Markov Decision Processes

A **Markov Decision Process** (**MDP**) is a classification of stochastic Dynamic Programming models.

An **MDP** consists of a finite set ***S*** of **states** and, for each state $s \in S$, a finite set A_s of alternative **actions**.

When in state s at **stage** n and action $a \in A_s$ is selected, a **reward** r(s,a) is earned or a **cost** c(s,a) is incurred.

The system then makes a **transition** into another state s' with probability $p^a_{s,s'} \equiv P\left\{X_{n+1} = s' \mid X_n = s \text{ \& action } a \in A_s \text{ is selected}\right\}$.

Notes:

- *we assume stationary transition probabilities! That is, for a given state/action combination i & j, the value of p_{ij} is the same for all stages.*

- the "*Markov property*" is assumed, that is, the future behavior of the process is dependent *only* upon the current state (and the action selected), and not on any prior history.

- *MDPs may have either a **finite** number of stages or **infinitely** many stages.*

Examples

Maintenance planning

Inventory replenishment

No-claim limits for auto insurance

Objective criterion

- *Finite number* ***N*** *of stages*:
  - Total *expected cost* or return
  - Total *expected discounted cost* or return
- *Infinitely many stages*:
  - *Average cost per stage* (assuming steady state behavior)
  - total *expected discounted cost* or return

Finite horizon MDPs may be solved by

- *stochastic dynamic programming* (DP).

Infinite horizon MDPs may be solved by

- Value iteration (limit of DP solution as $N \to \infty$)
- Policy improvement algorithm
- Linear programming algorithm