# Markov Processes

## Markov Process

A Markov Process (MP), or Markov Chain, is a directional node-edge graph which satisfies the Markov property. The Markov property requires that the probability of reaching any successor state depends ONLY on the current state (i.e. the states visited prior to the current state have no impact on future outcomes). Each node in the MP is a state , and each edge is some state transition that occurs with probability . MPs have a start state, a terminal state, and any number of intermediate states. Each state transition has a unique probability, and the sum of probabilities of all state transitions from any given state must equal 1. State transitions can be self-targeting.

Therefore, an MP sets forth the fundamental framework that we can use to establish a state navigation task with probabilistic state transitions. However, in an MP there is no maximizing objective per se (other than an objective related to network traversal, e.g. reaching the terminal state in the fewest number of steps).

## Markov Reward Process

An MRP has all of the properties of an MP, with two additions: First, we introduce the concept of a reward function which gives rewards to our agent for arriving at a particular state:

We also introduce a discount factor which reflects the time-value of the reward, i.e. a reward realized now is more valuable than an equivalent reward realized later in time.

In this formulation, the reward system now provides an objective function which can be maximized, that being the present value of all rewards accrued over a given simulation rollout.

## Markov Decision Process

A Markov decision process is a further extension of the MRP, in which actions are added to each given state. In other words, each state has a corresponding set of actions that can be taken from that state, and each state-action pair has a set of transition probabilities associated with a set of successor states.

NOTE: An MDP subject to a given policy (i.e. a set of prescribed actions for any possible state) simply reduces to an MRP!

## Value Function

Generally, the value function is meant to capture the expected value of the total discounted reward starting from any given state. The solution to the value function is obtained by decomposing the value function into the immediate reward, and the sum of all future rewards starting from the successor state. We note that the second part is in fact the value function for the successor state, and so solving this recurrence is the foundation of the Bellman equation.

In the case of an MRP, the actions are all prescribed, and so at any point the value function can be computed analytically by solving a system of matrix equations. In the case of an MDP, one must sum over all possible actions, and so the value function must be computed using iterative methods.