FACULTÉ DES SCIENCES APPLIQUÉES

MATH0488-1 ÉLÉMENTS DE PROCESSUS STOCHASTIQUES

Méthodes de Monte Carlo par chaînes de Markov pour la détection de communautés dans un graphe

---

Professeur :
Pierre GEURTS

Groupe :
Maxime FIRRINCIELI
Romain LAMBERMONT
Arthur LOUIS

4 avril 2022

Table des matières

Table des figures

Liste des tableaux

Introduction

Ce projet a pour but de détecter les communautés dans un graphe grâce aux chaînes de Markov et par les méthodes de Monte-Carlo. L'utilisation de ces outils nous permettra de détecter efficacement les communautés dans un graphe. Le projet sera constitué en deux parties. Une première se focalisera sur la familiarisation avec les chaînes de Markov et les méthodes de Monte-Carlo, plus précisement l'algorithme de Metropolis-Hastings, tandis que la seconde sera le cœur du projet avec la recherche de communautés dans un graphe.

1 Première partie : chaînes de Markov et algorithme MCMC

1.1 Chaînes de Markov

Rappelons d'abord brievement ce qu'est une chaîne de Markov. Une chaîne de Markov est un outil mathématique stochastique, qui utilise un principe de "non-mémoire". Tout état d'un système est simplement calculé à partir du précédent, ce qui en facilite l'analyse.

Ces chaînes sont simplement décrites mathématiquement comme suit avec $X_1, X_2, ..., X_t$ une suite de variables aléatoires qui définit une chaîne de Markov si (pour t $>$ 1) elle suit cette relation :

$$\mathbb{P}(X_1, X_2, ..., X_t) = \mathbb{P}(X_1) \prod_{l=2}^{t} \mathbb{P}(X_l |\ X_{l-1})$$

1.1.1

Nous calculons donc pour des valeurs de t croissantes les différentes valeurs demandées, ici avec $t = 20$ (suffit pour avoir convergé) :

— Cas de base distribué uniformément : $\mathbb{P}(X_t = x) = \begin{pmatrix} 0.3488 & 0.0698 & 0.2326 & 0.3488 \end{pmatrix}$ avec $x = 1, 2, 3, 4$

— Cas de base fixé : $\mathbb{P}(X_t = x) = \begin{pmatrix} 0.3488 & 0.0698 & 0.2326 & 0.3488 \end{pmatrix}$ avec $x = 1, 2, 3, 4$

— $Q^t = \begin{pmatrix} 0.3488 & 0.0698 & 0.2325 & 0.3488 \\ 0.3488 & 0.0698 & 0.2326 & 0.3488 \\ 0.3488 & 0.0698 & 0.2326 & 0.3488 \\ 0.3488 & 0.0698 & 0.2326 & 0.3488 \end{pmatrix}$

On remarque donc bien une convergence vers des probabilités et ceci peut importe le cas de départ, on peut montrer cette convergence sur les figures ci-dessous :

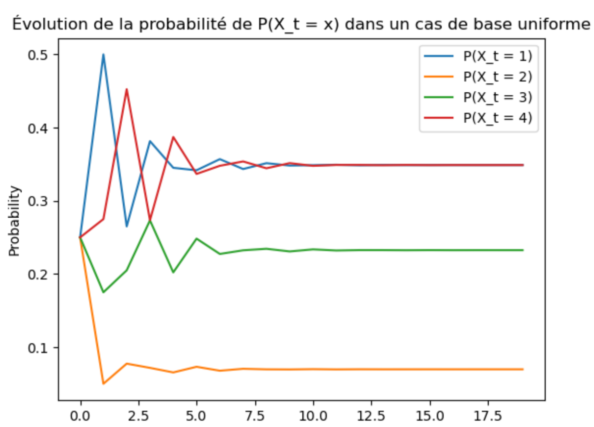

Figure 1 – Évolution des probabilités dans une distribution de départ uniforme

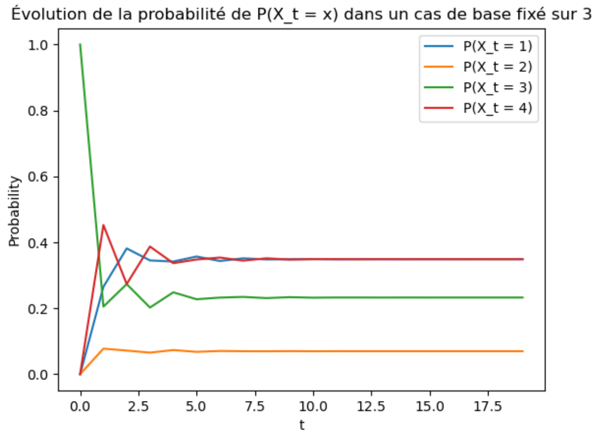

FIGURE 2 – Évolution des probabilités dans une distribution de départ fixée sur 3

1.1.2

Afin de déduire la distribution stationnaire π_∞ de notre chaîne qui est décrite comme suit :

$$[\pi_\infty]_j = \lim_{t\to\infty} \mathbb{P}(X_t = j)$$

Nous allons simplement calculer $\mathbb{P}(X_t)$ avec un grand t ce qui nous donne :

$$\pi_\infty = \begin{pmatrix} 0.3488 & 0.0698 & 0.2326 & 0.3488 \end{pmatrix}$$

1.1.3

Afin de vérifier les résultats obtenus, nous effectuons des réalisations de notre chaîne de Markov. Nous pouvons mettre en tableau la proportion de réalisation de chaque état lors de tests avec un nombre de pas croissant. Nous utilisons un point de départ distribué uniformément entre les 4 états car il a été prouvé plus haut que ça n'avait pas d'influence.

# Pas / État	1	2	3	4
100	0.35	0.05	0.21	0.39
1000	0.351	0.075	0.23	0.344
10000	0.346	0.064	0.236	0.354
100000	0.3501	0.0722	0.228	0.3497
1000000	0.34902	0.6757	0.23422	0.34919

TABLE 1 – Proportion des états obtenus en fonction de nombre de pas

En comparant ce tableau avec les résultats obtenus avec les valeurs précédement obtenus, on remarque bien que les proportions sont respectées et font sens, on converge bien vers les mêmes valeurs.

1.1.4

/!\ TODO /!\

1.2 Méthode MCMC : analyse théorique dans le cas fini

Nous allons maintenant nous attarder sur l'aspect théorique de l'algorithme de Metropolis-Hastings avant de passer à l'aspect pratique.

1.2.1

Soit π_0 une distribution initiale d'une chaîne de Markov invariante dans le temps et Q la matrice de transition :

$$\forall i, j \in \{1, ..., N\} : \pi_0(i)[Q]_{i,j} = \pi_0(j)[Q]_{j,i}$$

Si π_0 est une distribution stationnaire de notre chaîne de Markov alors, avec Q la matrice de transition :

$$[\pi_0 Q]_j = [\pi_0]_j \tag{1}$$

Partons du membre de gauche :

$$\begin{aligned}
[\pi_0 Q]_j &= \sum_{k=1}^{N} \pi_k Q_{k,j} && \text{(définition du produit scalaire)} \\
&= \sum_{k=1}^{N} \pi_j Q_{j,k} && \text{(par hypothèse des équations de balance détaillées)} \\
&= [\pi_0]_j \underbrace{\sum_{k=1}^{N} Q_{j,k}}_{=1} && \text{(définition de la matrice de transition)} \\
&= [\pi_0]_j && (\pi \text{est donc bien une distribution stationnaire})
\end{aligned}$$

On peut finalement conclure sur l'unicité de π_0. Pour que la distribution stationnaire soit unique, il faut que le chaîne de Markov soit irréductible, c'est-à-dire que si on représente notre chaîne selon un graphe avec un poids par arête représentant les probabilités, celui-ci doit être connexe.

1.2.2

1.3 Méthode MCMC : illustration sur un exemple simple

1.3.1

1.3.2

1.3.3

2 Deuxième partie : détection de communautés dans un graphe par algorithmes MCMC

2.1 Etude théorique

2.1.1

2.1.2

2.1.3

2.1.4

2.2 Analyse expérimentale

2.2.1

2.2.2

2.3 Application à un grand graphe