Théorie algorithmique des jeux et apprentissage par renforcement: Examen

Lundi 11 décembre 2023

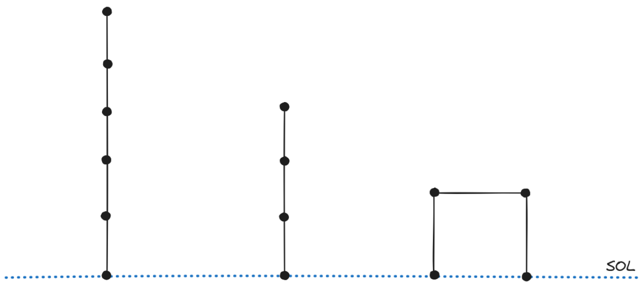
Durée : 2 heures.
Tous les documents sur papier sont autorisés. Les appareils électroniques sont interdits. Le barème est indicatif. Les réponses doivent être justifiées.

Exercice 1 (4 points)

On rappelle les règles du jeu combinatoire du Hackenbush qui se joue sur un dessin constitué de segments :

- tant que le dessin contient un segment, le joueur à qui c'est le tour de jouer choisit un segment restant de la figure;
- ce segment est supprimé du dessin, ainsi que tous les segments qui ne sont plus connectés à la ligne représentant le sol;
- le premier joueur qui ne peut plus jouer a perdu.

On considère l'instance du jeu de Hackenbush de la figure suivante. Les segments sont les traits entre 2 points : il y en a 11 sur la figure. Le sol est représenté par la ligne en pointillé. Montrer qu'il existe une stratégie gagnante pour le joueur qui joue en premier, et donner un coup optimal depuis la position de la figure.



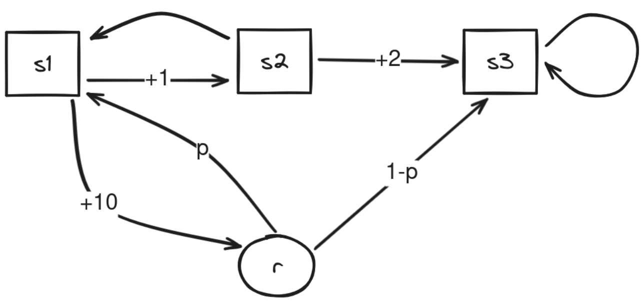
Exercice 2 (5 points)

Soit le processus de décision Markovien (MDP dans la suite) avec facteur d'actualisation γ représenté sur la figure ci-dessous. Les états décisionnels sont $s1$, $s2$ et $s3$. L'état $s3$ n'a qu'une seule action possible, et peut être considéré comme un puits du MDP. L'état r est un état aléatoire. Les probabilités de transitions sont données par le paramètre p. Les récompenses sont données par les valeurs écrites sur les arcs. Lorsqu'elle n'est pas spécifiée, la récompense est nulle. On cherche à maximiser le gain actualisé (tenant compte du paramètre γ) espéré.

a) Calculer la valeur des états décisionnels en fonction des paramètres p et γ lorsque la stratégie choisit la transition vers r pour l'état $s1$ et la transition vers $s1$ pour l'état $s2$.

b) Donner les équations d'optimalité de Bellman pour la valeur optimale du MDP.

c) A quelles conditions sur γ et p la stratégie précédente est-elle optimale?



Exercice 3 (5 points)

Considérons la figure suivante.

Deux jetons, symbolisés par des losanges, sont initialement positionnés aux emplacements 2 et i. Ils se trouvent sur un chemin menant à un trésor. À chaque tour, un joueur choisit l'un des deux jetons. Celui-ci se déplace ensuite selon la règle suivante : s'il n'y a qu'un seul sommet voisin (positions 2 et iii), il se déplace vers ce voisin; sinon, s'il y a deux voisins, le déplacement est sélectionné de manière aléatoire (avec une probabilité de 0.5) vers l'un de ces deux voisins. En particulier, le trésor est voisin des positions i et 1. L'objectif du joueur est, compte tenu de la position des deux jetons, de choisir celui à déplacer afin de **minimiser le nombre moyen d'étapes** avant qu'un des deux jetons n'atteigne le trésor (seul le premier compte, les deux n'ont pas besoin d'y parvenir).

a) Déterminez le nombre moyen d'étapes nécessaires pour atteindre le trésor si seul le jeton de gauche est déplacé.

b) Même question si seul le jeton de droite est déplacé.

c) Modéliser ce jeu sous forme de processus de décision markovien en spécifiant l'ensemble des états, les probabilités de transition et les récompenses. Vous pouvez vous aider d'un graphique représentant partiellement ce MDP (*indication* : l'espace d'état est le produit des espaces d'état des chaînes de Markov engendrées par les chemins à gauche et à droite du trésor).

d) Montrer que la stratégie consistant à ne déplacer que le jeton de gauche n'est pas optimale (*indication* : utiliser la position des jetons telle que donnée dans la figure).

Exercice 4 (3 points)

On considère le jeu sous forme normale à 2 joueurs suivant :

2,4	6,2	3,1
4,0	0,6	1,3
1,5	5,1	2,2

La première valeur correspond au gain du joueur qui choisit la ligne, la deuxième valeur est celle du joueur qui choisit le colonne. Les joueurs cherchent à maximiser leur gain.

a) Éliminer tant que c'est possible les stratégies strictement dominées du jeu.

b) Dans ce jeu réduit, représenter graphiquement les fonctions multivaluées de réponse optimale de chaque joueur, et en déduire les équilibres de Nash du jeu.

Exercice 5 (3 points)

On considère le jeu à deux joueurs suivant : les deux joueurs posent simultanément et en secret une pièce de monnaie côté Pile ou Face sur la table, puis ils révèlent ces pièces simultanément. Si les pièces sont de côté différent, alors personne ne gagne ni ne perd. Si les pièces sont du même côté alors si ce côté est Pile le joueur 1 prend 1 euro au joueur 2, et si ce côté est Face le joueur 1 prend $k > 0$ euros au joueur 2.

1) Modéliser ce jeu sous forme matricielle. De quel type de jeu s'agit-t'il?

2) Ecrire le programme linéaire associé à ce jeu et le résoudre. On pourra procéder numériquement ou graphiquement mais on attend une justification.

3) En déduire la valeur du jeu et calculer les stratégies d'équilibre des joueurs en fonction du paramètre k.