Lecture 14: Introduction to Decision Making

- Preferences
- Utility functions
- Maximizing expected utility
- Value of information

Actions and consequences

- So far, we have focused on ways of modeling a stochastic, uncertain world
- But intelligent agents should be not only observers, but *actors*
  I.e. they should choose *actions* in a *rational way*
- Most often, actions produce as a *consequence* changes in the world
  Pearl example: buying a baseball ticket
- How should we choosing between buying and not buying a ticket???

Preferences

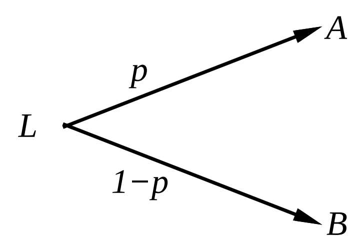
- A rational method would be to evaluate the *benefit* (desirability, or value) of each consequence and weigh these by the probabilities of the consequences
- We will call the consequences of an action **payoffs** or **prizes**
- In order to compare different actions we need to know, for each one, the set of consequences $C = \{c_1, \ldots c_n\}$ and a probability distribution over the consequences, P, s.t. $\sum_i P(c_i) = 1$.
- A pair $L = (C, P)$ is called a **lottery** (Luce and Raiffa, 1957)
- So choosing between actions amounts to choosing between lotteries

Lotteries

- A lottery can be represented as a list of pairs, e.g.

$$L = [A, p; B, (1-p)]$$

  or as a tree-like diagram:



- Agents have preferences over payoffs:
  - $A \succ B$ - A preferred to B
  - $A \sim B$ - indifference between A and B
  - $A \succsim B$ - B not preferred to A
- In order for an agent to act rationally, its preferences have to obey certain constraints

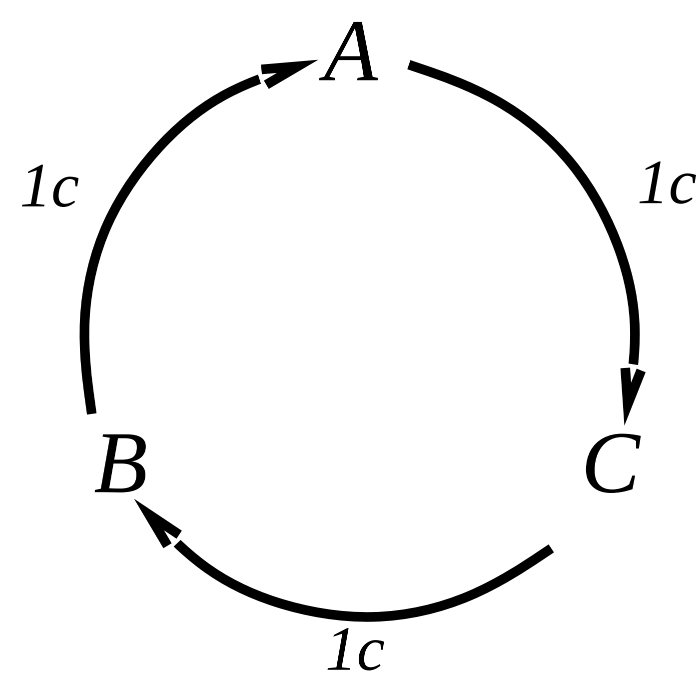
Example: Transitivity

Suppose an agent has the following preferences: $B \succ C$, $A \succ B$, $C \succ A$, and it owns C.

If $B \succ C$, then the agent would pay (say) 1 cent to get B

If $A \succ B$, then the agent, who now has B would pay (say) 1 cent to get A

If $C \succ A$, then the agent (who now has A) would pay (say) 1 cent to get C



The agent looses money forever!

The Axioms of Utility Theory

These specify constraints over the preferences that a rational agent can have:

1. *Orderability*: A linear and transitive preference relation must exist between the prizes of any lottery
   - Linearity: $(A \succ B) \vee (B \succ A) \vee (A \sim B)$
   - *Transitivity*: $(A \succ B) \wedge (B \succ C) \Rightarrow (A \succ C)$
2. *Continuity*: If $A \succ B \succ C$, then there exists a lottery L with prizes A and C that is equivalent to receiving B for sure:

$$\exists p, L = [p, A;\ 1 - p, C] \sim B$$

The probability p at which equivalence occurs can be used to compare the merit of B w.r.t A and C

3. *Substitutability*: Adding the same prize with the same probability to two equivalent lotteries does not change the preference between them:

   For any $L_1, L_2, L_3, 0 < p \leq 1, L_1 \sim L_2 \Leftrightarrow [p, L_1; (1-p), L_3] \sim [p, L_2; (1-p), L_3]$

4. *Monotonicity*: If two lotteries have the same prizes, the one producing the best prize most often is preferred

   $$A \succ B \Rightarrow [p, A; (1-p), B] \succsim [p', A; (1-p'), B] \text{ iff } p \geq p'$$

5. *Reduction of compound lotteries* ("No fun in gambling"): For any lotteries L_1 and $L_2 = [p, C_1; (1-p), C_2]$,

   $$[p, L_1; (1-p), L_2] \sim [p, L_1; (1-p)q, C_1; (1-p)(1-q)C_2]$$

Maximizing expected utility (MEU)

If an agent has rational preferences, his behavior is describable as *maximization of expected utility*

Theorem: (Ramsey, 1931; von Neumann and Morgenstern, 1944): Given a preference relation over lotteries satisfying the axioms of utility theory, there exists a real-valued function U on the set of prizes C such that

$$L_1 \succsim L_2 \text{ iff } U(L_1) \geq U(L_2)$$

where

$$U([p_1, C_1;\ \ldots\ ;\ p_n, C_n]) = \sum_i p_i U(C_i)$$

Proof: see Pearl book and board

Acting under Uncertainty

- **MEU principle**: Choose the action that maximizes expected utility. Most widely accepted as a standard for rational behavior
- Note that an agent can be entirely rational (consistent with MEU) without ever representing or manipulating utilities and probabilities

  E.g., a lookup table for perfect tic-tac-toe
- *Random choice models*: choose the action with the highest expected utility most of the time, but keep non-zero probabilities for other actions as well
  - Avoids being too predictable
  - If utilities are not perfect, allows for *exploration*
- *Minimizing regret*

Utilities

- Utilities map states to real numbers. How do we get these numbers?
- The proof of the utility theorem suggests that a way to obtain these numbers is by comparing a given prize A with a *standard (calibration) lottery* L_p that has
  - "best possible prize" $u_\top$ with probability p
  - "worst possible catastrophe" $u_\bot$ with probability $(1-p)$

  Usually utilities are *normalized*: $u_\top = 1.0$, $u_\bot = 0.0$
- Adjust lottery probability p until $A \sim L_p$. Then p is used as the utility of A.

Utility scales

- Note that given a preference behavior, the utility function is **NOT unique**
- E.g. behavior is invariant w.r.t. additive linear transformations:

$$U'(x) = k_1 U(x) + k_2 \quad \text{where } k_1 > 0$$

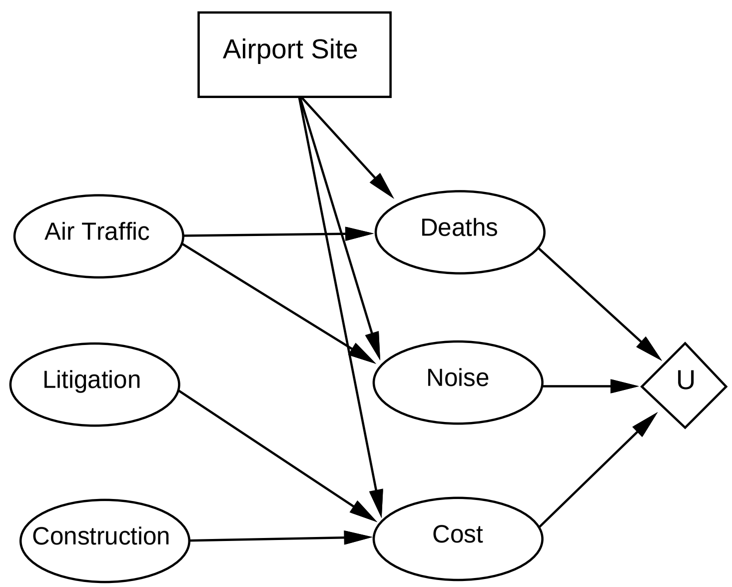
- With deterministic prizes only (no lottery choices), only **ordinal utility** can be determined, i.e., total order on prizes

Money

- Suppose you had to choose between two lotteries:
  - L_1: win $1 million for sure
  - L_2: 5 million w.p. 0.1, 1 million w.p. 0.89 and nothing w.p. 0.01
- Suppose you had to choose between two lotteries:
  - L_3: 5 million w.p. 0.1, nothing w.p. 0.9
  - L_4: 1 million w.p. 0.11, nothing w.p. 0.89
- See also Bernoulli's paradox
- People are *risk-averse*

Decision networks

Add *action nodes* (rectangles) and *utility nodes* (diamonds) to belief networks to enable rational decision making



1. For each value of action node:

   compute expected value of utility node given action, evidence
2. Return MEU action

Example: Value of Information

Buying oil drilling rights:

- Two blocks A and B, exactly one has oil, worth k
- Prior probabilities 0.5 each, mutually exclusive
- Current price of each block is $k/2$
- Consultant offers accurate survey of A

What is a fair price for the survey?

Solution for the example

Compute expected value of information = expected value of best action given the information minus expected value of best action without information

Survey may say “oil in A” or “no oil in A”, with probability 0.5 each:

= [$0.5 \times$ value of “buy A” given “oil in A”

+ $0.5 \times$ value of “buy B” given “no oil in A”]

– 0 = $(0.5 \times k/2) + (0.5 \times k/2) - 0 = k/2$

Value of Perfect Information (VPI)

Current evidence E, current best action α

Possible action outcomes c_i, potential new evidence X

$$EU(\alpha|E) = \max_a \sum_i U(c_i) P(c_i|E, a)$$

Suppose we knew $X = x$. Then we would choose α_x s.t.

$$EU(\alpha_x|E, X = x) = \max_a \sum_i U(c_i) P(c_i|E, a, X = x)$$

X is a random variable whose value is *currently* unknown

$\Rightarrow$ we must compute expected gain over all possible values:

$$VPI_E(X) = \left(\sum_k P(X = x|E) EU(\alpha_x|E, X = x) \right) - EU(\alpha|E)$$

Properties of VPI

- **Nonnegative**: $\forall X, EVPI_E(X) \geq 0$

  Note that VPI is an *expectation*! Depending on the actual value we find for X, there can actually be a loss post-hoc
- **Nonadditive**—e.g. consider obtaining X twice

$$VPI_E(X,Y) \neq VPI_E(X) + VPI_E(Y)$$

- **Order-independent**

$$VPI_E(X,Y) = VPI_E(X) + VPI_{E,X}(Y) = VPI_E(Y) + VPI_{E,Y}(X)$$

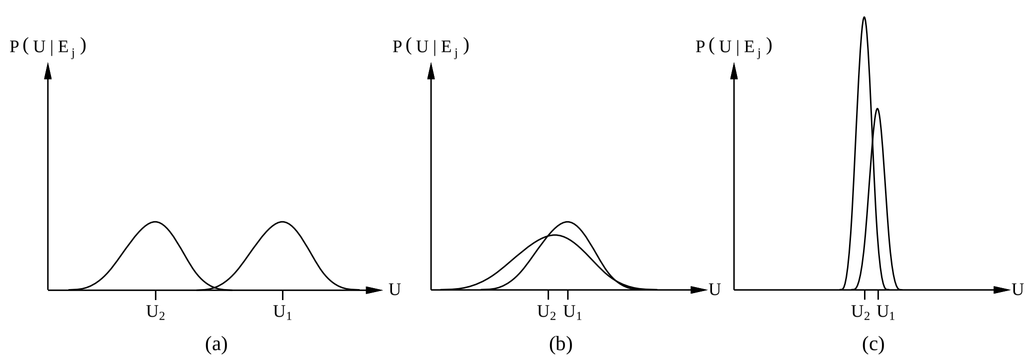
Note: when more than one piece of evidence can be gathered, maximizing VPI for each to select one *myopic strategy* is not always optimal

$\Rightarrow$ evidence-gathering becomes a **sequential** decision problem

Qualitative behaviors

There are three possible cases:

- Choice is obvious, information worth little
- Choice is nonobvious, information worth a lot
- Choice is nonobvious, information worth little



Information has value to the extent that it is likely to cause a change in plan, and the new plan is significantly better than the old one

Summary: Decision making under uncertainty

- To make decisions under uncertainty, we need to know the likelihood (probability) of different possible outcomes, and have preferences among the outcomes:

  Decision Theory = Probability Theory + Utility Theory
- An agent with consistent preferences has a utility function, which associates a real number to each possible state
- Rational agents try to maximize their expected utility.
- Utility theory allows us to determine whether gathering more information is valuable.
- Next time: sequential decision making (Markov Decision Processes)