Uncertainty Analysis Can Help in Explaining Kahneman and Tversky’s Empirical Decision Weights

Joe Lorkowski

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

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

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

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

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:

<table>
<thead>
<tr>
<th>probability</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td>0</td>
<td>5.5</td>
<td>8.1</td>
<td>13.2</td>
<td>18.6</td>
<td>26.1</td>
<td>42.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>probability</th>
<th>80</th>
<th>90</th>
<th>95</th>
<th>98</th>
<th>99</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td>60.1</td>
<td>71.2</td>
<td>79.3</td>
<td>87.1</td>
<td>91.2</td>
<td>100</td>
</tr>
</tbody>
</table>

- There are qualitative explanations for this phenomenon.
- We propose a quantitative explanation based on uncertainty analysis.
Idea: "Distinguishable" Probabilities

- For decision making, most people do not estimate probabilities as numbers.
- Most people estimate probabilities with “fuzzy” concepts like \( \text{low}, \text{medium}, \text{high} \).
- The discretization converts a possibly infinite number of probabilities to a finite number of values.
- The discrete scale is formed by probabilities which are \textit{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 $\sigma$ for $\sigma'$ 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: \( E[w_i | 0 < w_1 < \ldots < w_m = 1] \).
- These averages are known: \( w_i = \frac{i}{m} \).
- So, to probability \( p_i \), we assign weight \( g(p_i) = \frac{i}{m} \).
- For \( p' \approx p + 2k_0 \cdot \sqrt{\frac{p \cdot (1 - p)}{n}} \), we have
  \[
  g(p) = \frac{i}{m} \quad \text{and} \quad g(p') = \frac{i + 1}{m}.
  \]
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 \( g'(p) = \frac{dg}{dp} \) denotes the derivative of the function \( g(p) \).
  - Thus, \( g(p') - g(p) = \frac{1}{m} = (p' - p) \cdot g'(p) \).

- 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 \( g(p) = \frac{2}{\pi} \cdot \arcsin(\sqrt{p}) \).
Assigning Weights to Probabilities: First Try

- For each probability $p \in [0, 1]$, assign the weight
  \[ g(p) = \frac{2}{\pi} \cdot \arcsin(\sqrt{p}) \]

- Results:
  $p_i$ are original probabilities,
  $\tilde{w}_i$ are Kahneman’s empirical weights, and
  $w_i = g(p_i)$ were computed with the above formula.

<table>
<thead>
<tr>
<th>$p_i$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{w}_i$</td>
<td>0</td>
<td>5.5</td>
<td>8.1</td>
<td>13.2</td>
<td>18.6</td>
<td>26.1</td>
<td>42.1</td>
</tr>
<tr>
<td>$w_i = g(p_i)$</td>
<td>0</td>
<td>6.4</td>
<td>9.0</td>
<td>14.4</td>
<td>20.5</td>
<td>29.5</td>
<td>50.0</td>
</tr>
</tbody>
</table>
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 $\lambda$ using Least Squares such that $\sum_i \left( \frac{\lambda \cdot w_i - \tilde{w}_i}{w_i} \right)^2$ is the smallest possible.
- Differentiating with respect to $\lambda$ and equating to zero:
  $$\sum_i \left( \lambda - \frac{\tilde{w}_i}{w_i} \right) = 0,$$
  so $\lambda = \frac{1}{m} \cdot \sum_i \frac{\tilde{w}_i}{w_i}$. 

For the values being considered, $\lambda = 0.910$

For $w'_i = \lambda \cdot w_i = \lambda \cdot g(p_i)$

| $\tilde{w}_i$ | 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 |

For most probabilities, the difference between the uncertainty-motivated weights $w'_i$ and the empirical weights $\tilde{w}_i$ is small.

**Conclusion:** Uncertainty analysis explains Kahneman and Tversky’s empirical decision weights.
Appendix: Derivations

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

\[
g(p) = \text{const} \cdot \int_0^p \frac{dq}{\sqrt{q \cdot (1 - q)}} = \\
\text{const} \cdot \int_0^{t_0} 2 \cdot \frac{\sin(t) \cdot \cos(t) \cdot dt}{\sin(t) \cdot \cos(t)} = \int_0^{t_0} 2 \cdot dt = \\
2 \cdot \text{const} \cdot t_0.
\]
Derivations (final)

- We know $t_0$ depends on $p$, so we get

\[ g(p) = 2 \cdot \text{const} \cdot \arcsin(\sqrt{p}) . \]

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

- Therefore, we conclude that

\[ g(p) = \frac{2}{\pi} \cdot \arcsin(\sqrt{p}) . \]