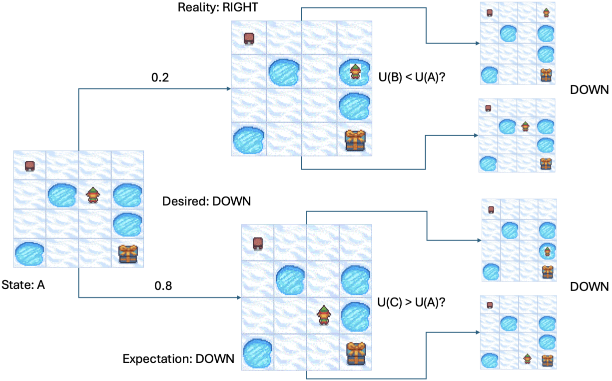
Artificial Intelligence Week 5: Utility Theory, Max-Expected Utility (MEU), Value of Information

Debasis Ganguly
University of Glasgow, Glasgow, UK

October 20, 2025

Overview

Maximization of Expected Utility



A rational agent

- **Accumulates the expected values** along each path.
- Selects the action that **leads to the most favorable state**.

Challenges

- How to design the utility function?

Expected values of random variables

- Expectation of a random variable is a key concept for (rational) decision making.
- Let X be a random variable with a finite number of outcomes $x_1, \ldots, x_k$.
  - Is Next_State (s_{t+1}) a random variable in a stochastic environment?
- Probabilities of each $p_1, p_2, \ldots, p_k$.
  - In an AI environment each action a on a state s_t is a probability distribution.
- The expectation of X is defined as:
- $\mathbb{E}_{p(x)}[X] = \sum_{i=1}^{k} x_i \, p(x_i) = x_1 p_1 + \cdots + x_k p_k$.

Expectation example

- Let X represent the outcome of a roll of a fair six-sided dice.
- Possible outcomes: $X \in \{$⚀, ⚁, ⚂, ⚃, ⚄, and ⚅$\}$, all of which are equally likely with a probability of $1/6$.
- The expectation of X is:
$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$
- Consider a biased dice, e.g. $P = (1/12, 1/12, 2/12, 2/12, 3/12, 3/12)$
- Yields a different expected value:
$E[X] = 1 \cdot \frac{1}{12} + 2 \cdot \frac{1}{12} + 3 \cdot \frac{2}{12} + 4 \cdot \frac{2}{12} + 5 \cdot \frac{3}{12} + 6 \cdot \frac{3}{12} = 4.166$

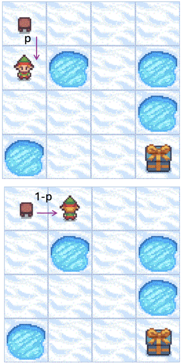
Expectations of functions of random variables

Functions of random variables

$$E[f(x)] = \sum_{i=1}^{k} f(x_i) p(x_i)$$

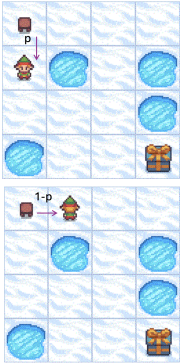
- With a fair die, the expectation of X^2 is:
- $E[X^2] = 1 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = 15.16.$
- $E[f(x)] \neq f(E[x])$.
- For the dice example:
  - What is $E[X]^2$?
  - What is $E[X^2] - E[X]^2$?

Maximization of Expected Utility



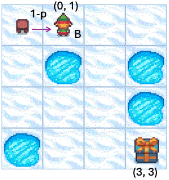
- Decision Theory: Choose actions based on desirability of certain outcomes.
- $P(Result(a) = s' | a, \mathbf{e})$ - $Result(a)$ is a **conditional random variable**.
- **Utility function**: Represents the *desirability* of a state.
  - $EU(a|\mathbf{e}) = \sum_{s}^{\prime} P(Result(a) = s' | a, \mathbf{e}) U(s')$
  - To be **rational**: Choose actions that maximizes the agent's **expected utility**.
  - $action = \mathrm{argmax}_a \, EU(a|\mathbf{e})$

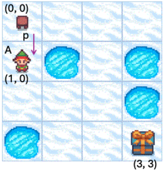
Maximization of Expected Utility



- MEU principle: (Almost!) the *theory of everything* in AI.
  - $EU(a|\mathbf{e}) = \sum_{s'} P(Result(a) = s'|a, \mathbf{e})U(s')$
  - $\mathbf{e}$: Captures the current knowledge that the agent has on the environment.
- The utility function has to accurately represent the performance measure (P) in an environment (E).
- Not always easy to define such a utility function (Why?).
  - Difficult to compute $U(s')$ unless we know where does s' lead to?
  - Closer to a goal?
  - Closer to balancing trade-off between conflicting goals?

Axioms of Preferences





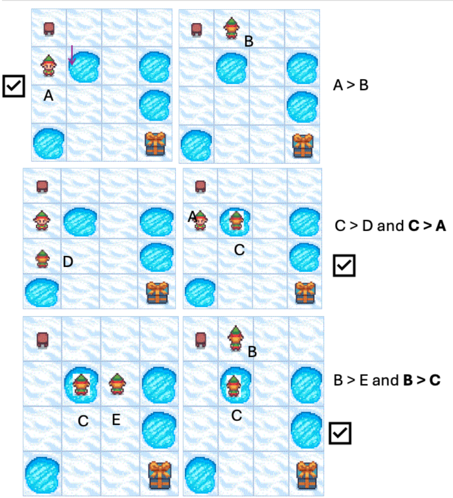
Operators

- $A \succ B$: An agent prefers A over B.
- $A \sim B$: An agent is indifferent between A and B.
- Is $A \succ B$ or $B \succ A$ or $A \sim B$? for the example shown?

What are the elements (As and Bs)?

- Results of an action in a stochastic environment.
- Distribution over states - $P(s|a)$
- As per formalism of utility theory:
  - $L = [p_1, S_1; p_2, S_2; \ldots, p_n, S_n]$.
  - Each S_i is also a lottery.
  - A certain event E can also be expressed as a lottery $L_E = [1, E]$.
    - Example: In the FLP, some action could be certain.

Axioms of Rationality in Utility Theory

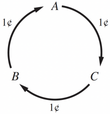


Orderability:

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

- No relative preferences $\implies$ No progress in an AI environment.
- **Mentimeter Quiz**

Transitivity:



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

- If not constrained, can create a cycle. Again - no progress.

Axioms of Rationality in Utility Theory

Continuity

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

- An agent needs to gamble going to a worse state (C) to make progress.
- Otherwise it gets stuck in state B.

- Let's say $U(B) = 100$. For what range of p, will an agent gamble in this example?
  - $p = 0.8$, and $U(A) = 200, U(C) = -10$?
  - $p = 0.2$ and $U(A) = 200, U(C) = -80$?
  - $p = 0.5$ and $U(A) = 200, U(C) = 0$?

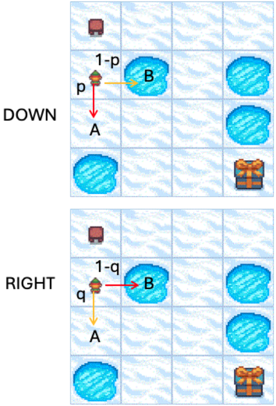
Axioms of Rationality in Utility Theory

Substitutability

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

- $U(A) = -20, U(B) = -15.$
- Attempting to go down can make the agent either fall on the left or the right holes.
- Both gambles are equally good (or bad).

Axioms of Rationality in Utility Theory



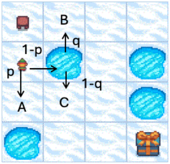
- Can reach A either by attempting to go DOWN or attempting to go RIGHT.
- Which uncertain path (lottery) is more preferable?

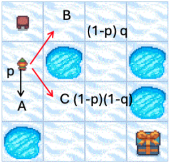
Monotonicity

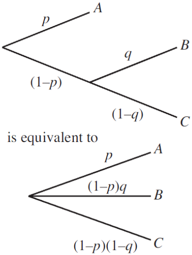
$$A \succ B \implies (p > q \iff [p, A; 1-p, B] \succ [q, A; 1-q, B])$$

- Two lotteries have the same possible outcome.
- State A is more rewarding.
- Agent should prefer the lottery which yields a higher probability for A.

Axioms of Rationality in Utility Theory







Decomposability

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

- This compresses a sequence of lotteries into a single lottery.
- What are the implications if this axiom is not satisfied?
  - We can't make sequential decision choices!

Preferences lead to Utility

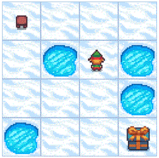
- Violation of axioms lead to irrational behaviour, e.g., we saw how violating transitivity can lead to irrationality.
  - **Mentimeter Quiz**: What about violating Continuity?
  - **Mentimeter Quiz**: What about violating Monotonicity?

Outcome from the axioms

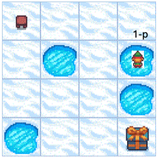
- Existence of a **utility function** $U(X)$ for each lottery X.
  - $A \succ B \iff U(A) > U(B)$
  - $A \sim B \iff U(A) = U(B)$
  - Utility of a lottery is the expected values (rewards) of its constituent states (lotteries).
  - $U([p_1, S_1; \ldots, p_n, S_n]) = \sum_{i=1}^{n} p_i U(S_i)$

Utility scales

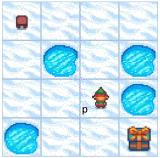
S = (2, 3)



U(2,3) = 0.8 U(3,3) + 0.2 U(2,4)



Agent moves falls into lake with probability = 1-*p*

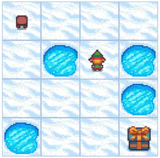


Agent moves closer to the goal with probability = *p*

- A utility function maps lotteries (states in a stochastic AI environment) to real numbers.
- **Preference elicitation**: how to choose this function for an environment?
- Compare a given state S to a standard lottery $L_p = [p, u_{max}; 1 - p, u_{min}]$
  - 'best possible prize' $u_{max} = 1$ with probability p
  - 'worst possible catastrophe' $u_{min} = 0$ with probability $(1 - p)$.
- Adjust lottery probability p until $S \sim L_p$.
- Finally: $U(S) = p$.

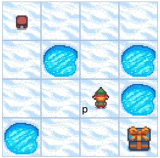
Revisit MEU Principle for Frozen Lake

S = (2, 3)



U(2,3) = 0.8 U(3,3) + 0.2 U(2,4)

Agent moves falls into lake with probability = 1-*p*



Agent moves closer to the goal with probability = *p*

- Choose the action that maximizes expected utility.
  - $EU(a|\mathbf{e}) = \sum_{s'} P(Result(a) = s'|a, \mathbf{e})U(s')$
  - $action = \operatorname{argmax}_a EU(a|\mathbf{e})$
- $EU(Down|S_{2,3}) = pU(S_{3,3}) + (1-p)U(S_{4,2})$
- $EU(Up|S_{2,3}) = qU(S_{1,3}) + (1-q)U(S_{2,2})$
- Is $action = Down|S = (2,3)$ rational?
  - **Mentimeter Quiz**:
  - When $p = 0.3, q = 0.3$?
  - When $p = 0.4, q = 0.9$?
- MEU $\rightarrow$irrational if utility function doesn't address the performance measure.
- E.g., for frozen lake: not enough high reward for getting close to the goal; not enough penalization for falling into a hole.

MEU Numerical Exercise (To be worked out in class)

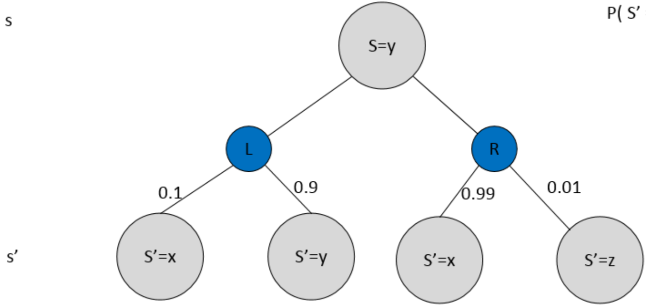
x	y	z

U(S=x)=10
U(S=y)=100
U(S=z)=1000

Transition model P(s'|s,a):

P(S' = x | A=Left, S = y) = 0.1
P(S' = y | A=Left, S = y) = 0.9

P(S' = x | A=Right, S = y) = 0.99
P(S' = z | A=Right, S = y) = 0.01



Utility: Root in Economics

- Suppose you have just won $1M\$$ in a TV game show. The host now offers you a choice: either you can take the $1M\$$ prize or you can gamble it on the flip of a coin.
  - Heads: end up with nothing,
  - Tails: you get $2.5M\$$.
- Will you gamble? **Most humans** won't.

Rational by MEU principle? No.

- $EU(No\_Gamble) = 1M$
- $EU(Gamble) = 0.5 \times 2.5M = 1.25M$

- But surely, accepting $1M$ sounds like a rational choice. So, what's wrong?
- Utility has to be computed **relative to** a person's (an agent's) current money (utility).

Utility: Root in Economics

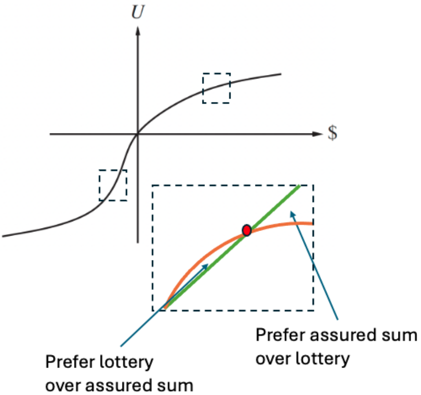
- Suppose you have just won $1M\$$ in a TV game show. The host now offers you a choice: either you can take the $1M\$$ prize or you can gamble it on the flip of a coin.
  - Heads: end up with nothing,
  - Tails: you get $2.5M\$$.
- Let S_k denote an agent's total wealth.

Expected Utilities

- $EU(No\_Gamble) = U(S_{k+10^6})$
- $EU(Gamble) = 0.5 \times U(S_{k+2.5\times10^6}) + 0.5 \times U(S_k)$

- For a sufficiently large value of k (e.g., $k = 1B$), the rational choice is to take the risk.
- Utility in this case is not just the wealth value; because the utility of the 1st million is high, whereas the utility of the 2nd million is less and so on.

Utility Curve



- Usually, the relation between current wealth and utility is **logarithmic** on both sides of the x-axis.
- $U(L) < U(S_{EMV}(L))$ – **risk-averse** behaviour.
- $U(L) > U(S_{EMV}(L))$ – **risk-seeking** behaviour.

Human Decision making isn't always rational!

Allais Paradox

Choose between (A, B) and then between (C, D).

A: 80% chance of \$4000	**C**: 20% chance of \$4000
B: 100% chance of \$3000	**D**: 25% chance of \$3000

$A \succ B$? $C \succ D$?

$EU(A) = 3200 > 3000 = EU(B)$

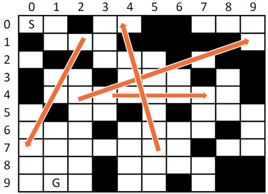
$EU(C) = 800 > 750 = EU(D)$

- Most people choose A over B (sure thing), and C over D (higher EMV) $\implies$ Inconsistency.

So, what's happening?

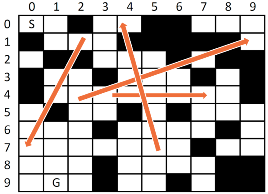
- A cognitive bias called **certainty effect** - People are strongly attracted to gains that are certain.
- People reduce cognitive load (of dealing with probabilities).
- **Minimize regret** (which can be **incorporated into the utility function**).

MEU for the Maze with Wormholes



- Different to a standard maze, it is a stochastic environment.
  - Action a (e.g., $a = R$) from each cell is a lottery $L_{i,j,R} = [p, L_{i,j+1}; 1 - p, L_{i',j'}]$
  - Each state's utility can be its distance from the goal: $U_{S_{i,j}} = |i - goal.i| + |j - goal.j|$
- What is the role of $\mathbf{e}$?
  $EU(a|\mathbf{e}) = \sum_{s'} P(Result(a) = s'|a, \mathbf{e})U(s')$
  - $P(Result(a) = s'|a, \mathbf{e})$ changes if the agent has experienced taking a particular wormhole before!
  - For $\mathbf{e} = e_1$ it can decide to gamble by MEU
  - Later on for $\mathbf{e} = e_2$ MEU makes it avoid a bad wormhole, and also makes it take a good wormhole.

Value of Information (VoI)



- New evidence e_{new} can alter an agen'ts belief of $P(Result(a)) = s'|\mathbf{e}) \neq P(Result(a)) = s'|\mathbf{e}, e_{new})$
- What kind of new evidences can change such beliefs in the maze with wormholes?
  - Prior discovery of wormholes.
  - The agent now knows which wormhole leads where.
- What does these changed beliefs do?
  - They change $EU(a|\mathbf{e})$ values, and (potentially) the optimal action taken
    - $a^* = \text{argmax}\, EU(a|\mathbf{e})$.
  - Thus MEU principle may lead to a different action taken.

Value of Information

Intended	Desired	Deflected
↓	↓	→
→	→	↑
↑	↑	←
←	←	↓

The agent can now check if A or B has now turned into safe ice! Which one is more valuable?

VoI Properties

- Agent's initial evidence: $\mathbf{e}$.
- Value of current best action (α):
  - $EU(\alpha|\mathbf{e}) = \max_a \sum_{s'} P(Result(a) = s'|a, \mathbf{e})U(s')$
- E_j: some environment variable, the value of which e_j is now known to us.
  - Example: $S_{i,j} = 1$ (whether a state is an entry to a wormhole or a hole got frozen)
- Value of new best action:
  - $EU(\alpha|\mathbf{e}, e_j) = \max_a \sum_{s'} P(Result(a) = s'|a, \mathbf{e}, e_j)U(s')$
- To compute the information that this value provides, we marginalize over other possible values (say k of them) of E_j.

Value of Information

VOI Definition

$$VOI_{\mathbf{e}}(E_j) = (\sum_k P(E_j = e_{j,k}|\mathbf{e})EU(\alpha|\mathbf{e}, E_j = e_{j,k})) - EU(\alpha|\mathbf{e})$$

- In simple words:
  - Compute expected value of information = expected value of best action given the information minus expected value of best action without information.
- Allows the agent to decide to explore or not:
  - the gain in obtaining the evidence
  - just sticking with the current evidence (i.e. not obtaining new evidence)

Value of Information - Example

- Consider n bins out of which **exactly one** contains a prize worth $C\$$ with uniform probability $P_i = 1/n$.
- Cost of opening a bin: $C/n\$$.
- Expected gains $= (1/n)C + (n - 1/n)0 = C/n$.
- Expected profit $= C/n - C/n = 0$.

Risk-neutral behaviour

You are indifferent to the lottery.

Additional information from an expert

- **Intuitive reasoning**: May be I should take "expert help" because either I know the prize or my chances of winning increases $(1/(n-1)$ - one less uncertain factor)!
- But what is the exact value of the information?

Value of Information - Example

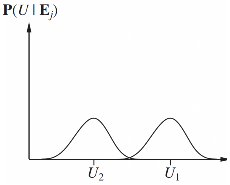
- Best case: With probability $1/n$ the opened bin "b" would contain the prize. In that case you will bet on b and win C.
  - $Profit_{best} = C - C/n = (n-1)C/n$
- Worst case: With probability $(n-1)/n$, the bin won't contain the prize. You then will bet on some other bin.
  - New Prob of winning prize $= 1/(n-1)$ (**beliefs changed**)
  - $Profit_{worst} = C/(n-1) - C/n = C/(n(n-1))$
- Expected profit $= \frac{1}{n} Profit_{best} + \frac{n-1}{n} P_{worst} = C/n$
- The value of the information is as much as the cost of the lottery ticket.
- Value of expert information in this case is 0!

Another VoI example

Two routes in winter:

- straight highway, low pass
- winding dirt road over top

- **Additional satellite image information** before deciding what route to take?
  - Consider: Winter with high chance of road blocks due to avalanches.



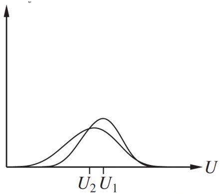
- Regardless of adverse weather conditions, $U(a_1) > U(a_2)$

Another VoI example

Two routes in winter:

- two winding roads of slightly different lengths

- **Additional satellite image information** before deciding what route to take?
  - Consider: Winter and you're **carrying an injured person**.

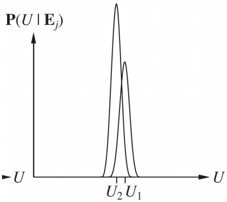


- Both $U(a_1)$ and $U(a_2)$ exhibit greater uncertainty (high variance).
- Additional information is **valuable**.

Another VoI example

Two winding routes in summer:

- **Additional satellite image information** before deciding what route to take?
  - Consider: Summer, so **no almost 0 chance of road blocks due to avalanche**.



- Both $U(a_1)$ and $U(a_2)$ are similar (I don't mind driving through either of them).
- Additional information is also **less valuable** because most likely **it won't change my plan**.

Summary

- The real-world is uncertain but we still need to make decisions and take actions based on uncertain outcomes and inputs.
- **Maximum Expected Utility (MEU)** allows rational, decisions when the outcomes of actions are uncertain.
- **Next week**: **Sequential decision**, i.e. how do you make a decision today knowing that you need to make new decision tomorrow.
- Week 10 (**Nov 24th**): Guest Lecture by Aritra Chakravarty, Head of the **Agentic AI Group** at **Lloyds Bank**.
- Please make sure that you **attend the lecture in-person**!