduced to the interval one. Once we know how to propagate interval uncertainty, then, to propagate the fuzzy uncertainty, we can consider, for each α, the fuzzy set y with the α-cuts

$$\mathbf{y}(\alpha) = f(\mathbf{x}_1(\alpha), \ldots, \mathbf{x}_1(\alpha)); \tag{16}$$

see, e.g., Dubois and Prade (1978); Klir and Yuan (1995); Moore and Lodwick (2003); Nguyen and Kreinovich (1996); Nguyen and Walker (2005). So, from the computational viewpoint, the problem of propagating fuzzy uncertainty can be reduced to several interval propagation problems.

Because of this reduction, in the following text, we will mainly concentrate on algorithms for the interval case.

6 How to Take Uncertainty Into Account

Let us analyze how we can take into account these different types of uncertainties.

Average utility function: general situation. Let us first consider the case when instead of the *individual* benefit functions $f_1(x_1), \ldots, f_n(x_n)$, we only know the *average* function $a(x)$. In this case, for a combination of levels $x_1, \ldots, x_n$, the resulting value of the objective function is

$$f(x_1, \ldots, x_n) = a(x_1) + \ldots + a(x_n). \tag{17}$$

Smooth utility functions. Usually, the utility function is reasonably smooth. In this case, if (hopefully) all the levels are close, we can expand the function $a(x)$ in Taylor series around the average level, and keep only quadratic terms in this expansion. The general form of this quadratic approximation is

$$a(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2, \tag{18}$$

for some coefficients a_0, a_1, and a_2. For this function, the expression (17) for the objective function takes the form

$$f(x_1, \ldots, x_n) = n \cdot a_0 + a_1 \cdot \sum_{i=1}^{n} x_i + a_2 \cdot \sum_{i=1}^{n} x_i^2, \tag{19}$$

i.e., the form

$$f(x_1, \ldots, x_n) = b_0 + b_1 \cdot E + b_2 \cdot M, \tag{20}$$

where $b_0 \stackrel{\text{def}}{=} n \cdot a_0$, $b_1 \stackrel{\text{def}}{=} n \cdot a_1$, $b_2 \stackrel{\text{def}}{=} n \cdot a_2$, E is the average (1), and M is the second sample moment:

$$M \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2. \tag{21}$$

Thus, for smooth benefit functions $a(x)$, to estimate the utility corresponding to a given combination of levels $x_1, \ldots, x_n$, it is not necessary to know all these n levels, it is sufficient to know the average level and the mean squared level (or, equivalently, the standard deviation of the levels).

Comment. In general, the benefit function $a(x)$ is increasing with x_i. However, it is worth mentioning that the above conclusion holds for every quadratic function $a(x)$, not necessarily a function which is increasing for all the values $x_1, \ldots, x_n$.

Case of interval uncertainty. Until now, we assumed that we know the exact values $x_1, \ldots, x_n$ of the participants' levels. What will happen if instead, we only know intervals $[\underline{x}_i, \overline{x}_i]$ of possible values of x_i?

Since the benefit function $a(x)$ is increasing (the higher levels the better),

- the largest possible value $\overline{f}$ of the objective function $f(x_1, \ldots, x_n)$ is attained when the values x_i are the largest possible $x_i = \overline{x}_i$, and
- the smallest possible value $\underline{f}$ of the objective function is attained when the values x_i are the smallest possible $x_i = \underline{x}_i$.

In other words, we get the following interval $[\underline{f}, \overline{f}]$ of possible values $f(x_1, \ldots, x_n)$ of the objective function:

$$[\underline{f}, \overline{f}] = \left[\sum_{i=1}^{n} a(\underline{x}_i), \sum_{i=1}^{n} a(\overline{x}_i) \right]. \tag{22}$$

Comment. We mentioned that for the case of smooth (quadratic) utility function and exactly known x_i, we do not need to keep all n levels – it is sufficient to keep only the first and second sample moments of these grades. A natural question is: in the case of interval uncertainty, do we need to keep n intervals, or can we use a few numbers instead? In the Appendix, we show that under interval uncertainty, in the general case, all n values are needed.

How to solve the corresponding optimization problem. To solve the resulting optimization problem, we need to know, for each participant i, the utility resulting from investing effort e in this participant. In practice, we only know this value with (interval) uncertainty. So, for each distribution of efforts, instead of a single value of the group utility, we only have an interval of possible values.

To compare such intervals, we can use Hurwicz optimism-pessimism criterion well justified in decision making Hurwicz (1951); Luce and Raiffa (1989): namely, we select a value $\alpha_{\text{opt}} \in [0, 1]$ describing our degree of optimism, and use the value $\alpha_{\text{opt}} \cdot \overline{f} + (1 - \alpha_{\text{opt}}) \cdot \underline{f}$ as the objective function.

7 Beyond Utility-Motivated Linear Combination: On the Example of Teaching

Alternative combination rules are sometimes used. In the above text, we considered utility-motivated linear combinations of utility functions. In practice, other combination rules $f(x_1, \ldots, x_n)$ are also used. Let us give education-related examples of such rules.

Minimizing failure rate. The objective of minimizing the failure rate means that we minimize the number of students whose grade is below the passing threshold x_0:

$$f(x_1, \ldots, x_n) = \#\{i : x_i < x_0\}. \tag{23}$$

Comment. Since the general objective is to *maximize* the value of the objective function $f(x_1, \ldots, x_n)$, we can reformulate the criterion (23) as a maximization one: namely, minimizing (23) is equivalent to maximize the number of students whose grade is above (or equal to) the passing threshold x_0:

$$f(x_1, \ldots, x_n) = \#\{i : x_i \geq x_0\}. \tag{24}$$

No Child Left Behind. Other criteria are also possible. For example, the idea that no child should be left behind means, in effect, that we gauge the quality of a school by the performance of the worst student – i.e., of the student with the lowest grade $\min(x_1, \ldots, x_n)$. Thus, the corresponding objective is to maximize this lowest grade:

$$f(x_1, \ldots, x_n) = \min(x_1, \ldots, x_n). \tag{25}$$

Explicit solution to the optimization problem: "No Child Left Behind" case. In the No Child Left Behind case, we maximize the lowest grade. For this objective function, there is also an explicit solution. Since our objective is to maximize the lowest grade, there is no sense to use the effort to get one of the student grades better than the lowest grade – because the lowest grade will not change. From the viewpoint of the objective function, it is more beneficial to use the same efforts to increase the grades of all the students at the same time – this will increase the lowest grade.

In this case, the common grade x_c that we can achieve can be determined from the condition (11), i.e., from the equation

$$e_1(x_c) + \ldots + e_n(x_c) = e. \tag{26}$$

Comment. A slightly more complex situation occurs when we take into account that even before the class, some students already have some knowledge about the class material. Let us denote the starting grades by $x_i^{(0)}$. Without losing generality, let us assume that the students are sorted in the increasing order of their grades, i.e., that $x_1^{(0)} \leq \ldots \leq x_n^{(0)}$. In this case, the optimal effort distribution aimed at maximizing the lowest grade is as follows:

- first, all the efforts must go into increasing the original grade $x_1^{(0)}$ of the worst student to the next level $x_2^{(0)}$;
- if this attempt to increase consumes all available effort, then this is what we got;
- otherwise, if some effort is left, we raise the grades of the two lowest-graded students x_1 and x_2 to the yet next level $x_3^{(0)}$, etc.

In precise terms, the resulting optimal distribution of efforts can be described as follows. First, we find the largest value k for which all the grades $x_1, \ldots, x_k$ can be raised to the k-th original level $x_k^{(0)}$. In precise terms, this means the largest value k for which

$$e_1(x_k^{(0}) + \ldots + e_k(x_k^{(0)}) \leq e. \tag{27}$$

This means that for the criterion $\min(x_1, \ldots, x_n)$, we can achieve the value $x_k^{(0)}$, but we cannot achieve the value $x_{k+1}^{(0)}$.

Then, we find the value $x \in [x_k^{(0)}, x_{k+1}^{(0)})$ for which

$$e_1(x) + \ldots + e_{k-1}(x) + e_k(x) = e. \tag{28}$$

This value x is the optimal value of the criterion $\min(x_1, \ldots, x_n)$.

Maximizing success rate. The quality of a high school is often gauged by the number of alumni who get into prestigious schools. In terms of the grades x_i, this means, crudely speaking, that we maximize the number of students whose grade exceeds the minimal entrance grade e_0 for prestigious schools:

$$f(x_1, \ldots, x_n) = \#\{i : x_i \geq e_0\}. \tag{29}$$

From the mathematical viewpoint, this criterion is equivalent to minimizing the number of students whose grade is below e_0 – and is, thus, equivalent to criterion (23), with $x_0 = e_0$,

Best school to get in. There is a version of the above criterion which is not equivalent to (23), when the quality of a high school is gauged by the success of the best alumnus: e.g., "one of our alumni got into Harvard". In terms of the grades x_i, this means, crudely speaking, that we maximize the highest of the grades $\max(x_1, \ldots, x_n)$, i.e., that we take

$$f(x_1, \ldots, x_n) = \max(x_1, \ldots, x_n). \tag{30}$$

Explicit solution to the optimization problem: "Best School to Get In" case. If the criterion is the Best School to Get In, i.e., in terms of grades, the largest possible grade x_i, then the optimal use of effort is, of course, to concentrate on a single individual and ignore the rest. Which individual to target depends on how much gain we will get. In other words,

- first, for each i, we find x_i for which $e_i(x_i) = e$, and then
- we choose the student with the largest value of x_i as a recipient of all the efforts.

Criteria combining mean and variance. Another possible approach comes from the fact that the traditional criterion – that only takes into account the average (mean) grade E – is not always adequate. The reason for inadequacy is that the mean does not provide us any information about the "spread" of the grades, i.e., the information about how much the grades deviate from the mean. This information is provided by the standard deviation σ, or, equivalently, the sample variance $V = \sigma^2$. Thus, we arrive at criteria of the type $f(E, V)$.

When the mean is fixed, usually, we aim for the smallest possible variation – unless we gauge a school by its best students. Similarly, when the variance is fixed, we aim for the largest possible mean.

Thus, it is reasonable to require that the objective function $f(E, V)$ is an increasing function of E and a decreasing function of V.

Estimating $f(E, V)$ under interval uncertainty. Let us consider the case when the objective function has the form $f(E, V)$, where $f(E, V)$ increases as a function of E and decreases as a function of V. How can we estimate the range $[\underline{f}, \overline{f}]$ of the values of this objective function under interval uncertainty $x_i \in [\underline{x}_i, \overline{x}_i]$?

In general, this range estimation problem is NP-hard already for the case $f(E, V) = -V$; see, e.g., Kreinovich et al. (2006). This means, crudely speaking, that unless P=NP (and most computer scientists believe that P≠NP), no efficient (polynomial time) algorithm can always compute the exact range.

The maximum of the expression $f(E, V)$ can be found efficiently. For that, it is sufficient to consider all $2n + 2$ intervals $[\underline{r}, \overline{r}]$ into which the values $\underline{x}_i$ and $\overline{x}_i$ divide the real line, and for each of these intervals, and for each $r \in [\underline{r}, \overline{r}]$, take the values

- $x_i = \overline{x}_i$ when $\overline{x}_i \le \underline{r}$;
- $x_i = r$ when $[\underline{r}, \overline{r}] \subseteq [\underline{x}_i, \overline{x}_i]$; and
- $x_i = \underline{x}_i$ when $\overline{r} \le \underline{x}_i$.

(The proof is similar to the ones given in Kreinovich et al. (2006).)

For the minimum of $f(E, V)$, for reasonable cases, efficient algorithms are also possible. One such case is when none of the intervals $[\underline{x}_i, \overline{x}_i]$ is a proper subset of another one, i.e., to be more precise, when $\underline{x}_i, \overline{x}_i \not\subseteq (\underline{x}_j, \overline{x}_j)$.

In this case, a proof similar to the one from Kreinovich et al. (2006) shows that if we sort the intervals in lexicographic order

$$[\underline{x}_1, \overline{x}_1] \le [\underline{x}_2, \overline{x}_2] \le \ldots \le [\underline{x}_n, \overline{x}_n], \tag{31}$$

where

$$[\underline{a}, \overline{b}] \le [\underline{b}, \overline{b}] \leftrightarrow \underline{a} < \underline{b} \vee (\underline{a} = \underline{b} \,\&\, \overline{a} \le \overline{b}), \tag{32}$$

then the minimum of f is attained at one of the combinations

$$(\underline{x}_1, \ldots, \underline{x}_{k-1}, x_k, \overline{x}_{k+1}, \ldots, \overline{x}_n) \tag{33}$$

for some $x_k \in [\underline{x}_k, \overline{x}_k]$. Thus, to find the minimum, it is sufficient to sort the values, and then find the smallest possible value of $f(E, V)$ for each of $n + 1$ such combinations.

8 Conclusion

In many application problems, we need to optimally distribute limited efforts between several participants. In education, we need to make sure that all the students acquired the desired knowledge and skills. In economic planning, we need to make sure that all the regions prosper. In engineering, we need to optimally distribute servicing efforts between numerous customers. To formulate and solve the corresponding optimization problem, we need to select an expression for the objective function in terms of utilities of different participants.

We start by formulating a reasonable requirement: that if two actions have the same quality for all the participants, they should have the same quality for the group as a whole. It is easy to see that this requirement is satisfied when, as the objective function, we use a linear combination of the participants' utilities. We prove that such linear combinations are the only objective functions that satisfy the above requirement – i.e., that for every non-linear combination function, this requirement is sometimes violated. We then show how, for such linear objective functions, we can efficiently solve the problem of optimal effort distribution.

In practice, the utility is often known with uncertainty. We show how we can take this uncertainty into account when trying to optimally distribute efforts.

In addition to the (ideal) linear objective functions, we also consider objective functions which are currently used in optimizing effort distribution – e.g., criteria like minimum of the utilities which corresponds to the "No child left behind" approach to education. For these objective functions, we also describe efficient optimization algorithms.

Acknowledgments

This work was supported in part by the National Science Foundation grant Cyber-ShARE Center of Excellence: A Center for Sharing Cyberresources to Advance Science and Education (HRD-0734825) and by Grant 1 T36 GM078000-01 from the National Institutes of Health.

The authors are thankful to all the participants of the 4th International Workshop on Reliable Engineering Computing REC'2010 (Singapore, March 3–5, 2010) for valuable suggestions, and to the anonymous referees for useful recommendations.

References

Dubois, D., Prade, H. (1978) Operations on fuzzy numbers, *International Journal of Systems Science*, vol. 9, pp.613–626.

Fishburn, P.C. (1969) *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York.

Fishburn, P.C. (1988) *Nonlinear preference and utility theory*, The John Hopkins Press, Baltimore, MD.

Hurwicz, L. (1951) *Optimality Criteria for Decision Making Under Ignorance*, Cowles Commission Discussion Paper, Statistics, No. 370.

Klir, G., Yuan, B. (1995) *Fuzzy sets and fuzzy logic: theory and applications.* Prentice Hall, Upper Saddle River, New Jersey.

Kosheleva, O.M., Kreinovich. V. (2009) What is the Best Way to Distribute Efforts Among Students: Towards Quantitative Approach to Human Cognition. *Proceedings of the 28th North American Fuzzy Information Processing Society Annual Conference NAFIPS'09*, Cincinnati, Ohio, June 14–17, 2009.

Kreinovich, V., Xiang, G., Starks, S.A., Longpré, L., Ceberio, M., Araiza, R., Beck, J., Kandathi, R., Nayak, A., Torres, R., Hajagos, J. (2006) Towards combining probabilistic and interval uncertainty in engineering calculations: algorithms for computing statistics under interval uncertainty, and their computational complexity, *Reliable Computing*, vol. 12, no. 6, pp.471–501.

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

Moore, R.E., Lodwick, W. (2003) Interval Analysis and Fuzzy Set Theory, *Fuzzy Sets and Systems*, vol. 135, no. 1, pp.5–9.

Nguyen, H.T., Kreinovich, V. (1996) Nested Intervals and Sets: Concepts, Relations to Fuzzy Sets, and Applications, In: Kearfott, R.B., Kreinovich, V., eds., *Applications of Interval Computations*, Kluwer, Dordrecht, pp.245–290.

Nguyen, H.T., Walker, E.A. (2005) *A first course in fuzzy logic*, CRC Press, Boca Raton, Florida.

Sheskin, D.J. (2007) *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC Press, Boca Raton, Florida.

A Proof that under Reasonable Requirements, the Only Combination Functions are Linear Ones

For a linear function f, the property (6) is easy to derive. Indeed, due to the properties of expected value, we have

$$E[f(u_1,\ldots,u_n)] = E\left[w_0 + \sum_{i=1}^{n} w_i \cdot u_i\right] = w_0 + \sum_{i=1}^{n} w_i \cdot E[u_i] =$$

$$f(E[u_1], \ldots, E[u_n]).$$

Vice versa, let us assume that the function $f(u_1, \ldots, u_n)$ satisfies the property (6). For simplicity, let us restrict ourselves to the case $n = 2$; the case $n > 2$ can be handled similarly. Let us show that in the case $n = 2$, the function is indeed linear.

Indeed, without loss of generality, let us consider values $u_1 \in [0,1]$ and $u_2 \in [0,1]$ for which $u_1 \geq u_2$. Let us consider the situation with three possible outcomes:

- an outcome ω_1 whose probability is $p(\omega_1) = u_2$;
- an outcome ω_2 whose probability is $p(\omega_2) = u_1 - u_2$; and
- an outcome ω_1 whose probability is $p(\omega_3) = 1 - u_1$.

One can easily see that

$$p(\omega_1) + p(\omega_2) + p(\omega_3) = 1.$$

We then select the individual utility functions as follows:

- we take $u_1(\omega_1) = 1$ and $u_2(\omega_1) = 1$;
- we take $u_1(\omega_2) = 1$ and $u_2(\omega_2) = 0$; and
- we take $u_1(\omega_3) = 0$ and $u_2(\omega_3) = 0$.

In this case, the expected values of individual participants are equal to:

$$E[u_1] = p(\omega_1) \cdot u_1(\omega_1) + p(\omega_2) \cdot u_1(\omega_2) + p(\omega_3) \cdot u_1(\omega_3) =$$

$$u_2 \cdot 1 + (u_1 - u_2) \cdot 1 + (1 - u_1) \cdot 0 = u_1$$

and

$$E[u_2] = p(\omega_1) \cdot u_2(\omega_1) + p(\omega_2) \cdot u_2(\omega_2) + p(\omega_3) \cdot u_2(\omega_3) =$$

$$u_2 \cdot 1 + (u_1 - u_2) \cdot 0 + (1 - u_1) \cdot 0 = u_2.$$

In this situation, the group utility $u(\omega) = f(u_1(\omega), u_2(\omega))$ is equal to:

- for the outcome ω_1, to $u(\omega_1) = f(u_1(\omega_1), u_2(\omega_1)) = f(1,1)$;
- for the outcome ω_2, to $u(\omega_2) = f(u_1(\omega_2), u_2(\omega_2)) = f(1,0)$;
- for the outcome ω_3, to $u(\omega_3) = f(u_1(\omega_3), u_2(\omega_3)) = f(0,0)$.

Thus, the expected value of the group utility is equal to:

$$E[u] = p(\omega_1) \cdot u(\omega_1) + p(\omega_2) \cdot u(\omega_2) + p(\omega_3) \cdot u(\omega_3) =$$

$$u_2 \cdot f(1,1) + (u_1 - u_2) \cdot f(1,0) + (1 - u_1) \cdot f(0,0).$$

On the other hand, due to the property (6), this expected value should be equal to $f(E[u_1], E[u_2]) = f(u_1, u_2)$. Thus, we conclude that

$$f(u_1, u_2) = u_2 \cdot f(1,1) + (u_1 - u_2) \cdot f(1,0) + (1 - u_1) \cdot f(0,0). \quad (34)$$

So, for $u_1 \geq u_2$, the function $f(u_1, u_2)$ can indeed be described by an expression which is linear in u_1 and u_2.

In particular, for $u_1 = u_2 = 0.5$, we conclude that

$$f(0.5, 0.5) = 0.5 \cdot f(1,1) + 0.5 \cdot f(0,0). \tag{35}$$

Similarly, we can prove that another linear expression describes the function $f(u_1, u_2)$ for the case when $u_1 \leq u_2$:

$$f(u_1, u_2) = u_1 \cdot f(1,1) + (u_2 - u_1) \cdot f(0,1) + (1 - u_2) \cdot f(0,0). \tag{36}$$

To complete our proof, we must show that these two linear expressions coincide, i.e., that the same linear expression is applicable both:

- in the case of $u_1 \geq u_2$ and
- in the case of $u_2 \geq u_1$.

For that, let us consider a new situation with two possible outcomes:

- an outcome ω_1 whose probability is $p(\omega_1) = 0.5$, and
- an outcome ω_2 whose probability is $p(\omega_2) = 0.5$.

We then select the individual utility functions as follows:

- we take $u_1(\omega_1) = 1$ and $u_2(\omega_1) = 0$, and
- we take $u_1(\omega_2) = 0$ and $u_2(\omega_2) = 1$.

In this case, the expected values of individual participants are equal to:

$$E[u_1] = p(\omega_1) \cdot u_1(\omega_1) + p(\omega_2) \cdot u_1(\omega_2) = 0.5 \cdot 1 + 0.5 \cdot 0 = 0.5$$

and

$$E[u_2] = p(\omega_1) \cdot u_2(\omega_1) + p(\omega_2) \cdot u_2(\omega_2) = 0.5 \cdot 0 + 0.5 \cdot 1 = 0.5.$$

In this situation, the group utility $u(\omega) = f(u_1(\omega), u_2(\omega))$ is equal to:

- for the outcome ω_1, to $u(\omega_1) = f(u_1(\omega_1), u_2(\omega_1)) = f(1, 0)$, and
- for the outcome ω_2, to $u(\omega_2) = f(u_1(\omega_2), u_2(\omega_2)) = f(0, 1)$.

Thus, the expected values of the group utility is equal to:

$$E[u] = p(\omega_1) \cdot u(\omega_1) + p(\omega_2) \cdot u(\omega_2) = 0.5 \cdot f(1,0) + 0.5 \cdot f(0,1).$$

On the other hand, due to the property (6), this expected value should be equal to

$$f(E[u_1], E[u_2]) = f(0.5, 0.5).$$

Thus, we conclude that

$$f(0.5, 0.5) = 0.5 \cdot f(1,0) + 0.5 \cdot f(0,1). \tag{37}$$

By comparing (37) and (35), we conclude that

$$0.5 \cdot f(1,1) + 0.5 \cdot f(0,0) = 0.5 \cdot f(1,0) + 0.5 \cdot f(0,1), \tag{38}$$

hence

$$f(1,1) + f(0,0) = f(1,0) + f(0,1), \tag{39}$$

and, thus,

$$f(1,0) = f(1,1) + f(0,0) - f(0,1). \tag{40}$$

Substituting the expression (40) into the formula (34), we get the expression (36). Thus, the two linear formulas (34) and (36) are indeed identical.

The statement is proven.

B Interval Uncertainty, Smooth Benefit Function: Analysis

Informal description of our result. In the main text, we mentioned that for the case of smooth (quadratic) utility function $a(x)$ and exactly known x_i, we do not need to keep all n levels, it is sufficient to keep only the first and second sample moments of these levels. Let us show that for interval uncertainty, all n bounds are needed.

Specifically, we will prove the following.

Precise formulation of the result. Suppose that we have n intervals

$$[\widetilde{x}_i - \varepsilon_i, \widetilde{x}_i + \varepsilon_i].$$

We will consider a *non-degenerate* case when all the grades $\widetilde{x}_i$ are different.

Let us assume that for every quadratic function $a(x)$, we know the range $[\underline{f}, \overline{f}]$ of the function $a(x_1) + \ldots + a(x_n)$ over the intervals $[\widetilde{x}_i - \varepsilon_i, \widetilde{x}_i + \varepsilon_i]$. Then, based on the ranges corresponding to different quadratic functions $a(x)$, we can uniquely reconstruct the original collection of intervals.

In other words, if two different non-degenerate collections of intervals lead to exact same ranges for every quadratic function, then these collections coincide – i.e., they differ only by permutations.

Comment. It is not known whether the same is true if we allow arbitrary – not necessarily non-degenerate – collections of intervals.

Proof. For every quadratic function $a(x)$, the largest possible value $\overline{f}$ of the sum $\sum\limits_{i=1}^{n} a(x_i)$ is attained when each of the terms $u(x_i)$ is the largest possible, and is equal to the sum of the corresponding n largest values:

$$\overline{f} = \overline{f}_1 + \ldots + \overline{f}_n. \tag{41}$$

For every real number α, the quadratic function $a(x) = (x - \alpha)^2$ attains its largest value on the interval $[\widetilde{x}_i - \varepsilon_i, \widetilde{x}_i + \varepsilon_i]$ at one of the endpoints $\widetilde{x}_i - \varepsilon_i$ or $\widetilde{x}_i + \varepsilon_i$. One can easily check that:

- when $\alpha \le \widetilde{x}_i$, then the largest possible value $\overline{f}_i$ of $a(x)$ on the interval

$$[\widetilde{x}_i - \varepsilon_i, \widetilde{x}_i + \varepsilon_i]$$

 is attained when $x_i = \overline{x}_i = \widetilde{x}_i + \varepsilon_i$ and is equal to $\overline{f}_i = (\overline{x}_i - \alpha)^2$;

- when $\alpha \ge \widetilde{x}_i$, then the largest possible value $\overline{f}_i$ of $a(x)$ on the interval

$$[\widetilde{x}_i - \varepsilon_i, \widetilde{x}_i + \varepsilon_i]$$

 is attained when $x_i = \underline{x}_i = \widetilde{x}_i - \varepsilon_i$ and is equal to $\overline{f}_i = (\underline{x}_i - \alpha)^2$.

Let us use this fact to describe the dependence of $\overline{f}$ on the parameter α.

When $\alpha \ne \widetilde{x}_i$, the value $\overline{f}$ is the sum of n smooth expressions.

At each point $\alpha = \overline{x}_i$, all the terms $\overline{f}_j$ in the sum $\overline{f}$ are smooth except for the term $\overline{f}_i$ that turns from $(\overline{x}_i - \alpha)^2$ to $(\underline{x}_i - \alpha)^2$. The derivative of $\overline{f}_i$ with respect to α changes from $2 \cdot (\alpha - \overline{x}_i)$ to $2 \cdot (\alpha - \underline{x}_i)$, i.e., increases by

$$2 \cdot (\alpha - \underline{x}_i) - 2 \cdot (\alpha - \overline{x}_i) = 2 \cdot (\overline{x}_i - \underline{x}_i) = 4 \cdot \varepsilon_i. \quad (42)$$

Since all the other components $\overline{f}_j$ are smooth at $\alpha = \widetilde{x}_i$, at $\alpha = \widetilde{x}_i$, the derivative of the sum $\overline{f}(\alpha)$ also increases by $4\varepsilon_i$.

Thus, once we know the value $\overline{f}$ for all α,

- we can find the values $\widetilde{x}_i$ as the values at which the derivative is discontinuous; and
- we can find each value ε_i as 1/4 of the increase of the derivative at the corresponding point $\widetilde{x}_i$.

The statement is proven.