

TD4. Chaînes de Markov contrôlées

Exercice 1. Une chaîne de Markov contrôlée $(X_n)_{n \geq 0}$ à valeurs dans $M = \mathbb{R}$ évolue selon la récurrence aléatoire contrôlée

$$X_{n+1} = X_n + U_n + \varepsilon_{n+1}$$

où $U_n = u_n(X_0, \dots, X_n)$, $u : \mathbb{C}_k$ est un contrôle à valeurs dans $A = \mathbb{R}$ et $(\varepsilon_n)_{n \geq 1}$ est une suite des v.a. iid de moyenne nulle et variance $\sigma^2 > 0$. On se fixe un horizon $T > 0$ et une constante $\gamma \in]0, 1[$. On veut trouver un contrôle u qui minimise le coût moyen (actualisé)

$$W_T^u(t, x) = \mathbb{E}_{(t, x)}^u \left[\sum_{k=t}^{T-1} \gamma^{k-t} C(X_k, U_k) + \gamma^{T-t} R(X_T) \right]$$

où $C(x, u) = -u^2 + ax^2/2$ et $R(x) = a_0 x^2/2 + b_0$ avec a, a_0, b_0 constantes réelles.

a) Montrer que la fonction $W_T(t, x) = \inf_{u \in \mathcal{C}_t} W_T^u(t, x)$ satisfait l'équation

$$W_T(t, x) = \inf_{u \in \mathbb{R}} \{C(x, u) + \mathbb{E}[W_T(t+1, x+u+\varepsilon_1)]\}.$$

b) Montrer par récurrence rétrograde que $W_T(t, x)$ est de la forme $W_T(t, x) = \frac{1}{2} a_T x^2 + b_T$ avec $(a_j)_{j=0}$ et $(b_j)_{j=0}$ des constantes.

c) Montrer que le contrôle optimal u est Markovien et tel que $u_t(x) = k_T x$ pour une certaine suite $(k_j)_{j=0}$ de constantes.

d) Donner la récurrence satisfaite par les constantes a_j, b_j, k_j pour $j \geq 0$.

e) Montrer que l'équation

$$f(z) = a + \frac{z^2}{1+z} = z$$

a une unique solution positive $z = a$ et que la suite $(a_j)_{j \geq 0}$ converge vers a .

f) Montrer que les limites $b = \lim_{j \rightarrow \infty} b_j$, $c = \lim_{j \rightarrow \infty} c_j$ existent.

g) Que peut-on dire sur le problème de minimiser le coût moyen en horizon infini :

$$W^u(t, x) = \mathbb{E}_{(t, x)}^u \left[\sum_{k=t}^{\infty} \gamma^{k-t} C(X_k, U_k) \right].$$

Exercice 2. (Contrôle optimale déterministe) Soit M l'espace d'états, A un espace des actions. Une dynamique contrôlée (homogène par simplicité) avec espace d'états M et espace des actions A est la donnée d'une fonction $f : M \times A \rightarrow M$, d'un état initial $x \in M$ et d'un temps initial k . Fixé un contrôle (déterministe) $a : \mathbb{N} \rightarrow A$ on considère la suite d'états $(x_n)_{n \geq 1}$ tel que $x_k = x$ et

$$x_{n+1} = f(n, x_n, a_n) \quad n \geq k.$$

Soit $r: \mathbb{N} \times M \times A \rightarrow \mathbb{R}_+$ une fonction de gain positive, on considère la fonction

$$V^a(k, x) = \sum_{n=k}^{\infty} r(n, x_n, a_n)$$

qui associe à chaque contrôle a et chaque état initial x le gain total réalisé par le système contrôlé partant de l'état x et utilisant le contrôle a . Soit $V(k, x) = \sup_{a: \mathbb{N} \rightarrow A} V^a(k, x)$ la fonction valeur du problème d'optimisation.

- a) Montrer que V est solution de l'équation de la programmation dynamique (équation de Bellman)

$$V(n, x) = \max_{z \in A} \{r(n, x, z) + V(n+1, f(n, x, z))\} \quad n \leq k, x \in M.$$

- b) Considérons le problème d'horizon fini soit $r(k, x, a) = 0$ si $k \geq N$, avec $N > 0$. Montrer que un contrôle $a: \mathbb{N} \rightarrow A$ tel que

$$V(n, x_n) = r(n, x_n, a_n) + V(n+1, f(n, x_n, a_n)) \quad (1)$$

pour tout $n \leq k$ est optimale.

Exercice 3. (Absence d'une stratégie optimale) On considère la dynamique contrôlée $f(x, a) = a(x + 1_{x=1})$ sur l'espace d'états \mathbb{N} et espace d'actions $A = \{0, 1\}$. La fonction de gain soit $r(x, a) = (1-a)(1/x)1_{x=1}$. Dans l'état $x = 1$ on peut choisir de monter de 1 ou de descendre jusqu'à 0 et ceci faisant on gagne une récompense de $1/x$. Une fois qu'on ne peut plus rien gagner. Il n'y a pas des variables aléatoires en jeux et le problème est déterministe. Soit

$$V(x) = \sup_{a: \mathbb{N} \rightarrow A} \sum_{n=0}^{\infty} r(x_n, a_n)$$

où $x_{n+1} = f(x_n, a_n)$ pour $n \geq 0$ et $x_0 = x$.

- a) Justifier le fait que la fonction V est la plus petite solution de l'équation

$$V(x) = \max \{1/x, V(x+1)\} \quad x \geq 1 \quad (2)$$

avec $V(0) = 0$.

- b) Montrer que $V(x) = 1/x$ est solution de l'équation (2) pour tout $x \geq 1$.
 c) Montrer que il n'y a pas d'autres solutions de cet équation et donc que la solution du problème d'optimisation est la plus petite des $V, c \leq V_1$.
 d) Montrer que, pour tout $x \geq 1$, il n'y a pas de contrôle optimale pour cet solution, i.e. que il n'existe pas une fonction $a: \mathbb{N} \rightarrow A$ tel que

$$f(x_n, a_n) = V_1(x).$$

Cet exercice montre que en général la condition (1) n'est pas suffisante pour garantir l'optimalité d'un contrôle.

Exercice 4. (Parier de façon optimale) On possède 1 euro et on voudrait les jouer pour obtenir N euros. On mise sur une suite de jeux favorable avec probabilité $p > 1/2$ de gagner chaque fois et tels que les jeux successifs sont indépendants. Je peut miser un nombre entier de euros et chaque fois je ne peut miser que ce que je possède. Quel stratégie doit-je adopter pour atteindre mon but de gagner N euros?

Pour modéliser cette situation on considère l'espace d'états $M = N \cup \{0\}$ qui donne la fortune accumulée pendant le jeu où si je décide de jouer, les actions possibles sont toutes les mises : $A = N$. Pour tenir compte de la contrainte de ne pouvoir pas miser plus que ma fortune courante on considère la fonction de gain

$$r(n, x, a) = (1/p)1_{x < a} + 1_{x \leq N} \quad \text{pour } x \in \mathbb{R}$$

et $r(n, 0, a) = 0$. La dynamique est donnée par $X_{n+1} = G(X_n, U_n, Z_{n+1})$ avec $(Z_n)_{n \geq 1}$ une suite iid de Bernoulli de paramètre p et $G(x, a, z) = (x + a)1_{z=1} + (x - a)1_{z=0}$ si $0 \leq x < N$ et $G(x, a, z) = x$ si $x \geq N$ (car je décide de m'arrêter de jouer une fois que j'ai gagné N).

- a) Montrer que l'équation d'optimalité est donnée par

$$V(x) = \max_{a \geq x} \{pV(x+a) + (1-p)V(x-a)\} \quad 1 \leq x < N$$

et $V(0) = 0$, $V(x) = 1$ pour tout $x \geq N$.

- b) Soit u la stratégie de parier chaque fois un euro. Montrer que

$$V^u(x) = pV^u(x+1) + (1-p)V^u(x-1), \quad 1 \leq x < N$$

avec $V^u(0) = 0$ et $V^u(x) = 1$ pour tout $x \geq N$.

- c) Montrer que cette équation a une unique solution $V^u = (1 - p^x)/(1 - p^N)$ avec $p = (1-p)/p \in]0, 1[$.
- d) Montrer que V^u est concave et que cela implique que V^u satisfait l'équation d'optimalité et donc que u est optimal.