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

Olga Kosheleva[1] and Vladik Kreinovich[2]

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

1. Formulation of the Problem

- *Need for effort distribution:* we often need to take care of several (reasonably) independent participants:
  - we want all *economic* regions to prosper;
  - we want all geographic regions to have healthy *environment*;
  - we want all *students* to learn the knowledge and skills.
- *Fact:* amount of resources is limited.
- *Problem:* how to best distribute these resources?
- *Additional problem:* uncertainty – we do not know the exact results of different actions.

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

- *Situation:* we have two strategies T and T' leading to success values $x_1, \ldots, x_n$ and $x'_1, \ldots, x'_n$.
- *Example:* x_i and x'_i are grades of i-th student.
- *Case when comparison is easy:* $x_i \le x'_i$ for all i.
- *Traditional idea:* $E = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} x_i$ vs. $E' = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} x'_i$.
- *Specifics* – use t-test: T' better if $\dfrac{E' - E}{\sqrt{V/n + V'/n}} \ge t_\alpha$.
- *Teaching example:* $x_1 = x_2 = 70$, $x'_1 = 60$, and $x'_2 = 90$.
- *Recommendation:* $E' = 75 > E = 70$, so T' is better.
- *Problem:* for T, both students pass, but for T' one fails.

3. Towards a More Adequate Approach

- *More adequate approach:* use *utilities* $u_i(a)$ to describe consequences of each action.
- *What is utility:* we pick $A_0 \ll A_1$; $u(E)$ is the probability p for which $E \equiv$"A_1 w/prob. p else A_0."
- *Decision making theory:* we prefer an action a for which $E[u_i(a)]$ is the largest.
- *Question:* how to combine values $u_i(a)$ into a single value $u(a) = f(u_1(a), \ldots, u_n(a))$.
- *Meaning:* the larger $u(a)$, the better the alternative a for the group as a whole.
- *Natural idea:* if alternatives are equivalent for all participants, they should be equivalent for the group:

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

- *Result:* only linear functions f satisfy this property.

4. Towards the Optimal Effort Distribution: Constraint Optimization Problem

- $e_i(x_i) \stackrel{\text{def}}{=}$ effort needed for i-th person to reach level x_i.
- *Max* $f(x_1, \ldots, x_n) = w_0 + w_1 \cdot u_1(x_1) + \ldots + w_n \cdot u_n(x_n)$ under the constraint $e_1(x_1) + \ldots + e_n(x_n) \le e$.
- *Simplification:* re-scale utilities to $f_i(x_i) \stackrel{\text{def}}{=} w_i \cdot u_i(x_i)$, then we maximize $f_1(x_1) + \ldots + f_n(x_n)$.
- *Comment:* gain most when all effort used $\sum\limits_{i=1}^{n} e_i(x_i) = e$.
- *Lagrange multiplier:* $\sum\limits_{i=1}^{n} f_i(x_i) + \lambda \cdot \sum\limits_{i=1}^{n} e_i(x_i) \to \max$,

  hence $f_i'(x_i) + \lambda \cdot e_i'(x_i) = 0$, and $-\dfrac{f_i'(x_i)}{e_i'(x_i)} = \lambda$.
- *Comment.* Once we know λ, we can find all x_i.
- λ can be found from the condition $\sum\limits_{i=1}^{n} e_i(x_i) = e$.

5. Need to Take Uncertainty Into Account

- *Ideal situations:* we know the exact utility $f_i(x_i)$ and effort $e_i(x_i)$ for each participant i.
- *In practice:* we usually know the average benefit function $a(x)$ and the average effort function $e(x)$.
- *Additional uncertainty:* we only have approximate estimates $\widetilde{x}_i$ of the levels x_i.
- *Crisp uncertainty:* we know the upper bound ε_i on the the approximation error $|x_i - \widetilde{x}_i|$.
- *Intervals:* after measuring $\widetilde{x}_i$, we only know that

$$x_i \in \mathbf{x}_i \stackrel{\text{def}}{=} [\widetilde{x}_i - \varepsilon_i, \widetilde{x}_i + \varepsilon_i].$$

- *Estimate:* instead of a single value $f(x_1, \ldots, x_n)$, we get an *interval* of possible values

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

6. Case of Fuzzy Uncertainty

- *Situation:* we have expert estimates of the accuracy of $\widetilde{x}_i$.
- *Natural description:* we can represent this expert information as a fuzzy number $\mu_i(x_i)$.
- *Equivalent formulation:* we have different intervals (α-cuts) $\mathbf{x}_i(\alpha)$ corresponding to different $\alpha \in [0,1]$:

$$\mathbf{x}_i(\alpha) \stackrel{\text{def}}{=} \{x_i \mid \mu_i(x_i) \ge \alpha\}.$$

- *Processing fuzzy estimates:* for every α, we have

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

- *Conclusion:* from the computational viewpoint, it is sufficient to consider interval uncertainty.

7. How to Take Uncertainty Into Account

- *Average utility function – reminder:* we only know the average utility function $a(x)$.
- *Resulting formula:* $f(x_1, \ldots, x_n) = a(x_1) + \ldots + a(x_n)$.
- *Usual case:* $a(x)$ is smooth.
- *Idea:* expand $a(x)$ in Taylor series around the average level, and keep only quadratic terms:

$$a(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2.$$

- $f = a_0 + a_1 \cdot S_1 + a_2 \cdot S_2$, where $S_1 = \sum\limits_{i=1}^n x_i$, $S_2 = \sum\limits_{i=1}^n x_i^2$.
- *Interval case:* $[\underline{f}, \overline{f}] = \left[\sum\limits_{i=1}^n a(\underline{x}_i), \sum\limits_{i=1}^n a(\overline{x}_i)\right]$ (since $a\uparrow$).
- *How to solve the corresponding optimization problem:* use Hurwicz criterion and optimize $\alpha_{\rm opt} \cdot \overline{f} + (1 - \alpha_{\rm opt}) \cdot \underline{f}$.

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

- *So far:* we considered utility-motivated linear combinations $f(x_1, \ldots, x_n)$ of utility functions.
- *In practice:* other functions $f(x_1, \ldots, x_n)$ are also used.
- *Examples:* from education.
- *Smallest failure rate:* $f(x_1, \ldots, x_n) = \#\{i : x_i < x_0\}$.
- *Equivalent reformulation:* maximizing the number of passing students $f(x_1, \ldots, x_n) = \#\{i : x_i \geq x_0\}$.
- *No Child Left Behind:* we gauge the quality of a school by the performance of the worst student

$$f(x_1, \ldots, x_n) = \min(x_1, \ldots, x_n).$$

- *Max success rate:* $f(x_1, \ldots, x_n) = \#\{i : x_i \geq x_0\}$.
- *Best school to get in:* $f(x_1, \ldots, x_n) = \max(x_1, \ldots, x_n)$.

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

- *Reminder:* we maximize the lowest grade.
- *Analysis:* there is no sense to use the effort to get one of the student grades better than the lowest grade.
- *Reason:* because the lowest grade will not change.
- *Conclusion:* use the efforts to increase the grades of all the students at the same time – this will increase the lowest grade.
- *Solution:* get the common grade x_c, where x_c can be determined from the condition $e_1(x_c)+\ldots+e_n(x_c) = e$.
- *More realistic situation:* students have some prior knowledge $x_i^{(0)}$, $i = 1, \ldots, n$.
- *Question:* what is the optimal effort allocation?

10. Explicit Solution: "No Child Left Behind" Case (cont-d)

- *Reminder:* we maximize the lowest grade.
- *Reminder:* students have prior knowledge $x_i^{(0)}$.
- *Solution:* sort students in order of prior knowledge

$$x_1^{(0)} \le \ldots \le x_n^{(0)};$$

  then:
  - first, 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 students w/$x_1^{(0)}$ and $x_2^{(0)}$ to the next level $x_3^{(0)}$;
  - etc.

11. Explicit Solution: "No Child Left Behind" Case (exact algorithm)

- First, we find the largest value k for which all the grades $x_1, \ldots, x_k$ can be raised to the k-th prior level $x_k^{(0)}$.
- In precise terms, this means the largest value k for which
$$e_1(x_k^{(0)}) + \ldots + e_k(x_k^{(0)}) \le e.$$
- This means that for the criterion $\min(x_1, \ldots, x_n)$, we can achieve the value $x_k^{(0)}$, but not $x_{k+1}^{(0)}$.
- Then, we find the value $x \in [x_k^{(0)}, x_{k+1}^{(0)})$ for which
$$e_1(x) + \ldots + e_{k-1}(x) + e_k(x) = e.$$
- This value x is the optimal value of the criterion
$$\min(x_1, \ldots, x_n).$$

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

- *Reminder:* maximize the largest possible grade x_i.
- *Meaning:* "one of our students went to Harvard".
- *Natural optimality idea:*
  - concentrate on a single individual, and
  - ignore the rest.
- Which individual to target depends on how much gain we will get.
- *Resulting solution:*
  - 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.

13. Criteria Combining Mean and Variance

- *Traditional approach (reminder):* we only take into account the average (mean) grade E.
- *Limitation:* the mean does not tell us how much the grades deviate from the mean.
- *Fact:* This information is provided by the variance V.
- *Idea:* use criteria of the type $f(E,V)$.
- When the mean E is fixed, usually, we aim for the smallest possible variation.
- *Comment:* unless we gauge a school by its best students.
- Similarly, when the variance V is fixed, we aim for the largest possible mean E.
- *Conclusion:* we require that $f(E,V)$ is increasing in E and decreasing in V.

14. Estimating $f(E, V)$ Under Interval Uncertainty

- *Problem:* compute the range $[\underline{f}, \overline{f}]$ of $f(E, V)$ when $x_i \in [\underline{x}_i, \overline{x}_i]$.
- *Alas:* computing $\underline{f}$ is NP-hard, even for $f(E, V) = -V$.
- *Meaning:* unless P=NP (and most computer scientists believe that P≠NP), no efficient (polynomial time) algorithm can always compute the exact range.
- *Good news:* $\overline{f}$ can be found efficiently:
  - consider all $2n + 2$ intervals $[\underline{r}, \overline{r}]$ into which the values $\underline{x}_i$ and $\overline{x}_i$ divide the real line;
  - compute $f(E, V)$ when $x_i = \overline{x}_i$ for $\overline{x}_i \le \underline{r}$; $x_i = r$ for $[\underline{r}, \overline{r}] \subseteq [\underline{x}_i, \overline{x}_i]$; and $x_i = \underline{x}_i$ for $\overline{r} \le \underline{x}_i$;
  - the largest of the resulting $2n + 2$ values $f(E, V)$ is $\overline{f}$.

15. Estimating $f(E, V)$ Under Interval Uncertainty (cont-d)

- *Problem:* computing the minimum $\underline{f}$ of $f(E, V)$ when $x_i \in [\underline{x}_i, \overline{x}_i]$ is NP-hard.
- *It is possible:* to efficiently compute $\underline{f}$ when none of the intervals $[\underline{x}_i, \overline{x}_i]$ is a proper subset of another one.
- Sort the intervals in lexicographic order

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

  where $[\underline{a}, \overline{b}] \le [\underline{b}, \overline{b}] \leftrightarrow \underline{a} < \underline{b} \vee (\underline{a} = \underline{b} \,\&\, \overline{a} \le \overline{b})$.
- The minimum of f is attained at one of the combinations $(\underline{x}_1, \ldots, \underline{x}_{k-1}, x_k, \overline{x}_{k+1}, \ldots, \overline{x}_n)$ for some $x_k \in [\underline{x}_k, \overline{x}_k]$;
- Thus, $\underline{f}$ is the smallest of the corresponding $n+1$ values of $f(E, V)$.

16. Acknowledgments

This work was supported in part by NSF grants:

- Cyber-ShARE Center of Excellence (HRD-0734825),
- Computing Alliance of Hispanic-Serving Institutions CAHSI (CNS-0540592),

and by NIH Grant 1 T36 GM078000-01.