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 Mathematics Education
[2]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
Emails: {olgak,vladik}@utep.edu

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.
- To perform a meaningful testing, we need to agree on the criterion.
- Once we have selected a criterion, a natural question is: what is the optimal way to teaching the students.

2. How Techniques Are Compared Now: A Brief Description

- 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 $x'_1, \ldots, x'_{n'}$.
- In pedagogical experiments, the decision is usually made based on the comparison of the average grades

$$E \stackrel{\text{def}}{=} \frac{x_1 + \ldots + x_n}{n} \text{ and } E' \stackrel{\text{def}}{=} \frac{x'_1 + \ldots + x'_{n'}}{n'}.$$

- *Example:* we had $x_1 = 60$, $x_2 = 90$, hence $E = 75$. Now, we have $x'_1 = x'_2 = 70$, and $E' = 70$. In T':
  - the average grade is worse, but
  - in contrast to T, no one failed.

3. Towards Selecting the Optimal Teaching Strategy: Possible Objective Functions

- *Fact:* the traditional approach – of using the average grade as a criterion – is not always adequate.
- *Conclusion:* other criteria $f(x_1, \ldots, x_n)$ are needed.
- *Maximizing passing rate:* $f = \#\{i : x_i \geq x_0\}$.
- *No child left behind:* $f(x_1, \ldots, x_n) = \min(x_1, \ldots, x_n)$.
- *Best school to get in:* $f(x_1, \ldots, x_n) = \max(x_1, \ldots, x_n)$.
- *Case of independence:* decision theory leads to $f = f_1(x_1) + \ldots + f_n(x_n)$ for some functions $f_i(x_i)$.
- *Criteria combining mean E and variance V to take into account that a larger mean is not always better:*

$$f(x_1, \ldots, x_n) = f(E, V).$$

- *Comment:* it is reasonable to require that $f(E, V)$ is increasing in E and decreasing in V.

4. 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 takes the following form:

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

under the constraint

$$e_1(x_1) + \ldots + e_n(x_n) \le e.$$

5. Explicit Solution: Case of Independent Students

- *Maximize:* $f_1(x_1) + \ldots + f_n(x_n)$ under the constraint

$$e_1(x_1) + \ldots + e_n(x_n) \le e.$$

- *Observation:* the more efforts, the better results, so we can assume $e_1(x_1) + \ldots + e_n(x_n) = e$.
- *Lagrange multiplier:* maximize

$$J = \sum_{i=1}^{n} f_i(x_i) + \lambda \cdot \sum_{i=1}^{n} e_i(x_i).$$

- Equation $\dfrac{\partial J}{\partial x_i} = 0$ leads to $f_i'(x_i) + \lambda \cdot e_i'(x_i) = 0$.
- Thus, once we know λ, we can find all x_i.
- λ can be found from the condition $\sum\limits_{i=1}^{n} e_i(x_i(\lambda)) = e$.

6. Explicit Solution: "No Child Left Behind" Case

- In the No Child Left Behind case, we maximize the lowest grade.
- There is no sense to use the effort to get one of the student grades better than the lowest grade.
- It is more beneficial to use the same efforts to increase the grades of all the students at the same time.
- In this case, the common grade x_c that we can achieve can be determined from the equation

$$e_1(x_c) + \ldots + e_n(x_c) = e.$$

- Students may already have knowledge $x^{(1)} \le x^{(2)} \le \ldots$
- In this case, we find the largest k for which $e_1(x^{(0)}_k) + \ldots + e_k(x^{(0)}_k) \le e$ and then $x \in [x^{(0)}_k, x^{(0)}_{k+1})$ s.t.

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

7. Explicit Solution: "Best School to Get In" Case

- Best School to Get In means maximizing the largest possible grade x_i.
- 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:
  - 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.

8. Need to Take Uncertainty Into Account

- We assumed that:
  - we know *exactly* the benefits $f(x_1, \ldots, x_n)$ of achieving knowledge levels x_i;
  - we know *exactly* how much effort $e_i(x_i)$ is needed for each level x_i, and
  - we know *exactly* the level of knowledge x_i of each student.
- In practice, we have *uncertainty*:
  - we only know the *average* benefit $u(x)$ of grade x to a student;
  - we only know the *average* effort $e(x)$ needed to bring a student to the level x; and
  - the grade $\widetilde{x}_i$ is only an approximate indication of the student's level of knowledge.

9. Average Benefit Function

- *Objective function:* $f(x_1, \ldots, x_n) = u(x_1) + \ldots + u(x_n)$.
- Usually, the benefit function is reasonably smooth.
- In this case, if (hopefully) all grades are close, we can keep only quadratic terms in the Taylor expansion:

$$u(x) = u_0 + u_1 \cdot x + u_2 \cdot x^2.$$

- So, the objective function takes the form

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

- *Fact:* $E = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$ and $M = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 = V + E^2$.
- *Conclusion:* f depends only on the mean E and on the variance V.

10. Case of Interval Uncertainty

- *Situation:* we only know intervals $[\underline{x}_i, \overline{x}_i]$ of possible values of x_i.
- *Fact:* the benefit function $u(x)$ is increasing (the more knowledge the better).
- *Conclusion:*
  - the benefit is the largest when $x_i = \overline{x}_i$, and
  - the benefit is the smallest when $x_i = \underline{x}_i$.
- *Resulting formula:* $[\underline{f}, \overline{f}] = \left[\sum_{i=1}^{n} u(\underline{x}_i), \sum_{i=1}^{n} u(\overline{x}_i)\right]$.
- *Reminder:* for quadratic $u(x)$ and exactly known x_i, we only need to know E and M.
- *New result:* under interval uncertainty, we need all n intervals.

11. Case of 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.
- A natural way to formalize such words is to use fuzzy logic:
  - for each possible value of $x_i \in [\underline{x}_i, \overline{x}_i]$,
  - we describe the degree $\mu_i(x_i)$ to which x_i is possible.
- Alternatively, we can consider α-cuts $\{x : \mu_i(x_i) \geq \alpha\}$.
- For each α, the fuzzy set $y = f(x_1, \ldots, x_n)$ has α-cuts

$$\mathbf{y}(\alpha) = f(\mathbf{x}_1(\alpha), \ldots, \mathbf{x}_1(\alpha)).$$

- So, the problem of propagating fuzzy uncertainty can be reduced to several interval propagation problems.

12. Acknowledgments

This work was supported in part:

- by NSF grant HRD-0734825, and
- by Grant 1 T36 GM078000-01 from the National Institutes of Health.