

# Notes for chapter 3: Markov Processes

## Formal Definitions for Markov Processes

**Definition** (Markov Process). A Markov Process consists of:

- A countable set of states  $\mathcal{S}$  and a set  $\mathcal{T} \subset \mathcal{S}$
- A time-indexed sequence of random states  $S_t \in \mathcal{S}$  for time steps  $t = 0, 1, 2, \dots$  where each random state satisfies:  $\mathbb{P}[S_t | S_{t-1}, S_{t-2}, \dots, S_0] = \mathbb{P}[S_t | S_{t-1}]$
- Termination: if an outcome for  $S_T$  is in  $\mathcal{T}$  then this sequence outcome terminates at time step  $T$

**Definition** (Time Homogeneous Markov Process). A Markov Process is Time Homogeneous when  $\mathbb{P}[S_t | S_{t-1}]$  is independent of  $t$ .

This property means that we can define the transition probability of a Time-Homogeneous Markov Process as follows:

$\mathcal{P} : (\mathcal{S} - \mathcal{T}) \times \mathcal{S} \rightarrow [0, 1]$ ,  $\mathcal{P}(s, s') = \mathbb{P}[S_{t+1} = s' | S_t = s]$ , for all  $t = 0, 1, 2, \dots, s \in (\mathcal{S} - \mathcal{T}), s' \in \mathcal{S}$ . Where  $s$  is the source state and  $s'$  the destination state.

Note that any Markov Process that is not Time Homogeneous can be converted to one that is, by augmenting all states that are dependent of time index  $t$ .

## Starting States

We'd like to specify a probability distribution of start states so we can perform simulations and compute the probability distributions of future states at specific time steps. For this end we separate the following 2 questions:

- Specification of transition probability  $\mathcal{P}$
- Specification of the probability distribution of start states  $\mu : \mathcal{N} \rightarrow [0, 1], \mathcal{N} = \mathcal{S} - \mathcal{T}$ .

Given  $\mu$  together with  $\mathcal{P}$  we can generate sampling traces of Markov Processes.

## Stationary Distribution of a Markov Process

**Definition** (Stationary Distribution). The Stationary Distribution of a Discrete-Time Time-Homogeneous Markov Process with state space  $\mathcal{S} = \mathcal{N}$  and transition probability function  $\mathcal{P} : \mathcal{N} \times \mathcal{N} \rightarrow [0, 1]$  is probability distribution  $\pi : \mathcal{N} \rightarrow [0, 1]$  s.t.  $\pi(s') = \sum_{s \in \mathcal{N}} \pi(s) \mathcal{P}(s, s'), \forall s \in \mathcal{S}$

The intuition behind the stationary distribution is that if we let the Markov Process run forever then in the long run the states occur with relative frequencies given by the distribution  $\pi$  independent of the time steps.

For Finite-States, Discrete-Time, Time-Homogeneous Markov Processes with state space  $\mathcal{S} = s_0, s_1, \dots, s_n = \mathcal{N}$  we can specialize the definition of the  $\pi$  distribution as follows:

$\pi(s_j) = \sum_{i=0}^n \pi(s_i) \mathcal{P}(s_i, s_j), \forall j = 0, 1, 2, \dots, n$ . Let  $\boldsymbol{\pi}$  be a column vector of length  $n$  and let  $\mathbf{P}$  be the  $n \times n$  transition probability matrix. Rows being the source states and Columns being the destination states with each row of course summing to 1. Then we can write:  $\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \mathbf{P} \iff \mathbf{P}^T \boldsymbol{\pi} = \boldsymbol{\pi}$ . That is  $\boldsymbol{\pi}$  is just an eigenvector of  $\mathbf{P}^T$  with eigenvalue 1.

## Formalism of Markov Reward Processes

The main purpose of Markov Reward Processes is to calculate how much reward would accumulate in expectation from each of the non-terminal states, if we let the process run indefinitely, bearing in mind that future rewards need to be discounted properly.

**Definition** (Markov Reward Processes). A Markov Reward Process is a Markov process along with a time-indexed sequence of reward variables  $R_t \in \mathcal{D}$ ,  $t = 0, 1, 2, \dots$  satisfying the Markov Property:  $\mathbb{P}[(S_{t+1}, R_{t+1})|S_t, S_{t-1}, \dots, S_0] = \mathbb{P}[(S_{t+1}, R_{t+1})|S_t]$ .

We will also assume that each Markov Reward Process is Time-Homogeneous. With this assumption we can define the transition reward probability function of Time-Homogeneous Markov Reward Processes as follows:

$\mathcal{P}_{\mathcal{R}} : \mathcal{N} \times \mathcal{D} \times \mathcal{S} \rightarrow [0, 1]$ ,  $\mathcal{P}_{\mathcal{R}}(s, r, s') = \mathbb{P}[(S_{t+1} = s', R_{t+1} = r)|S_t = s]$ , for all  $s \in \mathcal{N}$ ,  $r \in \mathcal{D}$ ,  $s' \in \mathcal{S}$ , such that  $\sum_{r \in \mathcal{D}} \sum_{s' \in \mathcal{S}} \mathcal{P}_{\mathcal{R}}(s, r, s') = 1$ ,  $\forall s \in \mathcal{N}$ .

The transition probability of the implicit Markov Process is defined as:

$$\mathcal{P} : \mathcal{N} \times \mathcal{S} \rightarrow [0, 1], \mathcal{P}(s, s') = \sum_{r \in \mathcal{D}} \mathcal{P}_{\mathcal{R}}(s, r, s')$$

The reward transition function is defined as:

$$\mathcal{R}_T : \mathcal{N} \times \mathcal{S} \rightarrow \mathbb{R}, \mathcal{R}_T(s, s') = \mathbb{E}[R_{t+1}|S_{t+1} = s', S_t = s] = \sum_{r \in \mathcal{D}} \frac{\mathcal{P}_{\mathcal{R}}(s, r, s')}{\mathcal{P}(s, s')} \cdot r = \sum_{r \in \mathcal{D}} \frac{\mathcal{P}_{\mathcal{R}}(s, r, s')}{\sum_{r' \in \mathcal{D}} \mathcal{P}_{\mathcal{R}}(s, r, r')} \cdot r$$

We can express  $\mathcal{R}_T$  into a more compact version sufficient for performing key calculations involving Markov Reward Processes. This reward function  $\mathcal{R}$  is defined as:

$$\mathcal{R} : \mathcal{N} \rightarrow \mathbb{R}, \mathcal{R}(s) = \mathbb{E}[R_{t+1}|S_t = s] = \sum_{s' \in \mathcal{S}} \mathcal{P}(s, s') \mathcal{R}_T(s, s') = \sum_{s' \in \mathcal{S}} \sum_{r \in \mathcal{D}} \mathcal{P}_{\mathcal{R}}(s, r, s') \cdot r$$

## Value function of a Markov Reward Process

We define the future random Return of our process at a specific time step  $t = 0, 1, 2, \dots$  as:  
 $G_t = \sum_{i=t+1}^{\infty} \gamma^{i-t-1} R_i = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots$ , for some discount factor  $\gamma \in [0, 1]$ .

And we define the Value Function as:

$$V : \mathcal{N} \rightarrow \mathbb{R}, V(s) = \mathbb{E}[G_t|S_t = s], \forall s \in \mathcal{N}, \forall t = 0, 1, 2, \dots$$

An important result found by Richard Bellman is that the Value Function has a recursive structure:  
 $V(s) = \mathcal{R}(s) + \gamma \sum_{s' \in \mathcal{N}} \mathcal{P}(s, s') V(s')$ ,  $\forall s \in \mathcal{N}$