Points importants du cours 6

Vincent Guillemot

19 février 2009

1 Introduction aux Chaînes de Markov

On revient, pour ce chapitre, à des notions fondamentales et générales, le chapitre suivant sera plus orienté sur des applications à la biologie. Nous allons définir des outils utiles dans l'analyse de séquences biologiques (nous nous contenterons de séquences ADN) : chaque oligo sera modélisé par une variable aléatoire. L'ordre d'apparition de ces oligos n'est pas dû au hasard, et il serait intéressant de disposer d'outils statistiques permettant de traduire le fait qu'une base particulière existe à la position k avec une certaine probabilité dépendant des bases qui l'entourent (par exemple la base en position $k-1$). Les applications de ces modèles se trouvent dans la découverte de gènes de façon automatique et dans la comparaison de séquences biologiques.

On se place dans un espace probabilisé $(\Omega, \mathcal{T}, P)$.

1.1 Exemple préliminaire

1.2 Définitions, notations

Définition 1 - *Probabilité conditionnelle*

Soit A et B deux événements de $\mathcal{T}$ tels que la probabilité de B ne soit pas nulle, la probabilité de l'événement A sachant l'événement B, ou encore la probabilité A conditionnellement à B est notée $P(A|B)$ et vaut

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Proposition 1 - *Probabilité conditionnelle*

(i) On définit ainsi bien une probabilité sur $(A, \mathcal{T}_A, P_A)$,

(ii) la probabilité conditionnelle de deux événements indépendants est nulle.

Définition 2

Un système complet d'événements est une famille dénombrable ou finie $(B_n)_n$ d'événements deux à deux disjoints et vérifiant $\sum_n P(B_n) = 1$.

Théorème 1 - *Formule des probabilités totales*

Soit $(B_n)_n$ un système complet d'événements et A un événement. On a

$$P(A) = \sum_n P(A|B_n)P(B_n).$$

Théorème 2 - *Formule de Bayes*

Soit $(B_n)_n$ un système complet d'événements et A un événement, les B_n et A sont de probabilité non nulle. On a

$$\forall i, P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}.$$

Trajet du matin : Pour venir en cours le matin, un élève a le choix entre la voiture (A), le train (B), les rollers (C) et le vélo (D). La probabilité de choisir la voiture (resp. le train, les rollers) est de 1/3 (resp. 1/4, 1/12). La probabilité d'arriver en retard en voiture est de 1/20, en train de 1/10, en rollers de 1/5. Il n'arrive jamais en retard en vélo.

1. Quelle est la probabilité que l'élève vienne en vélo ?
2. L'élève arrive en retard. Quelle est la probabilité qu'il soit venu en train ?
3. Quelle est la probabilité d'arriver en retard ?

2 Modélisation par Chaînes de Markov

Dans ce paragraphe, on manipule des variables aléatoires réelles discrètes plutôt que des événements, on précise ainsi implicitement que Ω est un ensemble dénombrable ou fini.

Définition 3 - *Indépendance de deux variables*

Si (X, Y) est un couple de variables aléatoires discrètes, on appelle loi conditionnelle de X sachant $Y = y$ la donnée de :

$$\{P((X = x|Y = y), x \text{ valeur possible de } X\},$$

avec

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

Soit une suite $X_0, ..., X_{n+1}$ de variables aléatoires discrètes de $\mathcal{T}$ à valeurs dans une certaine tribu $\mathcal{A}$ (très souvent on associe ainsi à un événement de l'expérience aléatoire un "gain", notion importante en théorie des jeux, et qui a son importance en génomique également).

Définition 4 - *Chaîne de Markov*

Cette suite $X_0, ..., X_{n+1}$ est une chaîne de Markov si et seulement si

$$P(X_{n+1} = x_{n+1} | X_n = x_n, ..., X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n).$$

Les chaînes de Markov sont parfois appelées « processus sans mémoire », car le futur (X_{n+1}) ne dépend que de l'état présent (X_n), et pas des états précédents.

Définition 5 - *Homogénéité*

Si $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov, on dit qu'elle est homogène si :

$$\forall i, j \in E, \forall n \in \mathbb{N}, P(X_{n+1} = j | X_n = i) = \pi(i, j).$$

Autrement dit, la probabilité de passer de l'état i à l'état j au nième coup ne dépend que de i et de j et pas de n.

Définition 6 - *Matrice de transition*

On appelle matrice de transition, et on note Π la matrice constituée des $\pi(i, j)$

Proposition 2

Une chaîne de Markov homogène à espace d'états dénombrable est entièrement caractérisée par P_0 et Π.

Doudou, le hamster paresseux : Doudou, le hamster paresseux, ne connaît que 3 endroits dans sa cage : les copeaux où il dort (entre autres), la mangeoire où il mange et la roue où il fait de l'exercice. Ses journées sont assez semblables les unes aux autres, et son activité se représente aisément par une chaîne de Markov. Toutes les minutes, il peut soit changer d'activité, soit continuer celle qu'il était en train de faire. L'appellation processus sans mémoire n'est pas du tout exagérée pour parler de Doudou.

- Quand il dort, il a 9 chances sur 10 de ne pas se réveiller la minute suivante.
- Quand il se réveille, il va soit manger, soit faire de l'exercice, au hasard.
- Le repas ne dure qu'une minute, après il fait autre chose.
- Après avoir mangé, il y a 3 chances sur 10 qu'il parte courir dans sa roue, mais surtout 7 chances sur 10 qu'il retourne dormir.

– Courir est fatigant ; il a donc 80% de chance de retourner dormir au bout d'une minute. Sinon il continue en oubliant qu'il est déjà un peu fatigué.

Les diagrammes peuvent montrer toutes les flèches, chacune représentant une probabilité de transition. Cependant, c'est plus lisible si on ne dessine pas les flèches de probabilité zéro (transition impossible) et si on ne dessine pas les boucles (flèche d'un état vers lui-même). Cependant elles existent ; leur probabilité peut être sous-entendue car on sait que la somme des probabilités des flèches partant de chaque état doit être égale à 1.

La matrice de transition de ce système est la suivante (les lignes et les colonnes correspondent dans l'ordre aux états *dormir*, *manger*, *courir*) :

$$P = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.7 & 0 & 0.3 \\ 0.8 & 0 & 0.2 \end{bmatrix}$$

Prenons l'hypothèse que Doudou dort lors de la première minute de l'étude.

$$\mathbf{x}^{(0)} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

Au bout d'une minute, on peut prédire :

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)}P = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0,9 & 0,05 & 0,05 \\ 0,7 & 0 & 0,3 \\ 0,8 & 0 & 0,2 \end{bmatrix} = \begin{bmatrix} 0.9 & 0.05 & 0.05 \end{bmatrix}$$

Ainsi, après une minute, on a 90% de chances que Doudou dorme encore, 5% qu'il mange et 5% qu'il courre.

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)}P = \mathbf{x}^{(0)}P^2 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0,9 & 0,05 & 0,05 \\ 0,7 & 0 & 0,3 \\ 0,8 & 0 & 0,2 \end{bmatrix}^2 = \begin{bmatrix} 0.885 & 0.045 & 0.07 \end{bmatrix}$$

Après 2 minutes, il y a 4.5% de chances que l'hamster mange.

De manière générale, pour n minutes

$$\mathbf{x}^{(n)} = \mathbf{x}^{(n-1)}P$$

$$\mathbf{x}^{(n)} = \mathbf{x}^{(0)}P^n$$

La théorie montre qu'au bout d'un certain temps, la loi de probabilité est indépendante de la loi initiale. Notons la $\mathbf{q}$:

$$\mathbf{q} = \lim_{n\to\infty} \mathbf{x}^{(n)}$$

On obtient la convergence si et seulement si la chaine est apériodique et irréductible. C'est le cas dans notre exemple, on peut donc écrire :

$$\mathbf{q}(I - P) = \mathbf{0}$$

$$\begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} 0.1 & -0.05 & -0.05 \\ -0.7 & 1 & -0.3 \\ -0.8 & 0 & 0.8 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

Sachant que $q_1 + q_2 + q_3 = 1$, on obtient :

$$\begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} = \begin{bmatrix} 0.884 & 0.0442 & 0.0718 \end{bmatrix}$$

Doudou passe 88.4 % de son temps à dormir !

(source : http ://encyclopedie.snyke.com/articles/chaine_de_markov.html)