

Problème de Décision dans l'Incertain

CC 2

Philippe Preux

M2 Informatique

Tout document autorisé

Durée : 1h30.

Correction

14 octobre 2024

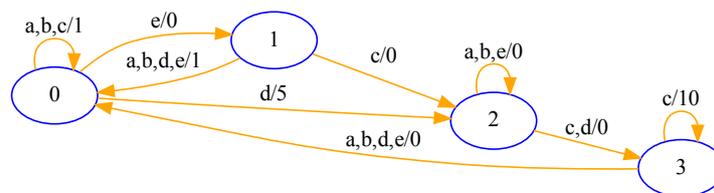
Prénom :

Nom :

À l'issue de l'épreuve, vous me rendez cette feuille.

Exercice 1

On considère le problème de décision de Markov dont l'ensemble des états est $\mathcal{S} = \{0, 1, 2, 3\}$, l'ensemble des actions est $\mathcal{A} = \{a, b, c, d, e\}$ et la fonction de transition est donnée sur la figure suivante :



Sur ce graphique, la notation $e/0$ entre l'état 0 et l'état 1 signifie que l'action e émise dans l'état 0 emmène dans l'état 1 et que le retour moyen perçu pendant cette transition est 0 ; la notation $a,b,d,e/1$ entre l'état 1 et l'état 0 signifie que si l'une des 4 actions a, b, d, e est émise dans l'état 1, le système passe dans l'état 0 et qu'un retour moyen de 1 est perçu ; *etc.*

Les transitions sont déterministes.

La fonction objectif est la fonction habituelle du cours $\zeta = \sum_t \gamma^t r_t$ que l'on cherche à maximiser.

Question 1 : combien y a-t-il de politiques déterministes possibles ? (Je dis bien « possibles », je ne dis pas « optimales ».)

Réponse :

5 actions, 4 états, donc $5^4 = 625$ politiques déterminites possibles.

Question 2 : juste en réfléchissant, pouvez-vous déterminer une politique optimale pour $\gamma = 0$?

Réponse :

Pour $\gamma = 0$, une politique optimale optimise le retour pour une seule transition. Dans l'état 0, on gagne 5 en allant dans l'état 0, donc en effectuant l'action d. Dans l'état 1 on gagne 1 en effectuant l'action a, b, d ou e. Dans l'état 2, on ne perçoit que des retours nuls, donc toutes les actions sont optimales. Dans l'état 3, la meilleure action est c.

Question 3 : juste en réfléchissant, pouvez-vous déterminer une politique optimale pour $\gamma = 0,99$?

Réponse :

Pour γ non nul, une politique optimale consiste à parcourir le chemin le plus court entre l'état courant et l'état 3 afin d'obtenir le retour 10. De l'état 0, le mieux est d'aller en 2 par l'action d. Dans l'état 1, le mieux est d'aller dans l'état 2 par l'action c. Dans l'état 2, le mieux est d'aller dans l'état 3 par l'action c ou d. Dans l'état 3, le mieux est d'y rester par l'action c.

Question 4 : avec le/les programmes écrits en TP, calculer la ou les politiques optimales pour $\gamma = 0$ et $\gamma = 0,99$. Sont-ce les mêmes que celles que vous avez indiquées aux questions 2 et 3 ?

Réponse :

Pour $\gamma = 0$:

état	0	1	2	3
π^*	d	a, b, d ou e	a, b, c, d ou e	c

Pour $\gamma = 0,99$:

état	0	1	2	3
π^*	d	c	c ou d	c

Question 5 : quelles sont les valeurs des politiques optimales pour $\gamma = 0$ et $\gamma = 0,99$?

Réponse :

Pour $\gamma = 0$:

état	0	1	2	3
V^*	5	1	0	10

Pour $\gamma = 0,99$:

état	0	1	2	3
V^*	985,1	980,1	990	1 000

Question 6 : calculer la politique optimale pour « chaque » valeur de $\gamma \in [0, 1[$.

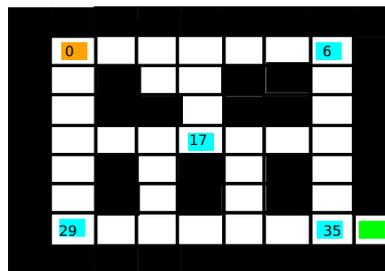
Réponse :

Exercice 2

On considère le problème consistant à sortir d'un labyrinthe. On prendra le labyrinthe ci-dessous. On suppose que l'on part de la case indiquée en orange et que l'on doit atteindre la case indiquée en vert ; à chaque itération, on peut se déplacer d'une case vers la gauche, ou vers la droite, ou vers le haut ou vers le bas. Il y a des couloirs et des murs : les couloirs sont indiqués en blanc, les murs en noir ; on ne peut pas traverser les murs, on doit toujours être localisé sur une case blanche. L'état est la case dans laquelle on se trouve ; cette numérotation est arbitraire mais pour simplifier les choses, on numérote les cases couloir de haut en bas, de gauche vers la droite. La case orange est donc la case 0. Dans la figure ci-dessous, j'ai indiqué le numéro de quelques états.

Partie 1

On suppose que les transitions sont déterministes : par exemple, si on décide d'aller à gauche, si la case à gauche n'est pas un mur (cellule noire), alors on se déplace bien d'une case sur la gauche. Les retours sont nuls sur toutes les transitions sauf quand on atteint la case verte : dans ce cas, le retour vaut 1.



Question 7 : combien y a-t-il d'états dans ce problème ?

Réponse :

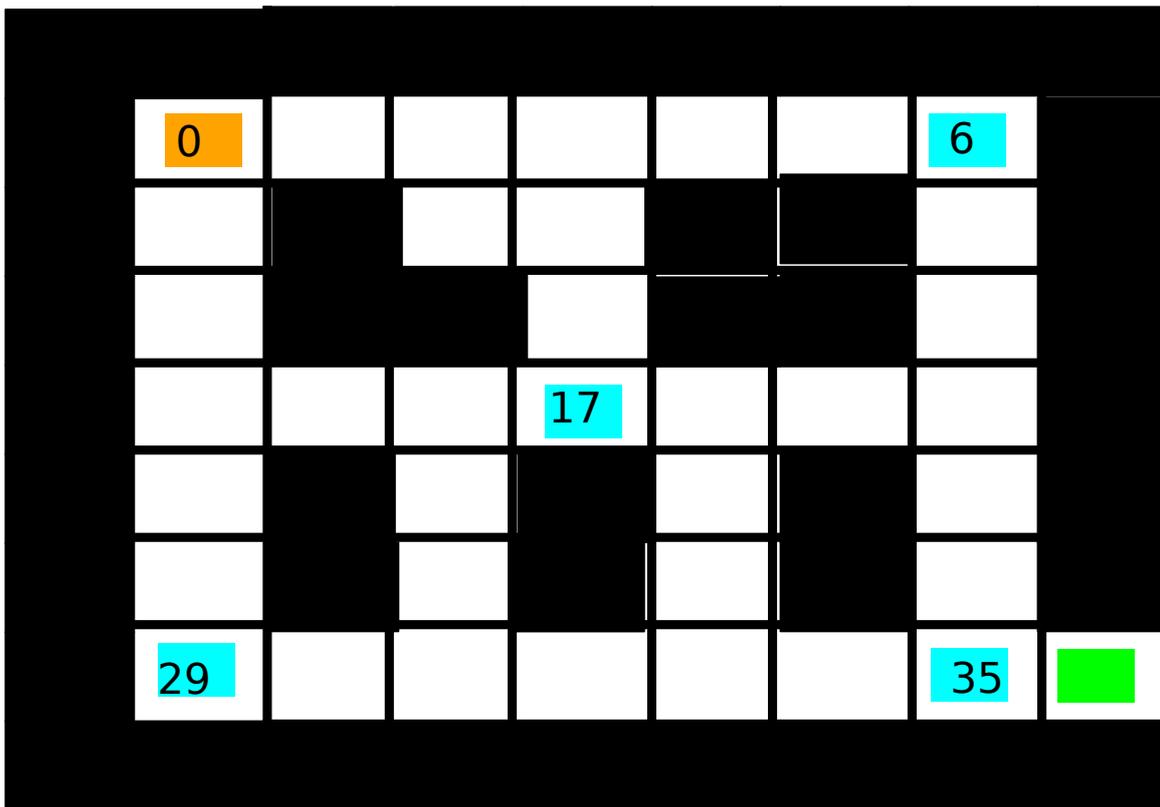
37 : un état par case du labyrinthe.

Question 8 : pour ce problème, pensez-vous qu'il y a une ou plusieurs politiques optimales ? Si vous pensez qu'il y en a plusieurs, combien y en a-t-il ?

Réponse :

Oui il y a plusieurs politiques optimales. La politique optimale consiste à aller de la case orange à la case verte par le chemin le plus court. Il y a plusieurs tels chemins. Par exemple dans l'état 0, on peut aller à droite ou vers le bas. De même dans les états 2, 3, 14, 16, 18. Dans les autres états, il n'y a qu'une seule action optimale. On compte 9 politiques optimales.

Dessinez cette, ou ces, politique(s) optimale(s) dans la figure ci-dessous. Faites en sorte que la figure soit lisible.



Question 9 : pour ce problème, pensez-vous que la réponse à la question 8 dépende de la valeur γ ? Répondez en argumentant.

Réponse :

Non. La valeur des états dépend de γ mais pas la/les politique(s) optimale(s). Ceci parce que la fonction de retour est nulle partout sauf sur la transition qui atteint la sortie du labyrinthe.

Je vous fournis la fonction de transition et la fonction de retour dans le fichier disponible à l'url <https://philippe-preux.github.io/ensg/m2-info/PDI/labyrinthe-P-et-R.txt>. Ce fichier est au format défini dans le TP de résolution de l'équation de Bellman.

Question 10 : quelle est la valeur d'une politique optimale pour les états indiqués en bleu : 0, 6, 17, 29 et 35? Indiquez ces valeurs dans la table ci-dessous.

Réponse :

Je prends $\gamma = 0,99$ et j'obtiens :

état	0	6	17	29	35
valeur	44,542	47,3106	47,3106	47,3106	50,2513

Partie 2

On suppose maintenant qu'un fort vent souffle depuis les murs à la droite du labyrinthe, rendant la progression horizontale aléatoire : parfois, on veut effectuer un pas vers la droite et une bourrasque nous laisse sur place ou nous fait reculer ; de même, quand on veut faire un pas vers la gauche, parfois le vent nous emporte plus loin que la case voisine.

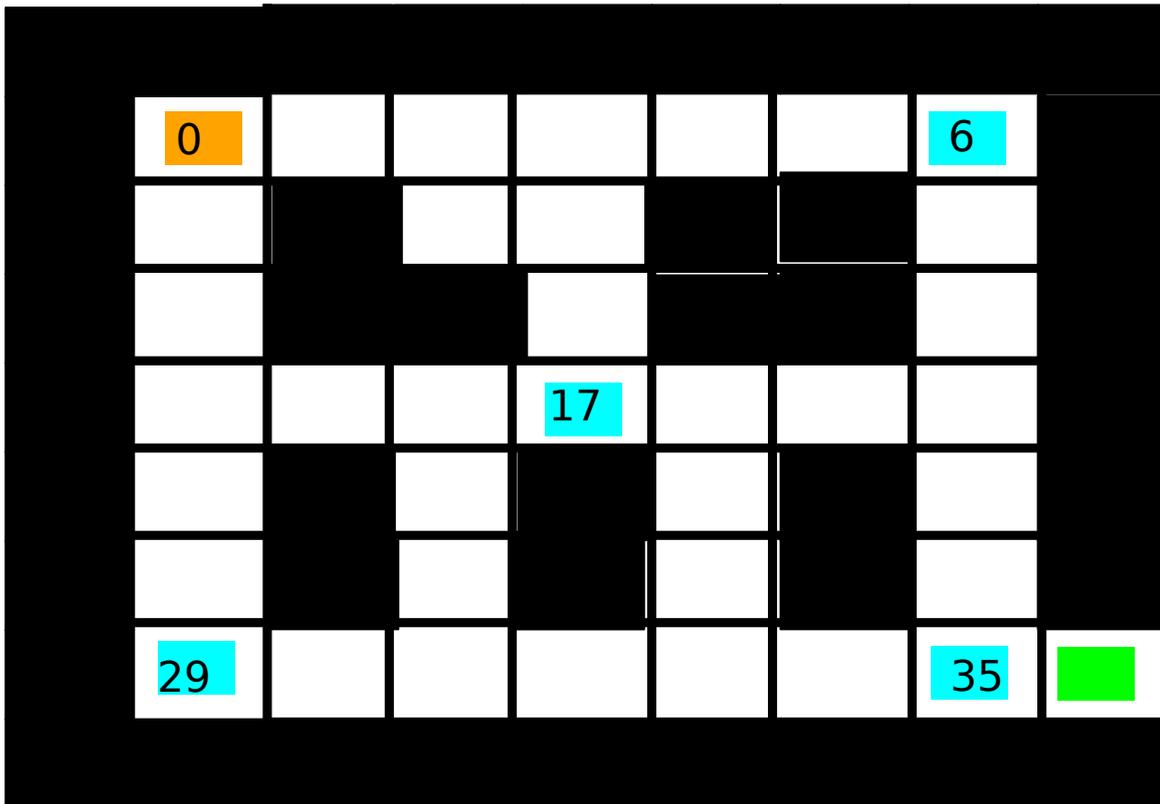
Question 11 : qu'est ce que cela implique dans la définition du problème de Markov? Rien, un changement de l'ensemble des états, des actions, de la fonction de transition, de la fonction de retour, de γ , ...? Justifier votre réponse.

Réponse :

Cela modifie la fonction de transition : alors que les transitions étaient déterministes dans la partie 1, elles sont maintenant stochastiques, dépendant de la force du vent qui varie de manière aléatoire.

Je vous fournis la fonction de transition et la fonction de retour dans le fichier disponible à l'url <https://philippe-preux.github.io/ensg/m2-info/PDI/labyrinthe-venteux-P-et-R.txt>. Ce fichier est au format défini dans le TP de résolution de l'équation de Bellman.

Question 12 : calculer la ou les politiques optimales et indiquez la ou les sur la figure ci-dessous. Je vous laisse choisir la valeur de γ . Quelle valeur avez-vous pris pour γ ?



Comme on l'a dit plus haut, la/les politiques optimales ne dépendent pas de γ pour ce problème.

Question 13 : pour ce labyrinthe venteux, quelle est la valeur d'une politique optimale pour les états indiqués en bleu : 0, 6, 17, 29 et 35? Indiquez ces valeurs dans la table ci-dessous.

Réponse :

La valeur de γ importe sur la valeur des états. On a pris $\gamma = 0,99$.

état	0	6	17	29	35
valeur	29,5244	32,1108	31,6827	31,2977	34,1067