# Algorithme d'évaluation de la politique : le problème du chauffeur de taxi ; résolution du système linéaire associé

## Master Recherche Informatique

Philippe Preux

19 octobre 2007

On veut calculer la valeur de chacun des états pour une politique aléatoire uniforme et $\gamma = 0{,}9$.

Pour cela, on utilise l'équation de Bellman que l'on rappelle ci-dessous :

$$V^\pi(x) = \Sigma_{a \in \mathcal{A}(x)} \pi(x,a) \Big\{ \Sigma_{x' \in \mathcal{X}} P(x,a,x')[R(x,a,x') + \gamma V^\pi(x')] \Big\} \tag{1}$$

On cherche donc $V^\pi(A)$, $V^\pi(B)$ et $V^\pi(C)$.

$\pi$ est définie par :

- $\pi(A,a_1) = \pi(A,a_2) = \pi(A,a_3) = 1/3$ puisque les trois actions sont possibles dans l'état A et que l'on suppose que la politique $\pi$ est aléatoire uniforme ;
- pour l'état C, c'est la même chose pour les mêmes raisons : $\pi(C,a_1) = \pi(C,a_2) = \pi(C,a_3) = 1/3$ ;
- en B, l'action $a_2$ est supposée impossible. Donc, on a $\pi(B,a_1) = \pi(B,a_3) = 1/2$.

$P$ et $R$ font partie de la définition du problème et ont été données en cours. On les rappelle ci-dessous :

| $P$ | état courant $x_t$ (ville courante) | action $a_t$ | état suivant $x_{t+1}$ ville suivante | | |
|---|---|---|---|---|---|
| | | | A | B | C |
| | 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 |

| $R$ | état courant $x_t$ (ville courante) | action $a_t$ | état suivant $x_{t+1}$ ville suivante | | |
|---|---|---|---|---|---|
| | | | A | B | C |
| | 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 |

Il n'y a donc plus qu'à écrire le système d'équations, remplacer les termes par leur valeur numérique et résoudre pour trouver les $V^\pi(A)$, $V^\pi(B)$ et $V^\pi(C)$. Pour simplifier les notations, on notera ici $A = V^\pi(A)$, $B = V^\pi(B)$ et $C = V^\pi(C)$ les inconnues.

On a donc :

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

Pour $A$, si on développe tout, on obtient :

$$\begin{aligned} A &= \pi(A,a_1)\{ & P(A,a_1,A)[R(A,a_1,A)+\gamma A] + \\ & & P(A,a_1,B)[R(A,a_1,B)+\gamma B] + \\ & & P(A,a_1,C)[R(A,a_1,C)+\gamma C]\} \\ & +\pi(A,a_2)\{ & P(A,a_2,A)[R(A,a_2,A)+\gamma A] + \\ & & P(A,a_2,B)[R(A,a_2,B)+\gamma B] + \end{aligned}$$

$$
\begin{aligned}
& P(A,a_2,C)[R(A,a_2,C)+\gamma C]\} \\
+\pi(A,a_3)\{\quad & P(A,a_3,A)[R(A,a_3,A)+\gamma A]+ \\
& P(A,a_3,B)[R(A,a_3,B)+\gamma B]+ \\
& P(A,a_3,C)[R(A,a_3,C)+\gamma C]\}
\end{aligned}
$$

que l'on peut ré-ordonner en :

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

et maintenant, il n'y a plus qu'à remplacer par le valeurs numériques :

$$
\begin{aligned}
A \quad = 1/3\{\quad & 1/2\times 10+ \\
& 1/4\times 4+ \\
& 1/4\times 8\} \\
+1/3\{\quad & 1/16\times 8+ \\
& 3/4\times 2+ \\
& 3/16\times 4\} \\
+1/3\{\quad & 1/4\times 4+ \\
& 1/8\times 6+ \\
& 5/8\times 4\} \\
+0{,}9A\{\quad & 1/3\times 1/2+1/3\times 1/16+1/3\times 1/4\}+ \\
+0{,}9B\{\quad & 1/3\times 1/4+1/3\times 3/4+1/3\times 1/8\}+ \\
+0{,}9C\{\quad & 1/3\times 1/4+1/3\times 3/16+1/3\times 5/8\}
\end{aligned}
$$

Des calculs élémentaires donnent :

$$A = 5 + 0{,}9 \times 13/48A + 0{,}9 \times 3/8B + 0{,}9 \times 17/48C$$

On procède de même pour $B$ et $C$ et on obtient le système d'équations linéaires :

$$\begin{cases} A = 5 + 0{,}9 \times 13/48A + 0{,}9 \times 3/8B + 0{,}9 \times 17/48C \\ B = 31/2 + 0{,}9 \times 9/32A + 0{,}9 \times 7/16B + 0{,}9 \times 9/32C \\ C = 31/6 + 0{,}9 \times 3/8A + 0{,}9 \times 17/48B + 0{,}9 \times 13/48C \end{cases}$$

Sa résolution ne pose pas de problème[1] : on le résout à la main ou avec l'aide d'un logiciel ; on examine pour cela plusieurs logiciels.

`maxima` est un logiciel libre de calcul formel : vous lui donnez simplement les équations et il vous fournit la solution. Ainsi, on entre les équations comme suit (les lignes commençant par `(%i` sont les lignes que vous tapez, celles commençant par `(%o` sont générées en réponse par maxima : elles permettent de repérer les expressions que vous entrez) :

```
(%i1) 5 + .9 * 13/48 * a + .9 * 9/24 * b + .9 * 17/48 * c - a = 0;
(%o1)                0.31875 c + 0.3375 b - 0.75625 a + 5 = 0
(%i2) 31/2 + .9 * 9/32 * a + .9 * 7/16 * b + .9 * 9/32 * c - b = 0;
                                                               31
(%o2)              0.253125 c - 0.60625 b + 0.253125 a + -- = 0
                                                               2
(%i3) 31/6 + .9 * 9/24 * a + .9 * 17/48 * b + .9 * 13/48 * c - c = 0;
                                                               31
(%o3)               - 0.75625 c + 0.31875 b + 0.3375 a + -- = 0
                                                               6
(%i4) solve ([%o1, %o2, %o3], [a, b, c]);
                             156420        5113540       13602460
(%o4)                  [[a = ------, b = -------, c = --------]]
                              1789          51881         155643
```

Cette dernière ligne demande à `maxima` de résoudre le système composé des équations `o1`, `o2` et `o3`, les inconnues étant `a`, `b` et `c`. `maxima` fournit :

$A = V^\pi(A) = \frac{156420}{1789} \approx$ 87,43, $B = V^\pi(B) = \frac{5113540}{51881} \approx$ 98,6 et $C = V^\pi(C) = \frac{13602460}{155643} \approx$ 87,40.

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

Libres également, `R`, `scilab`, `octave` résolvent eux-aussi ce genre de systèmes d'équations linéaires en un tour de main. Il faut juste leur expliquer un peu plus ce que l'on veut.

Ainsi, la résolution d'un système d'équations linéaires peut s'écrire sous forme matricielle $Mx = b$ où $M$ est une matrice, $b$ un vecteur et $x$ le vecteur inconnu. Dans le cas présent :

[1]elle ne doit pas en poser ; si elle en pose, me le dire.

$$x = \begin{pmatrix} A \\ B \\ C \end{pmatrix}$$

$$b = \begin{pmatrix} -5 \\ -31/2 \\ -31/6 \end{pmatrix}$$

et

$$M = \begin{pmatrix} 0{,}9 \times 13/48 - 1 & 0{,}9 \times 3/8 & 0{,}9 \times 17/48 \\ 0{,}9 \times 9/32 & 0{,}9 \times 7/16 - 1 & 0{,}9 \times 9/32 \\ 0{,}9 \times 3/8 & 0{,}9 \times 17/48 & 0{,}9 \times 13/48 - 1 \end{pmatrix}$$

Ces logiciels sont tous capables de trouver $x$ quand on leur fournit $M$ et $b$. En R, on fera comme suit :

```
m <- matrix (c (.9 * 13/48 - 1, .9 * 3/8,       .9 * 17/48,
                .9 * 9/32,       .9 * 7/16 - 1,  .9 * 9/32,
                .9 * 3/8,        .9 * 17/48,     .9 * 13/48 - 1),
            nrow = 3, ncol = 3, byrow = T)
b <- c (-5, -31/2, -31/6)
solve (m, b)
```

La dernière commande (`solve`) résout le système d'équations linéaires spécifié par `m` et `b`. On trouve la même réponse qu'avec `maxima`.

Avec `scilab` ou `octave`, seule la syntaxe change par rapport à R, le principe est exactement le même.

Il est extrêmement utile de connaître l'un ou l'autre de ces logiciels. Dans la suite, on utilisera fréquemment R pour implanter différents algorithmes (c'est un choix personnel ; `scilab` et `octave` ferait tout aussi bien l'affaire dans le cadre de ce cours). R est un langage interprêté, idéal pour un prototypage rapide ; il possède par ailleurs une très grande richesse de fonctions statistiques et numériques et est capable de générer de très nombreux graphiques.