

INTRODUCTION A LA PROGRAMMATION DYNAMIQUE STOCHASTIQUE

EXEMPLES

T. AL ANI
Laboratoire A²SI-ESIEE-Paris
e-mail: t.alani@esiee.fr

1. Supposons que vous avez 3 € pour faire une séquence de **quatre jeux**. A chaque jeu vous pouvez parier 0 €, 1 €, etc, avec des incréments de 1 € jusqu'à la totalité de la somme que vous avez. Quand vous pariez, **vous récupérez deux fois la somme que vous pariez avec la probabilité 0.6** et vous **perdez le pari de ce somme avec la probabilité 0.4**. Écrire une relation de récurrence qui peut être utilisée pour déterminer la stratégie de pari qui maximise la probabilité de **terminer le jeu avec au moins 5 €**. Utiliser cette relation pour **calculer la stratégie optimale et la probabilité maximale**.

Réponse : Soit $S(i,x)$ = la **probabilité maximale** de terminer avec au moins 5 € quand vous commencez parier par x_i (0, 1, 2,..., x €) partant d'un capital de x € à l'étape i (i=1, 2, 3, 4)

$$S(i, x) = \max_{x_i=0,1,\dots,x} [0.6S(i+1, 2x_i + (x-x_i)) + 0.4S(i+1, x-x_i)]$$

$$S(5, x) = 1 \text{ pour } x \geq 5, \quad S(5, x) = 0 \text{ autrement}$$

$$S(i, x) = 1 \text{ pour } x \geq 5 \quad \forall i.$$

23/01/2006

2

Soit $d(i,x)$ la décision optimale à l'étape i et en ayant un capital x.

i = 5 :

x=6 : $S(5,6) = 1$, x=5 : $S(5,5) = 1$, x=4 : $S(5,4) = 0$, x=3 : $S(5,3) = 0$, x=2 : $S(5,2) = 0$, x=1 : $S(5,1) = 0$

x=0 : $S(5,0) = 0$

i = 4 :

x = 5 : $S(4,5) = 1$ aucune décision à prendre (j'ai déjà gagné) !

x = 4 : $S(4,4) =$

$\max\{$

$x_4 = 0 : 0.6 S(5, 0+(4-0)) + 0.4S(5, 4) = S(5, 4) = 0$

$x_4 = 1 : 0.6 S(5, 2+(4-1)) + 0.4S(5, 3) = 0.6 S(5, 5) + 0.4S(5, 3) = 0.6+0 = 0.6$

$d(4,4) = 1$ ou

$x_4 = 2 : 0.6 S(5, 4+(4-2)) + 0.4S(5, 2) = 0.6 S(5, 6) + 0.4S(5, 2) = 0.6+0 = 0.6$

2 ou

$x_4 = 3 : 0.6 S(5, 6+(4-3)) + 0.4S(5, 1) = 0.6 S(5, 7) + 0.4S(5, 1) = 0.6+0 = 0.6$

3 ou

$x_4 = 4 : 0.6 S(5, 8+(4-4)) + 0.4S(5, 0) = 0.6 S(5, 8) + 0.4S(5, 0) = 0.6+0 = 0.6$

4;

$\}$

x=3 : $S(4,3) =$

$\max(0, 0, 0.6, 0.6) = 0.6,$

$d(4,3) = 2$ ou

3;

x=2 : $S(4,2) = 0,$

x=1 : $S(4,1) = 0.$


23/01/2006

3

$i = 4 :$
 $x=4 : S(4,4) = \max(0, 0.6, 0.6, 0.6, 0.6) = 0.6,$ $d(4,4) = 1, 2, 3, \text{ ou } 4;$
 $x=3 : S(4,3) = \max(0, 0, 0.6, 0.6) = 0.6,$ $d(4,3) = 2 \text{ ou } 3;$
 $x=2 : S(4,2) = 0,$
 $x=1 : S(4,1) = 0.$

$i = 3 :$
 $x=4 : S(3,4) = \max(0.6, 0.84, 0.6, 0.6, 0.6) = 0.84,$ $d(3,4) = 1;$
 $x=3 : S(3,3) = \max(0.6, 0.36, 0.6, 0.6) = 0.6,$ $d(3,3) = 0, 2 \text{ ou } 3;$
 $x=2 : S(3,2) = 0.36,$ $d(3,2) = 1 \text{ ou } 2;$
 $x=1 : S(3,1) = 0.$

$i = 2 :$
 $x=4 : S(2,4) = \max(0, 0.84, 0.84, 0.744, 0.6, 0.6) = 0.84,$ $d(2,4) = 0 \text{ ou } 1;$
 $x=3 : S(2,3) = \max(0.6, 0.648, 0.6, 0.6) = 0.648,$ $d(2,3) = 1;$
 $x=2 : S(2,2) = \max(0.36, 0.36, 0.504) = 0.504,$ $d(2,2) = 2;$
 $x=1 : S(2,1) = 0.216,$ $d(2,1) = 1;$

$i = 1 :$
 $S(1,3) = \max(0.648, 0.7056, 0.6864, 0.6) = 0.7056,$ $d(1,3) = 1.$ 

23/01/2006

4

L'exemple 2 suivant inclus un jeu compétitif, pour illustrer l'utilisation de la programmation dynamique dans les situations séquentielles où les joueurs alternent leurs mouvements.

23/01/2006

5

2. 2.1 **Cas déterministe** : N allumettes sont placées sur une table. Cédric se déplace d'abord et prend soit 1, 2, ou 3 allumettes. Ensuite Jacques prend soit 1, 2, ou 3 allumettes, ensuite Cédric prend soit 1, 2, ou 3 des allumettes restantes, ensuite Jacques, etc. **L'objectif est de forcer l'adversaire à prendre la dernière allumette.** Donner une formulation de programmation dynamique de ce problème, en supposant que les deux joueurs ne font jamais d'erreur.

Aide : Soit S_i la valeur du jeu pour le joueur dont le tirage commence quand i allumettes restent.

$S_1 = 0$ perdre.

23/01/2006

6

Réponse : Supposons que la valeur du jeu à Jacques est 1-(la valeur du jeu à Cédric). Ainsi, pour $i=4, 5, 6, \dots$

$$S_i = \max_{x_i=1,2,3} (1 - S_{i-x_i})$$

et pour $i=1, 2, 3$

$S_1=0$; $S_2=1$, $d_2=1$; $S_3=1$, $d_3=2$ où d_i est le tirage optimal pour le joueur qui vient de tirer une allumette. Il n'y a pas de tirage optimal si tous les tirages sont perdants. La relation ci-dessus signifie qu'un joueur gagne à moins que les trois S 's qui précèdent immédiatement soient toutes égales à 1.

23/01/2006

7

En appliquant la relation de récurrence, nous obtenons :

$$S_4=1, d_4=3;$$

$$S_5=0, S_6=1, d_6=1, S_7=1, d_7=2; S_8=1, d_8=3;$$

$$S_9=0, S_{10}=1, d_{10}=1, S_{11}=1, d_{11}=2; S_{12}=1, d_{12}=3; \text{ etc.}$$

Le joueur qui doit tirer essaie de jouer tel que le nombre d'allumettes restantes après le tirage est de la forme $4i+1$ pour un entier $i \geq 0$. Si la position initiale est de cette forme, le joueur qui tire le premier perd à moins que son adversaire ne joue pas d'une façon optimale (qui est le tirage de $(4-x)$ allumettes quand le joueur courant tire x allumettes).



2.2 **Cas stochastique** : Le même jeu précédent sauf qu'après chaque tirage d'un joueur, s'il existe encore des allumettes, une pièce de monnaie non biaisé est lancée et si le résultat est pile le joueur qui viens de jouer doit tirer une allumette de plus.

Réponse : Soit S_i = la valeur maximale espérée du jeu pour le joueur qui se prête à jouer quand il y a i allumettes restantes. Un gain a une valeur 1 et une perte a une valeur 0, ainsi la valeur espérée est égale à la probabilité de gagner.

$$S_i = \max_{x_i=1,2,3} \left[\frac{1}{2}(1-S_{i-x_i-1}) + \frac{1}{2}(1-S_{i-x_i}) \right], \quad i \geq 2,$$

et $S_{-2}=1, S_{-1}=1, S_0=1, S_1=0$.

23/01/2006

10

Ainsi

$$S_2 = \max (1/2, 0, 0) = 1/2, \quad d_2=1;$$

$$S_3 = \max (3/4, 1/2, 0) = 3/4, \quad d_3=1;$$

$$S_4 = \max (3/8, 3/4, 1/2) = 3/4, \quad d_4=2;$$

$$S_5 = \max (1/4, 3/8, 3/4) = 3/4, \quad d_5=3;$$

$$S_6 = \max (1/4, 1/4, 3/8) = 3/8, \quad d_6=3;$$

$$S_7 = \max (7/16, 1/4, 1/4) = 7/16, \quad d_7=1;$$

etc.

Le calcul peut continuer plus loin et la politique optimale peut être mémorisée.



23/01/2006

11

3. Pour un prix de 1 € / gallon (*), la chaîne de supermarchés SUPERM a acheté 6 gallons de lait d'une laiterie locale. Chaque gallon est vendu dans les trois magasins de la chaîne pour 2 € / gallon. La laiterie doit acheter pour 50 centimes / gallon n'importe quelle quantité de lait laissée à la fin du jour. La demande à chacun des trois magasins de la chaîne est incertaine. L'historique des données indique que la demande quotidienne à chaque magasin est comme indiquée dans le tableau suivant

(*) 1 gallon des USA = 3,7854118 gallons

1 gallon (BRITANNIQUE) impérial = approximativement 1,2 gallons des USA

23/01/2006

12

Magasin (m)	demande quotidienne (gallons)	Probabilité
Magasin 1	1	0.60
	2	0.00
	3	0.40
Magasin 2	1	0.50
	2	0.10
	3	0.40
Magasin 3	1	0.40
	2	0.30
	3	0.30

23/01/2006

13

SUPERM veut assigner les 6 gallons de lait aux trois magasins afin de maximiser le bénéfice net quotidien espéré (bénéfices moins coûts) gagné du lait.

Notez que puisque les coûts d'achat quotidiens de SUPERM sont toujours 6 €, nous pouvons concentrer notre attention sur le problème d'assigner le lait pour maximiser le bénéfice quotidien espéré gagné des 6 gallons.

Soit $r_m(g_m)$ = le bénéfice espéré gagné des g_m gallons assignés au magasin m

$f_m(x)$ = le bénéfice maximum espéré gagné de x gallons assignés aux magasins $m, m + 1, \dots, 3$

Puisque $f_3(x)$ doit être par définition le bénéfice gagné estimé d'assigner x gallons de lait au magasin 3, alors $f_3(x) = r_3(x)$. Pour $m = 1, 2$, nous pouvons écrire

$$f_m(x) = \max_{g_m} \{r_m(g_m) + f_{m+1}(x - g_m)\} \quad (1)$$

où g_m doit être un membre de $\{0, 1, 2, \dots, x\}$.

L'équation ci-dessus suit parce que pour n'importe quel choix de g_m (le nombre de gallons assignés à magasin m), le bénéfice estimé gagné de magasins $m, m + 1, \dots, 3$ sera la somme du bénéfice estimé gagné du magasin m si g_m gallons sont assignés au magasin m plus le bénéfice maximum estimé qui peut être gagné des magasins $m + 1, m + 2, \dots, 3$ quand $g_m - x$ gallons sont assignés à ces magasins.

Pour calculer l'attribution optimale du lait aux magasins, nous commençons par le calcul $f_3(0)$, $f_3(1)$, ..., $f_3(6)$. Alors nous employons l'équation ci-dessus pour calculer $f_2(0)$, $f_2(1)$, ..., $f_2(6)$. En conclusion, nous déterminons $f_1(6)$.

Nous commençons en calculant les $r_m(g_m)$. Notez que nous n'avons pas besoin d'assigner plus de 3 gallons à aucun magasin. Pour cette raison, nous calculons les $r_m(g_m)$ seuls pour $x = 0, 1, 2$, ou 3 .

Par exemple, nous calculons $r_3(2)$, le bénéfice estimé gagné si 2 gallons sont assignés au magasin 3.

- Si la demande au magasin 3 est pour 2 ou plus de gallons, les deux gallons assignés au magasin 3 seront vendus et 4 € de bénéfice seront gagnés.
- Si la demande au magasin 3 est 1 gallon, 1 gallon sera vendu à 2 € et un gallon sera retourné pour 50 centimes.

Ainsi, Si la demande au magasin 3 est 1 gallon, un bénéfice 2,50 € sera gagné. Puisqu'il y a une chance de 0,6 que la demande au magasin 3 soit supérieure ou égale à 2 gallons et une chance de 0,4 que la demande au magasin 3 soit égale à 1 gallon, alors

$$r_3(2) = (0,30+0,30)(4,00 \text{ €}) + 0,40(2,50 \text{ €}) = 3,40 \text{ €}$$

De la même façon,

$$r_3(0) = 0,00 \text{ €}, \quad r_2(0) = 0,00 \text{ €}, \quad r_1(0) = 0,00 \text{ €},$$

$$r_3(1) = 2,00 \text{ €}, \quad r_2(0) = 2,00 \text{ €}, \quad r_1(0) = 2,00 \text{ €},$$

$$r_3(2) = 3,40 \text{ €}, \quad r_2(0) = 3,25 \text{ €}, \quad r_1(0) = 3,10 \text{ €},$$

$$r_3(3) = 4,35 \text{ €}, \quad r_2(0) = 4,35 \text{ €}, \quad r_1(0) = 4,20 \text{ €},$$

Soit $g_m(x)$ une attribution de lait au magasin m qui atteint le $f_m(x)$, alors

$$f_3(0) = r_3(0) = 0,00 \quad g_3(0) = 0$$

$$f_3(1) = r_3(1) = 2,00, \quad g_3(1) = 1$$

$$f_3(2) = r_3(2) = 3,40, \quad g_3(2) = 2$$

$$f_3(3) = r_3(3) = 4,35, \quad g_3(3) = 3$$

Nous n'avons pas besoin de calcul $f_3(4)$, $f_3(5)$ et $f_3(6)$ parce qu'une attribution optimale n'aura jamais plus de 3 gallons à assigner à un seul magasin (la demande à n'importe quel magasin n'est jamais plus de 3 gallons).

23/01/2006

22

En utilisant l'équation de calcul en arrière
« Backward » (1),

$$f_2(0) = r_2(0) + f_3(0-0) = 0,00 \quad g_2(0) = 0$$

$$f_2(1) = \max\{r_2(0) + f_3(1-0); r_2(1) + f_3(1-1)\} \\ = \max\{2,00; 2,00\} = 2,00 \quad g_2(1) = 0 \text{ ou } 1$$

$$f_2(2) = \max\{r_2(0) + f_3(2-0); r_2(1) + f_3(2-1); r_2(2) + f_3(2-2)\} \\ = \max\{3,40; 4,00; 3,25\} = 4,00 \quad g_2(2) = 1$$

$$f_2(3) = \max\{r_2(0) + f_3(3-0); r_2(1) + f_3(3-1); r_2(2) + f_3(3-2); r_2(3) + f_3(3-3)\} \\ = \max\{4,35; 5,40; 5,25; 4,35\} = 5,40 \quad g_2(3) = 1$$

Notez qu'en calculant $f_2(4)$, $f_2(5)$ et $f_2(6)$, nous n'avons besoin de considérer aucune attribution pour plus que 3 gallons au magasin 2 ou aucune contribution qui laisse plus de 3 gallons pour le magasin 3.

23/01/2006

23

$$f_2(4) = \max\{r_2(1)+f_3(4-1); r_2(2)+f_3(4-2); r_2(3)+f_3(4-3)\}$$

$$= \max\{6,35; 6,65; 6,35\} = 6,65 \quad g_2(4) = 2$$

$$f_2(5) = \max\{r_2(2)+f_3(5-2); r_2(3)+f_3(5-3)\}$$

$$= \max\{7,60; 7,75\} = 7,75 \quad g_2(5) = 3$$

$$f_2(6) = r_2(3)+f_3(6-3) = 8,70 \quad g_2(6) = 3$$

Enfin

$$f_1(6) = \max\{r_1(0)+f_2(6-0); r_1(1)+f_2(6-1); r_1(2)+f_2(6-2); r_1(3)+f_2(6-3)\}$$

$$= \max\{8,70; 9,75; 9,75; 9,60\} = 9,75 \quad g_1(6) = 1 \text{ ou } 2$$

Ainsi, nous pouvons assigner 1 ou 2 gallons au magasin 1. Supposons que nous choisissons arbitrairement d'assigner 1 gallon au magasin 1. Alors nous avons le 6-1 = 5 gallons pour les magasins 2 et 3. Puisque $f_2(5)$ est atteint par $g_2(5) = 3$, nous assignons 3 gallons au magasin 2. Puis, le 5-3 = 2 gallons sont disponibles pour le magasin 3. Puisque $g_3(2) = 2$, nous assignons 2 gallons au magasin 3.

23/01/2006

24

Bien que cette politique obtienne le bénéfice maximum estimé, $f_1(6) = 9,75$ €; tout le bénéfice réellement reçu un jour donné peut être plus ou moins de 9,75 €.

Par exemple, si la demande à chaque magasin étaient de 1 gallon, le bénéfice total serait $3 \times (2,00 \text{ €}) + 3 \times (0,50 \text{ €}) = 7,50$ €, tandis que si la demande à chaque magasin étaient de 3 gallons, tout le lait serait vendu à 2 € /gallon et le bénéfice total estimé serait $6 \times (2,00 \text{ €}) = 12,00$ €.

23/01/2006

25