AI by Russel & Norvig

Making Simple Decisions

Decision Theoretic Agent

An appropriate action

- A decision theoretic agent is an agent that can make rational decisions based on its beliefs (what it believes) and desires (what it wants).

"correct"

- The situation can involve uncertainty or conflicting goals.

- The outcomes of actions can be judged with a continuous measure of quality.

Utility of an action

Combining Beliefs and Desires under Uncertainty

- Episodic Environment:

  - The agent's experience is divided into atomic episodes.

  - In each episode the agent receives a percept and then performs a single action.

  - The next episode does not depend on the actions taken in previous episodes.

- Deterministic Environment:

  Next state of the environment is completely determined by the current state and the action executed by the agent

Combining Beliefs and Desires under Uncertainty

- Non-Deterministic Environments

  Actions are characterized by their possible outcomes, but no probabilities are attached to them.

  - In a non-deterministic environment, the agent is required to succeed for all possible outcomes of its actions.

- An uncertain environment is the one which is partially observable or not deterministic.

Kriegspiel

Belief states

Combining Beliefs and Desires under Uncertainty

$s_0 \xrightarrow{a} s_1$

- State that is the deterministic outcome of taking action *a* in state s_0 is given by RESULT(s_0,*a*) .

(s_0, a) $s_1 = \text{Result}(s_0, a)$

- Here we consider the environment as non-deterministic and a partially observable environment

outcomes ?

exact state is not known

- We define RESULT(*a*) as a function that returns a random variable giving possible outcome states.

- The probability of outcome *s′*, given evidence observations **e**, is written

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

$e \quad ? \xrightarrow{a} ?$

$P[s' | a, e]$

Utility Function

- The utility function *U(s)* assigns a value to the state *s* which refers to its desirability

$$a \rightarrow s' \quad P. \quad U(s')$$

- The expected utility of a function is defined as

oodness of an action a

$$EU(a\,|\,\mathbf{e}) = \sum_{s'} \underbrace{P(\mathrm{RESULT}(a) = s'\,|\,\mathrm{a}, \mathbf{e})}_{\text{weight}}\, U(s')$$

weighted average of the utility of the outcome

- A rational agent should choose an action that maximises the agent's expected utility.

$$action = \underbrace{\operatorname*{argmax}_{a}}_{\text{Best Action}} EU\,(a|\mathbf{e})$$

Maximum Expected Utility.

MEU Principle

$$\underline{\underline{action}} = \operatorname*{argmax}_a EU(a|\mathbf{e})$$

Rational decision making

- This principle is called as the MEU (Maximum Expected Utility) principle
- Challenge lies in operationalising the MEU principle
  - Estimating the state of the world?
  - Computing $P(\textsc{Result}(a) = s' \mid a, \mathbf{e})$
  - Computing outcome utilities $U(s')$

Utility Theory

There is no universal utility function or preferences

$A > B > C > A$

$\uparrow X$

how to define the Utility function in a rational way.

Capture the preferences of the agent.

The Basis of Utility Theory

reflected in the utility function

- Is it important for a utility function to exist?

  Or can an agent act rationally by simply expressing preferences between states? Yes.

- What's the justification for the MEU principle… the way it is formulated?

?

Constraints on rational preferences

A, B : outcomes

- Notation for preferences

$A \succ B$	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.

- Here *A* and *B* are outcomes (states or *possible* states)
- An action leads to outcomes.
  - An outcome characterised by certainty are referred to as atomic state
  - An outcome characterised by uncertainty is referred to as lottery

Lottery

- A lottery defines a set of possible outcomes.
  - The possible outcomes can refer to an atomic state, or another lottery.
  - The uncertainty regarding the possible outcomes is described by associating a probability value with the outcomes.

$$L = [p_1, S_1;\ p_2, S_2;\ \ldots\ p_n, S_n]$$

??

Constraints on rational preferences

$A > B$. A ___ $S = [p, A, (1-p), B]$

- It is easy to define preference over states, but if the outcome is a lottery then how do we prefer one lottery over another?
- To help derive preference between lotteries, we require the preference relation to obey 6 constraints:
  - Orderability
  - Transitivity
  - Continuity
  - Substitutability
  - Monotonicity
  - Decomposability

rationality of the agent's preferences

Constraints on rational preferences

- Orderability

  Given any two lotteries A and B

  Exactly one of $(A \succ B)$, $(B \succ A)$, or $(A \sim B)$ holds.

  $\Rightarrow$ Agent must take a decision on the preference

- Transitivity

  Given any three lotteries

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

- Continuity

  $$A \succ B \succ C \Rightarrow \exists p \; \underbrace{[p, A;\ 1-p, C]}_{\text{Lottery}} \sim B$$

- Substitutability

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

  also holds if we substitute $\succ$ for $\sim$ in this axiom

Constraints on rational preferences

- Monotonicity

$$\underline{A \succ B} \Rightarrow (\underbrace{p > q}_{} \Leftrightarrow \underbrace{[p, A;\ 1-p, B]}_{p.} \succ \underbrace{[q, A;\ 1-q, B]}_{q.})$$

- Decomposability

Two consecutive lotteries can be compressed into a single equivalent lottery.

$$[\underbrace{p, A}_{\text{state}};\ 1-p, \underbrace{[q, B;\ 1-q, C]}_{\text{lottery}}] \sim [p, A;\ \underline{(1-p)}q, B;\ \underline{(1-p)}(1-q), C]$$

nested lottery

flattened lottery.

Constraints on rational preferences

- These 6 constraints are known as axioms of utility theory.
- If an agent violates any of these axioms, then it will exhibit irrational behaviour in some situations.
  - For example, an agent that shows non-transitive behaviour can go bankrupt.

$$A \succ B \succ C \succ A$$

-100 -100 -100

A

Preferences lead to utility

- The axioms of utility theory say something about preferences say nothing about the utility function. But, these axioms lead to following consequences:
- Existence of Utility Function:

  If an agent's preferences obey the axioms of utility, then there exists a function U such that

$$
\begin{aligned}
U(A) > U(B) &\Leftrightarrow A \succ B \\
U(A) = U(B) &\Leftrightarrow A \sim B
\end{aligned}
$$

- Expected Utility of a Lottery:

$$U([p_1, S_1; \ldots; p_n, S_n]) = \sum_i p_i U(S_i)$$

EU.

- With action outcome being a lottery, the agent can still choose the best action that maximises the EU $\Longrightarrow$ MEU principle.

- The utility function is not unique.

  The agent's behaviour would not change if its utility function $U(S)$ were transformed

$$U'(S) = aU(S) + b$$

$$EU(a \mid \mathbf{e}) = \sum_{s'} P(\text{RESULT}(a) = s' \mid \mathbf{a}, \mathbf{e})\ U(s')$$

- If an environment does not involve uncertainty, then we use a value function or <u>ordinal utility function</u> to just give the preference ranking on states.

Value function

Action → outcome utility value

can be a lottery

Simple decision → Choice of action

Utility Functions

Utility Functions

- Utility functions allow a decision theoretic agent to choose the rational action that obeys the preference constraints.
    - However, a preference can be arbitrary (without any associated rationality).

$$\begin{tikzcd} \text{Pref} \arrow[r] \arrow[rr, bend right] & \text{Utility} \arrow[r] & \text{Choices.} \end{tikzcd}$$

Utility assessment and utility scales

- The agent's utility function is worked out using a process of preference elicitation
- There is no absolute scale for utilities, because you can always apply an affine transform to the utility function and get the same preference.
- But it is good to establish some scale on which utilities can be recorded.

- To establish a scale, we select two particular outcomes (prizes), say, (i) the best one, (ii) the worst one.

  We assign utilities to them $u_\top$ and $u_\perp$ respectively.

best

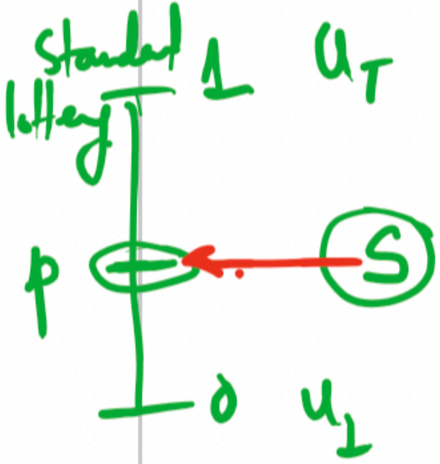
worst.

Lottery formulation helps to assign utility to various outcomes.

- Given the utility scale between $u_\top$ and $u_\perp$ we can assess the utility of any particular outcome S by asking the agent to choose between S and a standard lottery $[p, u_\top; (1-p), u_\perp]$
- The probability p is adjusted until the agent is indifferent between S and the standard lottery.
- If the utility scale is normalised (i.e. $u_\top = 1,\ u_\perp = 0$), then the utility of S is assigned as p .
- This process is repeated for each possible outcome.

$U_\top \quad U_\perp$

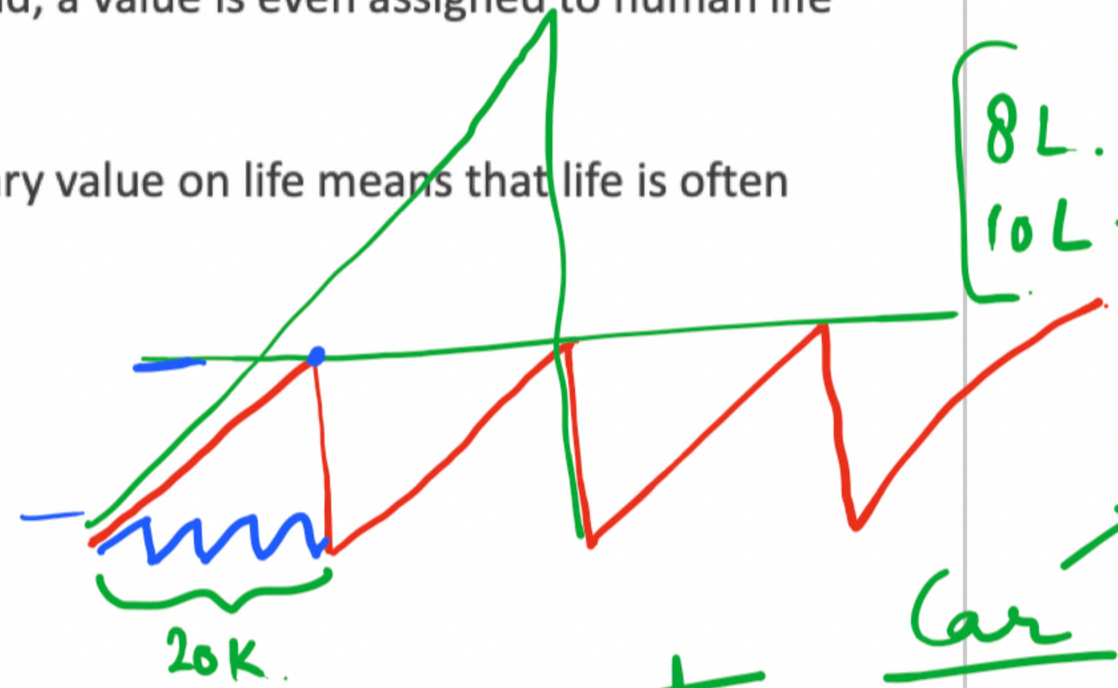
Outcome $\leftarrow$ p $[U_\top \ U_\perp]$



Value of human life !!

- When making decisions in the real world, a value is even assigned to human life while making tradeoffs.
- Paradoxically, a refusal to put a monetary value on life means that life is often undervalued.

Asbestos



8L.

10L ← ↓ 200 micromorts

1K per micromort

200 micromort

Car

400 km 1 micromort

1.5 Lac km

Russian Roulette

10^6 1 million

10^{-6} Micromort

10 Lacs.

The Utility of Money

Actions $\rightarrow$ Money

- Money can be considered as a candidate for a utility measure.

- An agent can exhibit a monotonic preference for more money.

- However, using money as a utility measure does not help in deciding preferences between lotteries involving money

Certain 10 L.

10 L.

30 L.

Expected Monetary value of the lottery

$\frac{0}{2} + \frac{30}{2} = 15L.$

The Utility of Money

- Let S_n denote the state when the total wealth possessed by the agent is \$$n$
- Consider that the current wealth of the agent is \$$k$ and hence the agent is in state S_k
- The expected utilities of Accept or Decline actions for the gamble are:

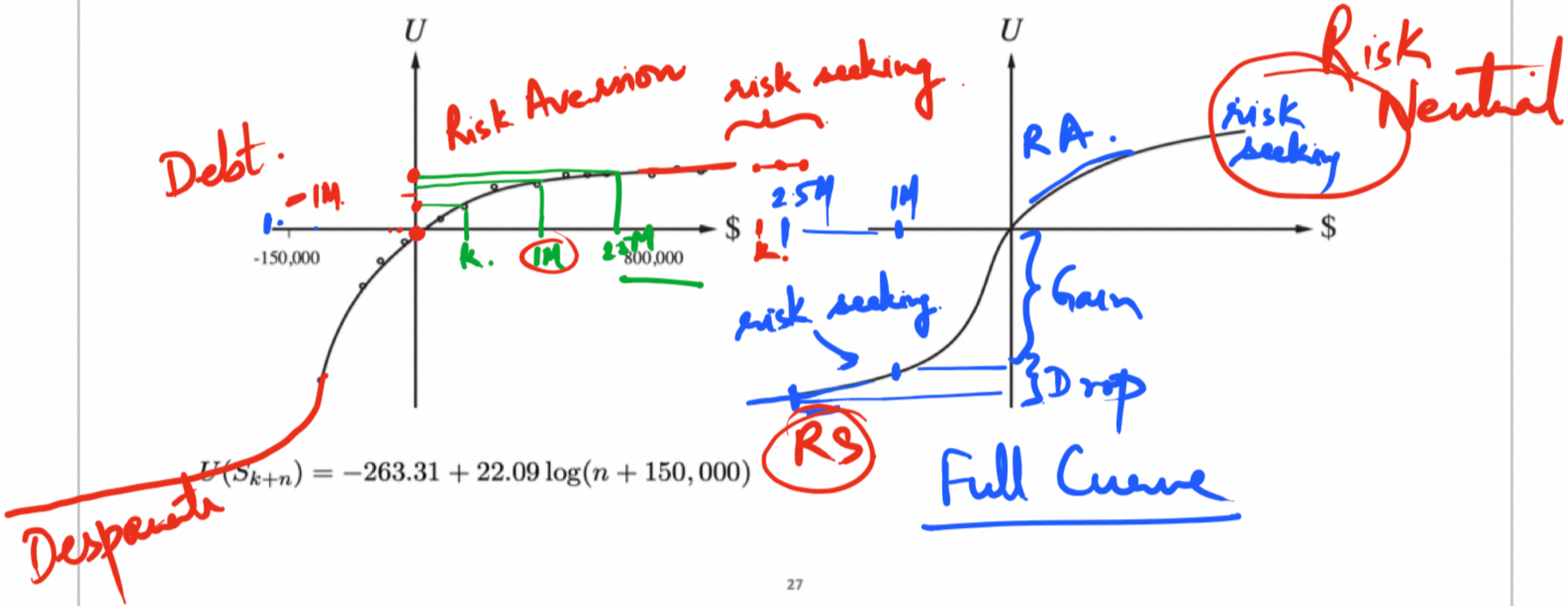
1M 2.5M

$$
\begin{aligned}
EU(Accept) &= \tfrac{1}{2}U(S_k) + \tfrac{1}{2}U(S_{k+2{,}500{,}000}) \,, \\
EU(Decline) &= U(S_{k+1{,}000{,}000}) \,.
\end{aligned}
$$

$U(S_{k+1M}) > \frac{1}{2}U(S_k) + \frac{1}{2}U(S_{k+2.5M})$

The Utility of Money

- Utility is not directly proportional to monetary value



$$U(S_{k+n}) = -263.31 + 22.09 \log(n + 150,000)$$

Insurance Premium for risk aversion

- Certainty equivalent of a lottery:

  The value an agent will accept in lieu of a lottery

- Insurance Premium:

  The difference between the EMV of a lottery and its certainty equivalent

EMV

House Theft.

0.01

10^6 worth

Expected Loss.

$= \underbrace{(0.01)\,10^6}_{} + \underbrace{(0.99)\,0}_{}$

EMV 10^4

Insure. certainty equivalent CE

.3 } Premium.

Working with estimates of expected utility

- A rational agent would choose the best action a^* as

$$a^* = \underset{a}{\operatorname{argmax}}\, EU(a|\mathbf{e})$$

outcome can be a lottery.

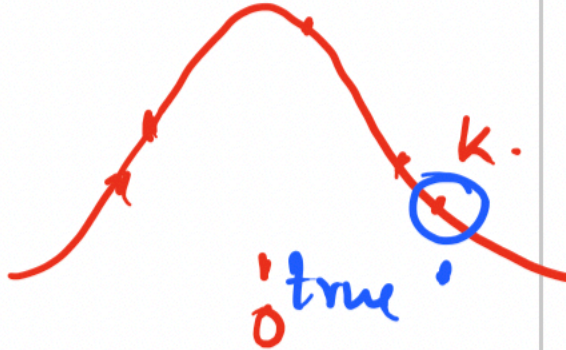
- We work with estimates of expected utility if the true expected utility is difficult to compute

$$\widehat{EU}(a|\mathbf{e})$$

- We conveniently assume that the estimates of expected utility are unbiased

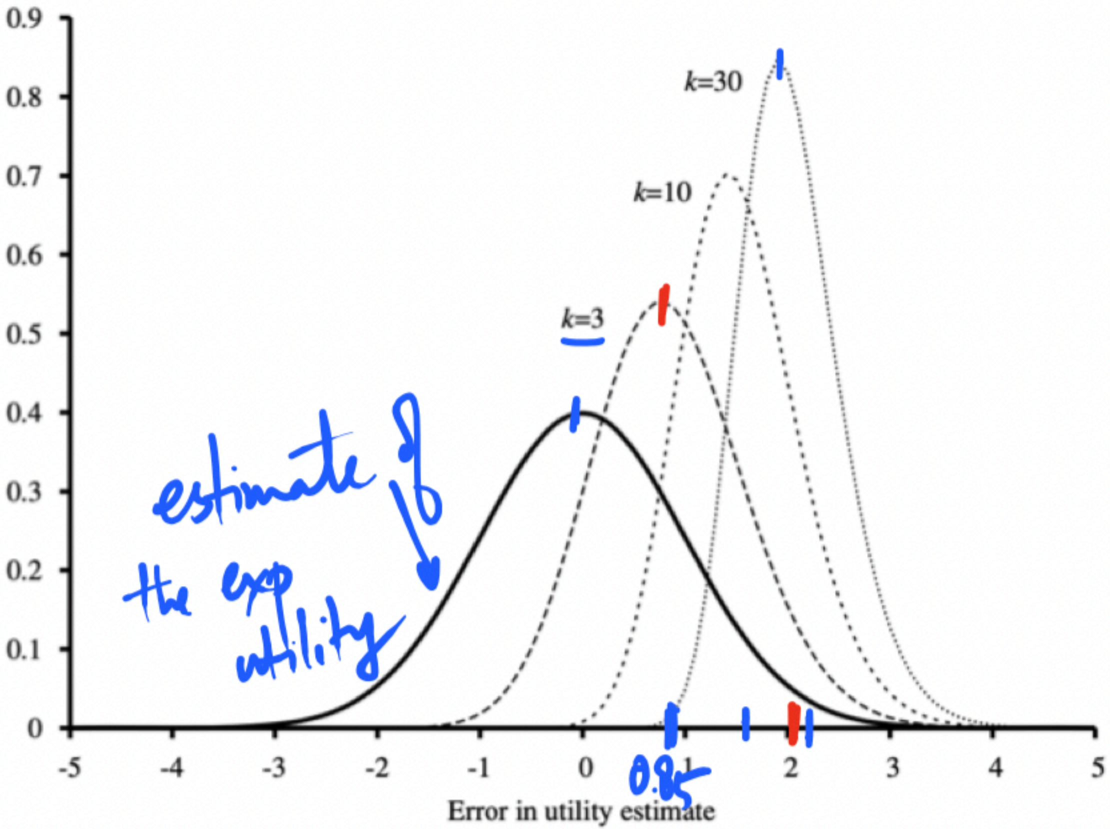
$$E(\underbrace{\widehat{EU}(a|\mathbf{e})}_{\text{estimated}} - \underbrace{EU(a|\mathbf{e})}_{\text{true}})) = 0$$

Disappointment due to overly optimistic $\widehat{EU}(a|\mathbf{e})$



- Consider a decision problem with k choices
    - Each choice has a true expected utility of 0
- Error in each utility estimate has 0 mean and a standard deviation of 1
- As we generate the estimates for the k choices, some of the errors will be negative (pessimistic estimate) and some will be positive (optimistic estimates).
- Since we select the action with the highest utility estimate, we introduce a bias by favouring the overly optimistic estimate.

Selection Bias: Making overly optimistic estimates



(max)

- Our optimistic selection eventually disappoints us.
- We can quantify the extent of our disappointment by observing the distribution of the maximum of the k estimates.
- Considering the k estimates to be the random variables $X_1, X_2, \ldots, X_k$, with density function $f(x)$,

  we compute the cumulative probability distribution for $\max\{X_1, X_2, \ldots . X_k\}$

$$P(\max\{X_1, \ldots, X_k\} \leq x) = P(X_1 \leq x, \ldots, X_k \leq x)$$

$$= P(X_1 \leq x) \ldots P(X_k \leq x)$$

Cumulative $$= F(x)^k$$

$F(x) \equiv P(X < x)$

- The density function for $\max\{X_1, \ldots, X_k\}$ is given by $\boxed{k\, f(x)\, (F(x))^{k-1}}$
- For $k = 3$, the mean is 0.85
- For $k = 30$, the mean is 2

- This tendency for the estimated expected utility of the best choice to be too high is called the optimiser's curse.

max

source

$\hat{EU}$ dont take this on face value.

$\left[P(\hat{EU} \mid EU) \quad P(EU) \right]$

0.85

Allais Paradox (Certainty effect)

- People are given choices between lotteries A and B and then between C and D.

$$\begin{array}{llll} A: & 80\% \text{ chance of } \$4000 & C: & 20\% \text{ chance of } \$4000 \\ B: & 100\% \text{ chance of } \$3000 & D: & 25\% \text{ chance of } \$3000 \end{array}$$

4 lotteries

$$U(3000) > U\left(\frac{80}{100} \times 4000\right).$$
$$U(3000) > U(3200)$$

$$U\left(\frac{20}{100}(4000)\right) > U\left(\frac{25}{100}\,3000\right)$$
$$U(800) > U(750).$$

- Most people prefer $\underline{B \text{ over } A}$ and C over D.
- This is irrational if we consider the expected monetary value (EMV) of the lotteries!

Ellsberg Paradox (Ambiguity Aversion)

- Here the prizes are fixed, but the probabilities are unconstrained. $\frac{1}{3}$ $\overbrace{\text{red, black, yellow}}^{2/3}$
- The prize depends on the colour of the ball chosen from an urn

$\frac{1}{3}$ A : \$100 for a red ball C : \$100 for a red or yellow ball $[\frac{1}{3}, 1]$

$[0, \frac{2}{3}]$ B : \$100 for a black ball D : \$100 for a black or yellow ball $\frac{2}{3}$

- The urn contains $\frac{1}{3}$ red balls and $\frac{2}{3}$ either black or yellow balls, but we do not know how many black or how many yellow.
- Most people prefer A over B, and also prefer D over C........... again irrational choices!

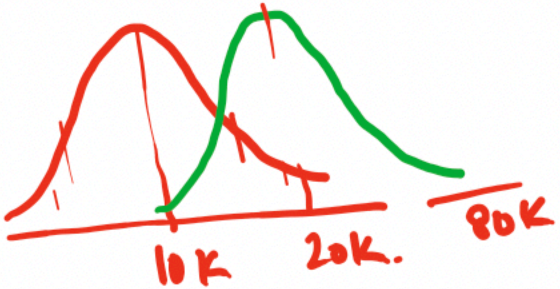
Other effects

- <u>Framing Effect</u>:

  Would you like to go for a medical procedure that has a "90% survival rate" or another one having a "10% death rate?"

- <u>Anchoring effect</u>:

  Would you like to buy a bicycle for Rs 50K ?



- Brain's decision making mechanism depends on the linguistic form (wordings) of the decision problem.
  - Brain's decision making mechanisms did not evolve to solve word problems with probabilities and prizes stated as decimal numbers.

Multi-attribute Utility Functions

Value function

Utility function

5 100

4 90

3 20

2 6

X

Multiattribute utility theory

- Multi-attribute utility theory is applicable to problems in which outcomes are characterised by two or more attributes.
- Selecting an airport site:

  - cost of land

  - distance from centers of population

  - noise of flight operations

  - safety issues due to local topology and weather

$\langle \#fan, \#AC, E_{rate} \rangle$.

Absence of Noise

Noise

- The attributes are $\mathbf{X} = X_1, \ldots X_n$

  Assigning attributes to these values gives us $\mathbf{x} = \langle x_1, \ldots, x_n \rangle$ where x_i can be numeric or discrete

  outcome.

- For simplicity of discussion we assume that higher values of an attribute correspond to higher utilities.

Simplified Scenario

- Can we make a decision without computing a single utility value as a function of the combined set of attributes?

$U(\langle \quad \rangle)$

- Yes, if there are specific preference structures present

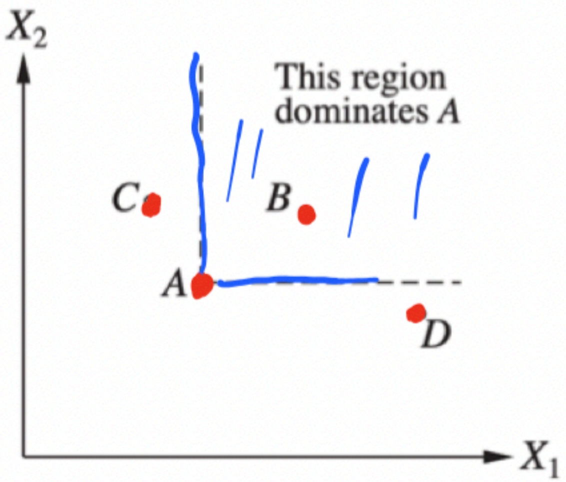
Check for dominance.

Regularity in the preferences.

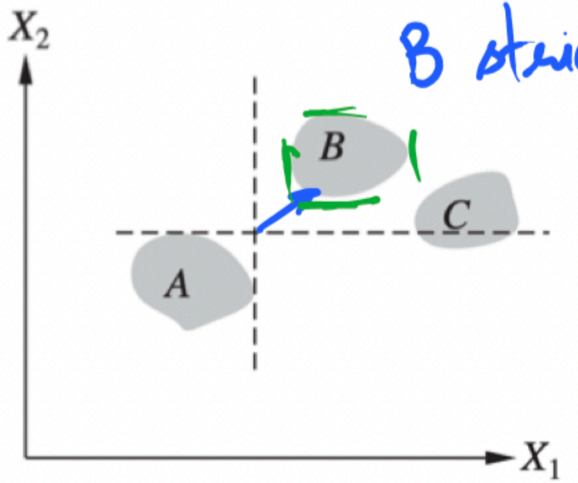
$$P(x_1, x_2 \ldots x_N) = \prod_i P(x_i \mid \text{Parents}(x_i))$$

Dominance

- Suppose that airport site S1 costs less, generates less noise pollution, and is safer than site S2.
- We can reject S2 since there is *strict dominance* of S1 over S2



Deterministic Case



Uncertain Case

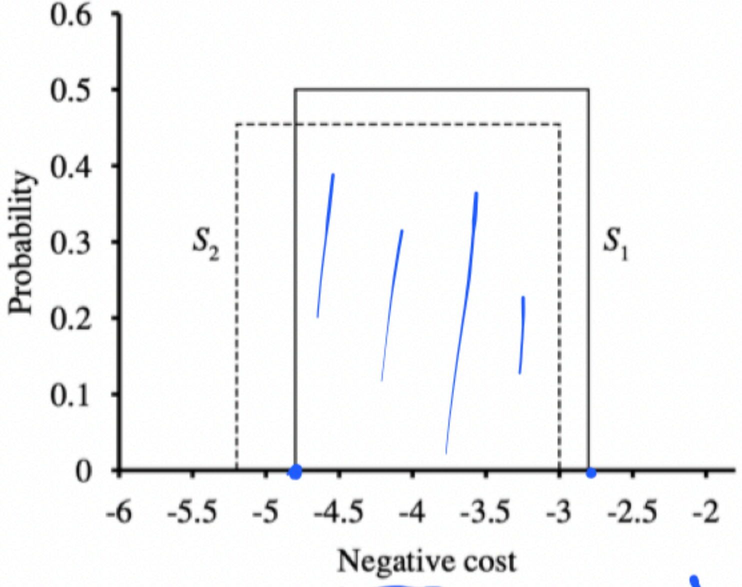
B strictly dominates A

Stochastic Dominance

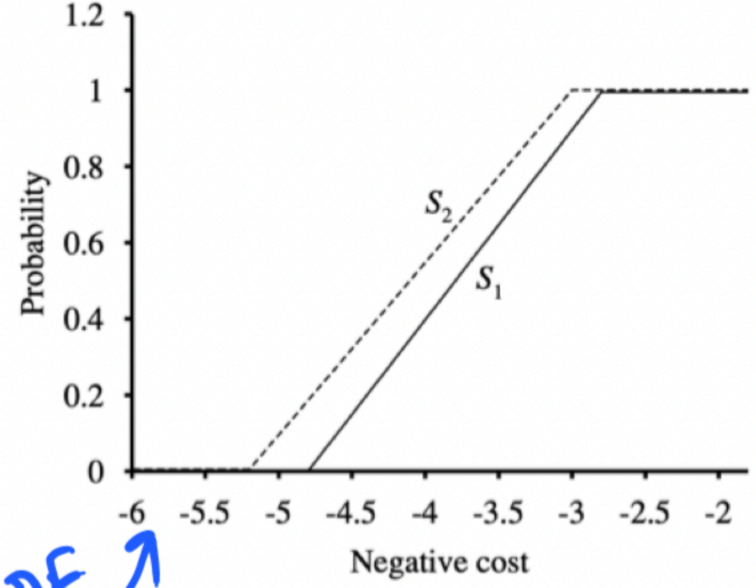
S_1 S_2

- The idea of dominance can be generalised to *Stochastic Dominance*

$[2.8, 4.8]$ $[3, 5.3]$



CDF



- An action A_1 stochastically dominates A_2, if for any monotonically nondecreasing utility function $U(x)$, the expected utility of A_1 is at least as high as the expected utility of A_2.

- If an action is stochastically dominated by another action <u>on all attributes</u>, then it can be discarded.

- Suppose, for example, that the construction transportation cost depends on the distance to the supplier.
- The cost itself is uncertain, but the greater the distance, the greater the cost.
- If S1 is closer than S2, then S1 will dominate S2 on cost.
- Rational decision can be made without using any numeric values.

Exploiting preference structure

- For certain regularities in the preference behaviour, the utility function can have a simplified structure

$$\boxed{U(x_1, \ldots, x_n)} = F[f_1(x_1), \ldots, f_n(x_n)]$$

add
multiply.

- We consider two cases here:
  - Preferences without uncertainty
  - Preferences with uncertainty

Lottery.

$\langle x_1 \; x_2 \; x_3 \rangle$.

$\langle x_1' \; x_2' \; x_3' \rangle$.

Preference Independence

- Two attributes X_1 and X_2 are preferentially independent of a third attribute X_3 if the preference between outcomes $\langle x_1, x_2, x_3 \rangle$ and $\langle x_1', x_2', x_3' \rangle$ does not depend on the particular value x_3 for attribute X_3

$\langle x_1, x_2\ 100 \rangle, \langle x_1', x_2'\ 100 \rangle$

Noise

Cost

Safety Level

A

B.

$\langle h_{fan}\ h_{AC}, r \rangle$

$\langle 8, 8, r \rangle \quad \langle 2, 14, r \rangle$

Preferences without uncertainty

- A set of attributes exhibits <u>mutual preferential independence</u> (MPI) when the value taken by an attribute does not affect the way in which one trades off the other attributes against each other.
- If attributes X1 , . . . , Xn are mutually preferentially independent, then the agent's preference behavior can be described as maximizing the function

$$\underbrace{V(x_1, \ldots x_n)} = \sum_i \underbrace{V_i(x_i)}$$

value function has been decomposed

$$V(noise, cost, deaths) = \underbrace{-noise \times 10^4} - \underbrace{cost} - \underbrace{deaths \times 10^{12}}$$

safety level

Preferences without uncertainty

- Example where (mutual preferential independence) MPI does not hold:
  - Consider that you want to purchase hunting dogs, chickens and cages
  - Tradeoff between dogs and chickens strongly depends on the number of cages.

100

Preferences with uncertainty

Preferences for lotteries

- With uncertainty in the picture, the value function is replaced with the utility function.

$U(x_1, x_2 \dots x_n)$.

- Preferences are now between lotteries instead of (certain) states.

- If preferences between lotteries as a function of the set of attributes $\mathbf{X}$ are independent of the particular values of attributes in the set $\mathbf{Y}$, then the attributes $\mathbf{X}$ and $\mathbf{Y}$ exhibit utility independence

Preference Independence $\longrightarrow$ Utility independence

$U(X_1, Y)$

$U(X_2, Y)$

$L(X = X_1)$

$L(X = X_2)$

Y.

- Mutual Utility Independence (MUI) X Y.

  A set of attributes exhibit MUI if each of its subsets is utility-independent of the remaining attributes.

- MUI implies that the agent's preferences can be generated using a utility function with terms containing multiples of single attribute utility functions.

$$\qquad x_1 \qquad x_2 \qquad x_3.$$

$$(x_1, x_2, x_3). \quad U = k_1U_1 + k_2U_2 + k_3U_3 + k_1k_2U_1U_2 + k_2k_3U_2U_3 + k_3k_1U_3U_1 + k_1k_2k_3U_1U_2U_3 \,.$$

Decision Networks

depicts the computations involved in decision making

Decision Network

- A decision network represents information about the agent's
  - Current state evidences.
  - Possible actions
  - State that will result from agent's actions outcome
  - Utility of that state

Decision Network

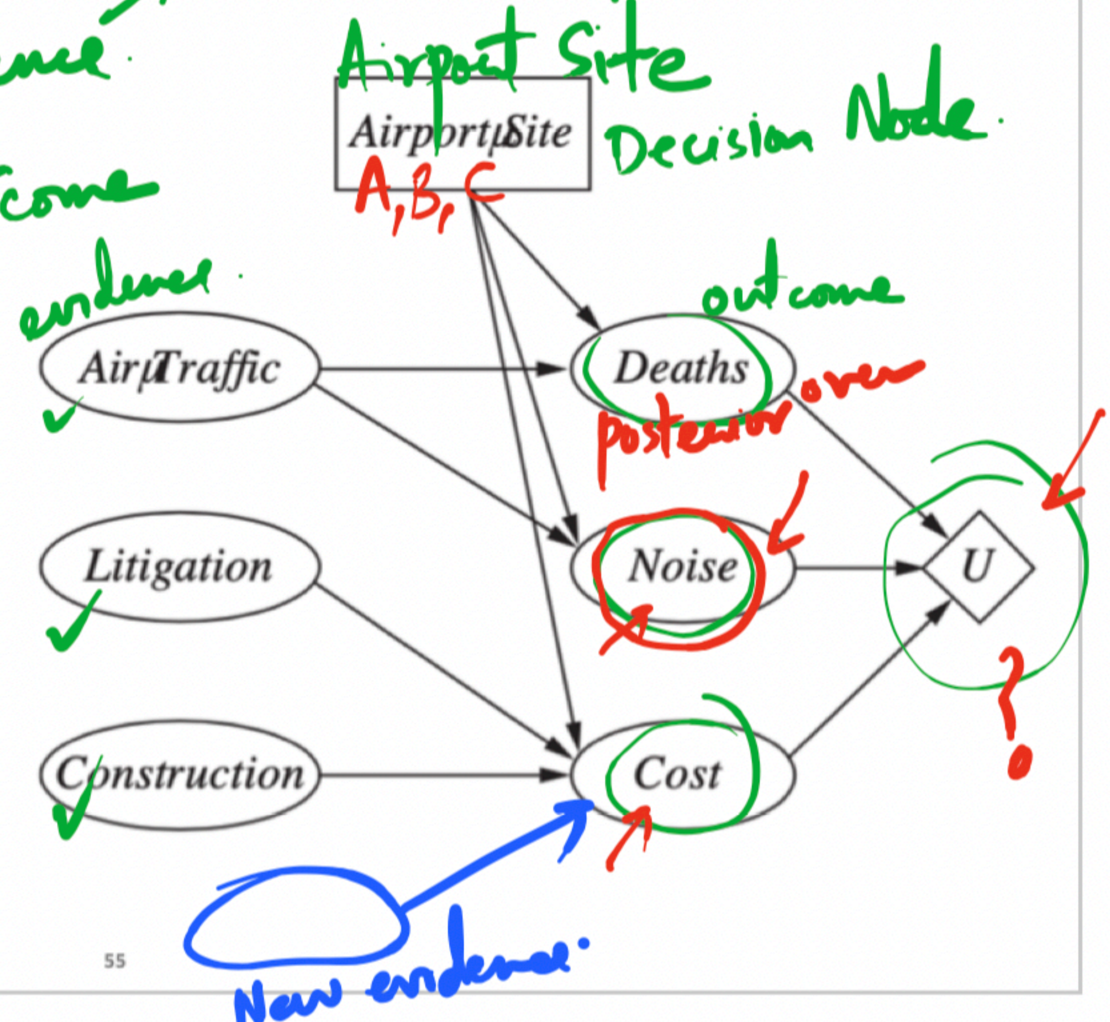
- A decision network has
  - Chance nodes
  - Decision nodes
  - Utility nodes

evidence

outcome

can be instantiated

diamond



- In a simplified form of the decision network, the utility node is directly connected to the action, instead of the outcome state.

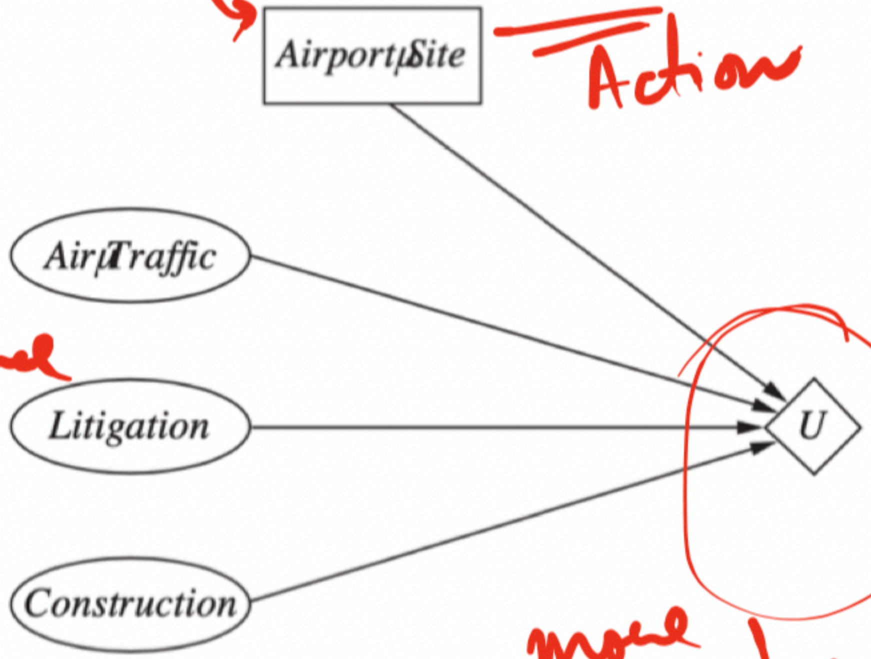
- The utility node implements the **action-utility function**

Airport Sited A, B, C.

Decision node

Action

evidence



more complex

Evaluating the decision network

- Set the evidence variables for the current state
- For each possible value of the decision node
  - Set the decision node to that value.
  - Calculate the posterior probabilities for the parent nodes of the utility node, using a standard probabilistic inference algorithm
  - Calculate the resulting utility for the action.

outcome nodes

- Return the action with the highest utility

What additional information, evidence. can improve the decision making

The value of Information

- More and more information helps in deciding for a better action

value

- The importance of a piece of information (evidence) depends on

  - whether it has the potential to change the physical action

  - the new action outcome will be significantly better than the previous one

- Additional information is incorporated in the decision network by instantiating one or more of the chance variables (nodes).

Example: Purchasing an oil field

exactly one block.

n

- An oil company is hoping to buy one of n indistinguishable blocks of ocean-drilling rights.
- Exactly one of the blocks contains oil worth C dollars, others are worthless.
- The asking price of each block is $\frac{C}{n}$ dollars.

$\frac{C}{n}$

- A seismologist offers the company the results of a survey of one of the blocks.
    - How much the company should be willing to pay for it?

There is oil. $\left(\frac{1}{n}\right)$ $\quad \frac{C}{n}$ $\quad$ Profit $C - \frac{C}{n} = \frac{(n-1)C}{n}$ P_{s1}

There is no oil. $\left(\frac{n-1}{n}\right)$ $\quad \left(\frac{1}{n-1}\right)$ $\quad$ Profit $C\left(\frac{1}{n-1}\right) - \frac{C}{n} = \frac{C}{n(n-1)}$ P_{s2}

Value of Perfect Information (VPI)

- Here "perfect information" means exact evidence
  - That is, some random variable E_j gets assigned a value $E_j = e_j$

e_j E_j ?

- Let the agent's initial evidence be $\mathbf{e}$.

Then the value of the current best action α is defined by

$$EU(\alpha|\mathbf{e}) = \max_a \sum_{s'} P(\text{RESULT}(a) = s' \mid a, \mathbf{e})\, U(s')$$

$\alpha = \arg\max_a \sum \ldots$

- The value of the new best action is after getting the evidence $E_j = e_j$

$$EU(\alpha_{e_j}|\mathbf{e}, e_j) = \max_a \sum_{s'} P(\text{RESULT}(a) = s' \mid a, \mathbf{e}, e_j)\, U(s')$$

EU of the new best action

$E_j \leftarrow e_j$ not fixed

- Since E_j is a random variable, we need to take an expectation over all possible values e_{jk} that can be observed for E_j

$$\sum_k P(E_j = e_{jk} | \mathbf{e})\ EU(\alpha_{e_{jk}} | \mathbf{e}, E_j = e_{jk})$$

- The value of information about E_j is

$$VPI_{\mathbf{e}}(E_j) = \underbrace{\left(\sum_k P(E_j = e_{jk} | \mathbf{e})\ EU(\alpha_{e_{jk}} | \mathbf{e}, E_j = e_{jk}) \right)}_{\text{Exp Utility of the best action after observing } E_j} - \underbrace{EU(\alpha | \mathbf{e})}_{\text{best action without any info about } E_j}$$