

## Markov Decision Processes: Making Decision in the Presence of Uncertainty

(some of) R&N 16.1-16.6  
R&N 17.1-17.4

## Decision Processes: General Description

- Suppose that you own a business. At any time, you know exactly the current state of the business (finances, stock, etc.).
- At any time, you have a choice of several possible actions (advertise, introduce new products,...).
- You cannot predict with certainty the consequence of these actions given the current state of the business, but you have a guess as to the likelihood of the possible outcomes.
- How can you define a policy that will guarantee that you always choose the action that maximizes *expected* future profits?

Note: Russel & Norvig, Chapter 17.

## Decision Processes: General Description

- Decide what action to take next, given:
    - A probability to move to different states
    - A way to evaluate the reward of being in different states
- Robot path planning  
Travel route planning  
Elevator scheduling  
Aircraft navigation  
Manufacturing processes  
Network switching & routing

## Example

- Assume that time is discretized into *discrete time steps*  $t = 1, 2, \dots$  (for example, fiscal years)
- Suppose that the business can be in one of a finite number of *states*  $s$  (this is a major simplification, but let's assume....)
- Suppose that for every state  $s$ , we can anticipate a reward that the business receives for being in that state:  $R(s)$  (in this example,  $R(s)$  would be the profit, possibly negative, generated by the business)
- Assume also that  $R(s)$  is bounded ( $R(s) < M$  for all  $s$ ), meaning that the business cannot generate more than a certain profit threshold
- Question: What is the total value of the reward for a particular configuration of states  $\{s_1, s_2, \dots\}$  over time?

## Example

- Question: What is the total value of the reward for a particular configuration of states  $\{s_1, s_2, \dots\}$  over time?
- It is simply the sum of the rewards (possibly negative) that we will receive in the future:

$$U(s_1, s_2, \dots, s_n, \dots) = R(s_1) + R(s_2) + \dots + R(s_n) + \dots$$

What is wrong with this formula???

## Horizon Problem

$$U(s_0, \dots, s_N) = R(s_0) + R(s_1) + \dots + R(s_N)$$

The sum may be arbitrarily large depending on  $N$

Need to know  $N$ , the length of the sequence (finite horizon)