8. Simple decision making agents

By: Alex S.

Apr, 2024

1. Making simple decisions

- The agent does not plan, but rather executes the action and sees the immediate outcome.
- Simple decisions can be done in the episodic environment.
- Complex decisions require a long chain of actions and only after them the outcome will be visible, e.g. in chess.
- The probability of the outcome s', given evidence observations e, is written:

$$P(\text{result}(a) = s'|a, e),$$

where a on the right-hand side of the conditioning bar stands for the event that action a is executed.[1] And result(a) is a *random variable* whose value are the possible outcome states, since we omit the current state.

The agent's preferences are captured by a **utility function**($U(s)$) which assigns a single number to express the desirability of a state.

The **expected utility** of an action given the evidence, $\text{EU}(a|e)$, is just the average utility value

---

[1]Classical decision theory leaves the current state S_0 implicit, but we could make it explicit by writing: $P(\text{result}(a) = s'|a, e) = \sum_s P(\text{result}(s, a) = s|a) \cdot P(S_0 = s|e)$

of the outcomes, weighted by the probability that the outcome occurs:

$$\text{EU}(a|e) = \sum_{s'} P(\text{result}(a) = s'|a, e) \cdot U(s')$$

The principle of **maximum expected utility**(MEU) says that a rational agent should choose the action that maximizes the agent's expected utility:

$$\text{action} = \underset{a}{\text{argmax}} \ \text{EU}(a|e)$$

2. Constraints on rational preferences

Agent's preferences can be described with the following notation:

- $A \succ B$ the agent prefers A over B.
- $A \sim B$ the agent is indifferent between A and B.
- $A \succsim B$ the agent prefers A over B or is indifferent between them.

A and B can be states of the world, but can be expressed as uncertainty. Set of outcomes for each action is a **lottery**, each action as a ticket. A lottery L with possible outcomes $S_1, ..., S_n$ with probabilities $p_1, ..., p_n$:

$$L = [A, 1]$$
$$L = [p, A, (1-p), B]$$
$$L = [p_1, S_1; ... p_n, S_n]$$

3. Constraints or axioms on preferences

1. **Orderability**: the agent cannot avoid the decision making, he must choose either the first or the second option.

$$(A \succ B) \vee (B \succ A) \vee (A \sim B)$$

2. **Transitivity**:

$$(A \succ B) \wedge (B \succ C) \rightarrow (A \succ C)$$

For example, $A \succ B \succ C \succ A$ does not make sense. Since the graph creates the cycle where agent cannot decide on one preference.

3. **Continuity**: If some lottery B is between A and C, them there is some probability p for which the rational agent will be indifferent between getting B for sure and the lottery that yields A with probability p and C with probability $1-p$. (The Monty Hall problem)

$$A \succ B \succ C \rightarrow \exists p \ [p, A; 1-p, C] \sim B$$

4. **Substitutability**: If an agent is indifferent between two lotteries A and B, then the agent is indifferent between two more complex lotteries that are the same except that B is substituted for A in one of them. Long story short: it does not matter what lottery to choose. This also holds for $\succ$:

$$A \sim B \rightarrow [p, A; 1-p, C] \sim [p, B; 1-p, C]$$

5. **Monotonicity**: Suppose two lotteries have the same two possible outcomes, A and B. If an agent prefers A to B, then the agent must prefer the lottery that has a higher probability for A:

$$A \succ B \rightarrow (p > q \equiv [p, A; 1-p, B] \succ [q, A; 1-q, B])$$

6. **Decomposability**: Compound lotteries can be reduced to simpler ones using the laws of probability. This has been called the *"no fun in gambling"* rule because it says that two

consecutive lotteries can be compressed into a single equivalent lottery:

$$[p, A; 1 - p, [q, B; 1 - q, C]] \sim$$
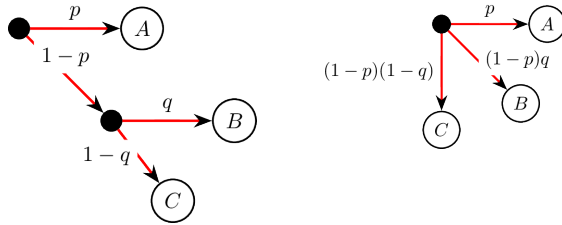$$\sim [p, A; (1-p)q, B; (1-p)(1-q), C]$$

Figure 1: **On the left**: A compound lottery. **On the right**: A decomposed lottery

4. Preferences lead to utility

- **Existence of Utility Function**: If an agent's preferences obey the axioms of utility, then there exists a function U such that:

$$U(A) > U(B) \equiv A \succ B$$
$$U(A) = U(B) \equiv A \sim B$$

- **Expected Utility of a Lottery**: The utility of a lottery is the sum of the probability of each outcome times the utility of that outcome:

$$U([p_1, S_1; ...; p_n, S_n]) = \sum_i p_i \cdot U(S_i) = \text{EU}(a|e)$$

The preceding theorems establish that a utility function exists for *any rational agent*, but they do not establish that it is *unique*.

5. Agent's utility function

- The best possible prize: $U(S) = u_\top$
- The worst possible prize: $U(S) = u_\bot$
- Normalized utilities: $u_\top = 1, u_\bot = 0$

- Standard lottery: $[p, u_\top; (1-p), u_\bot]$

6. Some paradoxes

It will usually be the case that an agent prefers more money to less, all other things being equal. We say that the agent exhibits a **monotonic preference** for more money.

6.1. The utility of money

Suppose a lottery: you can take a $1 million or flip a coin to get $2.5 million. Assuming the coin is fair we can write the **expected monetary value(EMV)**:

$$\text{EU}(\text{Accept}) = \frac{1}{2}U(S_k) + \frac{1}{2}U(S_{k+2.5})$$

$$\text{EU}(\text{Decline}) = U(S_{k+1})$$

where S_k is the current wealth in $ millions. Most people will choose the guaranteed option, even though it is not the most optimal(the first option EU is $1.25 million, the second one is $1 million). However, for a billionaire the gamble can be an option to do.
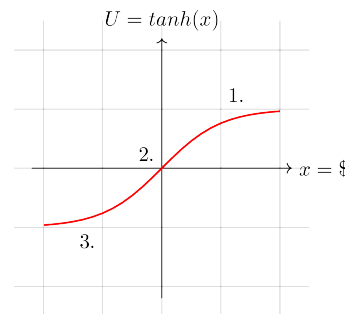

Figure 2: Not sure where this curve comes from and it is empirical curve, but it resembles $\tanh(x)$ function, but not exactly. Each number describes risk assessment.

In Figure 2 Grayson[2] or someone has created an empirical view on people behaviors depending on their monetary status:

1. **Risk-averse** - they prefer a sure thing than a gamble.
2. **Risk-neutral** - the agent gambles small sums to assess the probabilities.
3. **Risk-seeking** - the agent rather to play the gamble.

The difference between the Expected Monetary Value(EMV) of a lottery and its certainty equivalent is called the **insurance premium**, e.g. in the lottery above insurance premium $= 2.5 - 1 = 1.5 million.

6.2. Human judgment and irrationality

The evidence suggests that humans are "predictably irrational".

The best-known problem is the Allais paradox (Allais, 1953). People are given a choice between lotteries A and B and then between C and D, which have the following configuration:

Lottery	Conditions	EMV	Max choice	Human choice
A	80% $4k	$3200	$A \succ B$	$B \succ A$
B	100% $3k	$3000		
C	20% $4k	$800	$C \succ D$	$C \succ D$
D	25% $3k	$750		

[2]Grayson, C. J. (1960). Decisions under uncer- tainty: Drilling decisions by oil and gas operators. Tech. rep., Division of Research, Harvard Business School.

As seen people do not always choose the biggest EMV value. One explanation for the apparently irrational preferences is the **certainty effect** (Kahneman and Tversky, 1979): people are strongly attracted to gains that are certain.

6.3. Ellsberg paradox

Here the prizes are fixed, but the probabilities are underconstrained. Your payoff will depend on the color of a ball chosen from an urn. You are told that the urn contains 1/3 red balls, and 2/3 either black or yellow balls, but you don't know how many black and how many yellow. Again, you are asked whether you prefer lottery A or B; and then C or D:

Lottery	Cond-itions	Prob-ability	Human choice
A	$100 for a red ball	$\frac{1}{3}$	$A \succ B$
B	$100 for a black ball	$[0; \frac{2}{3}]$	
C	$100 for a red or yellow ball	$[\frac{1}{3}; \frac{3}{3}]$	$D \succ C$
D	$100 for a black or yellow ball	$\frac{2}{3}$	

Most people elect the known probability rather than the unknown unknowns.

7. Dominance

If there is an attribute set $X = (X_1, ..., X_n)$ then we can determine the dominance in Figure 3.
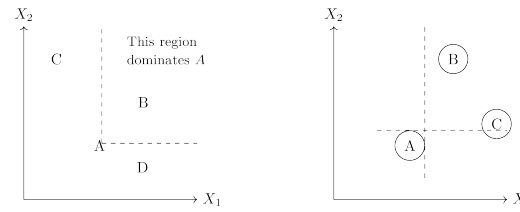

Figure 3: Right: Deterministic, option A is strictly dominated by B, but not by C or D. Left: Uncertain: A is strictly dominated by B, but not by C

8. Preference structure and multiattribute utility

Suppose we have n attributes, each of which has m distinct possible values. To specify the complete utility function $U(x_1, ..., x_n)$, we need m^n values in the worst case. Now, the worst case corresponds to a situation in which the agent's preferences have no regularity at all.

Representation theorems show that an agent with a certain kind of preference structure has a utility function:

$$U(x_1, ..., x_n) = F\left[f_1(x_1), ..., f_{n(x_n)}\right]$$

where F is, we hope, a simple function such as addition.

8.1. Preferences without uncertainty

For deterministic environments the agent has a value function $V(x_1, ..., x_n)$; the aim is to represent this function concisely.

Two attributes X_1 and X_2 are preferentially independent of a third attribute X_3 if the preference between outcomes (x_1, x_2, x_3) and (x'_1, x'_2, x_3) does not depend on the particular value x_3 for attribute X_3.

Consider the example of minimax in Figure 4. If node A value is changed to 40 instead of 4, it does not matter and the root value stays the same. The attributes exhibit **mutual preferential independence**.
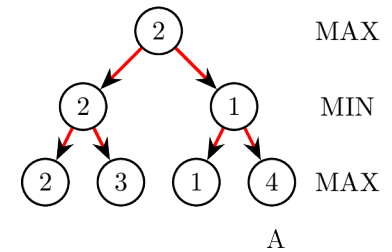

Figure 4: Minimax example

If attributes $X_1, ..., X_n$ are mutually preferentially independent, then the agent's preference behavior can be described as maximizing the **additive value function**:

$$V(x_1, ..., x_n) = \sum_i V_i(x_i)$$

The function with the state input:

$$V(S) = \sum_i V_i(x_i(S))$$

Need to remember that the function attributes must be expressed as the one unit of measurement and that the attributes most often have some sort of weight:

$$V(S) = \sum_i k_i \cdot V_i(x_i(S))$$

8.2. Preferences with uncertainty

A set of attributes is **mutually utility independent** (MUI) if each of its subsets is utility-independent of the remaining attributes.

MUI implies that the agent's behavior can be described using a multiplicative utility function. Example of three attributes:

$$U = k_1 U_1 + k_2 U_2 + k_3 U_3 +$$
$$+ k_1 k_2 U_1 U_2 + k_2 k_3 U_2 U_3 + k_3 k_1 U_3 U_1 +$$
$$+ k_1 k_2 k_3 U_1 U_2 U_3$$

For more mathematics reference Keeney K.L., Raiffa H. (1976) Decisions with multiple objectivesL preferences and value trade-off.

Example of uncertain preferences can be presented within the minimax example in Figure 5. Changing the leaf node from 4 to 40, does change the root result from 2.1 to 4.9.
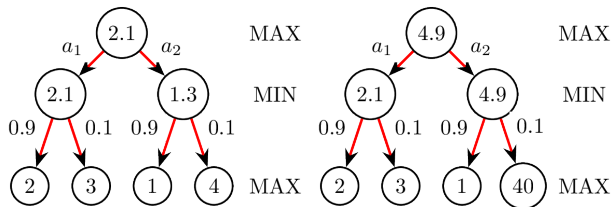

,

Figure 5: Minimax with uncertainty example

9. Decision networks

Figure 6 shows a decision network for the airport siting problem. It illustrates the three types of nodes used:
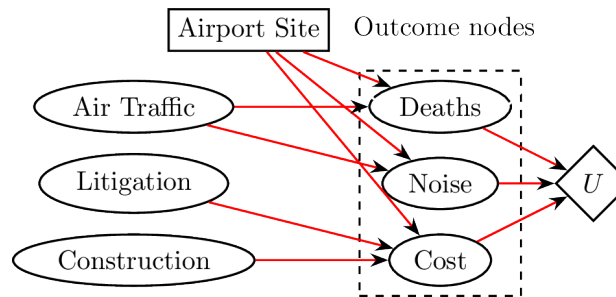

Figure 6: A simple decision network for the airport-siting problem

1. **Chance nodes**(ellipses) represents random variables. The agent could be uncertain about the construction cost, the level of air traffic and the potential for litigation, and the Deaths, Noise, and total Cost variables, each of which also depends on the site chosen.

2. **Decision node**(rectangles) represents points where the decision maker has a choice of actions.

3. **Utility node**(diamonds) represents the agent's utility function.

In Figure 6 outcome nodes can be omitted and the graph will be simplified.

Need to take into account:

- Decision maker forms the list of priorities and preferences
- Decision analyst creates the structure of priorities

Evaluating decision networks:

1. Set the evidence variables for the current state.

2. For each possible value of the decision node:
   1. Set the decision node to that value.
   2. Calculate the posterior probabilities for the parent nodes of the utility node, using a standard probabilistic inference algorithm.
   3. Calculate the resulting utility for the action.
3. Return the action with the highest utility.

10. Decision analysis methodology

1. Defines problem's scope
   1. What are the possible actions
   2. What are the possible states and outcomes
   3. What random variables affect the state
   4. What is the input data to calculate probabilities
2. Defines network's topology
3. Defines probabilities: $P(\text{effect}|\text{reason})$
4. Defines an utility function
5. Inputs an available example
6. Checks the decision network's actions
7. Tries to get a new example (cost of getting knowledge vs accuracy)
8. Sensitivity analysis is done