Introduction aux processus stochastiques

Chaînes de Markov

Warmup.

Soit la chaîne de Markov (π, Π) avec $\pi = [\ .1 \quad .2 \quad .7\ ]$ et $\Pi = \begin{bmatrix} .3 & 0 & .7 \\ .2 & .8 & 0 \\ .5 & .5 & 0 \end{bmatrix}$.

1. Tracer le diagramme de transition d'état de la chaîne de Markov.
   C'est un graphe dirigé, dont les noeuds sont les états de la chaîne et portent le numéro correspondant. On trouve un arc orienté du noeud i vers le noeud j si $\Pi_{ij} > 0$. Les arcs ij sont étiquetés avec la probabilité de transition Π_{ij}.
2. Déduire du graphe que la chaîne est irréductible et apériodique (définitions en annexe).
3. Donner une interprétation de l'élément ij de la matrice Π^k, pour k entier supérieur ou égal à 1.
4. Une chaîne de Markov irréductible et apériodique possède toujours une unique distribution stationnaire.
   Une distribution π_∞ stationnaire satisfait $\pi_\infty = \pi_\infty \Pi$.
   Calculer π_∞.

   (Solution : $\pi_\infty = [\ .2899 \quad .5072 \quad .2029\ ]$)

Acides aminés distincts présents dans une séquence.

On considère des séquences de longueur N d'acides aminés, avec M acides aminés différents possibles. On dispose d'un modèle chaîne de Markov (π, Π) qui reflète les règles de succession entre deux acides aminés de la séquence.
π est un vecteur ligne dont l'élément j vaut la probabilité que le premier acide aminé soit j, et Π une matrice dont l'élément ij vaut la probabilité que le $(n+1)$-ème acide aminé soit j si le n-ème acide aminé est i.
On veut connaître l'espérance du *nombre d'acides aminés distincts* qu'on peut trouver dans une séquence de longueur N.

1. Quelle est la probabilité $p_{\sim j}$ que l'acide aminé j n'apparaisse pas dans une séquence de longueur N ?
   En déduire la probabilité p_j que l'acide aminé j apparaisse dans une séquence de longueur N.

2. Que vaut l'espérance de la fonction indicatrice $\mathcal{X}_A$ d'un événement A ?
   ($\mathcal{X}_A$ est une variable aléatoire qui vaut 1 si l'événement survient, 0 sinon.)

3. Soit $\mathcal{Y} = \{1, 2, \ldots, M\}$ la variable aléatoire qui donne le nombre d'acides aminés distincts rencontrés sur une séquence de longueur $N \geq 1$. Quelle est son espérance $\mathbf{E}\{\mathcal{Y}\}$?

Satisfait ou remboursé.

Un constructeur automobile en difficulté veut proposer une garantie “satisfait ou remboursé” pour doper les ventes d’un de ses modèles, la Lemon.
Ce modèle introduit depuis 2 ans est critiqué par la presse spécialisée. Pour 1000 véhicules, on estime désormais que sur les 12 premiers mois d’utilisation 170 véhicules connaîtront une panne, 30 deux pannes, et 10 plus de deux pannes. Les pannes multiples sont toujours séparées d’au moins un mois.
La garantie consisterait à reprendre tout véhicule ayant connu sur les 36 premiers mois plus de deux pannes imputables au constructeur.
Le constructeur pense que la probabilité de panne est la même de mois en mois, mais que cette probabilité change à chaque fois qu’une nouvelle panne survient.
Pour quantifier le risque à 36 mois, le constructeur veut construire un modèle du processus de panne à partir des probabilités sur 12 mois, puis l’utiliser pour évaluer la probabilité de plus de 2 pannes au cours des premiers 36 mois.
Il veut aussi pouvoir calculer le prix de la garantie qu’il pourrait vendre à d’anciens clients, sur base de l’âge et du nombre de pannes déjà encourues du véhicule.

Pour fixer les idées, introduisons l’événement et les variables aléatoires suivants :

A : “Une panne se produit au cours du mois”.

$\mathcal{X}_t$: fonction indicatrice de A pour le mois t, qui vaut 1 si A se produit, 0 sinon.

$\mathcal{Y}_t = \sum_{\tau=1}^{t} \mathcal{X}_\tau$: nombre d’occurrences de A sur les t premiers mois.

1. Proposer un modèle type chaîne de Markov (π, Π) pour $\mathcal{Y}_t$ adapté au problème.
   Suggestion : identifier les états S_i pertinents pour le problème puis utiliser comme paramètres du modèle $p_i = P(A|\mathcal{Y}_t = S_i)$.
   Caractériser le modèle obtenu en répondant aux sous-questions suivantes (sans tenir compte d’éventuelles valeurs pathologiques des paramètres p_i) :
   (a) La chaîne de Markov du problème est-elle irréductible ? Qu’est-ce que cela signifie du point de vue des transitions entre états ?
   (b) On dit que S_i est un état récurrent si la chaîne de Markov y retourne infiniment souvent. Sinon on dit que S_i est un état transitoire, ce qui signifie que la chaîne de Markov ne peut y retourner qu’un nombre fini de fois. Caractériser les états de la chaîne de Markov du problème.
   (c) Pouvez-vous trouver une distribution stationnaire π_∞ relative à Π ? Est-elle unique ?
2. (a) Développer une méthode pour choisir les paramètres p_i tels que le modèle soit conforme aux données, i.e. $P(\mathcal{Y}_N = 1) = 0.17$, $P(\mathcal{Y}_N = 2) = 0.03$, $P(\mathcal{Y}_N > 2) = 0.01$ en $N = 12$. La valeur numérique des p_i n’est pas demandée.
   (b) Pour calculer numériquement A^k avec A diagonalisable , on utilise $A^k = VD^kV^{-1}$.
   Quelles sont les valeurs propres de la matrice de transition Π du problème ?
3. Etablir les formules qui donnent
   (a) La probabilité de plus de 2 pannes sur les 36 premiers mois,
   (b) La probabilité de plus de 2 pannes sur les 36 premiers mois, sachant qu’à ce jour (mois $t < 36$) aucune panne ne s’est produite.
   (c) La probabilité de plus de 2 pannes sur les 36 premiers mois, sachant qu’une première panne vient de se produire ce mois t.
   (d) La probabilité de plus de 2 pannes sur les 36 premiers mois, sachant qu’à ce jour (mois $t < 36$) seules deux pannes se sont produites aux mois t_1 et t_2 (avec $t_1 < t_2 \leq t$).

Annexe

Matrices non négatives, positives, irréductibles.

Une matrice A est dite non négative si ses éléments satisfont $A_{ij} \geq 0$.
Une matrice A est dite positive si $A_{ij} > 0$.
Une matrice non négative A est irréductible s'il existe un entier k tel que A^l est positive pour tout $l \geq k$, où A^l représente A élevée à la puissance l.
Ne pas confondre matrice non négative avec matrice semi-définie positive.
Ne pas confondre matrice positive avec matrice définie positive.

Chaînes de Markov irréductibles, apériodiques.

Une chaîne de Markov est irréductible s'il est possible de passer de tout état i à tout état j en un nombre fini d'étapes.
Une chaîne de Markov irréductible a sa matrice de transition Π irréductible (pourquoi ?).
Un état j est périodique de période $k \geq 2$ si l'ensemble des temps de retour (i.e. nombre de transitions nécessaires) d'un état i vers lui-même est un multiple de $k \geq 2$. Sinon $k = 1$ et l'état est apériodique.
Une chaîne de Markov est apériodique si tous ses états sont apériodiques.

Matrices diagonalisables.

Soit A une matrice carrée diagonalisable. On peut donc écrire $AV = VD$ avec D une matrice diagonale contenant les valeurs propres de A, et V une matrice carrée non singulière (i.e. V^{-1} existe) dont les colonnes sont les vecteurs propres correspondants.