Mathematics of Reinforcement Learning

Exercise Class 5

Exercise 1.
Let (S, A, D, p, r, γ) be a Markov Decision Model, π a policy and μ a probability measure on S. From the *Ionescu-Tulcea Theorem for Markov Decision Models*, let $\mathbb{P}_\mu^\pi$ be the unique probability measure on $(\Omega, \mathcal{A})$, with the property

$$\mathbb{P}_\mu^\pi\big[B \times \bigtimes_{k=n+1}^{\infty} (S \times A)\big] = \sum_{(s_0,a_0,\dots,s_n,a_n)\in B} \mu[\{s_0\}]\pi(a_0|s_0) \prod_{k=1}^{n} p(s_k|s_{k-1}, a_{k-1})\pi(a_k|s_k),$$

for all $B \subseteq (S \times A)^{n+1}$ and $n \in \mathbb{N}_0$.
Prove that $\mathbb{P}_\mu^\pi$ satisfies the following properties

1. For all $n \in \mathbb{N}_0$ and $(s, a) \in S \times A$

$$\mathbb{P}_\mu^\pi[A_n = a \mid S_n = s] = \pi(a \mid s).$$

2. For all $n \in \mathbb{N}_0$ and $(s_k, a_k)_{k=0,\dots,n+1} \in (S \times A)^{n+2}$

$$\begin{aligned}\mathbb{P}_\mu^\pi[S_{n+1} = s_{n+1}, &A_{n+1} = a_{n+1} \mid S_0 = s_0, A_0 = a_0, \dots, S_n = s_n, A_n = a_n] \\ &= \mathbb{P}_\mu^\pi[S_{n+1} = s_{n+1}, A_{n+1} = a_{n+1} \mid S_n = s_n, A_n = a_n].\end{aligned}$$

◇

Exercise 2 (Bellman equation)**.** In the setting of exercise class 2 task 4, show that for any deterministic policy $\pi\colon S \to A$, $t = 0, 1, 2$ and $s = (t, p, w) \in S$ the following equation holds

$$V^\pi(t, p, w) = \gamma \sum_{(t+1,p',w')\in S} p(t + 1, p', w' \mid s, \pi(s))V^\pi(t + 1, p', w').$$

Compute for all $s = (3, p, w) \in S$ the value $V^\pi(s)$. ◇

Exercise 3 (Programming exercises)**.**
Complete the programming exercises given in the Jupyter notebook file. ◇