Lecture 12: Introduction to reasoning under uncertainty

- Preferences
- Utility functions
- Maximizing expected utility
- Value of information
- Bandit problems and the exploration-exploitation trade-off

Actions and Consequences

- Probability allows us to model an uncertain, stochastic world
- But intelligent agents should be not only *observers*, but also *actors*
  I.e. they should choose actions in a rational way
- Most often, actions produce *consequences* which cause the world to change

Three Theories

- *Probability theory:*
  - Describes what the agent should believe based on the evidence
- *Utility theory:*
  - Describes what the agent wants
- *Decision theory:*
  - Describes what a rational agent should do (based on probability theory and utility theory)

Example: Buying a Football Ticket

- Possible consequences:
  - You start watching the game, but then it starts to rain and you catch pneumonia
  - You watch the game and get back home
  - You watch the game but when you get back home you find that the cat ate the parrot
  - You watch the game; when you want to get back home, the car won't start. But your favorite rock start passes by and gives you a ride.
- How should we choose between buying and not buying a ticket???

Preferences

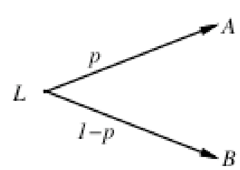
- A rational method would be to evaluate the *benefit* (desirability, value) of each consequence and *weigh* it by the *probabilities of consequences*.
- We will call the consequences of an action *payoffs* or *rewards*
- In order to compare different actions we need to know, for each one:
  - The *set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

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:

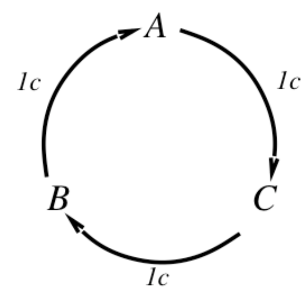


- Agents have preferences over payoffs:
  - $A \succ B$ - A preferred to B
  - $A \sim B$ - indifference between A and B
  - $A \succsim B$ - B not preferred to A
- For an agent to act rationally, its preferences have to obey certain constraints

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 are 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

The Axioms of Utility Theory (2)

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

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

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 $L_2 = [p, C_1; (1-p), C_2]$,

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

Utility Functions

Theorem (Ramsey, 1931; von Neumann and Morgenstern, 1944): Given preferences that satisfy these axioms, there exists at least one real-valued function U, called *utility function*, such that:

$$A \succsim B \text{ if and only if } U(A) \geq U(B)$$

and

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

Reminder: Expected value

- Suppose you have a discrete-valued random variable X, with n possible values $\{x_1, \dots x_n\}$, occurring with probabilities $p_1, \dots, p_n$ respectively. Then the *expected value (mean)* of X is:

$$E[X] = \sum_{i=1}^{n} p_i x_i$$

- Example: suppose you play a game in which your opponent tosses a fair coin. If it comes up heads, you get \$1, if it comes up tails, you get \$0. What is your expected profit?

  Answer: $(+1)\frac{1}{2} + (-1)\frac{1}{2} = 0$

Utilities

- Utilities map outcomes (or states) to real numbers
- Note that given a preference behavior, the utility function is *not unique*
- Eg., Behavior (action choice) is invariant with respect to 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:
    - win \$5 million w.p. 0.1
    - win \$1 million w.p. 0.89
    - win \$0 w.p. 0.01
- Which one would you choose?
- Which one *should* you choose?

Money (2)

- Suppose you had to choose between two lotteries:
  - L_1:
    - win \$1 million for sure
  - L_2:
    - win \$5 million w.p. 0.1
    - win \$1 million w.p. 0.89
    - lose \$1 million w.p. 0.01
- Which one would you choose?
- Which one *should* you choose?

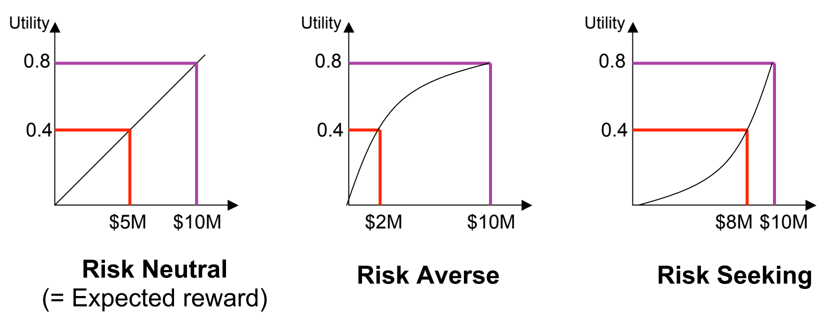
Money (3)

- Suppose you had to choose between two lotteries:
  - L_1:
    - $5 million w.p. 0.1
    - $0 w.p. 0.9
  - L_2:
    - $1 million w.p. 0.3
    - $0 w.p. 0.7
- Which one would you choose?
- Which one *should* you choose?

Utility Models

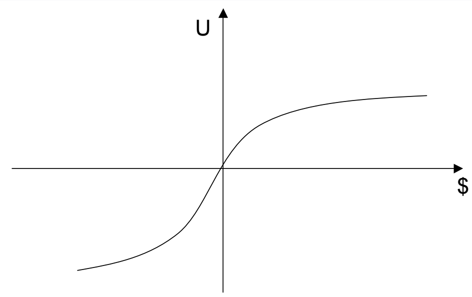
- Capture preferences towards rewards and resource consumption
- Capture risk attitudes

  E.g. if one is risk-neutral, getting $5 million has exactly half the utility of getting $ 10 million
- People are generally *risk-averse* when it comes to money



The Utility of Money

- Decision theory is *normative*: describes how *rational* agents should act
- People systematically violate the axioms of utility and decision theory, especially regarding money
  - Choose: 80% chance of $4000 or 100% chance of $3000
  - Choose: 20% chance of $4000 or 25% chance of $3000



Preference Elicitation

- An increasing number of applications require recommending something to a user or making a decision for them:
  - E.g. movie or book recommendation systems
  - E.g. deciding which cancer treatment to give to a patient (has to take into account chance of survival, cost, side effects)
  - E.g. deciding which ads to show on a dynamic web page
- For this, we need to know the utility that the user associates to different items
- But people are very bad at specifying utility values!
- Preference elicitation refers to finding out their preferences and translating them into utilities
- Very hard problem, lots of current research

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 (i.e. consistent with MEU) without ever representing or manipulating utilities and probabilities

  E.g., a lookup table for perfect tic-tac-toe

Acting under Uncertainty (2)

- Sometimes it can be advantageous to not always choose actions according to MEU, e.g. if the environment may change, or it is not fully known to the agent
- *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: consider the loss between current behavior and some "gold standard" and try to minimize it

Example: Single Stage Decision Making

- One random variable, X: does the kid have an ear infection or not?
- One decision, d: give antibiotic (yes) or not (no)
- The utility function associates a real value to possible states of the world and possible decisions

	$X =$ no	$X =$ yes
$d =$ no	0	-50
$d =$ yes	-100	10

- Unfortunately X is not directly observable!
- But we know $P(X = \text{yes}) = 0.1$, $P(X = \text{no}) = 0.9$.

Example: Maximizing Expected Utility

- In our case, U is:

	$X =$ no	$X =$ yes
$d =$ no	0	-50
$d =$ yes	-100	10

and $P(X = \text{yes}) = 0.1$, $P(X = \text{no}) = 0.9$. Compute:

$$\begin{aligned} EU(d = \text{no}) &= 0.9 \times 0 + 0.1 \times (-50) = -5 \\ EU(d = \text{yes}) &= 0.9 \times (-100) + 0.1 \times 10 = -8 \end{aligned}$$

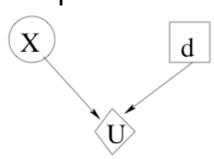
so according to MEU the best action is $d =$ no.

Some definitions

- *Utility function:* $U(x)$
  - Numerical expression of the desirability of a situation
- *Expected utility:* $EU(a|x) = \sum P(\mathit{Effect}(a)|x)U(\mathit{Effect}(a))$
  - Utility of each action outcome is weighted by the probability of that outcome
- *Maximum expected utility:* $\max_a EU(a|x)$
  - Best average payoff that can be achieved in situation x
- *Optimal action:* $\arg\max_a EU(a|x)$
  - Action chosen according to MEU principle
- *Policy:* a way of picking actions

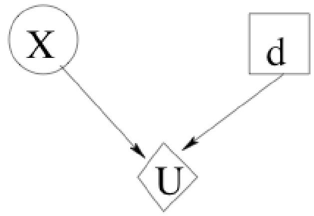
Decision Graphs

- We can represent the decision problem as a graphical model:



- Random variables are represented as oval nodes
  - Parameters associated with such nodes are *probabilities*
- Decisions are represented as rectangles
- Utilities are represented as diamonds
  - Parameters associated with such nodes are *utility values* for all possible values of the parents
- Restrictions on nodes:
  - Utility nodes have no out-going arcs
  - Decision nodes have no incoming arcs
- Computing the optimal action can be viewed as *inference*

Example



- Suppose we had evidence that $X =$ yes.
- We can set d to each possible value (yes/no)
- For each value, ask the utility node to give the utility of that situation, then pick d according to MEU
- If there is no evidence at X, we will have to *sum out* over all possible values of X, like in Bayes net inference
- This will give the expected utility at node U, for each choice of action d

Information Gathering

- In an environment with hidden information, an agent can choose to perform *information-gathering actions*
  - E.g., taking the kid to the doctor
  - E.g., scouting the price of a product at different companies
- Such actions take time, or have associated costs (e.g., medical tests). *When are they worth pursuing?*
- The *value of information* specifies the utility of every piece of evidence that can be acquired.

Example: 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?

Example: Solution

- Compute expected value of information as:
  expected value of best action given the information - expected value of best action without the information
- Survey may say "oil in A" or "no oil in A", with probability 0.5 each, so the value of the information is:
  $[0.5\times$ 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)

- Suppose you have current evidence E, current best action a^*, with possible outcomes c_i. Then the expected utility of a^* is:

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

- Suppose that you could gather further evidence about a variable X. Should you do it?

Value of Perfect Information

- Suppose we knew $X = x$. Then we would choose a^*_x s.t.

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

- X is a random variable whose value is unknown, so we must compute expected gain over all possible values:

$$VPI_E(X) = \left(\sum_x P(X = x|E) EU(a^*_x|E, X = x) \right) - EU(a^*|E)$$

This is the value of knowing X exactly

Properties of VPI

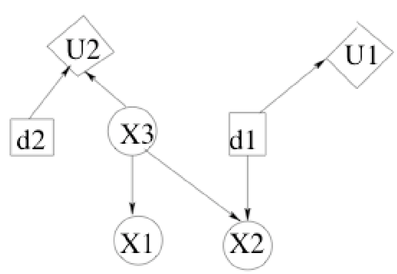
- *Nonnegative*: $\forall X, E \quad VPI_E(X) \geq 0$
  Note that VPI is an *expectation*! Depending on the actual value we find for X, there can actually be a loss post-hoc
- *Nonadditive*: E.g. consider obtaining X twice

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

- *Order-independent*

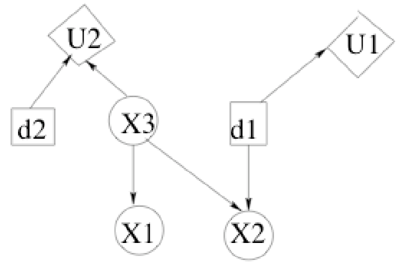
$$VPI_E(X,Y) = VPI_E(X) + VPI_{E,X}(Y) = VPI_E(Y) + VPI_{E,Y}(X)$$

A More Complex Example



- X1: Symptoms
- X3: is there infection
- d1: decision to go to the doctor
- X2: result of consultation
- d2: treatment or no treatment

Example continued



- Total utility is U1+U2
- X2 is only observed if we decide that d1= 1
- X3 is never observed

Now we have to optimize d1 and d2 together!

Summary

- To make decisions under uncertainty, we need to know the likelihood (probability) of different possible outcomes, and have preferences among 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 tell whether gathering more information is valuable.
- Decision graphs can be used to represent the decision problem
- An algorithm similar to variable elimination is useful to compute optimal decision, but this is very expensive in general