Markov Decision Processes with Applications to Finance

Exercises

Training Session 1

Exercise 1
Let $\mathcal{G}$ be a sub σ-field of $\mathcal{F}$ and X be an independent r.v. of $\mathcal{G}$ of law $\mathbb{P}_X$. We consider a $\mathcal{G}$-measurable r.v. Y, and define the r.v. Z by $Z = \Phi(X, Y)$ where Φ is a measurable function. Show that under suitable integrability conditions

$$\mathbb{E}[Z|\mathcal{G}] = \int_{\mathbb{R}} \Phi(t, Y)\, d\mathbb{P}_X(t).$$

Exercise 2
Assume that $X \sim \mathcal{N}(0, \sigma^2)$ and that $Z = \exp(-\frac{\sigma^2}{2}Y^2 + XY)$ with Y, X independent r.v..

1. *What is $\mathbb{E}[Z|Y]$?*
2. *Show that $\mathbb{E}[Z] = 1$.*

Exercise 3
Give the conditional law of Y given X if (X, Y) is a $\mathbb{R}^2$-random valued random variable whose law has a probability density with respect to the Lebesgue measure given by

$$f : (x, y) \in \mathbb{R}^2 \to \lambda^2 \exp(-\lambda y)\, \mathbf{1}_{0 \leq x \leq y}, \quad \lambda > 0.$$

Deduce then $\mathbb{E}(Y|X)$.

Exercise 4
Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean μ. We denote

$$S_N = \sum_{i=1}^{N} X_i.$$

1. *By using a symmetry argument show that for every $i \in \{1, \ldots, N\}$ we have that*

$$\mathbb{E}(X_i \mid S_N) = \frac{S_N}{N}$$

2. *Deduce* $\mathbb{E}(S_M \mid S_N)$ *for all* $M \leq N$.

Exercise 5

We consider two r.v. X *and* Y *such that* $Y \in \mathbb{L}^2$. *We define the conditional variance of* Y *given* X *as the r.v.* $\mathbb{V}(Y|X)$ *defined by*

$$\mathbb{V}(Y|\, X) \equiv \mathbb{E}\left(Y^2\middle|\, X\right) - (\mathbb{E}(Y|\, X))^2.$$

1. *Show that* $\mathbb{V}(Y|X)$ *is a non-negative r.v..*

2. *Show that*

$$\mathbb{E}(\mathbb{V}(Y|\, X)) = \mathbb{V}(Y) - \mathbb{V}(\mathbb{E}(Y|\, X)).$$

Exercise 6

A stochastic ARCH(1) process [1] *is a discrete time process* $(X_n)_{n\geq 0}$ *used to model financial time series (*X_n *is the price of the asset at time* n*). It is defined recursively by*

$$X_n = \varepsilon_n \sqrt{\alpha_0 + \alpha_1 X_{n-1}^2}$$

where $\alpha_0 \geq 0$, $\alpha_1 > 0$ *are given numbers and* $(\varepsilon_n)_{n\geq 1}$ *is an i.i.d. sequence of centred random variables with variance* 1 *independent of* X_0. *We denote* $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$.

1. *Is the process* $(X_n)_n$ *a Markov process?*

2. *Write a Python program to simulate samples trajectories of* $(X_n)_n$.

3. *Show that* $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = 0$.

4. *Let* $Y_n^2 = \mathbb{V}(X_n \mid \mathcal{F}_{n-1})$. *Show that*

$$Y_n^2 = \alpha_0 + \alpha_1 X_{n-1}^2.$$

5. *Show that the volatility* $\mathbb{V}(X_n)$ *of the asset price at time* n *satisfies*

$$\mathbb{V}(X_n) = \alpha_0 + \alpha_1 \mathbb{V}(X_{n-1}).$$

   What is the behaviour of $\mathbb{V}(X_n)$ *as* n *goes to infinity ?*

6. *Use a Python simulation to check this theoretical argument.*

[1] For "Auto-Regressive Conditional Heteroscedasticity".

Exercise 7 (Insurance Risk Management)

An insurance company has a wealth Y_n composed of financial assets at the beginning of the year number $n \geq 0$. Each year it gets a fixed prime $P > 0$ but has to pay $C_n \geq 0$ to reimburse the policyholders . We assume that $C_1, C_2, \dots$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$, $0 < \mu < P$, $\sigma^2 > 0$ r.v..

1. *Show that $Y_{n+1} = Y_n + P - C_{n+1}$.*
2. *Explain why the insurance company cannot survive if $\mu > P$.*
3. *Find $t \in \mathbb{R}$ such that $\mathbb{E}[e^{t(P-C_1)}] = 1$. We recall that if $Z \sim \mathcal{N}(0,1)$ then $\mathbb{E}(e^{tZ}) = e^{t^2/2}$. We fix such a t in the following.*
4. *Let $Z_n = \min\{e^{tY_n}, 1\}$.*
5. *Show that $(Z_n)_n$ is a non-negative $(\mathcal{F}_n^C)_n$-supermartingale.*
6. *Let $T = \inf\{n \ ; \ Y_n \leq 0\}$. Is T a $(\mathcal{F}_n^C)_n$ stopping time? Can we apply Doob's optional stopping theorem with T and the supermartingale $(Z_n)_n$? Justify (Hint:Apply Doob's optional stopping theorem with the bounded stopping time $\inf(T, m)$, $m \geq 1$, and let $m \to \infty$).*
7. *Deduce that*

$$\mathbb{P}(Y_n \leq 0 \text{ for a } n \geq 0) \leq \exp\Big(- \frac{2(P-\mu)Y_0}{\sigma^2}\Big).$$

Training Session 2

Exercise 8 *1. We consider a discrete time Markov Chain $(X_n)_{n\geq 0}$ on $\{0,1\}$ with transition matrix*

$$P = \begin{pmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{pmatrix}.$$

Show that this Markov chain is irreducible, positive recurrent, has a unique invariant probability measure μ. Compute μ.

2. *Recall the ergodic theorem for Markov chains.*

3. *Write a Python function that returns the realisations of a Markov chain with state space $\{0,\dots,d-1\}$, transition matrix P of size $d\times d$ with initial condition $x\in\{0,\dots,d-1\}$ on the time interval $\{0,1,\dots,n-1\}$.*

4. *Plot the graph of the realisations of the Markov chain of question 1.*

5. *Illustrate numerically the ergodic theorem for the Markov chain of question 1.*

6. *Write a Python function which returns the invariant probability measure μ of an irreducible positive recurrent Markov chain with state space $\{0,\dots,d-1\}$, transition matrix P of size $d\times d$ with initial condition $x\in\{0,\dots,d-1\}$. Check the consistence of this function for the Markov chain of question 1.*

Exercise 9
Write a Python function which returns the realisations of a Markov Decision Process (starting in the state x) with state space $\{0,\dots,d-1\}$ and possible matrix transitions $P[0], P[1],\dots,P[K-1]$ (hence with K possibles actions – not necessary to specify) during the time window $\{0,\dots,n-1\}$ by following the strategy imposed by the vector a $(a[0],\dots,a[n-2])$. In the case of a given Markov Decision Process with 4 states and 2 possible actions, plot the trajectories of this Markov Decision Process under different strategies.

Exercise 10
Une personne possède $r\geq 1$ parapluies, qu'elle utilise uniquement entre son domicile

et son bureau. Lorsqu'il pleut à son départ (du domicile ou du bureau), et seulement dans ce cas, elle prend un parapluie, à condition qu'il y en ait un. La probabilité qu'il pleuve est $p \in]0,1[$, indépendament du passé, chaque fois qu'elle sort de son bureau ou de son domicile.

1. *Montrer que l'on peut définir une chaîne de Markov $(X_n)_{n\geq 1}$, $X_n \in \{0,1,\ldots,r\}$, à $r+1$ états en considérant le nombre X_n de parapluies disponibles juste avant le n-ième départ du bureau.*

2. *Ecrire la matrice de transition de cette chaîne et montrer qu'elle est de la forme*

$$\begin{pmatrix} \beta & 1-\beta & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \alpha & 1-2\alpha & \alpha & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \alpha & 1-2\alpha & \alpha & 0 & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \alpha & 1-2\alpha & \alpha \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1-\gamma & \gamma \end{pmatrix}$$

   pour des valeurs de $\alpha \in]0,1/2[, \beta \in]0,1[, \gamma \in]0,1[$ que l'on exprimera en fonction de p. La matrice ci-dessus est carrée de taille $r+1$.

3. *Dire si cette chaîne est irréductible ou pas et dans le deuxième cas, préciser ses classes de communications. Existe-t-il des états absorbants ?*

4. *Rechercher toutes les probabilités stationnaires de la chaîne.*

5. *Supposons qu'il n'y ait initialement tous les parapluies au bureau. Quel est le temps moyen que la personne devra attendre pour de nouveau retrouver tous les parapluies au bureau ?*

Training Session 3

Exercise 11 (Forward Bellman equation and Applications)

A) *We consider a stationary MDP with a finite state space with n elements and a finite actions space with m elements. A state is denoted by x and an action by a. We assume that in each state, all actions are allowed. The reward matrix $(r(x,a))_{x,a}$ is of size (n,m). We denote the value function in the forward Bellman equation by $J := (J(x))_{x\in S}$ is a vector of size n indexed by $x \in S$. The transition kernel Q is represented by a matrix of size (mn,n) and in blocks columns, it contains successively the transition matrices $Q(\cdot|\cdot,a)$ for $a \in A$. The dicount factor is denoted by β.*

1. *Write a Python function which gives gives the argmax and the max of $J_{new}(x) = \max_{a\in A}\left\{r(x,a) + \beta\sum_{y\in S} Q(y|x,a)J(y)\right\}$.*
2. *Write a Python function which computes the optimal policy as well as the value functions by following this policy in the forward Bellman algorithm. The time horizon is N and the initial value function are given by $g := (g(x))_{x\in S}$ which is a column vector of size n.*

B) **Application:** *In order to survive, an animal must forage for food in one of m distinct areas. In area $k \in \{0,\dots,m-1\}$, the animal survives predation with probability p_k, finds food with probability q_k, and, if it finds food, gains e_k energy units. The animal expends one energy unit every period and has a maximum energy carrying capacity $\bar{s}$. If the animal's energy stock drops to zero, it dies. The goal is to find the foraging pattern which maximizes the animal's probability of surviving to reproduce after period N?*

1. *Define the Stationary Markov Decision Process associated to this problem which permits to solve the problem (one time period reward is 0).*

2. *Write the corresponding Bellman equation.*
3. *Use the previous Python functions defined to plot the optimal strategy and the probability of procreating.*

Training Session 4

Exercise 12

A) *We consider a stationary MDP with a finite state space with n elements and a finite actions space with m elements. A state is denoted by x and an action by a. The matrix $(D(x,a))_{x,a}$ of size (n,m) is such that $D(x,a)=1$ if the action a is allowed when in state x and 0 otherwise. The reward matrix $(r(x,a))_{x,a}$ is of size (n,m). We denote the value function in the forward Bellman equation by $J := (J(x))_{x\in S}$ is a vector of size n indexed by $x \in S$. The transition kernel Q is represented by a matrix of size (mn,n) and in blocks columns, it contains successively the transition matrices $Q(\cdot|\cdot,a)$ for $a \in A$. The discount factor is denoted by β.*

1. *Show that if we define $r_{inf}(x) = \inf_{\substack{a \text{ s.t.} \\ D(x,a)=1}} r(x,a)$, $J_{sup} = \sup_{x\in S} J(x)$, $J_{inf} = \inf_{x\in S} J(x)$ and $\tilde{r}(x,a) = r(x,a)$ if $D(x,a)=1$ and $\tilde{r}(x,a) = r_{inf}(x) + \beta[J_{inf} - J_{sup}] - 1$ otherwise then we have*

$$\max_{\substack{a \text{ s.t.} \\ D(x,a)=1}} \left\{ r(x,a) + \beta \sum_{y\in S} Q(y|x,a)J(y) \right\} = \max_{a\in A} \left\{ \tilde{r}(x,a) + \beta \sum_{y\in S} Q(y|x,a)J(y) \right\}$$

2. *Write a Python function which gives gives the argmax and the max of $J_{new}(x) = \max_{\substack{a \text{ s.t.} \\ D(x,a)=1}} \left\{ r(x,a) + \beta \sum_{y\in S} Q(y|x,a)J(y) \right\}$.*
3. *Write a Python function which computes the optimal policy as well as the value functions by following this policy in the forward Bellman algorithm. The time horizon is N and the initial value function are given by $g := (g(x))_{x\in S}$ which is a column vector of size n.*

B) **Application:**

A mine operator must determine how to extract ore for a mine that will be shut down and abandoned at the end of the N^{th} year. The market price of one ton of ore is p and the total cost of extracting x tons of ore in any year is $c(s,x) = x^2/(1+s)$ where s is the stock of ore available at the beginning of the year in tons. The mine currently contains $\bar{s}$ tons of ore. We assume the ore extracted per year is measured only in tons. The discount factor is denoted by β.The goal is to find the best strategy to extract ore.

1. *Define the Markov Decision Process associated to this problem.*
2. *Write the corresponding Bellman equation.*
3. *Define a Python function which computes numerically the optimal strategy and the value of the mine during the time evolution. Plot the optimal strategy and the value of the mine.*

Training Session 5 & 6

Exercise 13 (Stochastic Linear Quadratic Models)
We consider a MDP with state space $E = \mathbb{R}^m$, with actions space $A = \mathbb{R}^d$ and we assume that for each $x \in E$, $D_n(x) = A$ (all decision rules are admissible). The disturbances space is $M_{m,m}(\mathbb{R}) \times M_{m,d}(\mathbb{R})$ and disturbances $Z_n = (A_n, B_n)$ are supposed to be independent but not necessarily identically distributed and have finite variance. We assume that for positive definite symmetric matrix Q it holds that $\mathbb{E}\left[B_n^\dagger Q B_n\right]$ is also symmetric positive definite. When at time n, the state is X_n and the action $f_n(X_n) \in \mathbb{R}^d$ is taken, the state at time $n+1$ is $X_{n+1} = A_{n+1}X_n + B_{n+1}f_n(X_n)$, i.e. the system transition functions are given by

$$T_n(x, a, Z) = Ax + Ba, \quad Z = (A, B).$$

We assume that the one-stage reward (independent of the action) is given by $r_n(x, a) = x^\dagger Q_n x$ and the terminal reward (independent of the action) is $g_N(x, a) = x^\dagger Q_N x$. Here $Q_0, Q_1, \ldots, Q_N$ are deterministic symmetric positive definite matrices of size m. The discount factor is 1. The aim is to minimize

$$\mathbb{E}_x^\pi\left[\sum_{k=0}^{N} X_k^\dagger Q_k X_k\right].$$

1. *Show that the Integrability Assumption is satisfied.*

2. *Show that the Structure Assumption is satisfied by choosing accordingly the sets*

   $$\mathbb{M}_n = \{v : \mathbb{R}^m \to [0, \infty)\ ;\ v(x) = x^\dagger Q x \quad \text{with } Q \text{ symmetric positive definite }\}$$

   and

   $$\Delta_n := \Delta = \{f : E \to A\ ;\ f(x) = Cx \quad \text{for some } C \in M_{d,m}(\mathbb{R})\ \}$$

   and compute the minimal operator $T_n v$.

3. *Solve the stochastic linear quadratic problem by showing that at any time n, the value function is given by*

   $$V_n(x) = x^\dagger \tilde{Q}_n x, \quad x \in \mathbb{R}^m,$$

where the matrix $\{\tilde{Q}_n \ ; \ 0 \leq n \leq N\}$ *are defined recursively backward with* $\tilde{Q}_N = Q_N$. *Provide also an optimal policy defined in terms of* $\{\tilde{Q}_n, A_n, B_n \ ; \ 0 \leq n \leq N\}$.

4. *Write a Python function providing an optimal policy as well as the sequence of matrices* $(\tilde{Q}_n)_n$ *during the time evolution.*

Training Session 7&8

Exercise 14
Consider the consumption problem with horizon time N and utility function at stage n given by $U : a \in (0, +\infty) \to U(a) \in \mathbb{R}$. We recall that X_n represents the capital at time n and that an action $f_n(X_n) \in (0, X_n]$ at time n consists to consume a part of the capital. Using a representation of the corresponding MDP we assume that $(Z_n)_n$ is a sequence of i.i.d. positive real random variables with finite expectation and that

$$X_{n+1} = (X_n - f_n(X_n))Z_{n+1}.$$

The aim is thus to maximize the value function at time 0 given by

$$V_0(x) = \sup_{\pi=(f_0,\ldots,f_{N-1})} \mathbb{E}_x^{\pi}\left[\sum_{k=0}^{N-1} U(f_k(X_k)) + U(X_N)\right] .$$

1. *Show that if U is increasing and concave then the integrability assumption is satisfied.*

2. *Show that the maximal operator (independent of n) is given by*

$$\mathcal{T}v(x) = \sup_{a\in(0,x]} \{U(a) + \mathbb{E}\left[v\left((x-a)Z_1\right)\right]\} .$$

3. *We assume that Z_1 takes value $1/2$ with probability $1/2$ and value 1 with probability $1/2$. We fix som interval $[0, A]$ and $n \geq 1$ and divide $[0, A] = [x_0 = 0, x_1] \cup [x_1, x_2] \cup \ldots \cup [x_{n-1}, x_n] = A$ into n consecutive intervals of the same length. Assuming that v is numerically approximated at the points x_i's, propose a Python function which provides an approximation of $\mathcal{T}v$ at the same points and compute an approximation of the optimizer a in the variational formula.*

4. *We assume now that $U = \log$. Show that the Structure Assumption SA_N is satisfied (define spaces $\mathbb{M}_n$ and Δ_n) and find an optimal N-stage policy provided by the Backward Induction Algorithm.*

Training Session 9

Exercise 15 (European option pricing and hedging)
Consider the Cox-Ross-Rubinstein model with horizon time $N \geq 1$. There is one bond with interest rate i and one stock where the relative price changes can only take two values u (with probability $p \in (0,1)$) and d (with probability $1-p$). We can define it on the filtered probability space $(\Omega, (\mathcal{F}_n)_{0 \leq n \leq N}, \mathbb{P})$ with $\Omega = \{u, d\}^N$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_N = \mathcal{P}(\Omega)$. If $\omega = (\omega_1, \ldots, \omega_N) \in \Omega$, the relative price change at time n is $\tilde{R}_n(\omega) = \omega_n$ and the price $S_n(\omega)$ of the stock is $S_0\omega_1 \ldots \omega_n$. Here S_0 is the initial price of the stock at time 0. The price of the bond at time n is $(1+i)^n$ (we assume the initial price of the bond to be 1 at time 0).

1. *Write a Python function which plots the evolution of the prices.*
2. *Show that a necessary and sufficient condition for the absence of arbitrage is $d < 1 + i < u$. We assume now this condition holds.*
3. *Show that there exists a unique probability $\mathbb{P}^*$ equivalent to $\mathbb{P}$ such that under $\mathbb{P}^*$ the sequence of relative price changes $(\tilde{R}_n)_n$ are i.i.d. of mean $1+i$. This probability is called a risk-neutral probability.*
4. *Show that if φ is a self-financing strategy then under $\mathbb{P}^*$ the discounted values of the portfolio $(1+i)^{-n} X^{\varphi}_{n,-}$ is a martingale with respect to the filtration $(\mathcal{F}_n)_n$.*
5. *Let h be a non-negative function such that $\mathbb{E}[h(S_N)] < \infty$. We say that a strategy φ is an admissible strategy replicating h if φ is self-financing, such that $X^{\varphi}_{n,-} \geq 0$ with probability 1 (w.r.t. $\mathbb{P}$ or $\mathbb{P}^*$) and $X^{\varphi}_{N,-} = h(S_N)$. By using the previous question prove there exists a unique admissible strategy $\bar{\varphi}$ replicating h and define it explicitly. Give the value of $X^{\bar{\varphi}}_{0,-}$.*
6. *Write a Python function which plots the replicating strategy.*