Introduction aux processus stochastiques

Travail Matlab : Chaînes de Markov

Consignes de style pour le travail

Vous pouvez remettre un document rédigé grosso modo comme lors d'un examen pour lequel vous disposeriez de MATLAB *et d'une imprimante. Autrement dit, présentez vos réponses de façon claire et synthétique, faites des remarques quand vous le jugez utile, et commentez vos codes s'ils font partie de votre réponse.*

Contrôle d'un processus stochastique discret à temps discret.

©Edward Burtynsky

La commande des systèmes est un riche terrain pour les ingénieurs. Tous les systèmes ont en pratique une composante stochastique, plus ou moins prépondérante.
L'exercice vous propose une façon de formaliser le problème de la commande optimale d'un système dont la dynamique est fortement stochastique. L'exemple est réduit à sa plus simple expression, mais les notions abordées resteront valides pour des applications plus stimulantes.

But de l'exercice

On considère un système dynamique stochastique à 4 états, numérotés de 1 à 4.
On note $\mathcal{X}_1, \mathcal{X}_2, \ldots$ la trajectoire du système dans l'espace d'état : $\mathcal{X}_t = \{1, 2, 3, 4\}$ est une variable aléatoire qui représente l'état du système au pas de temps t, pour $t \in \{1, 2, \ldots\}$.
La dynamique est markovienne (d'ordre 1) : $P(\mathcal{X}_{t+1}|\mathcal{X}_1, \ldots, \mathcal{X}_t) = P(\mathcal{X}_{t+1}|\mathcal{X}_t)$.

Le système est contrôlé au moyen d'une action $\mathcal{U}_t = \{1, 2, 3\}$ au pas de temps t, capable d'influer sur $\mathcal{X}_{t+1}$. L'action peut être choisie sur base de l'observation de l'état $\mathcal{X}_t$.

Une *politique de décision* (déterministe) $\mathcal{A}_t : \{1, 2, 3, 4\} \to \{1, 2, 3\}$ consiste à spécifier, pour chaque valeur possible de $\mathcal{X}_t$, l'action $\mathcal{U}_t$ sélectionnée au pas de temps t.
Une *politique de décision stationnaire* (déterministe) $\mathcal{A} : \{1, 2, 3, 4\} \to \{1, 2, 3\}$ est une politique de décision indépendante de t : l'arrivée dans un état i provoque toujours le choix d'une certaine action $a(i) \in \{1, 2, 3\}$.

Aux états $s \in \{1, 2, 3, 4\}$ du système est associée une certaine *utilité* $\mathcal{R}_s \in \mathbb{R}$. Une trajectoire dans l'espace d'état, qui dépend d'une politique de décision $\mathcal{A}$, induit donc une trajectoire $\mathcal{R}_{\mathcal{X}_1}, \mathcal{R}_{\mathcal{X}_2}, \ldots$ dans $\mathbb{R} \times \mathbb{R} \times \ldots$

Aux actions $u \in \{1, 2, 3\}$ est associé un certain *coût* $\mathcal{C}_u \in \mathbb{R}$. Une politique de décision $\mathcal{A}$ et une trajectoire dans l'espace d'état induisent donc une trajectoire $\mathcal{C}_{a(\mathcal{X}_1)}, \mathcal{C}_{a(\mathcal{X}_2)}, \ldots$ dans $\mathbb{R} \times \mathbb{R} \times \ldots$

Dans cet exercice, on s'intéresse à l'espérance de l'utilité moyenne récoltée au cours d'une trajectoire arbitrairement longue, nette du coût moyen des actions, lorsqu'une certaine politique $\mathcal{A}$ est mise en oeuvre :

$$f(\mathcal{A}) = \lim_{T \to \infty} \mathbb{E}\{\frac{1}{T} \sum_{t=1}^{T} (\mathcal{R}_{\mathcal{X}_t} - \mathcal{C}_{a(\mathcal{X}_t)})\}, \tag{1}$$

l'espérance étant prise par rapport au processus $\mathcal{X}_1, \mathcal{X}_2, \ldots$ contrôlé par $\mathcal{A}$.

Remarquez bien qu'en spécifiant $f(\cdot)$, on ordonne les politiques de décision. Aux comportements les plus désirables du système sont associés des valeurs de f plus élevées, et ceci induit un ordre de préférence entre les différentes politiques de décision $\mathcal{A}$ possibles.

Le but de l'exercice est de rechercher une politique de décision stationnaire optimale, c'est-à-dire ici une politique de décision stationnaire $\mathcal{A}^\star$ qui maximise f.

Remarque : Les conditions techniques pour que (1) existe seront vérifiées pour le système et la structure d'utilités et de coûts étudiés.

Spécification du système, des utilités et des coûts

Les trajectoires dans l'espace d'état, conditionnellement aux actions choisies, sont décrites au travers de probabilités $P(\mathcal{X}_{t+1}|\mathcal{X}_t,\mathcal{U}_t)$ invariantes dans le temps, i.e. ne dépendant pas de t.

Les probabilités sont définies via 3 *matrices de transition* $\mathbf{\Pi}^{(u=1)}, \mathbf{\Pi}^{(u=2)}, \mathbf{\Pi}^{(u=3)}$.
L'élément $\Pi_{ij}^{(u=k)}$ vaut $P(\mathcal{X}_{t+1} = j|\mathcal{X}_t = i,\mathcal{U}_t = k)$, la probabilité de passer de l'état i à l'état j lorsque l'action k est prise.

$$\mathbf{\Pi}^{(u=1)} = \begin{bmatrix} 0.37 & 0.12 & 0.34 & 0.17 \\ 0.29 & 0.14 & 0.24 & 0.33 \\ 0.27 & 0.27 & 0.31 & 0.15 \\ 0.19 & 0.26 & 0.32 & 0.23 \end{bmatrix}$$

$$\mathbf{\Pi}^{(u=2)} = \begin{bmatrix} 0.07 & 0.61 & 0.06 & 0.26 \\ 0.23 & 0.00 & 0.20 & 0.57 \\ 0.40 & 0.07 & 0.44 & 0.09 \\ 0.30 & 0.13 & 0.30 & 0.27 \end{bmatrix}$$

$$\mathbf{\Pi}^{(u=3)} = \begin{bmatrix} 0.13 & 0.08 & 0.25 & 0.54 \\ 0.04 & 0.53 & 0.16 & 0.27 \\ 0.17 & 0.17 & 0.50 & 0.16 \\ 0.28 & 0.24 & 0.47 & 0.01 \end{bmatrix}$$

On définit un vecteur d'utilité des états $\boldsymbol{r}$, dont l'élément r_i représente l'utilité de se trouver dans l'état i :

$$\boldsymbol{r} = \begin{bmatrix} -2 \\ -1 \\ -1 \\ +3 \end{bmatrix}.$$

(Une valeur négative pour v_i signifie qu'on voudrait éviter de se trouver dans l'état i !)

Pour les coûts, on définit un vecteur de coût des actions $\boldsymbol{c}$, dont l'élément c_k représente le coût de l'action k. Pour certaines des questions, on utilisera un vecteur nul,

$$\boldsymbol{c}^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Pour d'autres, on utilisera

$$\boldsymbol{c} = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.1 \end{bmatrix}.$$

Représentation d'une politique de décision

On peut représenter une politique de décision par une matrice $\boldsymbol{A}$ de dimension 4×3, dont l'élément A_{ik} donne la probabilité de choisir l'action k lorsqu'on se trouve dans l'état i.
Comme on ne considére que des politiques déterministes, seul un élément par ligne sera non nul et vaudra 1.

Questions

1. A partir d'une politique stationnaire (déterministe) représentée par $\boldsymbol{A}$ et des matrices de transition $\boldsymbol{\Pi}^{(u=k)}$, déterminer comment construire ou calculer une matrice de transition $\boldsymbol{\Pi}$ dont l'élément Π_{ij} représente la probabilité
$$P(\mathcal{X}_{t+1} = j | \mathcal{X}_t = i, \mathcal{U}_t = a(i)).$$
2. Déterminer l'expression de $P(\mathcal{X}_t = j)$ pour un $t > 1$ fixé, en fonction de $\boldsymbol{\Pi}$ et d'un vecteur ligne $\boldsymbol{\pi}$ dont l'élément i représente $P(\mathcal{X}_1 = i)$.
Remarque : Point intéressant en soi mais non utilisé dans la suite (cf. point 3).
3. Parce que les matrices $\boldsymbol{\Pi}$ satisfont à 2 propriétés (irréductibilité et apériodicité) garanties par le choix des matrices $\boldsymbol{\Pi}^{(u=k)}$ de l'énoncé, quelle que soit la politique $\mathcal{A}$, $P(\mathcal{X}_t = j)$ converge vers une valeur unique lorsque $t \to \infty$.
On appelle la distribution résultante la *distribution stationnaire*.
On la représente par un vecteur ligne $\boldsymbol{\pi}^\infty$, avec $\pi_j^\infty = \lim_{t\to\infty} P(\mathcal{X}_t = j)$.
Connaissant $\boldsymbol{\Pi}$, on trouve $\boldsymbol{\pi}^\infty$ au moyen de 4 équations linéairement indépendantes choisies parmi les relations
$$\boldsymbol{\pi}^\infty \boldsymbol{\Pi} = \boldsymbol{\pi}^\infty \qquad \text{(prendre 3 des 4 équations)}$$
$$\sum_{i=1}^{4} \pi_i^\infty = 1 \qquad \text{(1 équation).}$$
Ecrire une fonction MATLAB prenant en argument une matrice de transition $\boldsymbol{\Pi}$ (supposée irréductible et apériodique) et renvoyant la distribution stationnaire en vecteur ligne :
```
function pi_inf = dist_stat(trans)
```
4. Etablir l'expression de $f(\mathcal{A}) = \lim_{T\to\infty} \mathbb{E}\{\frac{1}{T}\sum_{t=1}^{T}(\mathcal{R}_{\mathcal{X}_t} - \mathcal{C}_{a(\mathcal{X}_t)})\}$ donnée en (1) en fonction de $\boldsymbol{\pi}^\infty$, $\boldsymbol{A}$, $\boldsymbol{r}$ et $\boldsymbol{c}$.
Indications :
   - Réfléchir à ce que signifie "physiquement" $\frac{1}{T}\sum_{i=1}^{T} \mathcal{R}_{\mathcal{X}_i}$, et vers quoi ceci devrait tendre lorsque $T \to \infty$.
   - $\lim_{T\to\infty} \mathbb{E}\{\frac{1}{T}\sum_{i=1}^{T} g_i\} = \lim_{T\to\infty} \mathbb{E}\{\frac{1}{T-N}\sum_{i=N+1}^{T} g_i\}$
   lorsque $g_i \le K < \infty$, avec K une constante, pour tout N fini.
5. On considère une politique de décision $\mathcal{A}_k$ consistant à choisir systématiquement l'action k. Calculer $f(\mathcal{A}_k)$ pour $k = 1, 2, 3$:
   (a) Dans le cas du vecteur de coût $\boldsymbol{c}^{(0)}$ (ce qui élimine le terme $\mathcal{C}_{a(\mathcal{X}_i)}$ dans (1)) ;
   (b) Dans le cas du vecteur de coût $\boldsymbol{c}$.
Présenter les résultats ($\boldsymbol{A}$, $\boldsymbol{\pi}^\infty$, $f(\mathcal{A})$) sous forme d'un tableau,
avec $\boldsymbol{A}$ la représentation matricielle de la politique considérée.
6. Déterminer une politique stationnaire optimale $\mathcal{A}^\star$, en opérant une recherche systématique, grâce à MATLAB, de la valeur $f(\mathcal{A})$ sur toutes les politiques stationnaires (déterministes) possibles $\mathcal{A}$:
   (a) Dans le cas du vecteur de coût nul $\boldsymbol{c}^{(0)}$;
   (b) Dans le cas du vecteur de coût $\boldsymbol{c}$.
Présenter les résultats ($\boldsymbol{A}^\star, \boldsymbol{\Pi}^\star, \boldsymbol{\pi}^{\infty\star}, f(\mathcal{A}^\star)$) sous forme d'un tableau,
avec $\boldsymbol{A}^\star$ la représentation matricielle de la politique optimale trouvée.

Bon travail.