Artificial Intelligence Foundation – JC3001

Lecture 34: Stochastic Planning - I

Prof. Aladdin Ayesh (aladdin.ayesh@abdn.ac.uk)

Dr. Binod Bhattarai (binod.bhattarai@abdn.ac.uk)

Dr. Gideon Ogunniye, (g.ogunniye@abdn.ac.uk)

October 2025

Material adapted from:
Russell and Norvig (AIMA Book): Chapter 17
Sutton and Barto (Reinforcement Learning: An Introduction 2nd ed.)
Sebastian Thrun (Stanford University / Udacity)

Course Progression

o

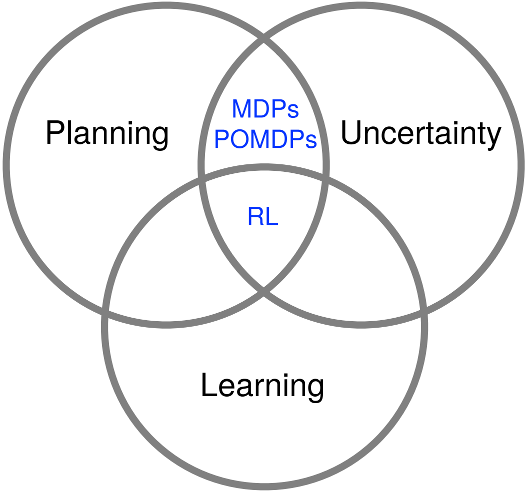
- Part 1: Introduction
  1. Introduction to AI ✓
  2. Agents ✓
- Part 2: Problem-solving
  1. Search 1: Uninformed Search ✓
  2. Search 2: Heuristic Search ✓
  3. Search 3: Local Search ✓
  4. Search 4: Adversarial Search ✓
- Part 3: Reasoning and Uncertainty
  1. Reasoning 1: Constraint Satisfaction ✓
  2. Reasoning 2: Logic and Inference ✓
  3. Probabilistic Reasoning 1: BNs ✓
  4. Probabilistic Reasoning 2: HMMs ✓
- Part 4: Planning
  1. Planning 1: Intro and Formalism ✓
  2. Planning 2: Algorithms & Heuristics ✓
  3. Planning 3: Hierarchical Planning ✓
  4. **Planning 4: Stochastic Planning**
- Part 5: Learning
  1. Learning 1: Intro to ML
  2. Learning 2: Regression
  3. Learning 3: Neural Networks
  4. Learning 4: Reinforcement Learning
- Part 6: Conclusion
  1. Ethical Issues in AI
  2. Conclusions and Discussion

Objectives

- Planning with Utilities
- Stochastic Planning Formalisms
- Dynamic Programming Algorithms for Stochastic Planning

- Recap
- Utilities and Expectation

- We know how to get from a current state to a goal state
  - Via domain specific search
  - Via domain independent planning
- We also know how to calculate the probability of future events via probability theory
- But, how can we get the agent to decide between different actions, particularly in the face of uncertainty?


Planning
MDPs
POMDPs
Uncertainty
RL
Learning

	Deterministic	Stochastic
Fully Observable		
Partially Observable		

	Deterministic	Stochastic
Fully Observable	A^*, DFS BFS, GP	
Partially Observable		

	Deterministic	Stochastic
Fully Observable	A^*, DFS BFS, GP	MDP
Partially Observable		

	Deterministic	Stochastic
Fully Observable	A^*, DFS BFS, GP	MDP
Partially Observable		POMDP

Outline

- Recap
- Utilities and Expectation

How Good is an action

- Consider an agent capable of planning and executing different actions to achieve different goals
- Example you can either:
  - Stay home and study
  - Go out and have fun
- Which should you do, and why?

Rationality

- A rational agent should pursue the goal that is, in some sense, "best for it"
- How can we identify whether the achievement of one goal is "better" for the agent than another?

- A rational agent should pursue the goal that is, in some sense, "best for it"
- How can we identify whether the achievement of one goal is "better" for the agent than another?
- Utility theory
  - We capture preferences between world states via a utility function $U(S)$

$$U : States \rightarrow \mathbb{R}$$

  - In other words, we assign a single number to a state to express the **desirability** of the state

- A rational agent should pursue the goal that is, in some sense, “best for it”
- How can we identify whether the achievement of one goal is “better” for the agent than another?
- Utility theory
  - We capture preferences between world states via a utility function $U(S)$

$$U : States \rightarrow \mathbb{R}$$

  - In other words, we assign a single number to a state to express the **desirability** of the state
    Winning the lottery $U(w) = 1000$

- A rational agent should pursue the goal that is, in some sense, “best for it”
- How can we identify whether the achievement of one goal is “better” for the agent than another?
- Utility theory
  - We capture preferences between world states via a utility function $U(S)$

$$U : States \rightarrow \mathbb{R}$$

  - In other words, we assign a single number to a state to express the **desirability** of the state
    Winning the lottery $U(w) = 1000$ Failing AI: $U(f) = -10$

- A rational agent should pursue the goal that is, in some sense, "best for it"
- How can we identify whether the achievement of one goal is "better" for the agent than another?
- Utility theory
  - We capture preferences between world states via a utility function $U(S)$

$$U : States \to \mathbb{R}$$

  - In other words, we assign a single number to a state to express the **desirability** of the state
    Winning the lottery $U(w) = 1000$ Failing AI: $U(f) = -10$
  - But what if actions have uncertain effects?
    Buying a lottery ticket does not give us a utility of 1000

- A nondeterministic action A could lead to several possible outcomes, e.g. $Result_1(A), Result_2(A), \ldots$
  $Result(BuyTicket) \in \{win, \neg win\}$
- We can associate a probability with the different outcomes

$$P(Result_i(A) \mid Do(A), E)$$

  - E captures the evidence we have about the world
  - $Do(A)$ is the proposition that action A is executed in the current state
- The expected utility of executing A given E is then

$$\mathbb{E}[U(A) \mid E] = \sum_i P(Result_i(A) \mid Do(A), E) U(Result_i(A))$$

- Principle of maximum expected utility (MEU):

  A rational agent should choose the action
  that maximises its expected utility

More generally

- We can use the expectation operator ($\mathbb{E}[X]$) for **any random variable** X

$$\mathbb{E}[X] \doteq \sum_{x \in X} P(x)x$$

- Example:
  Expected value of a six sided die:

More generally

- We can use the expectation operator ($\mathbb{E}[X]$) for **any random variable** X

$$\mathbb{E}[X] \doteq \sum_{x \in X} P(x)x$$

- Example:

  Expected value of a six sided die:

x	P(X=x)	x*P(X=x)
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	6/6
	Total	21/6 = 3.5

Maximum Expected Utility

- An agent following the principle of MEU is rational, acting optimally in any situation

- An agent following the principle of MEU is rational, acting optimally in any situation
- But choosing the best sequence of actions requires enumerating all action sequences; infeasible for long sequences

Maximum Expected Utility

- An agent following the principle of MEU is rational, acting optimally in any situation
- But choosing the best sequence of actions requires enumerating all action sequences; infeasible for long sequences
- Computations are often too expensive:
  - $P(Result_i(A) \mid Do(A), E)$ requires a complete causal model of the world. Even if all information is available, this is computationally intractable
  - Computing $U(Result_i(A))$ requires searching or planning as we do not know how good a state is until we know what goals we can reach from it (and how expensive they are)

- MEU is no silver bullet for the problems of AI
  - But does give us a nice framework for building an agent,

- MEU is no silver bullet for the problems of AI
  - But does give us a nice framework for building an agent,
  - An MEU maximising agent will achieve the highest possible performance averaged over all environments in which the agent can be placed

- MEU is no silver bullet for the problems of AI
  - But does give us a nice framework for building an agent,
  - An MEU maximising agent will achieve the highest possible performance averaged over all environments in which the agent can be placed
  - This moves from a performance measure in the external environment, to an internal criteria of utility maximisation

- A lottery is a probability distribution over a set of actual outcomes

Lotteries

- A lottery is a probability distribution over a set of actual outcomes
- A lottery L with possible outcomes $C_1, \ldots, C_n$ and associated probabilities $p_1, \ldots, p_n$ is written as

$$L = [p_1, C_1; \ldots p_n, C_n]$$

Lotteries

- A lottery is a probability distribution over a set of actual outcomes
- A lottery L with possible outcomes $C_1, \ldots, C_n$ and associated probabilities $p_1, \ldots, p_n$ is written as

$$L = [p_1, C_1; \ldots p_n, C_n]$$

- Each outcome C_i is an atomic state, <u>or another lottery</u>

Lotteries

- A lottery is a probability distribution over a set of actual outcomes
- A lottery L with possible outcomes $C_1, \ldots, C_n$ and associated probabilities $p_1, \ldots, p_n$ is written as

$$L = [p_1, C_1; \ldots p_n, C_n]$$

- Each outcome C_i is an atomic state, or another lottery
- Example: There is a 30% likelihood I will go out tonight.
  If I do so, there is a 60% likelihood I will do badly in my test.
  If I do not go out, I will do well in my test with a 90% likelihood

$$L = [0.3, [0.6, f; 0.4, p]; 0.7, [0.9, p; 0.1, f]]$$

- A lottery is a probability distribution over a set of actual outcomes
- A lottery L with possible outcomes $C_1, \ldots, C_n$ and associated probabilities $p_1, \ldots, p_n$ is written as

$$L = [p_1, C_1; \ldots p_n, C_n]$$

- Each outcome C_i is an atomic state, or another lottery
- Example: There is a 30% likelihood I will go out tonight.
  If I do so, there is a 60% likelihood I will do badly in my test.
  If I do not go out, I will do well in my test with a 90% likelihood

$$L = [0.3, [0.6, f; 0.4, p]; 0.7, [0.9, p; 0.1, f]]$$

- Our goal is to understand the relationships between preferences between lotteries and preferences between states

Properties of Preferences

- Preferences should be transitive

$$A \succ B, B \succ C \rightarrow A \succ C$$

- Proof:
  - Suppose $A \succ B \succ C \succ A$
  - And assume A, B, C are exchangeable goods
  - Then if the agent has A, we could offer it C for A and a bit of money
  - Since the agent prefers C, they would agree to this deal
  - Repeat, trading C for B, B for A and we are back at the start, with the agent having less money
  - Clearly the agent has not acted rationally

- Utility Principle: If an agent's preferences obey the axioms of utility, then there is a real-valued function U that operates on states such that $U(A) > U(B)$ if and only if A is preferred to B, and $U(A) = U(B)$ if and only if the agent is indifferent between A and B

$$U(A) > U(B) \leftrightarrow A \succ B \qquad U(A) = U(B) \leftrightarrow A \sim B$$

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

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

- Utility Principle: If an agent's preferences obey the axioms of utility, then there is a real-valued function U that operates on states such that $U(A) > U(B)$ if and only if A is preferred to B, and $U(A) = U(B)$ if and only if the agent is indifferent between A and B

$$U(A) > U(B) \leftrightarrow A \succ B \qquad U(A) = U(B) \leftrightarrow A \sim B$$

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

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

- Given probabilities and utilities of outcomes, we can determine the utility of a compound lottery

- Utility Principle: If an agent's preferences obey the axioms of utility, then there is a real-valued function U that operates on states such that $U(A) > U(B)$ if and only if A is preferred to B, and $U(A) = U(B)$ if and only if the agent is indifferent between A and B

$$U(A) > U(B) \leftrightarrow A \succ B \qquad U(A) = U(B) \leftrightarrow A \sim B$$

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

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

- Given probabilities and utilities of outcomes, we can determine the utility of a compound lottery
- Since outcomes of nondeterministic actions are lotteries, we obtain the MEU decision rule

- Utility Principle: If an agent's preferences obey the axioms of utility, then there is a real-valued function U that operates on states such that $U(A) > U(B)$ if and only if A is preferred to B, and $U(A) = U(B)$ if and only if the agent is indifferent between A and B

$$U(A) > U(B) \leftrightarrow A \succ B \qquad U(A) = U(B) \leftrightarrow A \sim B$$

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

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

- Given probabilities and utilities of outcomes, we can determine the utility of a compound lottery
- Since outcomes of nondeterministic actions are lotteries, we obtain the MEU decision rule
  - We now know what a utility function is
  - But without structure, there is little we can do with it

- A typical use of utility functions (stemming from economic theory) maps the amount of money an agent has to its utility
- Agents prefer having more money to less: such a preference is <u>monotonic</u>
- But what about lotteries about money?

- A typical use of utility functions (stemming from economic theory) maps the amount of money an agent has to its utility
- Agents prefer having more money to less: such a preference is <u>monotonic</u>
- But what about lotteries about money?
- You are a participant in a game show. You are in the final round, and can take $1,000,000, or gamble it on a coin flip.
  If the coin lands on heads, you will win $3,000,000, tails, you will take home nothing.
  What do you do?

- A typical use of utility functions (stemming from economic theory) maps the amount of money an agent has to its utility
- Agents prefer having more money to less: such a preference is monotonic
- But what about lotteries about money?
- You are a participant in a game show. You are in the final round, and can take \$1,000,000, or gamble it on a coin flip.
  If the coin lands on heads, you will win \$3,000,000, tails, you will take home nothing.
  What do you do?
- The expected monetary value (EMV) is \$1,500,000;
  this is more than the prize
- If \$1=1 utility, then a rational agent would take the gamble

- Instead, let us denote the state of having $\$n$ as S_n, and the your initial states as S_k. Then

$$EU(flip) = 0.5U(S_k) + 0.5U(S_{k+3000000}) \qquad EU(noflip) = U(S_{k+1000000})$$

- We now need to assign utilities to the different outcome states
  - Utility is typically not proportional to money:
    having a 3,000,000 is **not** worth 3 times as much as having 1,000,000.

- Instead, let us denote the state of having $\$n$ as S_n, and the your initial states as S_k. Then

$$EU(flip) = 0.5U(S_k) + 0.5U(S_{k+3000000}) \qquad EU(noflip) = U(S_{k+1000000})$$

- We now need to assign utilities to the different outcome states
  - Utility is typically not proportional to money:
    having a 3,000,000 is **not** worth 3 times as much as having 1,000,000.
  - So perhaps $S_k = 5, S_{k+3000000} = 10, S_{k+1000000} = 8$

- Instead, let us denote the state of having $\$n$ as S_n, and the your initial states as S_k. Then

$$EU(flip) = 0.5U(S_k) + 0.5U(S_{k+3000000}) \qquad EU(noflip) = U(S_{k+1000000})$$

- We now need to assign utilities to the different outcome states
  - Utility is typically not proportional to money: having a 3,000,000 is **not** worth 3 times as much as having 1,000,000.
  - So perhaps $S_k = 5, S_{k+3000000} = 10, S_{k+1000000} = 8$
  - What are the expected utilities? $EU(flip) =? \; EU(noflip) =?$

- Instead, let us denote the state of having $\$n$ as S_n, and the your initial states as S_k. Then

$$EU(flip) = 0.5U(S_k) + 0.5U(S_{k+3000000}) \qquad EU(noflip) = U(S_{k+1000000})$$

- We now need to assign utilities to the different outcome states
  - Utility is typically not proportional to money:
    having a 3,000,000 is **not** worth 3 times as much as having 1,000,000.
  - So perhaps $S_k = 5, S_{k+3000000} = 10, S_{k+1000000} = 8$
  - Then $EU(flip) = 7.5, EU(noflip) = 8$, and you would not flip the coin

- Instead, let us denote the state of having $\$n$ as S_n, and the your initial states as S_k. Then

$$EU(flip) = 0.5U(S_k) + 0.5U(S_{k+3000000}) \qquad EU(noflip) = U(S_{k+1000000})$$

- We now need to assign utilities to the different outcome states
  - Utility is typically not proportional to money: having a 3,000,000 is **not** worth 3 times as much as having 1,000,000.
  - So perhaps $S_k = 5, S_{k+3000000} = 10, S_{k+1000000} = 8$
  - Then $EU(flip) = 7.5, EU(noflip) = 8$, and you would not flip the coin
- Typically, for some lottery L, $U(L) < U(S_{EMV(L)})$. In other words, being given the cash without gambling is better than facing a lottery
  - Agents with this type of utility function are <u>risk averse</u>
  - If $U(S_{EMV(L)}) < U(L)$ the agent is <u>risk seeking</u>

- The amount an agent will accept instead of a lottery is called the certainty equivalent of the lottery
- The difference between the expected monetary value of a lottery and its certainty equivalent is called the insurance premium
- The insurance premium is only positive if an agent is risk averse
- If a utility curve is linear (i.e. $U(S_{EMV(L)}) = U(L)$), the agent is

- The amount an agent will accept instead of a lottery is called the certainty equivalent of the lottery
- The difference between the expected monetary value of a lottery and its certainty equivalent is called the insurance premium
- The insurance premium is only positive if an agent is risk averse
- If a utility curve is linear (i.e. $U(S_{EMV(L)}) = U(L)$), the agent is risk neutral
- Any affine transformation of a utility function is a utility function

To continue in the next session.