Fuzzy Logic Ideas Can Help in Explaining Kahneman and Tversky's Empirical Decision Weights

Joe Lorkowski and Vladik Kreinovich

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

World Conf. on Soft Computing, 2014

Introduction

In simple situations, an average person can easily make a decision.

- If the weather forecast predicts rain, take an umbrella.

In complex situations, even when we know all the possible consequences of each action, it is not easy to make a decision.

- medicines have side effects:
- surgery can have bad outcomes,
- immune system suppression can result in infections

It is not always easy to compare different actions and even skilled experts appreciate computer-based help.

Need to Analyze how People Make Decisions

- We don't know precisely what people need to make a decision.
- People cannot explain in precise terms why they selected an alternative.
  - We need analyze how people make decisions
  - and find a formal description to fit the observations.
- Start with the simplest case, full information:
  - we know all possible outcomes $o_1, \ldots, o_n$ of actions;
  - we know the exact value of each outcome o_i; and
  - we know the probability of each outcome $p_i(a)$.

Need to Analyze how People Make Decisions

- We know the same action 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:

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

  and 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

  so that people select the action a with the largest weighted gain

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

where $w_i(a) = f(p_i(a))$ for an appropriate function $f(x)$.

Empirical Results – Preferences for Gambles

Decision Weights for gains in gambles:

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 are qualitative explanations for this phenomenon.
- We propose a quantitative explanation based on fuzzy ideas.

Fuzzy 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.

Formalization of Distinguishable Probabilities

- Probabilities are often estimated from historical observations.
  - e.g. if rain falls in 40 days of some 100 day period, we estimate the probability for a similar period at 40%.
- In general, if some event occurs in m out or n observations, we estimate the probability as the ratio $\frac{m}{n}$
- If a value p corresponds to the i^{th} level out of n, the fuzzy truth value $\frac{i}{n}$
  then the next value p' corresponds to the $(i+1)^{st}$ level, the fuzzy truth value $\frac{i+1}{n}$.

Formalization of Distinguishable Probabilities

- So, if $g(p)$ and $g(p')$ denote fuzzy truth values, then
$$g(p) = \frac{i}{m} \text{ and } g(p') = \frac{i+1}{m}.$$
- When m is large – i.e. distance(p,p') is small – then $g(p') = g(p + (p' - p))$
- we can expand this expression in a Taylor series and keep only linear terms in $g(p') \approx g(p) + (p' - p) \cdot g'(p)$, where $g'(p)$ is the derivative of $g(p)$.
- After some derivations (described later in detail) we conclude
$$g(p) = \frac{2}{\pi} \cdot \arcsin\left(\sqrt{p}\right)$$

From Probabilities to Fuzzy Values: Main Idea

- Based on n samples, p is estimated as $p = \frac{i}{n}$.
- The st. dev. of this estimate is $\sigma = \sqrt{\frac{p\cdot(1-p)}{n}}$.
- This means that all values within a k_0-sigma interval $[p - k_0 \cdot \sigma, p + k_0 \cdot \sigma]$ are possible.
- The observed values $i_1 < i_2$ indicate different probabilities if the corresponding intervals are disjoint.
- The smallest such difference is when $p_1 + k_0 \cdot \sigma_1 = p_2 - k_0 \cdot \sigma_2$, i.e., when $\Delta p = p_2 - p_1 \approx 2k_0 \cdot \sqrt{\frac{p\cdot(1-p)}{n}} = \text{const} \cdot \sqrt{p \cdot (1-p)}$.
- For neighboring values on the scale from 0 to m, $g(p_2) - g(p_1) = \frac{1}{m}$, so $g'(p) \cdot \text{const} \cdot \sqrt{p \cdot (1-p)} = \frac{1}{m}$.
- This equation leads to $g(p) = \text{const} \cdot \arcsin\left(\sqrt{p}\right)$.

Resulting Discretization

- The resulting discretization is described by:
  - On a scale of 0 to m, $g(m) = \dfrac{i}{m}$
  - so, $i = m \cdot g(p)$.
- The desired discretization means, to each probability p:
  - we assign $i \approx m \cdot g(p)$ on the scale from 0 to m,
  - where $g(p)$ is described above.
- Then, we select equally distributed weights on the interval $[0, 1]$ resulting in

$$0, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{m-1}{m}, 1.$$

Assigning Weights to Probabilities

- We have a list of distinguishable probabilities $(0 =)p_0 < p_1 < \ldots < p_m(= 1)$ that correspond to the degree $g(p_i) = \dfrac{i}{m}$
- We need to assign an appropriate weight to each probability:
  - for $p_0 = 0$, we assign weight $w_0 = 0$
  - for p_1, we assign the next weight $w_1 = \dfrac{1}{m}$
  - for p_2, we assign the next weight $w_2 = \dfrac{2}{m}$
  - for p_{m-1}, we assign the next weight $w_{m-1} = \dfrac{m-1}{m}$
  - for $p_m = 1$, we assign the weight $w_m = 1$.

Assigning Weights to Probabilities

- For each probability $p \in [0, 1]$, assign the weight $g(p) = \frac{2}{\pi} \cdot \arcsin(\sqrt{p})$
- Result of the first try:
  p_i are original probabilities,
  $\widetilde{w}_i$ are Kahneman's empirical weights, and
  $w_i = g(p_i)$ were computed with the above formula.

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

Assigning Weights to Probabilities

- 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."
- So, 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:

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

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 probabilities, the difference between the fuzzy-motivated weights w'_i and the empirical weights $\widetilde{w}_i$ is small.
- **Conclusion:** Fuzzy-motivated ideas explain Kahneman and Tversky's empirical decision weights.

Appendix: Derivations

- 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 $(p - k_0 \cdot \sigma, p + k_0 \cdot \sigma)$,
- two probabilities p and p' are distinguishable if the corresponding intervals of possible values of frequency do not intersect

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

Appendix: Derivations

- 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

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

- If p corresponds to the fuzzy truth value $\dfrac{i}{m}$ then the next value p' corresponds to the truth value $\dfrac{i+1}{m}$.
- Let $g(p)$ denote the fuzzy truth value corresponding to the probability p. Then,

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

Appendix: Derivations

- 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, $g'(p) = \text{const} \cdot \frac{1}{\sqrt{p \cdot (1 - p)}}$.

Appendix: Derivations

- 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: Derivations

- 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: Derivations

- 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

$$g(p) = \frac{2}{\pi} \cdot \arcsin\left(\sqrt{p}\right). \tag{1}$$