Processus décisionnels markoviens

1 Le problème du parking

On veut se garer le plus près possible d'un cinéma afin d'y parvenir le plus rapidement possible (voir figure 1). Devant chaque allée perpenticulaire, on peut soit avancer tout droit, soit essayer de se garer dans la rangée. Si on essaye de se garer, on a 30% de chance de trouver un place libre dans la rangée. On a donc aussi 70% de chance de ne pas en trouver ce qui nous contraindra à nous rendre devant la rangée suivante. Si on arrive finalement au parking extérieur, on doit se garer (il y a toujours de la place).
La question est de savoir, lorsqu'on arrive devant une rangée, s'il faut essayer de se garer ou continuer d'avancer.

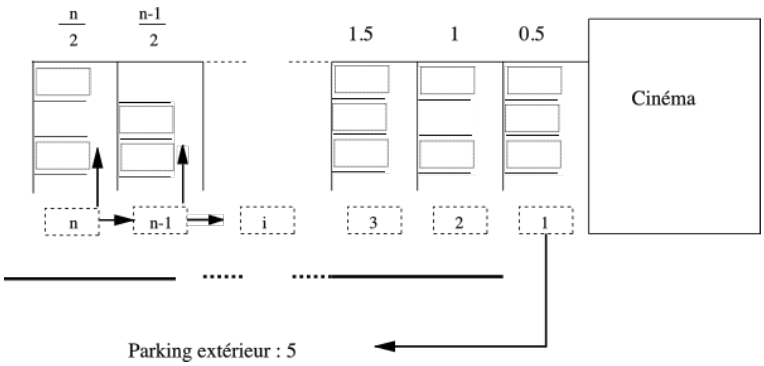


FIGURE 1 – Illustration du problème de parking. Les valeurs au dessus des rangées et pour le parking extérieur représentent les temps mis pour rejoindre le cinéma (en mn).

Question 1 : Définir les états dans lesquels l'agent peut se trouver, les actions possibles dans chaque état, les probabilités de transition entre les états et les coûts induits par les actions. Lorsque l'agent atteint un état terminal, il reste dans cet état avec une probabilité de 1 et un coût de transition nul.

POUR L'ENSEIGNANT SEUL (début)

Description formelle :

Etats : X = $\{1, ..., n, g_0, ..., g_n\}$ avec $1, ..., n$ les états où on doit prendre une décision. $g_0, ..., g_n$ représentent les états terminaux (une seule action est possible, celle qui nous renvoie dans le même état avec une probabilité de 1 et un coût de 0).

Actions possibles : pour tout état i, $A_i = \{\text{avancer, se garer}\}$

Probabilités de transition : $P_{i+1,i}(avancer) = 1 P_{i+1,i}(segarer) = 0.7 P_{i+1,g_{i+1}}(segarer) = 0.3$

Coûts induits : c(1,avancer)= 5, c(1,se garer)=0.3*(i+1)/2 + 0.7*5,c(i+1,avancer)=0, c(i+1,se garer) = 0.3*(i+1)/2 + 0.7*0 avec $1 < i \leq n$

POUR L'ENSEIGNANT SEUL (fin)

Question 2 : Représenter schématiquement le problème.

POUR L'ENSEIGNANT SEUL (début)

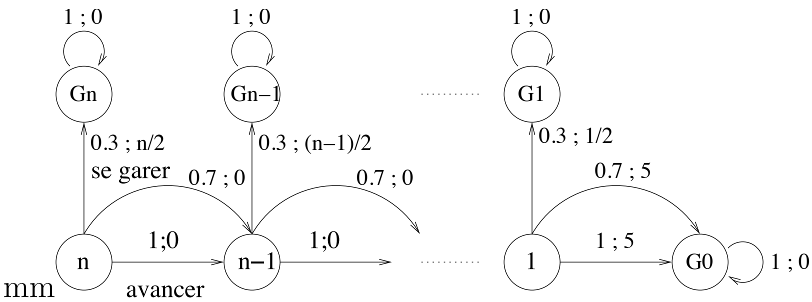


FIGURE 2 – Graphe de transition du parking

POUR L'ENSEIGNANT SEUL (fin)

Question 3 : Appliquer l'algorithme d'itérations sur les valeurs. Prendre $\gamma = 1$ (pour simplifier les calculs) et $n = 10$. On arrêtera l'itération quand $V_{t+1}(i) = V_t(i)$ pour tout état i.
Question 4 : Quelle est la politique optimale ?

POUR L'ENSEIGNANT SEUL (début)

t	s	V	
0		$V_0(s)$	$= 0$
1	1	$V_1(1)$	$= min(c(1, avancer) + \gamma P_{1,g_0}(avancer)V_0(g_0),$ $c(1, se\ garer) + \gamma P_{1,g_1}(se\ garer)V_0(g_1)) + \gamma P_{1,g_0}(se\ garer)V_0(g_0)$ $= min(5 + 1 * 1 * 0, (0.3 * 1/2 + 0.7 * 5) + 1 * 0.3 * 0 + 1 * 0.7 * 0)$ $= min(5, 3.65) = 3.65$
1	$2 \leq i \leq n$	$V_1(i)$	$= min(c(i, avancer) + \gamma P_{i,i-1}(avancer)V_0(i-1),$ $c(i, se\ garer) + \gamma P_{i,g_i}(se\ garer)V_0(g_i) + \gamma P_{i,i-1}(se\ garer)V_0(i-1))$ $= min(0 + 1 * 1 * 0, (0.3 * i/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 0)$ $= min(0, 0.3 * i/2) = 0$
1	g_i	$V_1(g_i)$	$=0 + 1.V_0(g_i) = 0$
2	1	$V_2(1)$	$= min(c(1, avancer) + \gamma P_{1,g_0}(avancer)V_1(g_0),$ $c(1, se\ garer) + \gamma P_{1,g_1}(se\ garer)V_1(g_1)) + \gamma P_{1,g_0}(se\ garer)V_1(g_0)$ $= min(5 + 1 * 1 * 0, (0.3 * 1/2 + 0.7 * 5) + 1 * 0.3 * 0 + 1 * 0.7 * 0)$ $= min(5, 3.65) = 3.65$
2	2	$V_2(2)$	$= min(c(2, avancer) + \gamma P_{2,1}(avancer)V_1(1),$ $c(2, se\ garer) + \gamma P_{2,g_2}(se\ garer)V_1(g_2)) + \gamma P_{2,1}(se\ garer)V_1(1)$ $= min(0 + 1 * 1 * 3, 65, (0.3 * 2/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 3.65)$ $= min(3.65, 2.855) = 2.855$
2	$3 \leq i \leq n$	$V_2(i)$	$= min(c(i, avancer) + \gamma P_{i,i-1}(avancer)V_1(i-1),$ $c(i, se\ garer) + \gamma P_{i,g_i}(se\ garer)V_1(g_i) + \gamma P_{i,i-1}(se\ garer)V_1(i-1))$ $= min(0 + 1 * 1 * 0, (0.3 * i/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 0)$ $= min(0, 0.3 * i/2) = 0$
2	g_i	$V_2(g_i)$	$=0 + 1.V_1(g_i) = 0$
3	1	$V_3(1)$	$= min(c(1, avancer) + \gamma P_{1,g_0}(avancer)V_2(g_0),$ $c(1, se\ garer) + \gamma P_{1,g_1}(se\ garer)V_2(g_1)) + \gamma P_{1,g_0}(se\ garer)V_2(g_0)$ $= min(5 + 1 * 1 * 0, (0.3 * 1/2 + 0.7 * 5) + 1 * 0.3 * 0 + 1 * 0.7 * 0)$ $= min(5, 3.65) = 3.65$
3	2	$V_3(2)$	$= min(c(2, avancer) + \gamma P_{2,1}(avancer)V_2(1),$ $c(2, se\ garer) + \gamma P_{2,g_2}(se\ garer)V_2(g_2)) + \gamma P_{2,1}(se\ garer)V_2(1)$ $= min(0 + 1 * 1 * 3, 65, (0.3 * 2/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 3.65)$ $= min(3.65, 2.855) = 2.855$

t	s	V	
3	3	$V_3(3)$	$= min(c(3, avancer) + \gamma P_{3,2}(avancer)V_2(2),$ $c(3, se\ garer) + \gamma P_{3,g_3}(se\ garer)V_2(g_3)) + \gamma P_{3,2}(se\ garer)V_2(2)$ $= min(0 + 1 * 1 * 2.855, (0.3 * 3/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 2.855)$ $= min(2.855, 2.4485) = 2.4485$
3	$4 \leq i \leq n$	$V_3(i)$	$= min(c(i, avancer) + \gamma P_{i,i-1}(avancer)V_2(i-1),$ $c(i, se\ garer) + \gamma P_{i,g_i}(se\ garer)V_2(g_i) + \gamma P_{i,i-1}(se\ garer)V_2(i-1))$ $= min(0 + 1 * 1 * 0, (0.3 * i/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 0)$ $= min(0, 0.3 * i/2) = 0$
3	g_i	$V_3(g_i)$	$=0 + 1.V_2(g_i) = 0$
4	4	$V_4(4)$	$= min(c(4, avancer) + \gamma P_{4,3}(avancer)V_3(3),$ $c(4, se\ garer) + \gamma P_{4,g_4}(se\ garer)V_3(g_4)) + \gamma P_{4,3}(se\ garer)V_3(3)$ $= min(0 + 1 * 1 * 2.4485, (0.3 * 4/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 2.4485)$ $= min(2.4485, 2.31395) = 2.31395$
4	$5 \leq i \leq n$	$V_4(i)$	$= min(c(i, avancer) + \gamma P_{i,i-1}(avancer)V_3(i-1),$ $c(i, se\ garer) + \gamma P_{i,g_i}(se\ garer)V_3(g_i) + \gamma P_{i,i-1}(se\ garer)V_3(i-1))$ $= min(0 + 1 * 1 * 0, (0.3 * i/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 0)$ $= min(0, 0.3 * i/2) = 0$
5	5	$V_5(5)$	$= min(c(5, avancer) + \gamma P_{5,4}(avancer)V_4(4),$ $c(5, se\ garer) + \gamma P_{5,g_5}(se\ garer)V_4(g_5)) + \gamma P_{5,4}(se\ garer)V_4(4)$ $= min(0 + 1 * 1 * 2.31395, (0.3 * 5/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 2.31395)$ $= min(2.31395, 2.369765) = 2.31395$
5	$6 \leq i \leq n$	$V_5(i)$	$= min(c(i, avancer) + \gamma P_{i,i-1}(avancer)V_4(i-1),$ $c(i, se\ garer) + \gamma P_{i,g_i}(se\ garer)V_4(g_i) + \gamma P_{i,i-1}(se\ garer)V_4(i-1))$ $= min(0 + 1 * 1 * 0, (0.3 * i/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 0)$ $= min(0, 0.3 * i/2) = 0$
6	6	$V_6(6)$	$= min(c(6, avancer) + \gamma P_{6,5}(avancer)V_5(5),$ $c(6, se\ garer) + \gamma P_{6,g_6}(se\ garer)V_5(g_6)) + \gamma P_{6,5}(se\ garer)V_5(5)$ $= min(0 + 1 * 1 * 2.31395, (0.3 * 6/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 2.31395)$ $= min(2.31395, 2.519765) = 2.31395$
7	7	$V_7(7)$	$= min(c(7, avancer) + \gamma P_{7,6}(avancer)V_6(6),$ $c(7, se\ garer) + \gamma P_{7,g_7}(se\ garer)V_6(g_7)) + \gamma P_{7,6}(se\ garer)V_6(6)$ $= min(0 + 1 * 1 * 2.31395, (0.3 * 7/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 2.31395)$ $= min(2.31395, 2.669765) = 2.31395$
8	8	$V_8(8)$	$= min(c(8, avancer) + \gamma P_{8,7}(avancer)V_7(7),$ $c(8, se\ garer) + \gamma P_{8,g_8}(se\ garer)V_7(g_8)) + \gamma P_{8,7}(se\ garer)V_7(7)$ $= min(0 + 1 * 1 * 2.31395, (0.3 * 8/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 2.31395)$ $= min(2.31395, 2.819765) = 2.31395$
9	9	$V_9(9)$	$= min(c(9, avancer) + \gamma P_{9,8}(avancer)V_8(8),$ $c(9, se\ garer) + \gamma P_{9,g_9}(se\ garer)V_8(g_9)) + \gamma P_{9,8}(se\ garer)V_8(8)$ $= min(0 + 1 * 1 * 2.31395, (0.3 * 9/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 2.31395)$ $= min(2.31395, 2.969765) = 2.31395$
10	10	$V_{10}(10)$	$= min(c(10, avancer) + \gamma P_{10,9}(avancer)V_9(9),$ $c(10, se\ garer) + \gamma P_{10,g_{10}}(se\ garer)V_9(g_{10})) + \gamma P_{10,9}(se\ garer)V_9(9)$ $= min(0 + 1 * 1 * 2.31395, (0.3 * 9/2 + 0.7 * 0) + 1 * 0.3 * 0 + 1 * 0.7 * 2.31395)$ $= min(2.31395, 2.969765) = 2.31395$

En conclusion, il faut essayer de se garer lorsque l'on est sur les emplacements 1, 2, 3 ou 4 et avancer sinon.
$\pi^*(1) = \pi^*(2) = \pi^*(3) = \pi^*(4) =$se garer
$\pi^*(5) = \pi^*(6) = \pi^*(7) = \pi^*(8) = \pi^*(9) = \pi^*(10) =$avancer

POUR L'ENSEIGNANT SEUL (fin)

2 Le problème du chauffeur de taxi

Un chauffeur de taxi travaille sur 3 quartiers A, B et C.
À chaque instant, il doit décider ce qu'il fait. Il peut :
- a1 : rouler en espérant être hélé par un client
- a2 : s'arrêter à une station de taxis dans le quartier où il se trouve et attendre un client
- a3 : stationner son taxi et attendre un appel.

Dans les quartiers A et C, les 3 actions sont possibles ; dans le quartier B, l'action a2 est impossible car il il n'y a pas de station de taxis dans B. Les probabilités de transition et les récompenses en Euro sont fournis par le tableau 1.

Table 1 – Probabilités de transition (à gauche) et Récompenses (à droite)

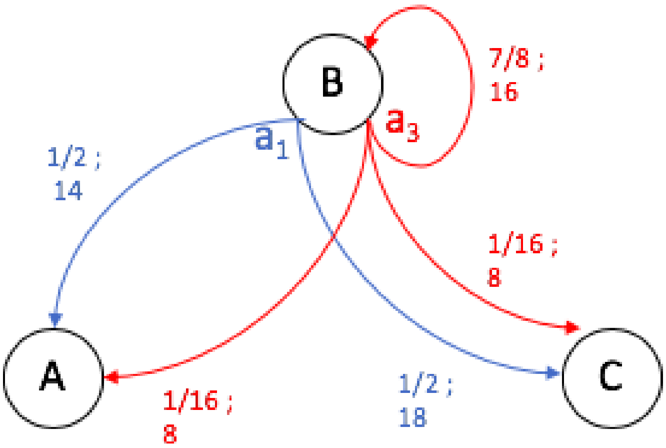
P			quartier suivant		
		action	A	B	C
quartier actuel	A	a_1	1/2	1/4	1/4
		a_2	1/16	3/4	3/16
		a_3	1/4	1/8	5/8
	B	a_1	1/2	0	1/2
		a_3	1/16	7/8	1/16
	C	a_1	1/4	1/4	1/2
		a_2	1/8	3/4	1/8
		a_3	3/4	1/16	3/16

R			quartier suivant		
		action	A	B	C
quartier actuel	A	a_1	10	4	8
		a_2	8	2	4
		a_3	4	6	4
	B	a_1	14	0	18
		a_3	8	16	8
	C	a_1	10	2	8
		a_2	6	4	2
		a_3	4	0	8

Question 1 : Quels sont les différents états du problème ?
Question 2 : Reproduire le graphe de transitions partant uniquement de l'état B résumant la dynamique.
Question 3 : Calculer la valeur des états $V^\pi(.)$ pour une politique aléatoire uniforme (c'est-à-dire que les actions applicables dans un état donné ont la même probabilité) avec $\gamma = 0.9$. On rappelle qu'il s'agit de résoudre un système linéaire d'équations de Bellman $V^\pi(x) = \sum_{a \in A(x)} \pi(x,a) \{ \sum_{x' \in X} P(x,a,x')[R(x,a,x') + \gamma V^\pi(x')] \}$.

POUR L'ENSEIGNANT SEUL (début)

1) Les états du problème sont les 3 quartiers dans lesquels peut se trouver le taxi.



2)

3) Pour simplifier l'écriture,

On a donc :

$$\begin{cases} V^\pi(A) = \Sigma_{a\in\{a_1,a_2,a_3\}}\pi(A,a)\{\Sigma_{x'\in\mathcal{X}}P(A,a,x')[R(A,a,x')+\gamma V^\pi(x')]\} \\ V^\pi(B) = \Sigma_{a\in\{a_1,a_3\}}\pi(B,a)\{\Sigma_{x'\in\mathcal{X}}P(B,a,x')[R(B,a,x')+\gamma V^\pi(x')]\} \\ V^\pi(C) = \Sigma_{a\in\{a_1,a_2,a_3\}}\pi(C,a)\{\Sigma_{x'\in\mathcal{X}}P(C,a,x')[R(C,a,x')+\gamma V^\pi(x')]\} \end{cases}$$

$$\begin{aligned} V^\pi(A) = \pi(A,a_1)\{ \quad & P(A,a_1,A)R(A,a_1,A)+ \\ & P(A,a_1,B)R(A,a_1,B)+ \\ & P(A,a_1,C)R(A,a_1,C)\} \\ +\pi(A,a_2)\{ \quad & P(A,a_2,A)R(A,a_2,A)+ \\ & P(A,a_2,B)R(A,a_2,B)+ \\ & P(A,a_2,C)R(A,a_2,C)\} \\ +\pi(A,a_3)\{ \quad & P(A,a_3,A)R(A,a_3,A)+ \\ & P(A,a_3,B)R(A,a_3,B)+ \\ & P(A,a_3,C)R(A,a_3,C)\} \\ +\gamma V^\pi(A)\{ \quad & \pi(A,a_1)P(A,a_1,A)+\pi(A,a_2)P(A,a_2,A)+\pi(A,a_3)P(A,a_3,A)\}+ \\ +\gamma V^\pi(B)\{ \quad & \pi(A,a_1)P(A,a_1,B)+\pi(A,a_2)P(A,a_2,B)+\pi(A,a_3)P(A,a_3,B)\}+ \\ +\gamma V^\pi(C)\{ \quad & \pi(A,a_1)P(A,a_1,C)+\pi(A,a_2)P(A,a_2,C)+\pi(A,a_3)P(A,a_3,C)\} \end{aligned}$$

Avec les valeurs numériques

$$
\begin{aligned}
V^\pi(A) = 1/3\{ \;& 1/2 \times 10 + \\
& 1/4 \times 4 + \\
& 1/4 \times 8\} \\
+1/3\{ \;& 1/16 \times 8 + \\
& 3/4 \times 2 + \\
& 3/16 \times 4\} \\
+1/3\{ \;& 1/4 \times 4 + \\
& 1/8 \times 6 + \\
& 5/8 \times 4\} \\
+0.9V^\pi(A)\{ \;& 1/3 \times 1/2 + 1/3 \times 1/16 + 1/3 \times 1/4\} + \\
+0.9V^\pi(B)\{ \;& 1/3 \times 1/4 + 1/3 \times 3/4 + 1/3 \times 1/8\} + \\
+0.9V^\pi(C)\{ \;& 1/3 \times 1/4 + 1/3 \times 3/16 + 1/3 \times 5/8\}
\end{aligned}
$$

Finalement

$$
\left\{
\begin{array}{lcl}
V^\pi(A) & = & 5 + 0.9(13/48V^\pi(A) + 3/8V^\pi(B) + 17/48V^\pi(C)) \\
V^\pi(B) & = & 31/2 + 0.9(9/32V^\pi(A) + 7/16V^\pi(B) + 9/32V^\pi(C)) \\
V^\pi(C) & = & 31/6 + 0.9(3/8V^\pi(A) + 17/48V^\pi(B) + 13/48V^\pi(C))
\end{array}
\right.
$$

$$
\left\{
\begin{array}{lcl}
V^\pi(A) & = & 156420/1789 = 87,43 \\
V^\pi(B) & = & 5113540/51881 = 98,6 \\
V^\pi(C) & = & 13602460/155643 = 87,40
\end{array}
\right.
$$

Question 4 : Conclure.

POUR L'ENSEIGNANT SEUL (début)

En conclusion, l'état le plus favorable si l'on suit une politique aléatoire uniforme est l'état B, le moins favorable est d'étre dans l'état C.

POUR L'ENSEIGNANT SEUL (fin)