# Which Value $\widetilde{x}$ Best Represents a Sample $x_1, \ldots, x_n$: Utility-Based Approach Under Interval Uncertainty

Andrzej Pownuk and Vladik Kreinovich

Computational Science Program
University of Texas at El Paso
500 W. University
El Paso, Texas 79968, USA
ampownuk@utep.edu, vladik@utep.edu

Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

# 1. Need to Combine Several Estimates

- In many practical situations, we have several estimates $x_1, \ldots, x_n$ of the same quantity $x$.
- In such situations, it is often desirable to combine this information into a single estimate $\widetilde{x}$.
- Sometimes, we know the probability distribution of the corresponding estimation errors $x_i - x$.
- Then, we can use known statistical techniques to find $\widetilde{x}$.
- E.g., we can use the Maximum Likelihood Method.
- In many cases, however, we do not have any information about the corresponding probability distribution.
- How can we then find $\widetilde{x}$?


Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

## 2. Utility-Based Approach

- According to the decision theory, decisions of a rational person are $\Leftrightarrow$ maximizing *utility value* $u$.
- Let us thus find the estimate $\widetilde{x}$ for which the utility $u(\widetilde{x})$ is the largest.
- We use a single value $\widetilde{x}$ instead of all $n$ values $x_i$; for each $i$:
  - if the actual estimate is $x_i$ and we use a different value $\widetilde{x} \neq x_i$ instead,
  - then we are not doing an optimal thing.
- For example:
  - if the optimal speed at which the car needs the least amount of fuel is $x_i$,
  - and we instead run it at a speed $\widetilde{x} \neq x_i$, we thus waste some fuel.

Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

# 3. Utility-Based Approach (cont-d)

- For each $i$, the disutility $d$ comes from the fact that the difference $\widetilde{x} - x_i$ is different from 0.
- There is no disutility if we use the actual value, so $d = d(\widetilde{x} - x_i)$ for some function $d(y)$.
- Here $d(0) = 0$ and $d(y) > 0$ for $y \neq 0$.
- The estimates are usually reasonably accurate, so the difference $x_i - \widetilde{x}$ is small.
- So, we can expand the function $d(y)$ in Taylor series and keep only the first few terms in this expansion:

$$d(y) = d_0 + d_1 \cdot y + d_2 \cdot y^2 + \ldots$$

- From $d(0) = 0$ we conclude that $d_0 = 0$.
- From $d(y) > 0$ for $y \neq 0$ we conclude that $d_1 = 0$ (else we would have $d(y) < 0$ for small $y$) and $d_2 > 0$, so

$$d(y) = d_2 \cdot y^2 = d_2 \cdot (\widetilde{x} - x_i)^2.$$

Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

# 4. Utility-Based Approach (final)

- The overall disutility $d(\widetilde{x})$ of using $\widetilde{x}$ instead of each of the values $x_1, \ldots, x_n$ can be computed as the sum:

$$d(\widetilde{x}) = \sum_{i=1}^{n} d(\widetilde{x} - x_i)^2 = d_2 \cdot \sum_{i=1}^{n} (\widetilde{x} - x_i)^2.$$

- $u(\widetilde{x}) \stackrel{\text{def}}{=} -d(\widetilde{x}) \to \max \Leftrightarrow d(\widetilde{x}) \to \min.$
- Since $d_2 > 0$, minimizing disutility is equivalent to minimizing re-scaled disutility:

$$D(\widetilde{x}) \stackrel{\text{def}}{=} \frac{d(\widetilde{x})}{d_2} = \sum_{i=1}^{n} (\widetilde{x} - x_i)^2.$$

- Equating the derivative to 0, we get the well-known sample mean: $\widetilde{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i.$


Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

# 5. Case of Interval Uncertainty

- In many practical situations, we only know the intervals $[\underline{x}_i, \overline{x}_i]$ that contain the unknown values $x_i$.
- For different $x_i \in [\underline{x}_i, \overline{x}_i]$, we get, in general, different values of utility

$$U(\widetilde{x}, x_1, \ldots, x_n) = -D(\widetilde{x}, x_1, \ldots, x_n), \text{ where}$$

$$D(\widetilde{x}, x_1, \ldots, x_n) = \sum_{i=1}^{n} (\widetilde{x} - x_i)^2.$$

- Thus, all we know is that the actual (unknown) value of the utility belongs to the interval

$$[\underline{U}(\widetilde{x}), \overline{U}(\widetilde{x})] = [-\overline{D}(\widetilde{x}), -\underline{D}(\widetilde{x})], \text{ where}$$

$$\underline{D}(\widetilde{x}) = \min D(\widetilde{x}, x_1, \ldots, x_n), \ \ \overline{D}(\widetilde{x}) = \max D(\widetilde{x}, x_1, \ldots, x_n).$$

- In such situations, decision theory recommends using Hurwicz optimism-pessimism criterion, i.e., maximize:

$$U(\widetilde{x}) \stackrel{\text{def}}{=} \alpha \cdot \overline{U}(\widetilde{x}) + (1 - \alpha) \cdot \underline{U}(\widetilde{x}).$$

Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

# 6. Case of Interval Uncertainty (cont-d)

- According to Hurwicz criterion, we maximize:

$$U(\widetilde{x}) \stackrel{\text{def}}{=} \alpha \cdot \overline{U}(\widetilde{x}) + (1-\alpha) \cdot \underline{U}(\widetilde{x}).$$

- The parameter $\alpha \in [0,1]$ describes the decision maker's degree of optimism.
- For $U = -D$, this is equivalent to minimizing the expression

$$D(\widetilde{x}) = -U(\widetilde{x}) = \alpha \cdot \underline{D}(\widetilde{x}) + (1-\alpha) \cdot \overline{D}(\widetilde{x}).$$

- In this paper, we describe an efficient algorithm for computing such $\widetilde{x}$.


Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

# 7. Analysis of the Problem

- Each term $(\widetilde{x}-x_i)^2$ in the sum $D(\widetilde{x}, x_1, \ldots, x_n)$ depends only on its own variable $x_i$. Thus, with respect to $x_i$:
  - the sum is the smallest when each term is min, and
  - the sum is the largest when each term is the largest.
- When $x_i \in [\underline{x}_i, \overline{x}_i]$, the max of $(\widetilde{x} - x_i)^2$ is attained:
  - at $x_i = \underline{x}_i$ when $\widetilde{x} \geq \widetilde{x}_i \stackrel{\text{def}}{=} \dfrac{\underline{x}_i + \overline{x}_i}{2}$ and
  - at $x_i = \overline{x}_i$ when $\widetilde{x} < \widetilde{x}_i$.
- Thus, $\overline{D}(\widetilde{x}) = \sum\limits_{i:\widetilde{x}<\widetilde{x}_i} (\widetilde{x} - \overline{x}_i)^2 + \sum\limits_{i:\widetilde{x}\geq\widetilde{x}_i} (\widetilde{x} - \underline{x}_i)^2.$
- Similarly, the minimum of the term $(\widetilde{x}-x_i)^2$ is attained:
  - for $x_i = \widetilde{x}$ when $\widetilde{x} \in [\underline{x}_i, \overline{x}_i]$ (in this case, min $= 0$);
  - for $x_i = \underline{x}_i$ when $\widetilde{x} < \underline{x}_i$; and
  - for $x_i = \overline{x}_i$ when $\widetilde{x} > \overline{x}_i$.


Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

## 8. Analysis of the Problem (cont-d)

- Thus, $\underline{D}(\widetilde{x}) = \sum\limits_{i:\widetilde{x}>\overline{x}_i} (\widetilde{x}-\overline{x}_i)^2 + \sum\limits_{i:\widetilde{x}<\underline{x}_i} (\widetilde{x}-\underline{x}_i)^2.$
- So, for $D(\widetilde{x}) = \alpha\cdot\underline{D}(\widetilde{x}) + (1-\alpha)\cdot\overline{D}(\widetilde{x})$, we get

$$D(\widetilde{x}) = \alpha\cdot\sum_{i:\widetilde{x}>\overline{x}_i}(\widetilde{x}-\overline{x}_i)^2 + \alpha\cdot\sum_{i:\widetilde{x}<\underline{x}_i}(\widetilde{x}-\underline{x}_i)^2 +$$

$$(1-\alpha)\cdot\sum_{i:\widetilde{x}<\widetilde{x}_i}(\widetilde{x}-\overline{x}_i)^2 + (1-\alpha)\cdot\sum_{i:\widetilde{x}\geq\widetilde{x}_i}(\widetilde{x}-\underline{x}_i)^2.$$

Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

# 9. Towards an Algorithm

- The terms depends on the relation between $\widetilde{x}$ and the values
$$\underline{x}_i, \overline{x}_i, \text{ and } \widetilde{x}_i.$$
- Let us sort these $3n$ values into a sequence
$$s_1 \le s_2 \le \ldots \le s_{3n}.$$
- Then on each interval $[s_j, s_{j+1}]$, the function $D(\widetilde{x})$ is simply a quadratic function of $\widetilde{x}$.
- A quadratic function attains min on an interval:
  - either at one of its midpoints,
  - or at a point when the derivative is equal to 0 (if this point is inside the given interval).


Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

## 10. Towards an Algorithm (cont-d)

- Equating the derivative $D(\widetilde{x})$ to 0, we get:

$$(\alpha\cdot\#\{i:\widetilde{x}<\underline{x}_i \text{ or } \widetilde{x}>\overline{x}_i\}+1-\alpha)\cdot\widetilde{x}=$$

$$\alpha\cdot\sum_{i:\widetilde{x}>\overline{x}_i}\overline{x}_i+\alpha\cdot\sum_{i:\widetilde{x}<\underline{x}_i}\underline{x}_i+(1-\alpha)\cdot\sum_{i:\widetilde{x}<\widetilde{x}_i}\overline{x}_i+(1-\alpha)\cdot\sum_{i:\widetilde{x}\geq\widetilde{x}_i}\underline{x}_i.$$

- $s_j$ is a listing of all thresholds values $\underline{x}_i$, $\overline{x}_i$, and $\widetilde{x}_i$.
- So, for $\widetilde{x}\in(s_j,s_{j+1})$, $\widetilde{x}<\underline{x}_i\Leftrightarrow s_{j+1}\leq\underline{x}_i$.
- Similarly, the inequality $\widetilde{x}>\underline{x}_i$ is equivalent to $s_j\geq\overline{x}_i$.
- In general, for $\widetilde{x}\in(s_j,s_{j+1})$, we get:

$$(\alpha\cdot\#\{i:\widetilde{x}<\underline{x}_i \text{ or } \widetilde{x}>\overline{x}_i\}+1-\alpha)\cdot\widetilde{x}=$$

$$\alpha\cdot\sum_{i:s_j\geq\overline{x}_i}\overline{x}_i+\alpha\cdot\sum_{i:s_{j+1}\leq\underline{x}_i}\underline{x}_i+(1-\alpha)\cdot\sum_{i:s_{j+1}\leq\widetilde{x}_i}\overline{x}_i+(1-\alpha)\cdot\sum_{i:s_j\geq\widetilde{x}_i}\underline{x}_i.$$

- We can thus find $\widetilde{x}$ at which the derivative is 0.
- Thus, we arrive at the following algorithm.


Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

## 11. Resulting Algorithm

- We want to minimize the expression

$$D(\widetilde{x}) = -U(\widetilde{x}) = \alpha \cdot \underline{D}(\widetilde{x}) + (1-\alpha)\cdot \overline{D}(\widetilde{x}), \text{ where}$$

$$D(\widetilde{x}) = \alpha \cdot \sum_{i:\widetilde{x}>\overline{x}_i} (\widetilde{x}-\overline{x}_i)^2 + \alpha \cdot \sum_{i:\widetilde{x}<\underline{x}_i} (\widetilde{x}-\underline{x}_i)^2 +$$

$$(1-\alpha)\cdot \sum_{i:\widetilde{x}<\widetilde{x}_i} (\widetilde{x}-\overline{x}_i)^2 + (1-\alpha)\cdot \sum_{i:\widetilde{x}\ge\widetilde{x}_i} (\widetilde{x}-\underline{x}_i)^2.$$

- First, for each interval $[\underline{x}_i, \overline{x}_i]$, we compute its midpoint

$$\widetilde{x}_i = \frac{\underline{x}_i + \overline{x}_i}{2}.$$

- Then, we sort the $3n$ values $\underline{x}_i$, $\overline{x}_i$, and $\widetilde{x}_i$ into an increasing sequence $s_1 \le s_2 \le \ldots \le s_{3n}$.
- To cover the whole real line, to these values, we add $s_0 = -\infty$ and $s_{3n+1} = +\infty$.

Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

## 12. Algorithm (cont-d)

- We compute the value of the objective function on each of the endpoints $s_1, \ldots, s_{3n}$.
- Then, for each interval $(s_i, s_{j+1})$, we compute $\widetilde{x}$ as:

$$\frac{\alpha \sum\limits_{i:s_j\ge\overline{x}_i} \overline{x}_i + \alpha \sum\limits_{i:s_{j+1}\le\underline{x}_i} \underline{x}_i + (1-\alpha) \sum\limits_{i:s_{j+1}\le\widetilde{x}_i} \overline{x}_i + (1-\alpha) \sum\limits_{i:s_j\ge\widetilde{x}_i} \underline{x}_i}{\alpha\cdot \#\{i : \widetilde{x} < \underline{x}_i \text{ or } \widetilde{x} > \overline{x}_i\} + 1 - \alpha}.$$

- If $\widetilde{x}$ is within $(s_i, s_{j+1})$, we compute $D(\widetilde{x})$.
- After that:
  - out of all the values $\widetilde{x}$ for which we computed $D(\widetilde{x})$,
  - we return $\widetilde{x}$ for which $D(\widetilde{x})$ is the smallest.


Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

# 13. What Is the Computational Complexity of This Algorithm

- Sorting $3n = O(n)$ values $\underline{x}_i$, $\overline{x}_i$, and $\widetilde{x}_i$ takes time

$$O(n \cdot \ln(n)).$$

- Computing each value $D(\widetilde{x})$ of the objective function requires $O(n)$ computational steps.
- We compute $D(\widetilde{x})$:
  - for $3n$ endpoints and
  - for $\le 3n + 1$ values at which the derivative is 0 at each of the intervals $(s_j, s_{j+1})$.
- Overall, we compute $D(\widetilde{x})$ at $O(n)$ values.
- Thus, overall, we need $O(n \cdot \ln(n)) + O(n) \cdot O(n) = O(n^2)$ computation steps.
- Hence, our algorithm runs in quadratic time.


Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments

# 14. Acknowledgments


- This work was supported in part by the National Science Foundation grants:
    - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
    - DUE-0926721, and
- by an award from Prudential Foundation.



Need to Combine . . .
Utility-Based Approach
Case of Interval . . .
Analysis of the Problem
Towards an Algorithm
Resulting Algorithm
What Is the . . .
Acknowledgments