Bounded Rationality in Decision Making Under Uncertainty: Towards Optimal Granularity

Joe Lorkowski

Department of Computer Science
University of Texas at El Paso
El Paso, Texas 79968, USA
lorkowski@computer.org

Overview

- Starting with Kahneman and Tversky, researchers found many examples when decision making seems irrational.
- In this dissertation, we show that:
  - this seemingly irrational decision making can be explained
  - if we take into account that human abilities to process information are limited.
- As a result of these limited abilities:
  - instead of the exact *values* of different quantities,
  - we operate with *granules* that contain these values.
- On several examples, we show that:
  - optimization under such granularity restriction
  - indeed leads to observed human decision making.
- Thus, granularity helps explain seemingly irrational human decision making.

Bad Decisions vs. Irrational Decisions

- Most economic models are based on the assumption that a rational person maximizes his/her “utility”.
- Some weird behaviors can be still explained this way – just utility is weird.
- For a drug addict, the utility of getting high is so large that it overwhelms any negative consequences.
- However, sometimes, people exhibit behavior which cannot be explained as maximizing utility.

Simple Example of Irrational Decision Making

- A customer shopping for an item has several choices a_i:
  - some of these choices have better quality $a_i \succ a_j$,
  - but are more expensive.
- When presented with three alternatives $a_1 \succ a_2 \succ a_3$, in most cases, most customers select a middle one a_2.
- This means that a_2 is better than a_3.
- However, when presented with $a_2 \succ a_3 \succ a_4$, the same customer selects a_3.
- This means that to him, a_3 is better than a_2 – a clear inconsistency.
- We show that granularity explains this behavior.

Part 0: Traditional Decision Theory

Traditional Decision Theory: Reminder

- *Main assumption* – for any two alternatives A and A':
  - either A is better (we will denote it $A' \prec A$),
  - or A' is better (we will denote it $A \prec A'$),
  - or A and A' are of equal value (denoted $A \sim A'$).
- *Resulting scale* for describing the quality of different alternatives A:
  - to define a scale, we select a very bad alternative A_0 and a very good alternative A_1;
  - for each $p \in [0, 1]$, we can form a lottery $L(p)$ in which we get A_1 with probability p and A_0 with probability $1 - p$;
  - for each reasonable alternative A, we have $A_0 = L(0) \prec A \prec L(1) = A_1$;
  - thus, for some p_0, we switch from $L(p) \prec A$ for $p < p_0$ to $L(p) \succ A$ for $p > p_0$, i.e., there exists a "switch" value $u(A)$ for which $L(u(A)) \equiv A$;
  - this value $u(A)$ is called the *utility* of the alternative A.

Utility Scale

- We have a lottery $L(p)$ for every probability $p \in [0, 1]$:
  - $p = 0$ corresponds to A_0, i.e., $L(0) = A_0$;
  - $p = 1$ corresponds to A_1, i.e., $L(1) = A_1$;
  - $0 < p < 1$ corresponds to $A_0 \prec L(p) \prec A_1$;
  - $p < p'$ implies $L(p) \prec L(p')$.
- There is a continuous monotonic scale of alternatives:

$$L(0) = A_0 \prec \ldots \prec L(p) \prec \ldots \prec L(p') \prec \ldots \prec L(1) = A_1.$$

- This *utility scale* is used to gauge the attractiveness of each alternative.

How to Elicit the Utility Value: Bisection

- We know that $A \equiv L(u(A))$ for some $u(A) \in [0, 1]$.
- Suppose that we want to find $u(A)$ with accuracy 2^{-k}.
- We start with $[\underline{u}, \overline{u}] = [0, 1]$. Then, for $i = 1$ to k, we:
  - compute the midpoint $u_{\rm mid}$ of $[\underline{u}, \overline{u}]$
  - ask the expert to compare A with the lottery $L(u_{\rm mid})$
  - if $A \preceq L(u_{\rm mid})$, then $u(A) \le u_{\rm mid}$, so we can take
  $$[\underline{u}, \overline{u}] = [\underline{u}, u_{\rm mid}];$$
  - if $A \succeq L(u_{\rm mid})$, then $u(A) \ge u_{\rm mid}$, so we can take
  $$[\underline{u}, \overline{u}] = [u_{\rm mid}, \underline{u}].$$
- At each iteration, the width of $[\underline{u}, \overline{u}]$ decreases by half.
- After k iterations, we get an interval $[\underline{u}, \overline{u}]$ of width 2^{-k} that contains $u(A)$.
- So, we get $u(A)$ with accuracy 2^{-k}.

Utility Theory and Human Decision Making

- Decision based on utility values
  - Which of the utilities $u(A')$, $u(A'')$, $\dots$, of the alternatives A', A'', $\dots$ should we choose?
  - By definition of utility, A' is preferable to A'' if and only if $u(A') > u(A'')$.
  - We should always select an alternative with the largest possible value of utility.
  - So, to find the best solution, we must solve the corresponding optimization problem.
- Our claim is that when people make *definite* and *consistent* choices, these choices *can* be described by probabilities.
  - We are *not* claiming that people always make rational decisions.
  - We are *not* claiming that people estimate probabilities when they make rational decisions.

Estimating the Utility of an Action a

- We know possible outcome situations $S_1, \ldots, S_n$.
- We often know the probabilities $p_i = p(S_i)$.
- Each situation S_i is equivalent to the lottery $L(u(S_i))$ in which we get:
  - A_1 with probability $u(S_i)$ and
  - A_0 with probability $1 - u(S_i)$.
- So, a is equivalent to a complex lottery in which:
  - we select one of the situations S_i with prob. $p_i = P(S_i)$;
  - depending on S_i, we get A_1 with prob. $P(A_1 | S_i) = u(S_i)$.
- The probability of getting A_1 is

$$P(A_1) = \sum_{i=1}^{n} P(A_1 | S_i) \cdot P(S_i), \text{ i.e., } \; u(a) = \sum_{i=1}^{n} u(S_i) \cdot p_i.$$

- The sum defining $u(a)$ is the expected value of the outcome's utility.
- So, we should select the action with the largest value of expected utility $u(a) = \sum p_i \cdot u(S_i)$.

Subjective Probabilities

- Sometimes, we do not know the probabilities p_i of different outcomes.
- In this case, we can gauge the *subjective* impressions about the probabilities.
- Let's fix a prize (e.g., \$1). For each event E, we compare:
  - a lottery ℓ_E in which we get the fixed prize if the event E occurs and 0 is it does not occur, with
  - a lottery $\ell(p)$ in which we get the same amount with probability p.
- Here, $\ell(0) \prec \ell_E \prec \ell(1)$; so for some p_0, we switch from $\ell(p) \prec \ell_E$ to $\ell_E \prec \ell(p)$.
- This threshold value $ps(E)$ is called the *subjective probability* of the event E: $\ell_E \equiv \ell(ps(E))$.
- The utility of an action a with possible outcomes $S_1, \ldots, S_n$ is thus equal to $u(a) = \sum\limits_{i=1}^{n} ps(E_i) \cdot u(S_i)$.

Traditional Approach Summarized

- We assume that
  - we know possible actions, and
  - we know the exact consequences of each action.
- Then, we should select an action with the largest value of expected utility.

Part 1: First Example of Seemingly Irrational Decision Making – Compromise Effect

Compromise Effect: Reminder

- A customer shopping for an item has choices: some cheaper, some more expensive but of higher quality.
- Examples: shopping for a camera, for a hotel room.
- Researchers asked the customers to select one of the three randomly selected alternatives.
- They expected all three to be selected with equal probability.
- Instead, in the overwhelming majority of cases, customers selected the intermediate alternative.
- The intermediate alternative provides a compromise between the quality and cost.
- So, this phenomenon was named *compromise effect*.

Why This Is Irrational?

- Selecting the middle alternative seems reasonable.
- But let's consider alternatives $a_1 < a_2 < a_3 < a_4$ sorted by price (and quality).
- If we present the user with three choices $a_1 < a_2 < a_3$, the user will select the middle choice a_2.
- This means that, to the user, a_2 is better than a_3.
- But if we present the user with three other choices $a_2 < a_3 < a_4$, the same user will select a_3.
- So, to the user, the alternative a_3 is better than a_2.
- If in a pair-wise comparison, a_3 is better, then the first choice is wrong, else the second choice is wrong.
- In both cases, one of the two choices is irrational.

This is Not Just an Experimental Curiosity, Customers' Have Been Manipulated This Way

- At first glance, this seems like an optical illusion or a logical paradox: interesting but not very important.
- Actually, it is important: customers have been manipulated into buying a more expensive product.
- If there are two types of a product, a company adds an even more expensive third option.
- Recent research shows the compromise effect only happens when a customer has no additional information.
- In situations when customers were given access to additional information, their selections were consistent.
- However, in situation when decisions need to be made under major uncertainty, this effect is clearly present.
- How to explain such a seemingly irrational behavior?

Symmetry Approach: Main Idea

- Main idea:
  - if the situation is invariant with respect to some natural symmetries,
  - then it is reasonable to select an action which is also invariant with respect to all these symmetries.
- This approach has indeed been helpful in dealing with uncertainty. In particular, it explains:
  - the use of a sigmoid activation function $s(z) = \dfrac{1}{1 + \exp(-z)}$ in neural networks,
  - the use of the most efficient t-norms and t-conorms in fuzzy logic,
  - etc.

What Do We Know About the Utility of Each Alternative?

- The utility of each alternatives comes from two factors:
  - the first factor u_1 comes from the quality: the higher the quality, the better – i.e., the larger u_1;
  - the second factor u_2 comes from price: the lower the price, the better – i.e., the larger u_2.
- We have alternatives $a < a' < a''$ characterized by pairs $u(a) = (u_1, u_2)$, $u(a') = (u'_1, u'_2)$, and $u(a'') = (u''_1, u''_2)$.
- We do not know the values of these factors, we only know that
$$u_1 < u'_1 < u''_1 \text{ and } u''_2 < u'_2 < u_2.$$
- Since we only know the order, we can mark the values u_i as L (Low), M (Medium), and H (High).
- Then $u(a) = (L, H)$, $u(a') = (M, M)$, $u(a'') = (H, L)$.

Natural Transformations and Symmetries

- We do not know a priori which of the utility components is more important.
- It is thus reasonable to treat both components equally.
- So, swapping the two components is a reasonable transformation:
  - if we are selecting an alternative based on the pairs

$$u(a) = (L, H), \;\; u(a') = (M, M), \text{ and } u(a'') = (H, L),$$

  - then we should select the exact same alternative based on the "swapped" pairs

$$u(a) = (H, L), \;\; u(a') = (M, M), \text{ and } u(a'') = (L, H).$$

Transformations and Symmetries (cont-d)

- Similarly, there is no reason to a priori prefer one alternative versus the other.
- So, any permutation of the three alternatives is a reasonable transformation.
- We start with

$$u(a) = (L, H), \quad u(a') = (M, M), \quad u(a'') = (H, L).$$

- If we rename a and a'', we get

$$u(a) = (H, L), \quad u(a') = (M, M), \quad u(a'') = (L, H).$$

- For example:
  - if we originally select an alternative a with

$$u(a) = (L, H),$$

  - then, after the swap, we should select the same alternative – which is now denoted by a''.

What Can We Conclude From These Symmetries

- We start with

$$u(a) = (L, H), \quad u(a') = (M, M), \quad u(a'') = (H, L).$$

- If we swap u_1 and u_2, we get

$$u(a) = (H, L), \quad u(a') = (M, M), \quad u(a'') = (L, H).$$

- Now, if we also rename a and a'', we get

$$u(a) = (L, H), \quad u(a') = (M, M), \quad u(a'') = (H, L).$$

- These are the same utility values with which we started.
- So, if originally, we select a with $u(a) = (L, H)$, in the new arrangements we should also select a.
- But the new a is the old a''.
- So, if we selected a, we should select a'' – a contradiction.

What Can We Conclude (cont-d)

- We start with

$$u(a) = (L, H), \;\; u(a') = (M, M), \;\; u(a'') = (H, L).$$

- If we swap u_1 and u_2, we get

$$u(a) = (H, L), \;\; u(a') = (M, M), \;\; u(a'') = (L, H).$$

- Now, if we also rename a and a'', we get

$$u(a) = (L, H), \;\; u(a') = (M, M), \;\; u(a'') = (H, L).$$

- These are the same utility values with which we started.
- So, if originally, we select a'' with $u(a'') = (H, L)$, in the new arrangements we should also select a.
- But the new a'' is the old a.
- So, if we selected a'', we should select a – a contradiction.

First Example: Summarizing

- We start with

$$u(a) = (L, H), \;\; u(a') = (M, M), \;\; u(a'') = (H, L).$$

- If we swap u_1 and u_2, we get

$$u(a) = (H, L), \;\; u(a') = (M, M), \;\; u(a'') = (L, H).$$

- Now, if we also rename a and a'', we get

$$u(a) = (L, H), \;\; u(a') = (M, M), \;\; u(a'') = (H, L).$$

- We cannot select a – this leads to a contradiction.
- We cannot select a'' – this leads to a contradiction.
- The only consistent choice is to select a'.
- This is exactly the compromise effect.

First Example: Conclusion

- Experiments show that:
  - when people are presented with three choices $a < a' < a''$ of increasing price and increasing quality,
  - and they do not have detailed information about these choices,
  - then in the overwhelming majority of cases, they select the intermediate alternative a'.
- This "compromise effect" is, at first glance, irrational:
  - selecting a' means that, to the user, a' is better than a'', but
  - in a situation when the user is presented with $a' < a'' < a'''$, the user prefers a'' to a'.
- We show that a natural symmetry approach explains this seemingly irrational behavior.

Part 2: Second Example of Seemingly Irrational Decision Making – Biased Probability Estimates

Second Example of Irrational Decision Making: Biased Probability Estimates

- We know an action a may have different outcomes u_i with different probabilities $p_i(a)$.
- By repeating a situation many times, the average expected gain becomes close to the mathematical expected gain:

$$u(a) \stackrel{\text{def}}{=} \sum_{i=1}^{n} p_i(a) \cdot u_i.$$

- We expect a decision maker to select action a for which this expected value $u(a)$ is greatest.
- This is close, but not exactly, what an actual person does.

Kahneman and Tversky's Decision Weights

- Kahneman and Tversky found a more accurate description is obtained by:
  - an assumption of maximization of a *weighted gain* where
  - the weights are determined by the corresponding probabilities.
- In other words, people select the action a with the largest weighted gain

$$w(a) \stackrel{\text{def}}{=} \sum_i w_i(a) \cdot u_i.$$

- Here, $w_i(a) = f(p_i(a))$ for an appropriate function $f(x)$.

Decision Weights: Empirical Results

- Empirical decision weights:

probability	0	1	2	5	10	20	50
weight	0	5.5	8.1	13.2	18.6	26.1	42.1

probability	80	90	95	98	99	100
weight	60.1	71.2	79.3	87.1	91.2	100

- There exist *qualitative* explanations for this phenomenon.
- We propose a *quantitative* explanation based on the granularity idea.

Idea: “Distinguishable" Probabilities

- For decision making, most people do not estimate probabilities as numbers.
- Most people estimate probabilities with “fuzzy” concepts like *(low, medium, high).*
- The discretization converts a possibly infinite number of probabilities to a finite number of values.
- The discrete scale is formed by probabilities which are *distinguishable* from each other.
  - 10% chance of rain is distinguishable from a 50% chance of rain, but
  - 51% chance of rain is not distinguishable from a 50% chance of rain.

Distinguishable Probabilities: Formalization

- In general, if out of n observations, the event was observed in m of them, we estimate the probability as the ratio $\frac{m}{n}$.
- The expected value of the frequency is equal to p, and that the standard deviation of this frequency is equal to $\sigma = \sqrt{\frac{p \cdot (1-p)}{n}}$.
- By the Central Limit Theorem, for large n, the distribution of frequency is very close to the normal distribution.
- For normal distribution, all values are within 2–3 standard deviations of the mean, i.e. within the interval $(p - k_0 \cdot \sigma, p + k_0 \cdot \sigma)$.
- So, two probabilities p and p' are distinguishable if the corresponding intervals do not intersect:

$$(p - k_0 \cdot \sigma, p + k_0 \cdot \sigma) \cap (p' - k_0 \cdot \sigma', p' + k_0 \cdot \sigma') = \emptyset$$

- The smallest difference $p' - p$ is when $p + k_0 \cdot \sigma = p' - k_0 \cdot \sigma'$.

Formalization (cont-d)

- When n is large, p and p' are close to each other and $\sigma' \approx \sigma$.
- Substituting σ for σ' into the above equality, we conclude

$$p' \approx p + 2k_0 \cdot \sigma = p + 2k_0 \cdot \sqrt{\frac{p \cdot (1-p)}{n}}.$$

- So, we have distinguishable probabilities $p_1 < p_2 < \ldots < p_m$, where $p_{i+1} \approx p_i + 2k_0 \cdot \sqrt{\frac{p_i \cdot (1-p_i)}{n}}$.
- We need to select a weight (subjective probability) based only on the level i.
- When we have m levels, we thus assign m probabilities $w_1 < \ldots < w_m$.
- All we know is that $w_1 < \ldots < w_m$.
- There are many possible tuples with this property.
- We have no reason to assume that some tuples are more probable than others.

Analysis (cont-d)

- It is thus reasonable to assume that all these tuples are equally probable.
- Due to the formulas for complete probability, the resulting probability w_i is the average of values w_i corresponding to all the tuples: $E[w_i \mid 0 < w_1 < \ldots < w_m = 1]$.
- These averages are known: $w_i = \dfrac{i}{m}$.
- So, to probability p_i, we assign weight $g(p_i) = \dfrac{i}{m}$.
- For $p_{i+1} \approx p_i + 2k_0 \cdot \sqrt{\dfrac{p \cdot (1-p)}{n}}$, we have

$$g(p_i) = \frac{i}{m} \text{ and } g(p_{i+1}) = \frac{i+1}{m}.$$

Analysis (cont-d)

- Since $p = p_i$ and $p' = p_{i+1}$ are close, $p' - p$ is small:
  - we can expand $g(p') = g(p + (p' - p))$ in Taylor series and keep only linear terms
  - $g(p') \approx g(p) + (p' - p) \cdot g'(p)$, where $g'(p) = \dfrac{dg}{dp}$ denotes the derivative of the function $g(p)$.
  - Thus, $g(p') - g(p) = \dfrac{1}{m} = (p' - p) \cdot g'(p)$.
- Substituting the expression for $p' - p$ into this formula, we conclude

$$\frac{1}{m} = 2k_0 \cdot \sqrt{\frac{p \cdot (1 - p)}{n}} \cdot g'(p).$$

- This can be rewritten as $g'(p) \cdot \sqrt{p \cdot (1 - p)} = \text{const}$ for some constant.
- Thus, $g'(p) = \text{const} \cdot \frac{1}{\sqrt{p \cdot (1 - p)}}$ and, since $g(0) = 0$ and $g(1) = 1$, we get $g(p) = \dfrac{2}{\pi} \cdot \arcsin\left(\sqrt{p}\right)$.

Assigning Weights to Probabilities: First Try

- For each probability $p_i \in [0, 1]$, assign the weight $w_i = g(p_i) = \frac{2}{\pi} \cdot \arcsin\left(\sqrt{p_i}\right)$
- Here is how these weights compare with Kahneman's empirical weights $\widetilde{w}_i$:

p_i	0	1	2	5	10	20	50
$\widetilde{w}_i$	0	5.5	8.1	13.2	18.6	26.1	42.1
$w_i = g(p_i)$	0	6.4	9.0	14.4	20.5	29.5	50.0

p_i	80	90	95	98	99	100
$\widetilde{w}_i$	60.1	71.2	79.3	87.1	91.2	100
$w_i = g(p_i)$	70.5	79.5	85.6	91.0	93.6	100

How to Get a Better Fit between Theoretical and Observed Weights

- All we observe is which action a person selects.
- Based on selection, we cannot uniquely determine weights.
- An empirical selection consistent with weights w_i is equally consistent with weights $w_i' = \lambda \cdot w_i$.
- First-try results were based on constraints that $g(0) = 0$ and $g(1) = 1$ which led to a perfect match at both ends and lousy match "on average."
- Instead, select λ using Least Squares such that $\sum_i \left(\dfrac{\lambda \cdot w_i - \widetilde{w}_i}{w_i} \right)^2$ is the smallest possible.
- Differentiating with respect to λ and equating to zero:

$$\sum_i \left(\lambda - \frac{\widetilde{w}_i}{w_i} \right) = 0, \text{ so } \lambda = \frac{1}{m} \cdot \sum_i \frac{\widetilde{w}_i}{w_i}.$$

Second Example: Result

- For the values being considered, $\lambda = 0.910$
- For $w'_i = \lambda \cdot w_i = \lambda \cdot g(p_i)$

$\widetilde{w}_i$	0	5.5	8.1	13.2	18.6	26.1	42.1
$w'_i = \lambda \cdot g(p_i)$	0	5.8	8.2	13.1	18.7	26.8	45.5
$w_i = g(p_i)$	0	6.4	9.0	14.4	20.5	29.5	50.0

$\widetilde{w}_i$	60.1	71.2	79.3	87.1	91.2	100
$w'_i = \lambda \cdot g(p_i)$	64.2	72.3	77.9	82.8	87.4	91.0
$w_i = g(p_i)$	70.5	79.5	85.6	91.0	93.6	100

- For most i, the difference between the granule-based weights w'_i and empirical weights $\widetilde{w}_i$ is small.
- *Conclusion:* Granularity explains Kahneman and Tversky's empirical decision weights.

Part 3: Third Example of Seemingly Irrational Decision Making – Use of Fuzzy Techniques

Third Example: Fuzzy Uncertainty

- Fuzzy logic formalizes imprecise properties P like "big" or "small" used in experts' statements.
- It uses the degree $\mu_P(x)$ to which x satisfies P:
  - $\mu_P(x) = 1$ means that we are confident that x satisfies P;
  - $\mu_P(x) = 0$ means that we are confident that x *does not* satisfy P;
  - $0 < \mu_P(x) < 1$ means that there is *some* confidence that x satisfies P, and some confidence that it doesn't.
- $\mu_P(x)$ is typically obtained by using a *Likert scale*:
  - the expert selects an integer m on a scale from 0 to n;
  - then we take $\mu_P(x) := m/n$;
- This way, we get values $\mu_P(x) = 0, 1/n, 2/n, \ldots, n/n = 1$.
- To get a more detailed description, we can use a larger n.

Fuzzy Techniques as an Example of Seemingly Irrational Behavior

- Fuzzy tools are effectively used to handle imprecise (fuzzy) expert knowledge in control and decision making.
- On the other hand, we know that rational decision makers should use the traditional utility-based techniques.
- To explain the empirical success of fuzzy techniques, we need to describe Likert scale selection in utility terms.

Likert Scale in Terms of Traditional Decision Making

- Suppose that we have a Likert scale with $n + 1$ labels 0, 1, 2, ..., n, ranging from the smallest to the largest.
  - We mark the smallest end of the scale with x_0 and begin to traverse.
  - As x increases, we find a value belonging to label 1 and mark this threshold point by x_1.
  - This continues to the largest end of the scale which is marked by x_{n+1}
- As a result, we divide the range $[\underline{X}, \overline{X}]$ of the original variable into $n + 1$ intervals $[x_0, x_1], \ldots, [x_n, x_{n+1}]$:
  - values from the first interval $[x_0, x_1]$ are marked with label 0;
  - ...
  - values from the $(n + 1)$-st interval $[x_n, x_{n+1}]$ are marked with label n.
- Then, decisions are based only on the label, i.e., only on the interval to which x belongs:

$$[x_0, x_1] \text{ or } [x_1, x_2] \text{ or } \ldots \text{ or } [x_n, x_{n+1}]$$

.

Which Decision To Choose?

- Ideally, we should make a decision based on the actual value of the corresponding quantity x.
- This sometimes requires too much computation, so instead of the actual value x we only use the label containing x.
- Since we only know the label k to which x belongs, we select $\widetilde{x}_k \in [x_k, x_{k+1}]$ and make a decision based on $\widetilde{x}_k$.
- Then, for all x from the interval $[x_k, x_{k+1}]$, we use the decision $d(\widetilde{x}_k)$ based on the value $\widetilde{x}_k$.
- We should select intervals $[x_k, x_{k+1}]$ and values $\widetilde{x}_k$ for which the expected utility is the largest.

Which Value $\widetilde{x}_k$ Should We Choose

- To find this expected utility, we need to know two things:
  - the probability of different values of x; described by the probability density function $\rho(x)$;
  - for each pair of values x' and x, the utility $u(x', x)$ of using a decision $d(x')$ when the actual value is x.
- In these terms, the expected utility of selecting a value $\widetilde{x}_k$ can be described as

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot u(\widetilde{x}_k, x)\, dx.$$

- For each interval $[x_k, x_{k+1}]$, we need to select a decision $d(\widetilde{x}_k)$ such that the above expression is maximized.
- Thus, the overall expected utility is equal to

$$\sum_{k=0}^{n} \max_{\tilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot u(\widetilde{x}_k, x)\, dx.$$

Equivalent Reformulation In Terms of Disutility

- In the ideal case, for each value x, we should use a decision $d(x)$, and gain utility $u(x, x)$.
- In practice, we have to use decisions $d(x')$, and thus, get slightly worse utility values $u(x', x)$.
- The corresponding decrease in utility $U(x', x) \stackrel{\text{def}}{=} u(x, x) - u(x', x)$ is usually called *disutility*.
- In terms of disutility, the function $u(x', x)$ has the form

$$u(x', x) = u(x, x) - U(x', x),$$

- So, to maximize utility, we select $x_1, \ldots, x_n$ for which the expected disutility attains its smallest possible value:

$$\sum_{k=0}^{n} \min_{\tilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\, dx \to \min .$$

Membership Function $\mu(x)$ as a Way to Describe Likert Scale

- As we have mentioned, fuzzy techniques use a membership function $\mu(x)$ to describe the Likert scale.
- In our n-valued Likert scale:
  - label 0 = $[x_0, x_1]$ corresponds to $\mu(x) = 0/n$,
  - label 1 = $[x_1, x_2]$ corresponds to $\mu(x) = 1/n$,
  - ...
  - label n = $[x_n, x_{n+1}]$ corresponds to $\mu(x) = n/n = 1$.
- The actual value $\mu(x)$ corresponds to the limit, when n is large, and the width of each interval is narrow.
- For large n, x' and x belong to the same narrow interval, and thus, the difference $\Delta x \stackrel{\text{def}}{=} x' - x$ is small.
- Let us use this fact to simplify the expression for disutility $U(x', x)$.

Using the Fact that Each Interval Is Narrow

- Thus, we can expand $U(x+\Delta x, x)$ into Taylor series in Δx, and keep only the first non-zero term in this expansion.

$$U(x+\Delta x, x) = U_0(x) + U_1(x)\cdot\Delta x + U_2(x)\cdot\Delta x^2 + \ldots,$$

- By definition of disutility, $U_0(x) = U(x,x) = u(x,x) - u(x,x) = 0$
- Simularly, since disutility is smallest when $x+\Delta x = x$, the first derivative is also zero.
- So, the first nontrivial term is $U_2(x)\cdot\Delta x^2 \approx U_2(x)\cdot(\widetilde{x}_k - x)^2$
- Thus, we need to minimize the expression

$$\sum_{k=0}^{n} \min_{\widetilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x)\cdot U_2(x)\cdot(\widetilde{x}_k - x)^2\,dx.$$

Resulting Formula

- Minimizing the above expression, we conclude that the membership function $\mu(x)$ corresponding to the optimal Likert scale is equal to

$$\mu(x) = \frac{\int_{\underline{X}}^{x} (\rho(t) \cdot U_2(t))^{1/3}\,dt}{\int_{\underline{X}}^{\overline{X}} (\rho(t) \cdot U_2(t))^{1/3}\,dt}, \text{ where:}$$

  where
  - $\rho(x)$ is the probability density describing the probabilities of different values of x,
  - $U_2(x) \stackrel{\text{def}}{=} \frac{1}{2} \cdot \frac{\partial^2 U(x+\Delta x, x)}{\partial^2(\Delta x)}$,
  - $U(x', x) \stackrel{\text{def}}{=} u(x, x) - u(x', x)$, and
  - $u(x', x)$ is the utility of using a decision $d(x')$ corresponding to the value x' in the situation in which the actual value is x.

Resulting Formula (cont-d)

- *Comment:*
  - The resulting formula only applies to properties like "large" whose values monotonically increase with x.
  - We can use a similar formula for properties like "small" which decrease with x.
  - For "approximately 0," we separately apply these formulas to both increasing and decreasing parts.
- The resulting membership degrees incorporate both probability and utility information.
- *This explains why fuzzy techniques often work better than probabilistic techniques without utility information*.

Additional Result: Why in Practice, Triangular Membership Functions are Often Used

- We have considered a situation in which we have full information about $\rho(x)$ and $U_2(x)$.
- In practice, we often do not know how $\rho(x)$ and $U_2(x)$ change with x.
- Since we have no reason to expect some values $\rho(x)$ to be larger or smaller, it is natural to assume that $\rho(x) = \text{const}$ and $U_2(x) = \text{const}$.
- In this case, our formula leads to the linear membership function, going either from 0 to 1 or from 1 to 0.
- This may explain why *triangular membership functions* – formed by two such linear segments – are often successfully used.

Part 4: Applications

Towards Applications

- Most of the above results deal with theoretical foundations of decision making under uncertainty.
- In the dissertation, we supplement this theoretical work with examples of practical applications:
  - in business,
  - in engineering,
  - in education, and
  - in developing generic AI decision tools.
- In *engineering*, we analyzed how quality design improves with the increased computational efficiency.
- This analysis is performed on the example of the ever increasing fuel efficiency of commercial aircraft.

Applications (cont-d)

- In *business*, we analyzed how the economic notion of a fair price can be translated into algorithms for decision making under interval and fuzzy uncertainty.
- In *education*, we explain the semi-heuristic Rasch model for predicting student success.
- In general *AI* applications, we analyze of how to explain:
  - the current heuristic approach
  - to selecting a proper level of granularity.
- Our example is selecting the basic concept level in concept analysis.

Computational Aspects

- One of the most fundamental types of uncertainty is interval uncertainty.
- In interval uncertainty, the general problem of propagating this uncertainty is NP-hard.
- However, there are cases when feasible algorithms are possible.
- Example: single-use expressions (SUE), when each variable occurs only once in the expression.
- In our work, we show that for double-use expressions, the problem is NP-hard.
- We have also developed a feasible algorithm for checking when an expression can be converted into SUE.

Acknowledgments

- My sincere appreciation to the members of my committee: Vladik Kreinovich, Luc Longpré, and Scott A. Starks.
- I also wish to thank:
  - Martine Ceberio and Pat Teller for advice and encouragement,
  - Olga Kosheleva and Christopher Kiekintveld for valuable discussions in decision theory,
  - Olac Fuentes for his guidance, and
  - all Computer Science Department faculty and staff for their hard work and dedication.
- Finally, I wish to thank my wife, Blanca, for all her help and love.

Appendix 1: Applications

Appendix 1.1
Application to Engineering

How Design Quality Improves with Increasing Computational Abilities: General Formulas and Case Study of Aircraft Fuel Efficiency

Outline

- It is known that the problems of optimal design are NP-hard.
- This means that, in general, a feasible algorithm can only produce close-to-optimal designs.
- The more computations we perform, the better design we can produce.
- In this paper, we theoretically derive the dependence of design quality on computation time.
- We then empirically confirm this dependence on the example of aircraft fuel efficiency.

Formulation of the Problem

- Since 1980s, computer-aided design (CAD) has become ubiquitous in engineering; example: Boeing 777.
- The main objective of CAD is to find a design which optimizes the corresponding objective function.
- Example: we optimize fuel efficiency of an aircraft.
- The corresponding optimization problems are non-linear, and such problems are, in general, NP-hard.
- So – unless P = NP – a feasible algorithm cannot always find the exact optimum, only an approximate one.
- The more computations we perform, the better the design.
- It is desirable to quantitatively describe how increasing computational abilities improve the design quality.

Because of NP-Hardness, More Computations Simply Means More Test Cases

- In principle, each design optimization problem can be solved by exhaustive search.
- Let d denote the number of parameters.
- Let C denote the average number of possible values of a parameter.
- Then, we need to analyze C^d test cases.
- For large systems (e.g., for an aircraft), we can only test *some* combinations.
- NP-hardness means that optimization algorithms to be significantly faster than exponential time C^d.
- This means that, in effect, all possible optimization algorithms boil down to trying many possible test cases.

Enter Randomness

- Increasing computational abilities mean that we can test more cases.
- Thus, by increasing the scope of our search, we will hopefully find a better design.
- Since we cannot do significantly better than with a simple search,
  - we cannot meaningfully predict whether the next test case will be better or worse,
  - because if we could, we would be able to significantly decrease the search time.
- The quality of the next test case cannot be predicted and is, in this sense, a random variable.

Which Random Variable?

- Many different factors affect the quality of each individual design.
- Usually, the distribution of the resulting effect of several independent random factors is close to Gaussian.
- This fact is known as the *Central Limit Theorem*.
- Thus, the quality of a (randomly selected) individual design is normally distributed, with some μ and σ.
- After we test n designs, the quality of the best-so-far design is $x = \max(x_1, \ldots, x_n)$.
- We can reduce the case of y_i with $\mu = 0$ and $\sigma = 1$: namely, $x_i = \mu + \sigma \cdot y_i$ hence $x = \mu + \sigma \cdot y$, where

$$y \stackrel{\text{def}}{=} \max(y_1, \ldots, y_n).$$

Let Us Use Max-Central Limit Theorem

- For large n, y's cdf is $F(y) \approx F_{EV}\left(\frac{y-\mu_n}{\sigma_n}\right)$, where:
  - $F_{EV}(y) \stackrel{\text{def}}{=} \exp(-\exp(-y))$ (*Gumbel distribution*),
  - $\mu_n \stackrel{\text{def}}{=} \Phi^{-1}\left(1-\frac{1}{n}\right)$, where $\Phi(y)$ is cdf of $N(0,1)$,
  - $\sigma_n \stackrel{\text{def}}{=} \Phi^{-1}\left(1-\frac{1}{n}\cdot e^{-1}\right) - \Phi^{-1}\left(1-\frac{1}{n}\right)$.
- Thus, $y = \mu_n + \sigma_n \cdot \xi$, where ξ is distributed according to the Gumbel distribution.
- The mean of ξ is the Euler's constant $\gamma \approx 0.5772$.
- Thus, the mean value m_n of y is equal to $\mu_n + \gamma \cdot \sigma_n$.
- For large n, we get asymptotically $m_n \sim \gamma \cdot \sqrt{2\ln(n)}$.
- Hence the mean value e_n of $x = \mu + \sigma \cdot y$ is asymptotically equal to $e_n \sim \mu + \sigma \cdot \gamma \cdot \sqrt{2\ln(n)}$.

Resulting Formula: Let Us Test It

- *Situation:* we test n different cases to find the optimal design.
- *Conclusion:* the quality e_n of the resulting design increases with n as
$$e_n \sim \mu + \sigma \cdot \gamma \cdot \sqrt{2\ln(n)}.$$
- *We test this formula:* on the example of the average fuel efficiency E of commercial aircraft.
- *Empirical fact:* E changes with time T as
$$E = \exp(a + b \cdot \ln(T)) = C \cdot T^b, \text{ for } b \approx 0.5.$$
- *Question:* can our formula $e_n \sim \mu + \sigma \cdot \gamma \cdot \sqrt{2\ln(n)}$ explain this empirical dependence?

How to Apply Our Theoretical Formula to This Case?

- The formula $q \sim \mu + \sigma \cdot \gamma \cdot \sqrt{2\ln(n)}$ describes how the quality changes with the # of computational steps n.
- In the case study, we know how it changes with time T.
- According to *Moore's law*, the computational speed grows exponentially with time T: $n \approx \exp(c \cdot T)$.
- Crudely speaking, the computational speed doubles every two years.
- When $n \approx \exp(c \cdot T)$, we have $\ln(n) \sim T$; thus,

$$q \approx a + b \cdot \sqrt{T}.$$

- This is exactly the empirical dependence that we actually observe.

Caution

- *Idea:* cars also improve their fuel efficiency.
- *Fact:* the dependence of their fuel efficiency on time is piece-wise constant.
- *Explanation:* for cars, changes are driven mostly by federal and state regulations.
- *Result:* these changes have little to do with efficiency of Computer-Aided design.

Appendix 1.2
Application to Business

Towards Decision Making under Interval, Set-Valued, Fuzzy, and Z-Number Uncertainty: A Fair Price Approach

Need for Decision Making

- In many practical situations:
  - we have several alternatives, and
  - we need to select one of these alternatives.
- *Examples:*
  - a person saving for retirement needs to find the best way to invest money;
  - a company needs to select a location for its new plant;
  - a designer must select one of several possible designs for a new airplane;
  - a medical doctor needs to select a treatment for a patient.

Need for Decision Making Under Uncertainty

- Decision making is easier if we know the exact consequences of each alternative selection.
- Often, however:
  - we only have an incomplete information about consequences of different alternative, and
  - we need to select an alternative under this uncertainty.

How Decisions Under Uncertainty Are Made Now

- Traditional decision making assumes that:
  - for each alternative a,
  - we know the probability $p_i(a)$ of different outcomes i.
- It can be proven that:
  - preferences of a rational decision maker can be described by *utilities* u_i so that
  - an alternative a is better if its expected utility $\overline{u}(a) \stackrel{\text{def}}{=} \sum_i p_i(a) \cdot u_i$ is larger.

Hurwicz Optimism-Pessimism Criterion

- Often, we do not know these probabilities p_i.
- For example, sometimes:
  - we only know the range $[\underline{u}, \overline{u}]$ of possible utility values, but
  - we do not know the probability of different values within this range.
- It has been shown that in this case, we should select an alternative s.t. $\alpha_H \cdot \overline{u} + (1 - \alpha_H) \cdot \underline{u} \to \max$.
- Here, $\alpha_H \in [0, 1]$ described the optimism level of a decision maker:
  - $\alpha_H = 1$ means optimism;
  - $\alpha_H = 0$ means pessimism;
  - $0 < \alpha_H < 1$ combines optimism and pessimism.

What If We Have Fuzzy Uncertainty? Z-Number Uncertainty?

- There are many semi-heuristic methods of decision making under fuzzy uncertainty.
- These methods have led to many practical applications.
- However, often, different methods lead to different results.
- R. Aliev proposed a utility-based approach to decision making under fuzzy and Z-number uncertainty.
- However, there still are many practical problems when it is not fully clear how to make a decision.
- In this talk, we provide foundations for the new methodology of decision making under uncertainty.
- This methodology which is based on a natural idea of a *fair price*.

Fair Price Approach: An Idea

- When we have a full information about an object, then:
  - we can express our desirability of each possible situation
  - by declaring a price that we are willing to pay to get involved in this situation.
- Once these prices are set, we simply select the alternative for which the participation price is the highest.
- In decision making under uncertainty, it is not easy to come up with a fair price.
- A natural idea is to develop techniques for producing such fair prices.
- These prices can then be used in decision making, to select an appropriate alternative.

Case of Interval Uncertainty

- *Ideal case:* we know the exact gain u of selecting an alternative.
- *A more realistic case:* we only know the lower bound $\underline{u}$ and the upper bound $\overline{u}$ on this gain.
- *Comment:* we do not know which values $u \in [\underline{u}, \overline{u}]$ are more probable or less probable.
- This situation is known as *interval uncertainty*.
- We want to assign, to each interval $[\underline{u}, \overline{u}]$, a number $P([\underline{u}, \overline{u}])$ describing the fair price of this interval.
- Since we know that $u \le \overline{u}$, we have $P([\underline{u}, \overline{u}]) \le \overline{u}$.
- Since we know that $\underline{u}$, we have $\underline{u} \le P([\underline{u}, \overline{u}])$.

Case of Interval Uncertainty: Monotonicity

- *Case 1:* we keep the lower endpoint $\underline{u}$ intact but increase the upper bound.
- This means that we:
  - keeping all the previous possibilities, but
  - we allow new possibilities, with a higher gain.
- In this case, it is reasonable to require that the corresponding price not decrease:

$$\text{if } \underline{u} = \underline{v} \text{ and } \overline{u} < \overline{v} \text{ then } P([\underline{u}, \overline{u}]) \le P([\underline{v}, \overline{v}]).$$

- *Case 2:* we dismiss some low-gain alternatives.
- This should increase (or at least not decrease) the fair price:

$$\text{if } \underline{u} < \underline{v} \text{ and } \overline{u} = \overline{v} \text{ then } P([\underline{u}, \overline{u}]) \le P([\underline{v}, \overline{v}]).$$

Additivity: Idea

- Let us consider the situation when we have two consequent independent decisions.
- We can consider two decision processes separately.
- We can also consider a single decision process in which we select a pair of alternatives:
  - the 1st alternative corr. to the 1st decision, and
  - the 2nd alternative corr. to the 2nd decision.
- If we are willing to pay:
  - the amount u to participate in the first process, and
  - the amount v to participate in the second decision process,
- then we should be willing to pay $u + v$ to participate in both decision processes.

Additivity: Case of Interval Uncertainty

- About the gain u from the first alternative, we only know that this (unknown) gain is in $[\underline{u}, \overline{u}]$.
- About the gain v from the second alternative, we only know that this gain belongs to the interval $[\underline{v}, \overline{v}]$.
- The overall gain $u + v$ can thus take any value from the interval

$$[\underline{u}, \overline{u}] + [\underline{v}, \overline{v}] \stackrel{\text{def}}{=} \{u + v : u \in [\underline{u}, \overline{u}], v \in [\underline{v}, \overline{v}]\}.$$

- It is easy to check that

$$[\underline{u}, \overline{u}] + [\underline{v}, \overline{v}] = [\underline{u} + \underline{v}, \overline{u} + \overline{v}].$$

- Thus, the additivity requirement about the fair prices takes the form

$$P([\underline{u} + \underline{v}, \overline{u} + \overline{v}]) = P([\underline{u}, \overline{u}]) + P([\underline{v}, \overline{v}]).$$

Fair Price Under Interval Uncertainty

- By a *fair price under interval uncertainty*, we mean a function $P([\underline{u}, \overline{u}])$ for which:
  - $\underline{u} \leq P([\underline{u}, \overline{u}]) \leq \overline{u}$ for all u (*conservativeness*);
  - if $\underline{u} = \underline{v}$ and $\overline{u} < \overline{v}$, then $P([\underline{u}, \overline{u}]) \leq P([\underline{v}, \overline{v}])$ (*monotonicity*);
  - (*additivity*) for all $\underline{u}$, $\overline{u}$, $\underline{v}$, and $\overline{v}$, we have

$$P([\underline{u} + \underline{v}, \overline{u} + \overline{v}]) = P([\underline{u}, \overline{u}]) + P([\underline{v}, \overline{v}]).$$

- *Theorem:* Each fair price under interval uncertainty has the form

$$P([\underline{u}, \overline{u}]) = \alpha_H \cdot \overline{u} + (1 - \alpha_H) \cdot \underline{u} \text{ for some } \alpha_H \in [0, 1].$$

- *Comment:* we thus get a new justification of Hurwicz optimism-pessimism criterion.

Proof: Main Ideas

- Due to monotonicity, $P([u, u]) = u$.
- Due to monotonicity, $\alpha_H \stackrel{\text{def}}{=} P([0, 1]) \in [0, 1]$.
- For $[0, 1] = [0, 1/n] + \ldots + [0, 1/n]$ (n times), additivity implies $\alpha_H = n \cdot P([0, 1/n])$, so $P([0, 1/n]) = \alpha_H \cdot (1/n)$.
- For $[0, m/n] = [0, 1/n] + \ldots + [0, 1/n]$ (m times), additivity implies $P([0, m/n]) = \alpha_H \cdot (m/n)$.
- For each real number r, for each n, there is an m s.t. $m/n \le r \le (m + 1)/n$.
- Monotonicity implies $\alpha_H \cdot (m/n) = P([0, m/n]) \le P([0, r]) \le P([0, (m + 1)/n]) = \alpha_H \cdot ((m + 1)/n)$.
- When $n \to \infty$, $\alpha_H \cdot (m/n) \to \alpha_H \cdot r$ and $\alpha_H \cdot ((m + 1)/n) \to r$, hence $P([0, r]) = \alpha_H \cdot r$.
- For $[\underline{u}, \overline{u}] = [\underline{u}, \underline{u}] + [0, \overline{u} - \underline{u}]$, additivity implies $P([\underline{u}, \overline{u}]) = \underline{u} + \alpha_H \cdot (\overline{u} - \underline{u})$. Q.E.D.

Case of Set-Valued Uncertainty

- In some cases:
  - in addition to knowing that the actual gain belongs to the interval $[\underline{u}, \overline{u}]$,
  - we also know that some values from this interval cannot be possible values of this gain.
- For example:
  - if we buy an obscure lottery ticket for a simple prize-or-no-prize lottery from a remote country,
  - we either get the prize or lose the money.
- In this case, the set of possible values of the gain consists of two values.
- Instead of a (bounded) *interval* of possible values, we can consider a general bounded *set* of possible values.

Fair Price Under Set-Valued Uncertainty

- We want a function P that assigns, to every bounded closed set S, a real number $P(S)$, for which:
  - $P([\underline{u}, \overline{u}]) = \alpha_H \cdot \overline{u} + (1 - \alpha_H) \cdot \underline{u}$ (*conservativeness*);
  - $P(S + S') = P(S) + P(S')$, where $S + S' \stackrel{\text{def}}{=} \{s + s' : s \in S, s' \in S'\}$ (*additivity*).
- *Theorem:* Each fair price under set uncertainty has the form $P(S) = \alpha_H \cdot \sup S + (1 - \alpha_H) \cdot \inf S$.
- *Proof: idea.*
  - $\{\underline{s}, \overline{s}\} \subseteq S \subseteq [\underline{s}, \overline{s}]$, where $\underline{s} \stackrel{\text{def}}{=} \inf S$ and $\underline{s} \stackrel{\text{def}}{=} \sup S$;
  - thus, $[2\underline{s}, 2\overline{s}] = \{\underline{s}, \overline{s}\} + [\underline{s}, \overline{s}] \subseteq S + [\underline{s}, \overline{s}] \subseteq [\underline{s}, \overline{s}] + [\underline{s}, \overline{s}] = [2\underline{s}, 2\overline{s}]$;
  - so $S + [\underline{s}, \overline{s}] = [2\underline{s}, 2\overline{s}]$, hence $P(S) + P([\underline{s}, \overline{s}]) = P([2\underline{s}, 2\overline{s}])$, and

$$P(S) = (\alpha_H \cdot (2\overline{s}) + (1 - \alpha_H) \cdot (2\underline{s})) - (\alpha_H \cdot \overline{s} + (1 - \alpha_H) \cdot \underline{s}).$$

Crisp Z-Numbers, Z-Intervals, and Z-Sets

- Until now, we assumed that we are 100% certain that the actual gain is contained in the given interval or set.
- In reality, mistakes are possible.
- Usually, we are only certain that u belongs to the interval or set with some probability $p \in (0, 1)$.
- A pair of information and a degree of certainty about this this info is what L. Zadeh calls a *Z-number*.
- We will call a pair (u, p) consisting of a (crisp) number and a (crisp) probability a *crisp Z-number*.
- We will call a pair $([\underline{u}, \overline{u}], p)$ consisting of an interval and a probability a *Z-interval*.
- We will call a pair (S, p) consisting of a set and a probability a *Z-set*.

Additivity for Z-Numbers

- *Situation:*
  - for the first decision, our degree of confidence in the gain estimate u is described by some probability p;
  - for the 2nd decision, our degree of confidence in the gain estimate v is described by some probability q.
- The estimate $u + v$ is valid only if both gain estimates are correct.
- Since these estimates are independent, the probability that they are both correct is equal to $p \cdot q$.
- Thus, for crisp Z-numbers (u, p) and (v, q), the sum is equal to $(u + v, p \cdot q)$.
- Similarly, for Z-intervals $([\underline{u}, \overline{u}], p)$ and $([\underline{v}, \overline{v}], q)$, the sum is equal to $([\underline{u} + \underline{v}, \overline{u} + \overline{v}], p \cdot q)$.
- For Z-sets, $(S, p) + (S', q) = (S + S', p \cdot q)$.

Fair Price for Z-Numbers and Z-Sets

- We want a function P that assigns, to every crisp Z-number (u, p), a real number $P(u, p)$, for which:
  - $P(u, 1) = u$ for all u (*conservativeness*);
  - for all u, v, p, and q, we have $P(u + v, p \cdot q) = P(u, p) + P(v, q)$ (*additivity*);
  - the function $P(u, p)$ is continuous in p (*continuity*).
- *Theorem:* Fair price under crisp Z-number uncertainty has the form $P(u, p) = u - k \cdot \ln(p)$ for some k.
- *Theorem:* For Z-intervals and Z-sets,

$$P(S, p) = \alpha_H \cdot \sup S + (1 - \alpha_H) \cdot \inf S - k \cdot \ln(p).$$

- *Proof:* $(u, p) = (u, 1) + (0, p)$; for continuous $f(p) \stackrel{\text{def}}{=} (0, p)$, additivity means $f(p \cdot q) = f(p) + f(q)$, so

$$f(p) = -k \cdot \ln(p).$$

Case When Probabilities Are Known With Interval Or Set-Valued Uncertainty

- We often do not know the exact probability p.
- Instead, we may only know the interval $\left[\underline{p}, \overline{p}\right]$ of possible values of p.
- More generally, we know the set $\mathcal{P}$ of possible values of p.
- If we only know that $p \in [\underline{p}, \overline{p}]$ and $q \in [\underline{q}, \overline{q}]$, then possible values of $p \cdot q$ form the interval

$$\left[\underline{p} \cdot \underline{q}, \overline{p} \cdot \overline{q}\right].$$

- For sets $\mathcal{P}$ and $\mathcal{Q}$, the set of possible values $p \cdot q$ is the set

$$\mathcal{P} \cdot \mathcal{Q} \stackrel{\text{def}}{=} \{p \cdot q : p \in \mathcal{P} \text{ and } q \in \mathcal{Q}\}.$$

Fair Price When Probabilities Are Known With Interval Uncertainty

- We want a function P that assigns, to every Z-number $\left(u, \left[\underline{p}, \overline{p}\right]\right)$, a real number $P\left(u, \left[\underline{p}, \overline{p}\right]\right)$, so that:
  - $P(u, [p, p]) = u - k \cdot \ln(p)$ (*conservativeness*);
  - $P\left(u + v, \left[\underline{p} \cdot \underline{q}, \overline{p} \cdot \overline{q}\right]\right) = P\left(u, \left[\underline{p}, \overline{p}\right]\right) + P\left(v, \left[\underline{q}, \overline{q}\right]\right)$ (*additivity*);
  - $P\left(u, \left[\underline{p}, \overline{p}\right]\right)$ is continuous in $\underline{p}$ and $\overline{p}$ (*continuity*).
- *Theorem:* Fair price has the form

$$P\left(u, \left[\underline{p}, \overline{p}\right]\right) = u - (k - \beta) \cdot \ln(\overline{p}) - \beta \cdot \ln\left(\underline{p}\right) \text{ for some } \beta \in [0, 1].$$

- For set-valued probabilities, we similarly have $P(u, \mathcal{P}) = u - (k - \beta) \cdot \ln(\sup \mathcal{P}) - \beta \cdot \ln(\inf \mathcal{P})$.
- For Z-sets and Z-intervals, we have $P(S, \mathcal{P}) =$

$$\alpha_H \cdot \sup S + (1 - \alpha_H) \cdot \inf S - (k - \beta) \cdot \ln(\sup \mathcal{P}) - \beta \cdot \ln(\inf \mathcal{P}).$$

Proof

- By additivity, $P(S, \mathcal{P}) = P(S, 1) + P(0, \mathcal{P})$, so it is sufficient to find $P(0, \mathcal{P})$.
- For intervals, $P(0, [\underline{p}, \overline{p}]) = P(0, \overline{p}) + P(0, [p, 1])$, for $p \stackrel{\text{def}}{=} \underline{p}/\overline{p}$.
- For $f(p) \stackrel{\text{def}}{=} P(0, [p, 1])$, additivity means
$$f(p \cdot q) = f(p) \cdot f(q).$$
- Thus, $f(p) = -\beta \cdot \ln(p)$ for some β.
- Hence, $P(0, [\underline{p}, \overline{p}]) = -k \cdot \ln(\overline{p}) - \beta \cdot \ln(p)$.
- Since $\ln(p) = \ln(\overline{p}) - \ln(\underline{p})$, we get the desired formula.
- For sets $\mathcal{P}$, with $\underline{p} \stackrel{\text{def}}{=} \inf \mathcal{P}$ and $\overline{p} \stackrel{\text{def}}{=} \sup \mathcal{P}$, we have $\mathcal{P} \cdot [\underline{p}, \overline{p}] = [\underline{p}^2, \overline{p}^2]$, so $P(0, \mathcal{P}) + P(0, [\underline{p}, \overline{p}]) = P(0, [\underline{p}^2, \overline{p}^2])$.
- Thus, from known formulas for intervals $[\underline{p}, \overline{p}]$, we get formulas for sets $\mathcal{P}$.

Case of Fuzzy Numbers

- An expert is often imprecise ("fuzzy") about the possible values.
- For example, an expert may say that the gain is small.
- To describe such information, L. Zadeh introduced the notion of *fuzzy numbers*.
- For fuzzy numbers, different values u are possible with different degrees $\mu(u) \in [0,1]$.
- The value w is a possible value of $u+v$ if:
  - for some values u and v for which $u+v=w$,
  - u is a possible value of 1st gain, and
  - v is a possible value of 2nd gain.
- If we interpret "and" as min and "or" ("for some") as max, we get *Zadeh's extension principle:*

$$\mu(w) = \max_{u,v:\, u+v=w} \min(\mu_1(u), \mu_2(v)).$$

Case of Fuzzy Numbers (cont-d)

- *Reminder:* $\mu(w) = \max\limits_{u,v:\, u+v=w} \min(\mu_1(u), \mu_2(v))$.
- This operation is easiest to describe in terms of α-cuts

$$\mathbf{u}(\alpha) = [u^-(\alpha), u^+(\alpha)] \stackrel{\text{def}}{=} \{u : \mu(u) \geq \alpha\}.$$

- Namely, $\mathbf{w}(\alpha) = \mathbf{u}(\alpha) + \mathbf{v}(\alpha)$, i.e.,

$$w^-(\alpha) = u^-(\alpha) + v^-(\alpha) \text{ and } w^+(\alpha) = u^+(\alpha) + v^+(\alpha).$$

- For product (of probabilities), we similarly get

$$\mu(w) = \max_{u,v:\, u\cdot v=w} \min(\mu_1(u), \mu_2(v)).$$

- In terms of α-cuts, we have $\mathbf{w}(\alpha) = \mathbf{u}(\alpha) \cdot \mathbf{v}(\alpha)$, i.e.,

$$w^-(\alpha) = u^-(\alpha) \cdot v^-(\alpha) \text{ and } w^+(\alpha) = u^+(\alpha) \cdot v^+(\alpha).$$

Fair Price Under Fuzzy Uncertainty

- We want to assign, to every fuzzy number s, a real number $P(s)$, so that:
  - if a fuzzy number s is located between $\underline{u}$ and $\overline{u}$, then $\underline{u} \le P(s) \le \overline{u}$ (*conservativeness*);
  - $P(u+v) = P(u) + P(v)$ (*additivity*);
  - if for all α, $s^-(\alpha) \le t^-(\alpha)$ and $s^+(\alpha) \le t^+(\alpha)$, then we have $P(s) \le P(t)$ (*monotonicity*);
  - if μ_n uniformly converges to μ, then $P(\mu_n) \to P(\mu)$ (*continuity*).
- *Theorem.* The fair price is equal to

$$P(s) = s_0 + \int_0^1 k^-(\alpha)\, ds^-(\alpha) - \int_0^1 k^+(\alpha)\, ds^+(\alpha) \text{ for some } k^\pm(\alpha).$$

Discussion

- $\int f(x)\cdot dg(x) = \int f(x)\cdot g'(x)\,dx$ for a *generalized function* $g'(x)$, hence for generalized $K^{\pm}(\alpha)$, we have:

$$P(s) = \int_0^1 K^-(\alpha)\cdot s^-(\alpha)\,d\alpha + \int_0^1 K^+(\alpha)\cdot s^+(\alpha)\,d\alpha.$$

- Conservativeness means that

$$\int_0^1 K^-(\alpha)\,d\alpha + \int_0^1 K^+(\alpha)\,d\alpha = 1.$$

- For the interval $[\underline{u},\overline{u}]$, we get

$$P(s) = \left(\int_0^1 K^-(\alpha)\,d\alpha\right)\cdot\underline{u} + \left(\int_0^1 K^+(\alpha)\,d\alpha\right)\cdot\overline{u}.$$

- Thus, Hurwicz optimism-pessimism coefficient α_H is equal to $\int_0^1 K^+(\alpha)\,d\alpha$.
- In this sense, the above formula is a generalization of Hurwicz's formula to the fuzzy case.

Proof

- Define $\mu_{\gamma,u}(0) = 1$, $\mu_{\gamma,u}(x) = \gamma$ for $x \in (0, u]$, and $\mu_{\gamma,u}(x) = 0$ for all other x.
- $\mathbf{s}_{\gamma,u}(\alpha) = [0, 0]$ for $\alpha > \gamma$, $\mathbf{s}_{\gamma,u}(\alpha) = [0, u]$ for $\alpha \leq \gamma$.
- Based on the α-cuts, one check that $s_{\gamma,u+v} = s_{\gamma,u} + s_{\gamma,v}$.
- Thus, due to additivity, $P(s_{\gamma,u+v}) = P(s_{\gamma,u}) + P(s_{\gamma,v})$.
- Due to monotonicity, $P(s_{\gamma,u}) \uparrow$ when $u \uparrow$.
- Thus, $P(s_{\gamma,u}) = k^+(\gamma) \cdot u$ for some value $k^+(\gamma)$.
- Let us now consider a fuzzy number s s.t. $\mu(x) = 0$ for $x < 0$, $\mu(0) = 1$, then $\mu(x)$ continuously $\downarrow 0$.
- For each sequence of values $\alpha_0 = 1 < \alpha_1 < \alpha_2 < \ldots < \alpha_{n-1} < \alpha_n = 1$, we can form an approximation s_n:
  - $s_n^-(\alpha) = 0$ for all α; and
  - when $\alpha \in [\alpha_i, \alpha_{i+1})$, then $s_n^+(\alpha) = s^+(\alpha_i)$.

Proof (cont-d)

- Here, $s_n = s_{\alpha_{n-1}, s^+(\alpha_{n-1})} + s_{\alpha_{n-2}, s^+(\alpha_{n-2}) - s^+(\alpha_{n-1})} + \ldots + s_{\alpha_1, \alpha_1 - \alpha_2}$.
- Due to additivity, $P(s_n) = k^+(\alpha_{n-1}) \cdot s^+(\alpha_{n-1}) + k^+(\alpha_{n-2}) \cdot (s^+(\alpha_{n-2}) - s^+(\alpha_{n-1})) + \ldots + k^+(\alpha_1) \cdot (\alpha_1 - \alpha_2)$.
- This is minus the integral sum for $\int_0^1 k^+(\gamma)\, ds^+(\gamma)$.
- Here, $s_n \to s$, so $P(s) = \lim P(s_n) = \int_0^1 k^+(\gamma)\, ds^+(\gamma)$.
- Similarly, for fuzzy numbers s with $\mu(x) = 0$ for $x > 0$, we have $P(s) = \int_0^1 k^-(\gamma)\, ds^-(\gamma)$ for some $k^-(\gamma)$.
- A general fuzzy number g, with α-cuts $[g^-(\alpha), g^+(\alpha)]$ and a point g_0 at which $\mu(g_0) = 1$, is the sum of g_0,
  - a fuzzy number with α-cuts $[0, g^+(\alpha) - g_0]$, and
  - a fuzzy number with α-cuts $[g_0 - g^-(\alpha), 0]$.
- Additivity completes the proof.

Case of General Z-Number Uncertainty

- In this case, we have two fuzzy numbers:
  - a fuzzy number s which describes the values, and
  - a fuzzy number p which describes our degree of confidence in the piece of information described by s.
- We want to assign, to every pair (s, p) s.t. p is located on $[p_0, 1]$ for some $p_0 > 0$, a number $P(s, p)$ so that:
  - $P(s, 1)$ is as before (*conservativeness*);
  - $P(u + v, p \cdot q) = P(u, p) + P(v, q)$ (*additivity*);
  - if $s_n \to s$ and $p_n \to p$, then $P(s_n, p_n) \to P(s, p)$ (*continuity*).

- *Thm* : $$P(s, p) = \int_0^1 K^-(\alpha) \cdot s^-(\alpha)\, d\alpha + \int_0^1 K^+(\alpha) \cdot s^+(\alpha)\, d\alpha +$$

$$\int_0^1 L^-(\alpha) \cdot \ln(p^-(\alpha))\, d\alpha + \int_0^1 L^+(\alpha) \cdot \ln(p^+(\alpha))\, d\alpha.$$

Conclusions and Future Work

- In many practical situations:
  - we need to select an alternative, but
  - we do not know the exact consequences of each possible selection.
- We may also know, e.g., that the gain will be *somewhat larger* than a certain value u_0.
- We propose to make decisions by comparing the *fair price* corresponding to each uncertainty.
- *Future work:*
  - apply to practical decision problems;
  - generalize to type-2 fuzzy sets;
  - generalize to the case when we have several pieces of information (s, p).

Appendix 1.3

Application to Education

How Success in a Task Depends on the Skills Level: Two Uncertainty-Based Justifications of a Semi-Heuristic Rasch Model

An Empirically Successful Rasch Model

- For each level of student skills, the student is usually:
  - very successful in solving simple problems,
  - not yet successful in solving problems which are – to this student – too complex, and
  - reasonably successful in solving problems which are of the right complexity.
- To design adequate tests, it is desirable to understand how a success s in a task depends:
  - on the student's skill level ℓ and
  - on the problem's complexity c.
- Empirical *Rasch model* predicts $s = \dfrac{1}{1 + \exp(c - \ell)}$.
- Practitioners, however, are somewhat reluctant to use this formula, since it lacks a deeper justification.

What We Do

- In this talk, we provide two possible justifications for the Rasch model.
- The first is a simple fuzzy-based justification which provides a good intuitive explanation for this model.
- This will hopefully enhance its use in teaching practice.
- The second is a somewhat more sophisticated explanation which is:
  - less intuitive but
  - provides a quantitative justification.

First Justification for the Rasch Model

- Let us fix c and consider the dependence $s = g(\ell)$.
- When we change ℓ slightly, to $\ell + \Delta\ell$, the success also changes slightly: $g(\ell + \Delta\ell) \approx g(\ell)$.
- Thus, once we know $g(\ell)$, it is convenient to store not $g(\ell + \Delta\ell)$, but the difference $g(\ell + \Delta\ell) - g(\ell) \approx \dfrac{dg}{d\ell} \cdot \Delta\ell$.
- Here, $\dfrac{dg}{d\ell}$ depends on $s = g(\ell)$: $\dfrac{dg}{d\ell} = f(s) = f(g(\ell))$.
- In the absence of skills, when $\ell \approx -\infty$ and $s \approx 0$, adding a little skills does not help much, so $f(s) \approx 0$.
- For almost perfect skills $\ell \approx +\infty$ and $s \approx 1$, similarly $f(s) \approx 0$.
- So, $f(s)$ is big when s is big ($s \gg 0$) but not too big ($1 - s \gg 0$).

First Justification for the Rasch Model (cont-d)

- Rule: $f(s)$ is big when:
  - s is big ($s \gg 0$) but
  - not too big ($1 - s \gg 0$).
- Here, "but" means "and", the simplest "and" is the product.
- The simplest membership function for "big" is $\mu_{\text{big}}(s) = s$.
- Thus, the degree to which $f(s)$ is big is equal to

$$s \cdot (1 - s): \quad f(s) = s \cdot (1 - s).$$

- The equation $\dfrac{dg}{d\ell} = g \cdot (1 - g)$ leads exactly to Rasch's model $g(\ell) = \dfrac{1}{1 + \exp(c - \ell)}$ for some c.

What If Use min for “and”?

- What if we use a different “and”-operation, for example, $\min(a, b)$?
- Let us show that in this case, we also get a meaningful model.
- Indeed, in this case, the corresponding equation takes the form $\dfrac{dg}{d\ell} = \min(g, 1 - g)$.
- Its solution is:
  - $g(\ell) = C_- \cdot \exp(\ell)$ when $s = g(\ell) \le 0.5$, and
  - $g(\ell) = 1 - C_+ \cdot \exp(-\ell)$ when $s = g(\ell) \ge 0.5$.
- In particular, for $C_- = 0.5$, we get a cdf of the Laplace distribution $\rho(x) = \dfrac{1}{2} \cdot \exp(-|x|)$.
- This distribution is used in many applications – e.g., to modify the data in large databases to promote privacy.

Towards a Second Justification

- The success s depends on how much the skills level ℓ exceeds the complexity c of the task: $s = h(\ell - c)$.
- For each c, we can use the value $h(\ell - c)$ to gauge the students' skills.
- For different c, we get different scales for measuring skills.
- This is similar to having different scales in physics:
  - a change in a measuring unit leads to $x' = a \cdot x$; e.g., 2 m = $100 \cdot 2$ cm;
  - a change in a starting point leads to $x' = x + b$; e.g., 20° C = $(20 + 273)^\circ$ K.
- In physics, re-scaling is usually linear, but here, $0 \to 0$, $1 \to 1$, so we need a non-linear re-scaling.

How to Describe Not-Necessarily-Linear Re-Scalings

- If we first apply one reasonable re-scaling, and after that another one, we still get a reasonable re-scaling.
- For example, we can first change meters to centimeters, and then replace centimeters with inches.
- Then, the resulting re-scaling from meters to inches is still a linear transformation.
- In mathematical terms, this means that the class of reasonable e-scalings is closed under composition.
- Also, if we have a re-scaling, e.g., from C to F, then the "inverse" re-scaling from F to C is also reasonable.
- In precise terms, this means that the class of all reasonable re-scalings is invariant under taking the inversion.

How to Describe Re-Scalings (cont-d)

- Thus, we can say that reasonable re-scalings form a transformation group.
- Our goal is computations.
- In a computer, we can only store finitely many parameters.
- Thus, each re-scaling must be determined by finitely many parameters.
- Such groups are called *finite-dimensional*.
- So, we need to describe all finite-dimensional transformation groups that contain all linear transformations.
- It is known that all functions from these groups are fractionally-linear $f(s) = \dfrac{a \cdot s + b}{c \cdot s + d}$.

Resulting Equation

- We consider a transformation $s' = f(s)$ between
$$s = h(\ell - c) \text{ and } s' = h(\ell - c').$$
- We showed that this transformation is fractionally-linear $f(s) = \dfrac{a \cdot s + b}{c \cdot s + d}$.
- When $s = 0$, we should have $s' = 0$, hence $b = 0$.
- We can now divide both numerator and denominator by d, then $f(s) = \dfrac{A \cdot s}{C \cdot s + 1}$.
- When $s = 1$, we should have $s' = 1$, so $A = C + 1$, and $f(s) = \dfrac{(1 + C) \cdot s}{C \cdot s + 1}$.
- For $c' = 0$, we thus get
$$h(\ell - c) = \frac{(1 + C(c)) \cdot h(\ell)}{C(c) \cdot h(\ell) + 1}.$$

Solving the Resulting Equation Explains the Rasch Model

- We know that

$$h(\ell - c) = \frac{(1 + C(c)) \cdot h(\ell)}{C(c) \cdot h(\ell) + 1}.$$

- Differentiating both sides w.r.t. c and taking $c = 0$, we get a differential equation whose general solution is

$$h(\ell) = \frac{1}{1 + \exp(k \cdot (c - \ell))}.$$

- By changing measuring units for ℓ and c to k times smaller ones, we get the Rasch model

$$h(\ell) = \frac{1}{1 + \exp(c - \ell)}.$$

Conclusion

- It has been empirically shown that,
  - once we know the complexity c of a task, and the skill level ℓ of a student attempting this task,
  - the student's success s is determined by Rasch's formula

$$s = \frac{1}{1 + \exp(c - \ell)}.$$

- In this talk, we provide two uncertainty-based justifications for this model:
  - a simpler fuzzy-based justification provides an intuitive semi-qualitative explanation for this formula;
  - a more complex justification provides a quantitative explanation for the Rasch model.

Appendix 3: Proofs

Appendix 3.0: Utility Value

- Let A be any alternative such that $A_0 < A < A_1$; then:
  - as p increases from 0, $L(p) < A$;
  - then, at some point, $L(p) > A$;
  - so, there is a threshold separating values for which $L(p) < A$ from the values for which $L(p) > A$;
  - this threshold is called the *utility* of alternative A:

$$u(A) \stackrel{\text{def}}{=} \sup\{p : L(p) < A\} = \inf\{p : L(p) > A\}$$

- Here, for every $\varepsilon > 0$, we have

$$L(u(A) - \varepsilon) < A < L(u(A) - \varepsilon).$$

- In this sense, the alternative A is (almost) equivalent to $L(u(A))$; we will denote this almost equivalence by

$$A \equiv L(u(A)).$$

Appendix 3.0: Almost Uniqueness of Utility

- The definition of utility u depends on the selection of two fixed alternatives A_0 and A_1.
- What if we use different alternatives A'_0 and A'_1?
- By definition of utility, every alternative A is equivalent to a lottery $L(u(A))$ in which we get A_1 with probability $u(A)$ and A_0 with probability $1 - u(A)$.
- For simplicity, let us assume that $A'_0 < A_0 < A_1 < A'_1$. Then, for the utility u', we get $A_0 \equiv L'(u'(A_0))$ and $A_1 \equiv L'(u'(A_1))$.

Appendix 3.0: Almost Uniqueness of Utility

- So, the alternative A is equivalent to a complex lottery in which:
  - we select A_1 with probability $u(A)$ and A_0 with probability $1 - u(A)$;
  - depending on which of the two alternatives A_i we get, we get A'_1 with probability $u'(A_i)$ and A'_0 with probability $1 - u'(A_i)$.
- In this complex lottery, we get A'_1 with probability $u'(A) = u(A) \cdot (u'(A_1) - u'(A_0)) + u'(A_0)$.
- Thus, the utility $u'(A)$ is related with the utility $u(A)$ by a linear transformation $u' = a \cdot u + b$, with $a > 0$.

Appendix 3.2: Derivations Related to the Second Example

- We have $g'(p) \cdot \sqrt{p \cdot (1 - p)} = \text{const}$ for some constant.
- Integrating with $p = 0$ corresponding to the lowest 0-th level – i.e., that $g(0) = 0$

$$g(p) = \text{const} \cdot \int_0^p \frac{dq}{\sqrt{q \cdot (1 - q)}}.$$

- Introduce a new variable t for which $q = \sin^2(t)$ and
  - $dq = 2 \cdot \sin(t) \cdot \cos(t) \cdot dt$,
  - $1 - p = 1 - \sin^2(t) = \cos^2(t)$ and, therefore,
  - $\sqrt{p \cdot (1 - p)} = \sqrt{\sin^2(t) \cdot \cos^2(t)} = \sin(t) \cdot \cos(t)$.

Appendix 3.2: Derivations (cont-d)

- The lower bound $q = 0$ corresponds to $t = 0$
- the upper bound $q = p$ corresponds to the value t_0 for which $\sin^2(t_0) = p$
  i.e., $\sin(t_0) = \sqrt{p}$ and $t_0 = \arcsin(\sqrt{p})$.
- Therefore,

$$g(p) = \text{const} \cdot \int_0^p \frac{dq}{\sqrt{q \cdot (1 - q)}} =$$

$$\text{const} \cdot \int_0^{t_0} \frac{2 \cdot \sin(t) \cdot \cos(t) \cdot dt}{\sin(t) \cdot \cos(t)} = \int_0^{t_0} 2 \cdot dt =$$

$$2 \cdot \text{const} \cdot t_0.$$

Appendix 3.2: Derivations (final)

- We know t_0 depends on p, so we get

$$g(p) = 2 \cdot \mathrm{const} \cdot \arcsin\left(\sqrt{p}\right).$$

- We determine the constant by
  - the largest possible probability value $p = 1$ implies $g(1) = 1$, and
  - $\arcsin\left(\sqrt{1}\right) = \arcsin(1) = \frac{\pi}{2}$

- Therefore, we conclude that

$$g(p) = \frac{2}{\pi} \cdot \arcsin\left(\sqrt{p}\right).$$

Appendix 3.3: Reformulation In Terms of Disutility

- In the ideal case, for each value x, we should use a decision $d(x)$, and gain utility $u(x,x)$.
- In practice, we have to use decisions $d(x')$, and get slightly worse utility values $u(x',x)$.
- The corresponding decrease in utility $U(x',x) \stackrel{\text{def}}{=} u(x,x) - u(x',x)$ is usually called *disutility*.
- In terms of disutility, the function $u(x',x)$ has the form

$$u(x',x) = u(x,x) - U(x',x),$$

Appendix 3.3: Reformulation In Terms of Disutility

- Thus, the optimized expression takes the form

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot u(x,x)\,dx - \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\,dx.$$

- The first integral does not depend on $\widetilde{x}_k$; thus, the expression attains its maximum if and only if the second integral attains its minimum.
- The resulting maximum thus takes the form

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot u(x,x)\,dx - \min_{\tilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\,dx.$$

Appendix 3.3: Reformulation In Terms of Disutility

- Thus, we get the form

$$\sum_{k=0}^{n} \int_{x_k}^{x_{k+1}} \rho(x) \cdot u(x, x)\, dx - \sum_{k=0}^{n} \min_{\tilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\, dx.$$

- The first sum does not depend on selecting the thresholds.
- Thus, to maximize utility, we should select the values $x_1, \ldots, x_n$ for which the second sum attains its smallest possible value:

$$\sum_{k=0}^{n} \min_{\tilde{x}_k} \int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\, dx \to \min.$$

Appendix 3.3: Membership Function

- In an n-valued scale:
  - the smallest label 0 corresponds to the value $\mu(x) = 0/n$,
  - the next label 1 corresponds to the value $\mu(x) = 1/n$,
  - ...
  - the last label n corresponds to the value $\mu(x) = n/n = 1$.
- Thus, for each n:
  - values from the interval $[x_0, x_1]$ correspond to the value $\mu(x) = 0/n$;
  - values from the interval $[x_1, x_2]$ correspond to the value $\mu(x) = 1/n$;
  - ...
  - values from the interval $[x_n, x_{n+1}]$ correspond to the value $\mu(x) = n/n = 1$.
- The actual value of the membership function $\mu(x)$ corresponds to the limit $n \to \infty$, i.e., in effect, to very large values of n.
- Thus, in our analysis, we will assume that the number n of labels is huge – and thus, that the width of each of $n + 1$ intervals $[x_k, x_{k+1}]$ is very small.

Appendix 3.3: Membership Function (cont-d)

- The fact that each interval is narrow allows simplification of the expression $U(x', x)$.
- In the expression $U(x', x)$, both values x' and x belong to the same narrow interval
- Thus, the difference $\Delta x \stackrel{\text{def}}{=} x' - x$ is small.
- So, we can expand $U(x', x) = U(x + \Delta x, x)$ into Taylor series in Δx, and keep only the first non-zero term.
- In general, we have

$$U(x + \Delta, x) = U_0(x) + U_1 \cdot \Delta x + U_2(x) \cdot \Delta x^2 + \ldots,$$

where

$$U_0(x) = U(x, x), \quad U_1(x) = \frac{\partial U(x + \Delta x, x)}{\partial(\Delta x)},$$

$$U_2(x) = \frac{1}{2} \cdot \frac{\partial^2 U(x + \Delta x, x)}{\partial^2(\Delta x)}.$$

Appendix 3.3: Membership Function (cont-d)

- Here, by definition of disutility, we get $U_0(x) = U(x, x) = u(x, x) - u(x, x) = 0$.
- Since the utility is the largest (and thus, disutility is the smallest) when $x' = x$, i.e., when $\Delta x = 0$, the derivative $U_1(x)$ is also equal to 0
- Thus, the first non-trivial term corresponds to the second derivative:
$$U(x + \Delta x, x) \approx U_2(x) \cdot \Delta x^2,$$
- reformulated in terms of $\widetilde{x}_k$ (that needs to be minimized)
$$U(\widetilde{x}_k, x) \approx U_2(x) \cdot (\widetilde{x}_k - x)^2,$$
- is substituted into the expression
$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot U(\widetilde{x}_k, x)\, dx$$

.

Appendix 3.3: Membership Function (cont-d)

- We need to minimize the integral

$$\int_{x_k}^{x_{k+1}} \rho(x) \cdot U_2(x) \cdot (\widetilde{x}_k - x)^2 \, dx$$

- by differentiating with respect to the unknown $\widetilde{x}_k$ and equating the derivative to 0.
- Thus, we conclude that $\int_{x_k}^{x_{k+1}} \rho(x) \cdot U_2(x) \cdot (\widetilde{x}_k - x) \, dx = 0$,
- i.e., that

$$\widetilde{x}_k \cdot \int_{x_k}^{x_{k+1}} \rho(x) \cdot U_2(x) \, dx = \int_{x_k}^{x_{k+1}} x \cdot \rho(x) \cdot U_2(x) \, dx,$$

- and thus, that

$$\widetilde{x}_k = \frac{\int_{x_k}^{x_{k+1}} x \cdot \rho(x) \cdot U_2(x) \, dx}{\int_{x_k}^{x_{k+1}} \rho(x) \cdot U_2(x) \, dx}$$

which can be simplified because the intervals are narrow.

Appendix 3.3: Membership Function (cont-d)

- We denote the midpoint of the interval $[x_k, x_{k+1}]$ by $\overline{x}_k \stackrel{\text{def}}{=} \dfrac{x_k + x_{k+1}}{2}$, and denote $\Delta x \stackrel{\text{def}}{=} x - \overline{x}_k$,
- then we have $x = \overline{x}_k + \Delta x$.
- Expanding into Taylor series in terms of a small value Δx and keeping only main terms, we get

$$\rho(x) = \rho(\overline{x}_k + \Delta x) = \rho(\overline{x}_k) + \rho'(\overline{x}_k) \cdot \Delta x \approx \rho(\overline{x}_k),$$

where $f'(x)$ denoted the derivative of a function $f(x)$, and

$$U_2(x) = U_2(\overline{x}_k + \Delta x) = U_2(\overline{x}_k) + U_2'(\overline{x}_k) \cdot \Delta x \approx U_2(\overline{x}_k).$$

Appendix 3.3: Membership Function (cont-d)

- Using these new $\rho(\overline{x}_k)$ and $U_2(\overline{x}_k)$, we get

$$\widetilde{x}_k = \frac{\rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot \int_{x_k}^{x_{k+1}} x\,dx}{\rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot \int_{x_k}^{x_{k+1}}\,dx} = \frac{\int_{x_k}^{x_{k+1}} x\,dx}{\int_{x_k}^{x_{k+1}}\,dx} =$$

$$\frac{\dfrac{1}{2} \cdot (x_{k+1}^2 - x_k^2)}{x_{k+1} - x_k} = \frac{x_{k+1} + x_k}{2} = \overline{x}_k.$$

- Substituting $\widetilde{x}_k = \overline{x}_k$ into the integral and understanding that, on the k-th interval, we have $\rho(x) \approx \rho(\overline{x}_k)$ and $U_2(x) \approx U_2(\overline{x}_k)$,
- we conclude that the integral takes the form

$$\int_{x_k}^{x_{k+1}} \rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot (\overline{x}_k - x)^2\,dx =$$

$$\rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot \int_{x_k}^{x_{k+1}} (\overline{x}_k - x)^2\,dx.$$

Appendix 3.3: Membership Function (cont-d)

- When x goes from x_k to x_{k+1}, the difference Δx between x and the interval's midpoint $\overline{x}_k$ ranges from $-\Delta_k$ to Δ_k, where Δ_k is the interval's half-width:

$$\Delta_k \stackrel{\text{def}}{=} \frac{x_{k+1} - x_k}{2}.$$

- In terms of the new variable Δx, the right-hand side of the integral has the form

$$\int_{x_k}^{x_{k+1}} (\overline{x}_k - x)^2\,dx = \int_{-\Delta_k}^{\Delta_k} (\Delta x)^2\,d(\Delta x) = \frac{2}{3} \cdot \Delta_k^3.$$

- Thus, the integral takes the form

$$\frac{2}{3} \cdot \rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot \Delta_k^3.$$

Appendix 3.3: Membership Function (cont-d)

- The problem of selecting the Likert scale thus becomes the problem of minimizing the sum

$$\frac{2}{3} \cdot \sum_{k=0}^{n} \rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot \Delta_k^3.$$

- Here,
  $\overline{x}_{k+1} = x_{k+1} + \Delta_{k+1} = (\overline{x}_k + \Delta_k) + \Delta_{k+1} \approx \overline{x}_k + 2\Delta_k$, so
  $\Delta_k = (1/2) \cdot \Delta\overline{x}_k$, where $\Delta\overline{x}_k \stackrel{\text{def}}{=} \overline{x}_{k+1} - \overline{x}_k$.
- Thus, we get the form

$$\frac{1}{3} \cdot \sum_{k=0}^{n} \rho(\overline{x}_k) \cdot U_2(\overline{x}_k) \cdot \Delta_k^2 \cdot \Delta\overline{x}_k.$$

Appendix 3.3: Membership Function (cont-d)

- In terms of the membership function, we have $\mu(\overline{x}_k) = k/n$ and $\mu(\overline{x}_{k+1}) = (k+1)/n$.
- Since the half-width Δ_k is small, we have

$$\frac{1}{n} = \mu(\overline{x}_{k+1}) - \mu(\overline{x}_k) = \mu(\overline{x}_k + 2\Delta_k) - \mu(\overline{x}_k) \approx \mu'(\overline{x}_k) \cdot 2\Delta_k,$$

- thus, $\Delta_k \approx \dfrac{1}{2n} \cdot \dfrac{1}{\mu'(\overline{x}_k)}$.
- Substituting this expression into the sum, we get $\dfrac{1}{3 \cdot (2n)^2} \cdot I$, where

$$I = \sum_{k=0}^{n} \frac{\rho(\overline{x}_k) \cdot U_2(\overline{x}_k)}{(\mu'(\overline{x}_k))^2} \cdot \Delta \overline{x}_k.$$

Appendix 3.3: Membership Function (cont-d)

- The expression I is an integral sum, so when $n \to \infty$, this expression tends to the corresponding integral

$$I = \int \frac{\rho(x) \cdot U_2(x)}{(\mu'(x))^2}\, dx.$$

- With respect to the derivative $d(x) \stackrel{\text{def}}{=} \mu'(x)$, we need to minimize the objective function

$$I = \int \frac{\rho(x) \cdot U_2(x)}{d^2(x)}\, dx$$

under the constraint that

$$\int_{\underline{X}}^{\overline{X}} d(x)\, dx = \mu(\overline{X}) - \mu(\underline{X}) = 1 - 0 = 1.$$

Appendix 3.3: Membership Function (cont-d)

- By using the Lagrange multiplier method, we can reduce to the unconstrained problem of minimizing the functional

$$I = \int \frac{\rho(x) \cdot U_2(x)}{d^2(x)}\,dx + \lambda \cdot \int d(x)\,dx.$$

- Differentiating with respect to $d(x)$ and equating the derivative to 0, we conclude that $-2 \cdot \dfrac{\rho(x) \cdot U_2(x)}{d^3(x)} + \lambda = 0$,
- i.e., that $d(x) = c \cdot (\rho(x) \cdot U_2(x))^{1/3}$ for some constant c.
- Thus, $\mu(x) = \int_{\underline{X}}^{x} d(t)\,dt = c \cdot \int_{\underline{X}}^{x} (\rho(t) \cdot U_2(t))^{1/3}\,dt$.
- The constant c must be determined by the condition that $\mu(\overline{X}) = 1$.
- Thus, we arrive at the resulting formula.

Appendix 3.3: Resulting Formula

- The membership function $\mu(x)$ obtained by using Likert-scale elicitation is equal to

$$\mu(x) = \frac{\int_{\underline{X}}^{x} (\rho(t) \cdot U_2(t))^{1/3}\, dt}{\int_{\underline{X}}^{\overline{X}} (\rho(t) \cdot U_2(t))^{1/3}\, dt},$$

where $\rho(x)$ is the probability density describing the probabilities of different values of x,

$U_2(x) \stackrel{\text{def}}{=} \dfrac{1}{2} \cdot \dfrac{\partial^2 U(x + \Delta x, x)}{\partial^2(\Delta x)}$,

$U(x', x) \stackrel{\text{def}}{=} u(x, x) - u(x', x)$, and

$u(x', x)$ is the utility of using a decision $d(x')$ corresponding to the value x' in the situation in which the actual value is x.