IV. Chaînes de Markov

I.1. Définition et classification

Marche aléatoire sur un graphe

Definition

Soit G un graphe orienté à n sommets dénotés $v_1, \ldots, v_n$ et dont les arêtes sont valuées par des poids p_{uv} tels que

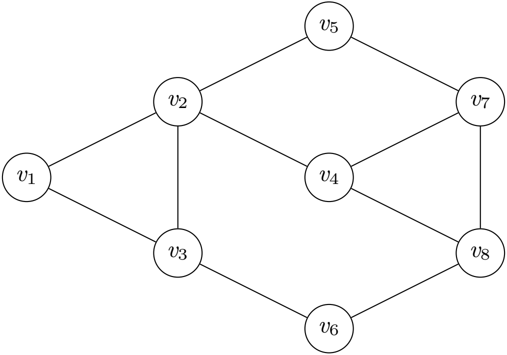
$$\forall u \in V(G), \quad \sum_{v \in N^+(u)} p_{uv} = 1$$

Une *marche aléatoire sans mémoire* sur G est une marche aléatoire définie de la façon suivante :

- on note x_0 la position initiale ;
- à l'instant i, la marche est en x_i et, pour tout $v \in N^+(x_i)$, $\mathbb{P}(x_{i+1} = v) = p_{x_i v}$.

En d'autres termes, p_{uv} est la probabilité d'emprunter l'arête (u, v) quand la marche est en u. Elle est dite sans mémoire car seule sa position à l'instant i détermine la position à l'instant $i + 1$.

Exemple



On considère une marche uniforme, c'est-à-dire que la marche choisit une des arêtes possibles avec équiprobabilité.

$$p_{v_2 v_4} = \frac{1}{4} \qquad \text{et} \qquad p_{v_4 v_2} = \frac{1}{3}$$

Marche aléatoire sur un graphe

- les points potentiellement visités sont ceux tels qu'il existe un chemin de probabilités positives depuis la source.
- on peut définir une notion de *popularité* d'un sommet x liée à la fréquence du passage de la marche en x.

Questions

- Quelle est la position 'moyenne' de la marche après un grand nombre de pas si on lance un grand nombre de marcheurs ? C'est la notion de loi de probabilité limite.
- Combien de temps faudra-t-il faire en moyenne avant de visiter un sommet donné ?
- Quelle est l'influence du choix du sommet de départ ?

Chaîne de Markov

De nombreuses questions peuvent s'écrire dans ce contexte. Il suffit pour cela que l'objet d'interêt soit une suite $(s_n)_{n\in\mathbb{N}}$ d'états tels que :

- le nombre d'états possibles $(v_1, \ldots, v_N)$ est fini ;
- il existe des probabilités de transition entre états $p_{ij} = \mathbb{P}(S_{n+1} = v_i | S_n = v_j)$ invariables dans le temps ;
- la chaîne est sans mémoire : soit S_n la variable aléatoire indiquant l'état dans lequel se trouve la chaîne après n étapes. Alors,

  $$\mathbb{P}(S_n = s_n | S_{n-1} = s_{n-1}, \ldots, S_0 = s_0) = \mathbb{P}(S_n = s_n | S_{n-1} = s_{n-1})$$

  Entre d'autres termes, la prochaine transition ne dépend que de l'état courant et pas du reste de l'histoire de la trajectoire suivie.

Definition

Un processus satisfaisant les contraintes ci-dessus est appelé *chaîne de Markov*.

Classification : état récurrent/transient

Soit v un état et p la probabilité, étant donné que le point de départ de la chaîne est v, de revenir en v.

- Si $p = 1$, v est un état récurrent.
- En particulier, si $p_{vv} = 1$, l'état est dit absorbant.
- Si $p < 1$, l'état est dit transient ou transitoire.
  Dans ce cas, le nombre de passages en v de la marche est presque sûrement fini et suit une loi géométrique de paramètre p.

Classification : état récurrent/état transient

- Une chaîne de Markov peut etre représentée par un graphe orienté G : (u, v) est une arête ssi $p_{uv} \neq 0$.
- Soit H un graphe ayant un sommet pour chaque composante fortement connexe de G et tel que $(u, v) \in E(H)$ s'il existe une arête de G allant de la composante correspondant à u à la composante correspondant à v. Alors le graphe H est acyclique. De plus, les états récurrents sont les états situés dans les composantes connexes dont le degré sortant dans H est nul.

Definition

Une chaîne de Markov est *irréductible* si le graphe associé est fortement connexe, ou autrement dit s'il existe un chemin entre toute paire d'états.

Classification : chaîne périodique

Définition

Un état x est un état *périodique* si le PGCD de l'ensemble $\{n|P^n(x,x)\}$ est supérieur à 1. Une chaîne est périodique si tous ses états sont périodiques. Sinon, elle est apériodique.

Remarque : En pratique, on construit toujours les chaînes de façon à ce qu'elles soient non périodiques.

I.2. Rappels d'algèbre linéaire

Valeurs et vecteurs propres

Definition

Soit A une matrice carrée de taille N. Soit X un vecteur de taille N et $\lambda \in \mathbb{R}$ tels que

$$AX = \lambda X$$

λ est une valeur propre de A et X un vecteur propre à droite associé.
Si

$$^{t}XA = \lambda\, ^{t}X$$

λ est une valeur propre de A et X un vecteur propre à gauche associé.

- Les valeurs propres sont les racines du polynôme $det(A - \lambda I)$.
- A et ^{t}A ont le mêmes valeurs propres.
- les vecteurs propres à gauche de A sont les vecteurs propres à droite de ^{t}A.

Matrices stochastiques

Definition

Une matrice P est une *matrice stochastique* si

1. $0 \leq p_{ij} \leq 1$ pour tout (i, j).
2. $\sum_j p_{ij} = 1$, pour tout i.

En particulier, la matrice des probabilités de transition d'une chaîne de Markov est une matrice stochastique.

Matrices stochastiques

Proposition

Soit P une matrice stochastique. Le vecteur $\begin{pmatrix} 1 \\ \cdots \\ 1 \end{pmatrix}$ *est un vecteur propre à droite associé à la valeur propre* 1 *pour* P*. De plus, toute autre valeur propre* λ *de* P *vérifie* $|\lambda| \leq 1$.

I.3. Convergence des chaînes de Markov

Distribution

- On note X_0 le vecteur de probabilités de la position initiale et X_n celui au bout de n pas (c'est-à-dire que X_n contient la distribution de la variable aléatoire S_n).
- X_n peut être vu comme la distribution de la position d'une trajectoire au bout de n pas ou comme la répartition d'un très grand nombre de trajectoires lancées an parallèle au bout de n pas.

Proposition

Soit P la matrice de transition de la chaîne de Markov. Pour tout n,

$$^{t}X_n =^{t} X_0 P^n$$

Distribution invariante

En passant à la limite dans l'égalité ${}^tX_{k+1} = {}^t X_kP$, on pressent que si la marche a une distribution limite quand le nombre de pas tend vers l'infini, cette distribution devra vérifier ${}^t\mu = {}^t \mu P$, c'est-à-dire être un vecteur propre à gauche associé à la valeur propre 1.

Definition

Une mesure invariante ${}^t\mu$ pour une chaîne de Markov de matrice de transition P est un vecteur vérifiant ${}^t\mu = {}^t \mu P$.

Questions

- une distribution limite existe-t-elle toujours ?
- est-elle unique ?
- la suite des distributions $(X_n)_n$ convergent-elle vers une telle mesure ?

Existence et unicité de la mesure invariante : Théorème de Perron-Frobenius

Théorème

Perron-Frobenius Soit P la matrice d'une chaîne de Markov **irréductible**. *Alors :*

1. 1 *est une valeur propre simple.*
2. *tout vecteur propre à gauche associé à 1 a toutes ses coordonnées de même signe. En particulier, celui de somme 1 correspond bien à une distribution de probabilités.*
3. *si la chaîne est apériodique, toute autre valeur propre λ vérifie $|\lambda| < 1$.*

En d'autres termes, toute chaîne de Markov irréductible admet une unique mesure invariante.

- Si la chaîne n'est pas irréductible, la partie concernant la monotonie du signe du vecteur propre est encore valable. Par contre, l'espace propre peut être de dimension supérieure : il n'y a plus unicité de la mesure invariante.

Convergence vers la mesure invariante

Théorème

Soit P la matrice d'une chaîne de Markov **irréductible et apériodique** *et μ l'unique mesure invariante associée. Alors, pour tout X_0, $\lim_{n\to+\infty}^t X_0 P^n =^t \mu$. De plus, la vitesse de convergence est en $|\lambda_2|^n$, où λ_2 est la valeur propre de valeur absolue maximale parmi les valeurs propres différentes de 1.*

En résumé

Cas possibles

- Si la chaîne est irréductible apériodique, la distribution de S_n tend vers l'unique mesure invariante, et ce quelle que soit la distribution de départ.
- Si la chaîne est irréductible mais périodique, la mesure invariante est unique mais suivant la distribution de départ, il peut ne pas y avoir convergence
- Si la chaîne n'est pas irréductible, il y a plusieurs mesures invariantes. Si elle est apériodique, il y aura bien convergence de la distribution, mais la limit dépend de la distribution de départ.

En pratique

Quand c'est possible, on considère des chaînes irréductibles apériodiques.
Exemple : PageRank et les algorithmes d'optimisation MCMC.

I.4. Un exemple : PageRank

Graphes de états

- On considère le graphe du web, les arêtes étant les pages référencées (dont un attribut est le texte qu'elles contiennent)
- On ajoute une arête orientée de i vers j si la page i pointe vers la page j.
- On considéère une marche aléatoire uniforme, c'est-à-dire que $p_{ij} = \frac{1}{d(i)}$ si l'arête (i, j) est présente.
- La popularité d'une page est définie par la mesure limite de la chaîne de Markov.

Graphes de états

- On considère le graphe du web, les arêtes étant les pages référencées (dont un attribut est le texte qu'elles contiennent)
- On ajoute une arête orientée de i vers j si la page i pointe vers la page j.
- On considéère une marche aléatoire uniforme, c'est-à-dire que $p_{ij} = \frac{1}{d(i)}$ si l'arête (i, j) est présente.
- La popularité d'une page est définie par la mesure limite de la chaîne de Markov.

Problème

Le graphe n'est pas fortement connexe : la chaîne n'est pas irréductible.

Chaîne de Markov modifiée

- n le nombre de sommets, P la matrice de transition définie précédemment, E la matrice dont tous les coefficients valent 1. On définit

$$A = \frac{1-d}{n}E + \frac{d}{n}P$$

- La chaîne définie par A a une probabilité non nulle d'aller de tout état à tout autre état : elle est irréductible apériodique.
- La deuxième valeur propre est de valeur absolue $\leq d$: la covergence se fait à la vitesse d^n.

En pratique

- d est choisi aux environs de 0.85.
- Simuler des marches est aisé car la multiplication par E est une somme et P est creuse
- La convergence des marches est rapide ($0.85^{100} < 1e-7$).
- Il suffit de classer les pages (à refaire de temps en temps pour se mettre à jour) puis de trier celle contenant la bonne chaîne de caractères.
- En réalité, seules les K premières du classement contenant la bonne chaîne sont gardées, puis d'autres sont recrutées par similarité + beaucoup d'autres subtilités publiques ou non !

III. Applications des chaînes de Markov

III.1 Partie publique de PageRank

Graphe considéré

- les noeuds sont les URLs
- une arête relie deux URLs si la première comporte un hyperlien vers la seconde

Principe

La distribution limite peut être considérée comme un classement des URLs.

Application du théorème principal des chaînes de Markov

La chaîne est-elle apériodique ? Oui

La chaîne est-elle irréductible ? Non.

Solution

Soit P la matrice d'adjacence du graphe, N le nombre d'URLs et Q la matrice de même taille dont tous les coefficients valent $\frac{1}{N}$.
Soit

$$P_\beta = (1 - \beta)P + \beta Q$$

Pour tout $\beta > 0$, la chaîne est irréductible et apériodique.

Application du théorème principal des chaînes de Markov

- Il suffit donc de choisir β faible (en pratique 0.15) et de déterminer la distribution limite.

  Problème Déterminer un vecteur propre sur un matrice de millions de noeuds est trop long.

  Solution Utiliser ${}^tX_n = {}^tX_0 P_\beta$ et que la convergence est exponentielle en $|\lambda_2|$ pour approximer la distribution limite par X_{50}.

- Il faut cependant calculer des produits $matrice \times vecteur$ pour des tailles de plusieurs millions. Ceci est possible car, si on pose X_0 le vecteur dont toutes les coordonnées valent $\frac{1}{N}$,

$$ {}^tX_n P_\beta = (1-\beta)\,{}^tX_n P + \beta\,{}^tX_0 $$

  Le premier produit peut être effectué efficacement car la matrice P est creuse.