

Problème de Décision dans l'Incertain

Philippe Preux

M2 Informatique

Tout document autorisé

Durée : 1h30.

correction

30 septembre 2024

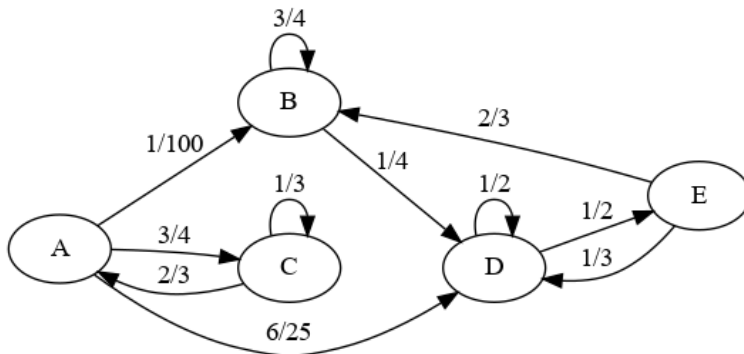
Prénom :

Nom :

Vous répondez aux questions directement dans le sujet. À l'issue de l'épreuve, vous me rendez ces feuilles et votre copie le cas échéant. Vous m'envoyez également le programme correspondant à l'exercice 6 à philippe.preux@univ-lille.fr dans un fichier nommé `PDI.cc1.mon-nom.py`.

Exercice 1

On considère ce graphe :



Les arcs sont étiquetés avec la probabilité de transition.

Estimer la probabilité d'occupation de chacun des états. Indiquer moyennes et écarts-types dans le tableau ci-dessous. Arrondir au centième.

Au centième près, la moyenne est la même quel que soit l'état de départ. À la cinquième décimale, ce n'est plus le cas mais on a demandé d'arrondir au centième.

De même, au centième, les écarts-types sont tous nuls.

état	A	B	C	D	E
moyenne	0	0,47	0	0,35	0,18
écart-type	0	0	0	0	0

Par ailleurs, comme il s'agit d'une probabilité d'occupation, que le graphe contient 5 états et qu'on ne peut pas quitter le graphe, la somme des moyennes doit être égale à 1 (aux arrondis près).

Exercice 2

On considère un problème de décision de Markov comme ceux que l'on a étudiés en cours, qui maximise la somme des retours pondérée par un facteur de dépréciation (γ). On suppose que les retours sont toujours positifs.

1. La valeur d'une politique dépend-elle de γ ?

☒ oui,

☐ non,

☐ parfois.

Justifier : La valeur V d'un état e est définie par $V(e) = \mathbb{E} \sum_t \gamma^t r_t$. Donc, $V(e)$ dépend de γ .

2. Si la valeur d'une politique dépend de γ , peut-on dire que pour deux valeurs $\gamma_1 < \gamma_2$, la valeur d'une politique fixée pour γ_1 est inférieure à sa valeur pour γ_2 ?

☒ oui,

☐ non,

☐ parfois.

Justifier : Puisque la valeur V d'un état e est définie par $V(e) = \mathbb{E} \sum_t \gamma^t r_t$, si $\gamma_1 < \gamma_2$, forcément $V(e, \gamma_1) < V(e, \gamma_2)$.

Exercice 3

On considère le problème du chauffeur de taxi. On considère la politique uniformément aléatoire.

Quelle est la probabilité de passer de l'état A à l'état B en effectuant l'action a_1 ? Détailler votre calcul ci-dessous.

Cette probabilité est : $Pr[A \rightarrow B|a_1] = \sum_{a \in \mathcal{A}} \pi(A \rightarrow B) \mathcal{P}(A, a_1, B) = \frac{1}{3} \mathcal{P}(A, a_1, B) = \frac{1}{3} \frac{1}{4} = \frac{1}{12}$.

Remarque : je numérote ici les actions à partir de 1 comme dans le polycopié. Certains les ont numérotées à partir de 0, j'en ai tenu compte dans la correction : dans ce cas la probabilité est $\frac{1}{4}$.

Exercice 4

On considère le problème du chauffeur de taxi et la politique suivante :

$$\pi(A, a_1) = 1/2, \pi(A, a_2) = 1/4, \pi(A, a_3) = 1/4,$$

$$\pi(B, a_1) = 1/4, \pi(B, a_3) = 3/4,$$

$$\pi(C, a_1) = 1/8, \pi(C, a_2) = 3/4, \pi(C, a_3) = 1/8,$$

Quelle est la valeur de cette politique ? Arrondir au centième près.

$$V^\pi(A) = 101,88$$

$$V^\pi(B) = 114,8$$

$$V^\pi(C) = 103,37$$

Exercice 5

On considère le problème du chauffeur de taxi.

Quelle est la valeur de la politique uniformément aléatoire pour les différentes valeurs de $\gamma \in \{0; 0,1; 0,5; 0,75\}$.

Remplir le tableau ci-dessous. Arrondir les valeurs au centième.

	$\gamma = 0$	$\gamma = 0,1$	$\gamma = 0,5$	$\gamma = 0,75$
$V^\pi(A)$	5	6,00	14,08	32,39
$V^\pi(B)$	15,5	16,57	24,92	43,40
$V^\pi(C)$	5,17	6,14	14,13	32,38

Le cas $\gamma = 0$ est un cas particulier. Si on applique sans réfléchir ce qui a été vu en cours, il est possible de tomber sur une division par zéro (puisque le critère d'arrêt est $\epsilon(1 - \gamma)/\gamma$). Cependant, pour $\gamma = 0$, l'équation de Bellman se simplifie en $V^\pi(e) = \sum_a \pi(e, a) \sum_{e'} \mathcal{P}(e, a, e') \mathcal{R}(e, a, e')$ qui se calcule très facilement.

Exercice 6

On considère l'évaluation de la valeur d'une politique par méthode de Monte Carlo.

Je rappelle que cela consiste à estimer la valeur en chaque état en simulant la dynamique à partir de cet état et en calculant $R(e_{t=0}) = \sum_{k \geq 0} \gamma^k r_k$, les r_k étant observés le long de la trajectoire. En simulant plusieurs trajectoires issues du même état e et en moyennant ces $R(e)$, on estime la valeur de cet état. On effectue ce même traitement à partir de chacun des états et on obtient une estimation de la valeur de la politique.

À faire :

1. Implanter cet algorithme.
2. Le mettre en œuvre sur le problème du chauffeur de taxi.
3. Quelle est la valeur estimée au centième près de la politique uniformément aléatoire et $\gamma = 0,9$? Indiquez-la ci-dessous, ainsi que l'écart-type et le nombre d'itérations effectuées pour l'estimer.

État	moyenne	écart-type	Nombre d'itérations
A	87,29	10,74	10^4
B	98,61	11,62	10^4
C	87,52	10,81	10^4

En faisant 100 itérations, j'obtiens respectivement 88,13, 99,26 et 87,52.

Avec 10^3 , j'obtiens respectivement 87,14, 99,14 et 87,45.

4. Comparer cette estimation à la valeur calculée avec les deux autres méthodes (résolution de système d'équations linéaires et itération de l'équation de Bellman).

On est à environ un dixième de la bonne valeur.