Feuille du Chapitre 3 : Modèle à n périodes

Exercice 1. Sur la modèle de la représentation de $U_0(i)$, donner une représentation de $U_n(i)$ pour chaque $n \in \{0, \cdots, N-1\}$. On sommera les coûts à partir de l'instant n.

On rappelle que

$$U_n(i) = \max_a \left(r(n,i,a) + \sum_{j \in \mathcal{S}} U_{n+1}(j) P(n,i,a,j) \right).$$

Par ailleurs, si on définit pour X_n une variable

$$\mathcal{R}_n(i) = \max \mathbb{E}\left[\sum_{k=n}^{N-1} r(k, X_k, \alpha_k) + g(X_N) \middle| X_n = i \right],$$

grâce au résultat du cours (principe de programmation dynamique), on a, si

$$\mathbb{P}\left(\bigcap_{k=n}^{N-1} (\alpha_k \in A_k(X_k)) \right) = 1, \quad A_k(x) = \mathrm{argmax}_a U_k(x),$$

pour tout $n \leq k \leq N-1$

$$\mathcal{R}_n(i) =$$

$$\mathcal{R}_n(i) = \max \mathbb{E}\left[\sum_{k=n}^{N-1} r(k, X_k, \alpha_k) + g(X_N) | X_n = i \right],$$

On rappelle également que le principe de programmation dynamique implique alors que

$$\mathcal{R}_n(i) = \mathbb{E}[U_{N-(N-n)}(X_0) | X_0 = i] = U_n(i),$$

et donc

$$U_n(i) = \max \mathbb{E}\left[\sum_{k=n}^{N-1} r(n, X_n, \alpha_n) + g(X_N) \middle| X_n = i \right]$$

Exercice 2. Montrer, pour une stratégie optimale comme donnée dans l'énoncé de la Proposition 2.6 du cours que la suite $(U_n(X_n))_{0 \leq n \leq N}$ est une martingale relativement à la filtration $(\mathcal{F}_n)_{0 \leq n \leq N}$ sous-jacente, si on suppose que le coût instantané $r \equiv 0$.
Indication : On écrira l'identité $U_n(i) = r(n,i,a_n(i)) + \sum_{j \in S} U_{n+1}(j) P(n,i,a_n(i),j)$ sous la forme d'une espérance conditionnelle sachant $\mathcal{F}_n$ calculée sur l'évènement $\{X_n = i\}$.

Soit, pour tout $n \in \mathbb{N}$ et $i \in \mathcal{S}$,

$$a_n(i) \in \mathrm{argmax}_a \left\{ r(n,i,a) + \sum_{j \in S} U_{n+1}(j) P(n,i,a,j) \right\}$$

de telle sorte que

$$U_n(i) = r(n,i,a_n(i)) + \sum_{j \in S} U_{n+1}(j) P(n,i,a_n(i),j) = \mathbb{E}[r(n,X_n,a_n(X_n)) + U_{n+1}(X_{n+1}) | X_n = i, \alpha_n = a_n(X_n)].$$

On remarque alors que

$$U_n(X_n) = \mathbb{E}[r(n,X_n,a_n(X_n)) + U_{n+1}(X_{n+1}) | X_n, \alpha_n = a_n(X_n)].$$

Ce qui donne

$$U_n(X_n) = r(n,X_n,a_n(X_n)) + \mathbb{E}[U_{n+1}(X_{n+1}) | X_n, \alpha_n = a_n(X_n)].$$

$$\begin{aligned}
\mathbb{E}[U_{n+1}(X_{n+1}) | \mathcal{F}_n] =& \mathbb{E}[\mathbb{E}[U_{n+1}(X_{n+1}) | X_n, \alpha_n = a_n(X_n)] | \mathcal{F}_n] \\
=& \mathbb{E}[U_n(X_n) - r(n,X_n,a_n(X_n)) | \mathcal{F}_n] \\
=& U_n(X_n) - r(n,X_n,a_n(X_n)).
\end{aligned}$$

Exercice 3. En choisissant $S = \{-1, 1\}$, $A = \{-1, 1\}$ $N = 1$, $r_0(s,a) = \frac{1}{2}a^2$, $g(s) = \frac{1}{2}s^2$, et $P(i,j,a) = \mathbf{1}_{\{j=a\}}$, étudier les stratégies optimales.

Exercice 4. On considère $S = \{-1, 1\}$, $A = \{0, 1\}$, $r(s,a) = s^2 + a$ et $P(s,1,1-s) = p$ et $P(s,0,s) = 1$. On pose $g(s) = s$.

(1) Écrire la récurrence de Bellman.

(2) Calculer la fonction valeur.

On pose $U_2(i) = i$, puis

$$U_n(i) = \max\big[1 + a + a\big(pU_{n+1}(1-i) + (1-p)U_{n+1}(i)\big) + (1-a)U_{n+1}(i)\big].$$

Exercice 5. On considère un système informatique, dont la protection est un entier entre 0 et 1 : 0 si la protection est nulle, 1 si la protection est présente. L'état est 0 ou 1 : 0 s'il est sain et 1 s'il est infecté.

A chaque instant, on peut payer un coût fixe c (et donc la récompense est négative $-c$) pour mettre la protection au niveau 1 et on paye $10 \times c$ si l'ordinateur est infecté (la récompense est alors de $-10 \times c$).

Si la protection est 1, la probabilité d'être infecté est p si l'état est 0 ou 1. Si la protection est 0, la probabilité d'être infecté est q si l'état 0 et est 1 si l'état est 1.

(1) En identifiant la protection à l'action, écrire le modèle.
(2) Donner la forme des accroissements dans Bellman.
(3) Résoudre Bellman sur $N = 2$ avec une récompense finale égale à αc si $X_2 = 0$ et βc si $X_2 = 1$.

On prend à nouveau $S = \{0,1\}$ et $A = \{0,1\}$ (1 si on paye la protection et 0 sinon).

Les transitions sont $P(0,1,0) = P(1,1,0) = p$, $P(0,0,0) = 1$ et $P(1,0,0) = q$.

L'équation de Bellman est

$$U_n(0) = -10c + \max\big[-c + pU_{n+1}(0) + (1-p)U_{n+1}(1), U_{n+1}(0)\big],$$
$$U_n(1) = \max\big[-c + pU_{n+1}(0) + (1-p)U_{n+1}(1), qU_{n+1}(0) + (1-q)U_{n+1}(1)\big].$$