What is the Best Way to Distribute Efforts Among Students: Towards Quantitative Approach to Human Cognition

Olga Kosheleva[1] and Vladik Kreinovich[2]

[1]Department of Teacher Education
[2]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract

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 should 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?

Even when we know the exact way each student learns, the answer depends on what is the objective of teaching the class. This can be illustrated on two extreme example: If the objective is to leave no student behind, then in the optimal resource arrangement all the effort goes to weak students who are behind, while more advanced students get bored. If the effort is to increase the school's rating by increasing the number of graduates who are accepted at top universities, then all the effort should go to the advanced students while weak students fail.

An additional difficulty is that in reality, we do not have exact information about the cognitive ability of each student, there is a large amount of uncertainty. In this paper, we analyze the problem of optimal resource distribution under uncertainty. We hope that the resulting algorithms will be useful in designing teaching strategies.

1 Deciding Which Teaching Method Is Better: Formulation of the Problem

Pedagogy is a fast developing field. New methods, new ideas and constantly being developed and tested. New methods and new idea may be different in many things:

- they may differ in the way material is presented,

- they may also differ in the way the teacher's effort is distributed among individual students: which of the students get more attention, etc.

Testing can also be different:

- Sometimes, the testing consists of comparing the new method with the method that is currently used.

- Sometimes, the testing consists of comparing two (or more) different versions of the same pedagogical technique; such testing is needed to decide which version is better.

To perform a meaningful comparison, we need to agree on the criterion: how do we decide which method is better? Once we have selected a criterion, and we have performed enough experiments to get a good idea of how students will learn under different version of the method, a natural question is: what is the optimal way to teaching the students – optimal with respect to the selected criterion?

2 How This Problem is Usually Solved Now: A Brief Description

How can we gauge the efficiency of different teaching techniques? The success of each individual student i can be naturally gauged by this student's grade x_i. So, for two different techniques T and T', we know the corresponding grades $x_1, \ldots, x_n$ and $x'_1, \ldots, x'_{n'}$. Which method is better?

In some cases, the answer to this question is straightforward. For example, when $n' = n$ and when we can rearrange the grades in such a way that $x_i \leq x'_i$ for all i and $x_i < x'_i$ for some i, then clearly the method T' is better.

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

In pedagogical experiments, 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., [4]) 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).

3   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 grade corresponding to C, and any 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.

4   Towards Selecting the Optimal Teaching Strategy: Possible Objective Functions

Motivations.   Since the traditional approach – of using the average grade as a criterion – is not always adequate, let us formulate the general problem of optimal teaching.

General description.   To formulate this problem, we must know how the relative "quality" of a given teaching strategy can be determined from the grades $x_1, \ldots, x_n$. In this section, we will denote the corresponding dependence by $f(x_1, \ldots, x_n)$.

The traditional approach.   In particular, the traditional approach corresponds to using the average

$$f(x_1, \ldots, x_n) = \frac{x_1 + \ldots + x_n}{n}. \tag{6}$$

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:

$$f(x_1, \ldots, x_n) = \#\{i : x_i < x_0\}. \tag{7}$$

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 (7) as a maximization one: namely, minimizing (7) is equivalent to maximize the number of students whose grade is above (or equal to) the passing threshold x_0:

$$f(x_1, \ldots, x_n) = \#\{i : x_i \geq x_0\}. \tag{8}$$

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:

$$f(x_1, \ldots, x_n) = \min(x_1, \ldots, x_n). \tag{9}$$

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:

$$f(x_1, \ldots, x_n) = \#\{i : x_i \geq e_0\}. \tag{10}$$

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 (7), with $x_0 = e_0$,

Best school to get in. There is a version of the above criterion which is not equivalent to (7), 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

$$f(x_1, \ldots, x_n) = \max(x_1, \ldots, x_n). \tag{11}$$

Case of independence. An important practical case is when students 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.,

$$f(x_1, \ldots, x_n) = f_1(x_1) + \ldots + f_n(x_n); \tag{12}$$

see, e.g., [1, 2].

In this case, increasing the grade of one of the students will make the situation better – so it is reasonable to assume that all the functions $f_i(x_i)$ are strictly increasing.

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.

5 Towards Selecting the Optimal Teaching Strategy: Formulation of the Problem

Let $e_i(x_i)$ denote the amount of effort (time, etc.) that is need for i-th student to achieve the grade x_i. Clearly, the better grade we want to achieve, the more effort we need, so each function $e_i(x_i)$ is strictly increasing.

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:

$$\text{Maximize } f(x_1, \ldots, x_n) \tag{13}$$

under the constraint

$$e_1(x_1) + \ldots + e_n(x_n) \leq e. \tag{14}$$

6 Explicit Solution: Case of Independent Students

For the case of independent students, when the objective function has the form (12), it is possible to derive an explicit solution to the corresponding constraint optimization problem (13), (14).

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 (12). In other words, the maximum is attained when all the effort is actually used, i.e., when we have the constraint

$$e_1(x_1) + \ldots + e_n(x_n) = e. \tag{15}$$

To maximize the objective function (12) under this constraint, we can use the Lagrange multiplier method. According to this method, the maximum of the function (12) under constraint (15) is attained when for some value λ, the auxiliary function

$$f_1(x_1) + \ldots + f_n(x_n) + \lambda \cdot (e_1(x_1) + \ldots + e_n(x_n)) \tag{16}$$

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

$$f_i'(x_i) + \lambda \cdot e_i'(x_i) = 0, \tag{17}$$

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

$$-\frac{f_i'(x_i)}{e_i'(x_i)} = \lambda. \tag{18}$$

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 (18) with a single variable x_i.

The value λ can be found from the formula (15), i.e., from the condition that for the resulting values x_i, we get $\sum\limits_{i=1}^{n} e_i(x_i) = e$.

7 Explicit Solution: "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 (15), i.e., from the equation

$$e_1(x_c) + \ldots + e_n(x_c) = e. \tag{19}$$

Comment. A slightly more complex situation occurs when we start not at the beginning, but at the intermediate situation when some students already have some knowledge. 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

$$e_1(x_k^{(0)}) + \ldots + e_k(x_k^{(0)}) \leq e. \tag{20}$$

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

$$e_1(x) + \ldots + e_{k-1}(x) + e_k(x) = e. \tag{21}$$

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

8   Explicit Solution: "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.

9   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 achieving the knowledge levels corresponding to the grades $x_1, \ldots, x_n$; for example, we know the exact expressions for the marginal functions $f_i(x_i)$;

- we know *exactly* how much effort $e_i(x_i)$ is needed to bring each student i to a given grade level x_i, and

- we know *exactly* the level of knowledge x_i of each student – it is exactly determined by the grade x_i.

In practice, we have *uncertainty*.

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

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

Interval uncertainty. Finally, the grade $\widetilde{x}_i$ is only an approximate indication of the student's level of knowledge. Once we know the grade $\widetilde{x}_i$, we cannot conclude that the level of knowledge x_i is exactly $\widetilde{x}_i$. At best, we know the accuracy ε_i of this representation. In this case, the actual (unknown) level of knowledge 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

$$[\underline{f}, \overline{f}] = f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{def}}{=} \{f(x_1, \ldots, x_n) \mid x_1 \in \mathbf{x}_1, \ldots, x_n \in \mathbf{x}_n\}. \tag{22}$$

10 How to Take Uncertainty Into Account

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

Average benefit 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* benefit function $u(x)$. In this case, for a combination of grades $x_1, \ldots, x_n$, the resulting value of the objective function is

$$f(x_1, \ldots, x_n) = u(x_1) + \ldots + u(x_n). \tag{23}$$

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

$$u(x) = u_0 + u_1 \cdot x + u_2 \cdot x^2, \tag{24}$$

for some coefficients u_0, u_1, and u_2. For this function, the expression (23) for the objective function takes the form

$$f(x_1, \ldots, x_n) = n \cdot u_0 + u_1 \cdot \sum_{i=1}^{n} x_i + u_2 \cdot \sum_{i=1}^{n} x_i^2, \tag{25}$$

i.e., the form

$$f(x_1, \ldots, x_n) = f_0 + f_1 \cdot E + f_2 \cdot M, \tag{26}$$

where $f_0 \stackrel{\text{def}}{=} n \cdot u_0$, $f_1 \stackrel{\text{def}}{=} n \cdot u_1, f_2 \stackrel{\text{def}}{=} n \cdot u_2$, E is the average (1), and M is the second sample moment:

$$M \stackrel{\text{def}}{=} \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2. \tag{27}$$

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

Comment. In general, the benefit function $u(x)$ is increasing with x_i. However, it is worth mentioning that this conclusion holds for every quadratic function $u(x)$, 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 students' knowledge 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 $u(x)$ is increasing (the more knowledge the better),

- the largest possible value $\overline{f}$ of the objective function 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:

$$[\underline{f}, \overline{f}] = \left[\sum_{i=1}^{n} u(\underline{x}_i), \sum_{i=1}^{n} u(\overline{x}_i) \right]. \tag{28}$$

Comment. We mentioned that for the case of smooth (quadratic) benefit function and exactly known x_i, we do not need to keep all n grades – 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.

References

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

[2] P.C. Fishburn, *Nonlinear preference and utility theory*, The John Hopkins Press, Baltimore, MD, 1988.

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

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

Appendix
Interval Uncertainty, Smooth Benefit Function: Analysis

Informal description of our result. In the main text, we mentioned that for the case of smooth (quadratic) benefit function $u(x)$ and exactly known x_i, we do not need to keep all n grades, it is sufficient to keep only the first and second sample moments of these grades. 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 $u(x)$, we know the range $[\underline{f}, \overline{f}]$ of the function $u(x_1) + \ldots + u(x_n)$ over the intervals $[\widetilde{x}_i - \varepsilon_i, \widetilde{x}_i + \varepsilon_i]$. Then, based on the ranges corresponding to different quadratic functions $u(x)$, 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 $u(x)$, the largest possible value $\overline{f}$ of the sum $\sum\limits_{i=1}^{n} u(x_i)$ 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:

$$\overline{f} = \overline{f}_1 + \ldots + \overline{f}_n. \tag{29}$$

For every real number a, the quadratic function $u(x) = (x - \alpha)^2$ 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 $a \leq \widetilde{x}_i$, then the largest possible value $\overline{f}_i$ of $u(x)$ 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 $\overline{f}_i = (\overline{x}_i - a)^2$;

- when $a \geq \widetilde{x}_i$, then the largest possible value $\overline{f}_i$ of $u(x)$ 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 $\overline{f}_i = (\underline{x}_i - a)^2$.

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

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

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

$$2 \cdot (a - \underline{x}_i) - 2 \cdot (a - \overline{x}_i) = 2 \cdot (\overline{x}_i - \underline{x}_i) = 4 \cdot \varepsilon_i. \tag{30}$$

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

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

- 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.

Estimating $f(E, V)$ under interval uncertainty.   Let us now 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., [3]. This means, crudely speaking, that unless P=NP (and most computer scientists believe that P$\neq$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 $\overline{x}_i \leq \underline{r}$;

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

- $x_i = \underline{x}_i$ when $\overline{r} \leq \underline{x}_i$.

(The proof is similar to the ones given in [3].)

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 [3] shows that if we sort the intervals in lexicographic order

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

where

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

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.

Received: August, 2009