Mathematics of Reinforcement Learning

Exercise Class 2

Exercise 1.
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a finite probability space, i.e. $\Omega = \{\omega_1, \ldots, \omega_n\}$. Show that there exists no sequence $\{Z_n\}_{n\in\mathbb{N}}$ of independent, almost surely non-constant random variables. ◇

Exercise 2.
Let $(\Omega, \mathcal{F})$ be a measurable space with $\Omega = \{\omega_1, \ldots, \omega_n\}$ finite. Show that there exists exactly one partition $\mathcal{P}$ of Ω which generates $\mathcal{F}$. ◇

Exercise 3.
Let $(\Omega, \mathcal{F})$ be a measurable space with $\Omega = \{\omega_1, \ldots, \omega_n\}$ finite. Let moreover $X\colon \Omega \to \mathbb{R}$ and denote by $\mathcal{P} = \{P_1, \ldots, P_j\}$ the unique partition of Ω which generates $\mathcal{F}$. Show that X is $\mathcal{F}$-measurable if and only if X is constant on each $P \in \mathcal{P}$.
Hint: Use Exercise 2. ◇

Definition 1 (CRR Model)**.** Let $T \in \mathbb{N}$ be a finite time horizon, $q \in (0,1)$ and $u, d > -1$ with $d < u$. We fix a sequence of independent Bernoulli random variables $R_1, \ldots, R_T$ taking values in $\{d, u\}$ such that

$$\mathbb{P}[R_t = u] = q \quad \text{and} \quad \mathbb{P}[R_t = d] = 1 - q, \qquad t = 1, \ldots, T.$$

The process $P = \{P_t\}_{t=0,1,\ldots,T}$ with $P_0 \in \mathbb{R}$ and

$$P_t \coloneqq (1 + R_t)P_{t-1}, \qquad t = 1, \ldots, T$$

models the price of a financial asset in the *Cox–Ross–Rubinstein* (CRR) model. The following figure shows the evolution of the risky asset in the CRR model.

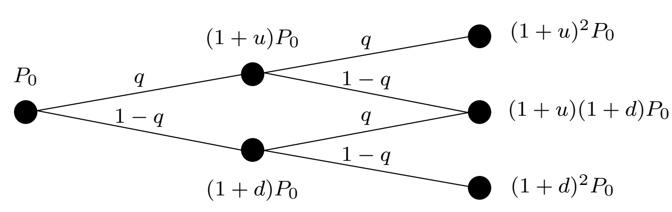

Figure 1: Evolution of P in the CRR model.

Exercise 4.
Let P be the price of a financial asset in the CRR model with time horizon $T = 3$, $P_0 = 8$, $d = -0.5$, $u = 2$, and $q = 0.5$. Define a Markov decision model that allows trading of the asset P.
For that, assume that the agent starts out with an initial wealth of $w_0 := 100$ and they are only allowed to hold integer amounts of P. Negative positions (short-selling) and positions exceeding the total wealth (borrowing) are not allowed. The objective of the agent is to maximize expected log-utility of terminal wealth, that is, the expectation of $\log(W_T)$ with W_T denoting the wealth at time T.
Hint: States should be three-dimensional and contain the current time, the financial asset price, and the agent's wealth. Actions can be modeled as the amount of wealth invested into financial asset. $\diamond$