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 Kahmenan and Tversky, researchers found many examples when decision making seems irrational.
- In this research, we plan to 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.

Overview (cont-d)

- 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 $a_i < a_j$,
  - but are more expensive.
- When presented with three alternatives $a_1 < a_2 < a_3$, 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 $a_2 < a_3 < a_4$, 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 (details if time allows).

Main 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 gained 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 $\dfrac{m}{n}$.
- The expected value of the frequency is equal to p, and that the standard deviation of this frequency is equal to $\sigma = \sqrt{\dfrac{p \cdot (1-p)}{n}}$.
- 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 $p_{i+1} \approx p_i + 2k_0 \cdot \sqrt{\dfrac{p_i \cdot (1-p_i)}{n}}$.
- 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 $p' \approx p + 2k_0 \cdot \sqrt{\dfrac{p \cdot (1-p)}{n}}$, we have

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

Analysis (cont-d)

- Since p and p' 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 $\sum_i \left(\frac{\lambda \cdot w_i - \widetilde{w}_i}{w_i}\right)^2$ is the smallest possible.
- Differentiating with respect to λ and equating to zero:

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

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.

Future Work

- Most of our results so far deal with theoretical foundations of decision making under uncertainty.
- We plan to supplement this theoretical work with examples of potential practical applications.
- We have already started working on some aspects of such applications.
- Another important aspect is computational:
  - once we describe our decisions in precise terms,
  - what is the most efficient way to compute the corresponding optimal decisions.

Applications: General Idea

- We plan to cover all aspects of decision making under uncertainty:
  - in business,
  - in engineering,
  - in education, and
  - in developing generic AI decision tools.
- In *engineering*, we started to analyze 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 to Business

- In *business*, we started to analyze how the economic notion of a fair price can be translated into algorithms for decision making under uncertainty.
- We have analyzed interval uncertainty from this viewpoint.
- We plan to extend this analysis to more complex types of uncertainty such as fuzzy.

Applications to Education and AI

- In *education*, we plan to analyze possible measures of gauging student success.
- We hope that these measures will be more adequate than the currently used ones such as GPA.
- In general *AI* applications, we start an analysis 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 levels 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 showed that for double-use expressions, the problem is NP-hard.
- Now, we are developing a feasible algorithm for checking when an expression can be converted into SUE.

Appendix: Derivations

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

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

- 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,
  - $\sqrt{p \cdot (1-p)} = \sqrt{\sin^2(t) \cdot \cos^2(t)} = \sin(t) \cdot \cos(t)$.

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,

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

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

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

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
  - $\arcsin\left(\sqrt{1}\right) = \arcsin(1) = \dfrac{\pi}{2}$
- Therefore, we conclude that

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