Chaînes de Markov Cachées

Faculté de Sciences et d'Ingénierie

Université Paul Sabatier

par

Chouman Ahmad et Haddadi Reda

Encadré par Berthet Philippe

Mai 2024

Remerciements

Nous tenions à remercier notre encadrant de projet M. Phillipe Berthet pour son accompagnement tout au long de la réalisation de ce projet.

Contents

Introduction

L'étude mathématique d'un phénomène se fait, à la différence de certaines sciences, par calcul. Pour certaines propositions, la démonstration et le calcul suffisent comme preuve (par exemple :

$$\forall x \in \mathbb{R}, x^2 \geq 0$$

est une assertion qui n'a pas besoin d'être observée, on n'essaye pas de mettre certains nombres au carré pour vérifier qu'ils sont bien positifs).

En probabilités, nous étudions le comportement du hasard, et le hasard n'est pas fixe et est imprévisible, malgré de fortes probabilités en faveur d'un certain événement. Donc il y a besoin d'observations pour voir son comportement.

Une chaîne de Markov est un modèle utilisé pour décrire un processus stochastique où l'évolution du système dépend uniquement de son état actuel et non des états précédents. Cela signifie qu'une chaine de Markov, à premier abord, n'est pas déterministe, donc que l'évolution à long terme du système ne dépend pas des "premiers pas".
Bien sûr, l'étude de l'évolution à long terme du système elle même est intéressante, car cela peut nous montrer le comportement du hasard en comparaison avec l'étude théorique par calculs, et nous montrer certaines nuances qu'il pourrait y avoir.

Ce projet a donc pour but d'étudier les différents aspects et applications des CMC à travers l'exemple d'un voilier qui navigue selon des vecteurs qui dépendent de la météo. Il sera intéressant d'étudier par exemple, à travers des simulations, la position du voilier sur le plan pour un temps donné ou même le temps nécessaire pour que le voilier soit dans une zone donnée avec un taux d'erreur faible.

Un fichier Jupyter Notebook est mis à disposition afin de pouvoir visualiser toutes les simulations nécessaires à une compréhension plus profonde des concepts étudiés.

1 Chaînes de Markov et propriétés

1.1 Définitions et propriétés des chaînes de Markov homogènes

Ici, E est un ensemble fini d'états, qu'on appelle espace d'états, et son cardinal est supérieur ou égal à 2.

Définition

Une chaîne de Markov est un processus stochastique à temps discret, un ensemble d'états avec un ensemble de transitions d'une certaine probabilité,avec un certain état initial i et à chaque temps n, on passe de l'état i à un autre état j avec une probabilité de transition $i - j$ $(i, j \in E)$. Une chaîne de Markov peut aussi être vu comme *une suite de variable aléatoires.*

Définition bis

Soient X_1,...,X_n une suite de variables aléatoires sur un espace probabilisé (Ω, $\mathbb{P}$). On dit que X_n est une chaîne de Markov si la probabilité de transition entre un état et un autre ne dépend pas du "passé". Autrement dit,

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

(Propriété de Markov)

On associe à la chaîne de Markov à une *matrice de transition*, que l'on note généralement P ou Q (ici, on l'appellera P), telle que $P = P(x, y), x, y \in E$ et $P(x, y)$ sont les coéfficients de la matrice, qui est une matrice carrée de taille k, avec k le nombre d'états possible. On dit aussi que la matrice de transition est *stochastique*, c'est à dire que :

- $\forall x, y \in$E,P(x,y)$\geqslant$ 0
- $\forall x \in$E, $\sum_{y=1}^{k} P(x, y)$

On appelle *trajectoire* une certaine réalisation d'une chaîne de Markov.

Exemple 1 : la chaîne à deux états

Si on considère $E = \{0, 1\}$, donc deux états 0 et 1 correspondant au beau temps et à la pluie, avec une probabilité de passer de 0 à 1 égale à la probabilité de passer de 1 à 0, c'est à dire une probabilité de $\frac{1}{2}$(la probabilité d'être en 0 (*resp en 1*) et de rester en 0 (*resp en 1*) est également de $\frac{1}{2}$).

Alors on peut définir X_n comme la chaîne de Markov liée à la météo, avec X_k l'état courant au jour k. De plus, la matrice de transition associée à cette chaîne de Markov est :

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Proposition

Si X_n est une chaîne de Markov sur E, avec $x, y \in E$ et $n \in \mathbb{N}$, et si P est la matrice de transition de X_n, alors $P^n(x, y)$ est la probabilité de passer de l'état x à l'état y en un temps n.

Démonstration :
Par la définition de multiplication de matrices, nous avons :

$$P^n(x, y) = \sum_{k \in E} P^{n-1}(x, k) \cdot P(k, y)$$

Cela signifie que la probabilité de passer de l'état x à l'état y en n transitions est la somme des probabilités de passer de l'état x à tous les états k en $n-1$ transitions, multipliée par la probabilité de passer de l'état k à l'état j en une transition supplémentaire. Mais par la définition de la matrice de transition P, $P(x, y)$ est la probabilité de passer de l'état k à l'état j en une seule transition.
Ainsi, la somme des termes $P^{n-1}(x, k) \cdot P(k, y)$ représente exactement la probabilité de passer de l'état x à l'état y en n transitions.
Par conséquent, $P^n(x, y)$ est la probabilité de passer de l'état x à l'état y en n transitions.

$\square$

Exemple :

Toujours dans la continuité de l'exemple du voilier et avec la même matrice de transition

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Nous voulons trouver la probabilité qu'il fasse beau le 3^{e} jour sachant qu'il a plu le 1er jour.

Nous cherchons $P^2(1, 0)$, où P^2 est la matrice de transition après 2 étapes (car nous passons du 1er jour au 3^{e} jour).
Calculons P^2 :

$$P^2 = P \times P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \times \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \left(\frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{2} \times \frac{1}{2}\right) & \left(\frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{2} \times \frac{1}{2}\right) \\ \left(\frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{2} \times \frac{1}{2}\right) & \left(\frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{2} \times \frac{1}{2}\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

On observe que $P^2 = P$. Cela signifie que la matrice de transition est la même après chaque étape, ce qui est typique pour une chaîne de Markov avec des transitions équiprobables.
Ainsi,

$$P^2(1, 0) = \frac{1}{2}$$

La probabilité qu'il fasse beau le 3^{e} jour sachant qu'il a plu le 1er jour est donc $\frac{1}{2}$.

(Voir Jupyter pour visualiser l'évolution du voilier sur le plan)

1.2 Récurrence, Classes, Périodicité, Irréductibilité

Dans cette section, X_n est une chaîne de Markov de matrice de transition P.

Définition

On dit que x *conduit à* y si on peut arriver de l'état x à l'état y en un nombre fini de transitions. Autrement dit, s'il existe $n \in \mathbb{N}$ tel que :

$$P^n(x, y) > 0$$

Définition

Si x conduit à y, et que y conduit à x, on dit que x et y *communiquent.*

Définition

Si E est de cardinal n, alors une *classe de communication* de la chaîne de Markov est un sous ensemble E' de E tel que tous les états de E' communiquent.

Remarque : La relation "communiquer avec" est une relation d'equivalence. En effet, si $x, y, z \in E$:

- x communique avec x
- Si x communique avec y, alors y communique avec x
- si x communique avec y, et que y communique avec z, alors x communique avec z

Les classes d'equivalences de cette relation sont les classes de communications

Définition

On dit que X_n est *irréductible* si tous les états communiquent.
Autrement dit, si la seule classe de communication de E est E lui-même.

Définition

On dit qu'un état $x \in E$ est *récurrent* ou *persistant* si en partant de x, on peut retourner en x en un nombre fini de transitions.
Dans le cas contraire, on dit que x est un état est transitoire.

Remarques :

- *Les états d'une même classe sont soit tous récurrents, soit tous transitoires.*
  On parlera alors de classes récurrentes et de classes transitoires.
- *Si X_n est irréductible et que tous ses états sont récurrents, alors on dit que X_n est réccurente*

Définition

X_n est dite *fortement irréductible* si :

$$\exists n \in \mathbb{N}, \forall x, y \in E, P^n(x, y) > 0$$

(Autrement dit, il existe une puissance de la matrice de transition qui a tous ses coéfficients positifs)

Définition

On appelle la *période* d'un état x le nombre suivant :

$$p = pgcd\{n \geqslant 1 | P^n(x, x) > 0\}$$

Remarque :
Si P(x,x) > 0, alors x est de période 1, puisque pgcd{1,n $\in \mathbb{N}$} = 1

Proposition

La périodicité est une période de classe. En effet, tous les états d'une même classe ont la même période *Corrolaire : Si X_n est irréductible, on parle alors de périodicité de classe*

Définition

Si X_n a une période égale à 1, on dit que X_n est *apériodique*.

Définition

Pour la suite, on définira la variable aléatoire $R_n(x)$ pour tout $n \geqslant 1$, qui est l'instant du retour en un certain état $x \in E$ apres n étapes. Autrement dit :

$$R_n(x) = min\{i > 0, X_i = x\}$$

Définition

On dit qu'un état $x \in E$ est *récurrent positif* si l'esperance de son instant de premier retour en x est fini. Autrement dit, si :

$$\mathbb{E}(R_1(x)) \in \mathbb{N}$$

Définition

Une mesure (probabilité) est dite strictement positive si tous ses coéfficients sont strictement positifs.

Proposition

On suppose que X_n est irréductible, alors les assertions suivantes sont équivalentes :

1. X_n a une unique probabilité invariante
2. X_n a un état récurrent positif
3. Tous les états de X_n sont récurrents positifs

De plus, on peut définir la probabilité invariante $\pi(x) = \mathbb{E}(R_1(x))^{-1}$
Démonstration :

- $3 \Rightarrow 2$: *trivial*
- $2 \Rightarrow 1$: Commençons par définir $m_x(y)$ comme le nombre moyen de visites de l'état y avant R_1.
  Soit x un état récurrent positif. On a que $\mathbb{E}(R_1(x)) = \sum_{y \in E} m_x(y)$. Cette quantité est finie, donc on définit π par:

  $$\pi(y) = \frac{m_x(y)}{\mathbb{E}(R_1(x))}, y \in E$$

  Or la moyenne qu'on a définie est strictement positive par hypothèse. Donc π est une mesure strictement positive. De plus, π est bien une probabilité puisque :

  $$\sum_{y \in E} \pi(y) = \frac{\sum_{y \in E} m_x(y)}{\mathbb{E}(R_1(x))} = \frac{\mathbb{E}(R_1(x))}{\mathbb{E}(R_1(x))} = 1$$

  π est unique car la moyenne qu'on a définit l'est à constante près.
  Remarquons que $m_x(x) = 1$. Ainsi,

  $$\frac{1}{\pi(y)} = \frac{\mathbb{E}(R_1(x))}{m_x(y)} = \mathbb{E}(R_1(x))$$

  Donc si un état est récurrent positif, X_n a une unique probabilité invariante.
- $1 \Rightarrow 3$: *Soit*π une probabilité invariante. C'est une probabilité donc la somme des coefficients est 1. Donc $\exists y_0 \in E, \pi(y_0) > 0$ De plus par irréductibilité, $\exists n > 0, P^n(y_0, x) > 0$

  Or π est invariante donc $\forall n > 0$:

  $$\pi(x) = \sum_{y \in E} \pi(y) P^n(y, x) > \pi(y_0) P^n(y_0, x) > 0.$$

  On pose pour $y \in E$, $\beta(y) = \frac{\pi(y)}{\pi(x)}$. On a $\beta(x) = 1$ et comme π est une probabilité invariante alors β est une mesure invariante. Par l'irréductibilité, si $y \in E$, $\exists n_y \geqslant 0$ tel que $P^{n_y}(x, y) > 0$. Or β est invariante pour P, alors elle est aussi invariante pour tout P^n. On a donc pour tout $y \in E$:

  $$\beta(y) = \sum_{z \in E} \beta(z) P^{n_y}(z, y) > \beta(x) P^{n_y}(x, y) = P^{n_y}(x, y) > 0$$

et donc β est strictement positive. Ainsi, a $\beta = m_x$. (la moyenne qu'on a définit ci-dessus).
On a alors :

$$\mathbb{E}_x[S_1(x)] = \sum_{y \in E} \gamma_x(y) = \sum_{y \in E} \beta(y) = \sum_{y \in E} \frac{\pi(y)}{\pi(x)} = \frac{1}{\pi(x)} < \infty$$

donc x est récurrent positif. Or on a choisit x de façon arbitraire. Donc on en déduit que l'assertion 3 est vraie.

1.3 Probabilité invariante

Définition

Soit X_n une chaîne de Markov de n états, de matrice de transition P (donc P de taille $n * n$). Soit,π une probabilité (c'est à dire que π peut etre vu comme un vecteur ligne de taille n, et que $\sum_{x \in E} \pi(x) = 1$, avec $\pi(x) \geqslant$ pour tout x) On dit que π est une *probabilité invariante* ou *mesure invariante* si :

$$\pi P = \pi$$

$\Leftrightarrow \forall y \in E, \pi(y) = \sum_{x \in E} \pi(x) P(x, y)$

Définition

Soit π une probabilité d'une chaîne de Markov X_n. On dit que π est *réversible* si :

$$\forall x, y \in E, \pi(x) P(x, y) = \pi(y) P(y, x)$$

Corrolaire

Si π est réversible, alors π est invariante

Proposition : Unicité de la probabilité invariante

Si X_n est irréductible et réccurente, alors elle a une unique probabilité invariante. De plus,

$$lim_{n \to \infty} \frac{1}{n} \sum_{k=0} n \pi_0 P^k = \pi$$

Démonstration :
Pour démontrer l'unicité, supposons qu'il existe deux distributions invariantes π et π'. Nous allons montrer qu'elles doivent être égales. Puisque π et π' sont des probabilités invariantes, nous avons :

$$\pi P = \pi \quad et \quad \pi' P = \pi'$$

Considérons la différence $\delta = \pi - \pi'$. Notez que $\delta P = \delta$ parce que :

$$\delta P = (\pi - \pi') P = \pi P - \pi' P = \pi - \pi' = \delta$$

De plus, comme π et π' sont des distributions de probabilité, nous avons :

$$\sum_{x \in E} \pi(x) = 1 \quad et \quad \sum_{x \in E} \pi'(x) = 1$$

donc,

$$\sum_{x \in E} \delta(x) = \sum_{x \in E} (\pi(x) - \pi'(x)) = 0$$

Puisque δ est invariante sous l'action de P et que la somme de ses composantes est nulle, nous allons montrer que δ doit être le vecteur nul. Supposons par l'absurde que $\delta \neq 0$. Cela signifie qu'il existe au moins un état $i \in E$ tel que $\delta(i) \neq 0$. Puisque X_n est *irréductible, il existe un entier k tel que $P^k(i,j) > 0$ pour tout $j \in E$. Considérons la somme :*

$$\sum_{j\in E} \delta(j) = \sum_{j\in E} \sum_{i\in E} \delta(i) P^k(i,j) = \sum_{i\in E} \delta(i) \sum_{j\in E} P^k(i,j) = \sum_{i\in E} \delta(i) = 0$$

Cependant, comme $\delta(i) \neq 0$ et $P^k(i,j) > 0$ pour tout j, cela implique que $\sum_{j\in E} \delta(j) \neq 0$, ce qui est une contradiction. Ainsi, δ doit être le vecteur nul, ce qui implique que $\pi = \pi'$. Donc, la probabilité invariante est unique.

(Voir Jupyter pour visualiser le calcul et l'utilité de la mesure invariante)

2 Théorèmes de convergence

Le but de cette partie est d'établir une loi des grands nombres et un théorème central limite pour les chaînes de Markov.

2.1 Théorèmes ergodiques

Théorème

Le théorème ergodique s'applique à une chaîne de Markov irréductible et apériodique avec une probabilité invariante unique π. Il affirme que, pour toute fonction test intégrable f et pour toute condition initiale $X_0 = x$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \sum_{y\in E} \pi(y) f(y) \quad \text{presque sûrement.}$$

Autrement dit, la moyenne des valeurs de $f(X_k)$ sur une longue période de temps converge vers l'espérance de f par rapport à la mesure invariante π.

Démonstration : Soit X_n une chaîne de Markov irréductible et apériodique avec matrice de transition P et une probabilité invariante unique π. Soit f une fonction test intégrable.
Étape 1: Moyenne temporelle
Considérons la moyenne temporelle de f le long de la trajectoire de la chaîne de Markov :

$$A_n = \frac{1}{n} \sum_{k=0}^{n-1} f(X_k)$$

Nous devons montrer que cette moyenne converge vers l'espérance de f par rapport à π, c'est-à-dire :

$$\lim_{n\to\infty} A_n = \sum_{y\in E} \pi(y) f(y)$$

Étape 2: Convergence en distribution
Puisque X_n est irréductible et apériodique, pour toute fonction g intégrable, la distribution de X_n converge vers π lorsque $n \to \infty$. En particulier,

$$\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \sum_{y\in E} \pi(y) g(y)$$

En appliquant cela à f, nous avons :

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \sum_{y\in E} \pi(y) f(y)$$

Étape 3: Loi des grands nombres
Utilisons la loi des grands nombres pour les chaînes de Markov. Puisque X_n est irréductible et apériodique, pour toute fonction f intégrable,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \mathbb{E}_\pi[f(X)] \quad \textit{presque sûrement}$$

où $\mathbb{E}_\pi$ est l'espérance par rapport à la mesure invariante π. Ainsi,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \sum_{y\in E} \pi(y) f(y) \quad \textit{presque sûrement}$$

Nous avons montré que la moyenne temporelle de $f(X_k)$ converge vers l'espérance de f par rapport à la mesure invariante π. C'est exactement ce que le théorème ergodique affirme.
(Voir Jupyter pour visualiser la simulation du théorème ergodique)

2.2 Théorème central limite

Théorème (Théorème central limite des statistiques)

Le théorème central limite (TCL) des statistiques est un résultat fondamental en théorie des probabilités. Il affirme que, pour une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) $\{X_i\}$ avec une espérance μ et une variance finie $\sigma^2 > 0$:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

La distribution de la variable aléatoire normalisée

$$Z_n = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$$

converge en distribution vers une loi normale standard $\mathcal{N}(0,1)$ lorsque $n \to \infty$.

Démonstration : La démonstration du TCL des statistiques repose sur la méthode des moments ou la méthode de la transformée de Fourier (fonction caractéristique). Voici une esquisse de la démonstration utilisant la méthode des moments. Soit $\{X_i\}$ une suite de variables aléatoires i.i.d. avec espérance $\mathbb{E}[X_i] = \mu$ et variance $Var(X_i) = \sigma^2$.

Étape 1: Espérance et Variance de $\bar{X}_n$

Calculons l'espérance et la variance de la moyenne empirique $\bar{X}_n$:

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] = \mu$$

$$Var(\bar{X}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$$

Étape 2: Normalisation Considérons la variable normalisée :

$$Z_n = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$$

Alors, $\mathbb{E}[Z_n] = 0$ et $Var(Z_n) = 1$

Étape 3: Convergence en distribution

En utilisant la méthode de la transformée de Fourier, montrons que la fonction caractéristique de Z_n converge vers celle d'une loi normale standard. La fonction caractéristique de Z_n est :

$$\varphi_{Z_n}(t) = \mathbb{E}[e^{itZ_n}] = \mathbb{E}\left[e^{it\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)}\right]$$

En utilisant le théorème de Lévy, on montre que $\varphi_{Z_n}(t) \to e^{-\frac{t^2}{2}}$, qui est la fonction caractéristique de $\mathcal{N}(0,1)$. Donc, par le théorème de Lévy, $Z_n \xrightarrow{d} \mathcal{N}(0,1)$.

2.3 Application des théorèmes de convergence aux chaînes de Markov

Théorème : (Théorème central limite pour les chaînes de Markov)

Le théorème central limite pour les chaînes de Markov s'applique à une chaîne de Markov irréductible et apériodique avec une probabilité invariante unique π. Si f est une fonction intégrable, avec la condition initiale de la chaîne de Markov $X_0 = x$, la somme normalisée :

$$S_n = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} (f(X_k) - \pi(f))$$

converge en distribution vers une loi normale $\mathcal{N}(0, \sigma_f^2)$, où $\pi(f) = \sum_{y \in E} \pi(y) f(y)$ et σ_f^2 est la variance asymptotique de f sous π.

Démonstration : Soit X_n une chaîne de Markov irréductible et apériodique avec matrice de transition P et une probabilité invariante unique π. Soit f une fonction test intégrable.

Étape 1: Théorème Ergodique

D'abord, par le théorème ergodique, nous savons que :

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \pi(f) \quad \textit{presque sûrement}$$

où $\pi(f) = \sum_{y \in E} \pi(y) f(y)$.

Étape 2: Centralisation

Considérons la somme centrée et normalisée :

$$S_n = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} (f(X_k) - \pi(f))$$

Par le théorème central limite classique, pour une suite de variables aléatoires i.i.d., nous avons vu que la somme centrée et normalisée converge en distribution vers une loi normale. Cependant, pour une chaîne de Markov, nous devons prendre en compte la dépendance temporelle.

Étape 3: Covariance et Variance Asymptotique

La variance asymptotique σ_f^2 est définie comme :

$$\sigma_f^2 = \lim_{n \to \infty} Var\left(\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} (f(X_k) - \pi(f)) \right)$$

En utilisant les propriétés de la chaîne de Markov et l'invariance de π, on peut montrer que σ_f^2 existe et est finie.

Ainsi, par un résultat avancé en théorie des processus stochastiques, nous obtenons que :

$$S_n = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} (f(X_k) - \pi(f)) \xrightarrow{d} \mathcal{N}(0, \sigma_f^2)$$

C'est le théorème central limite pour les chaînes de Markov.

(Voir Jupyter pour visualiser la simulation du théorème central limite)

3 Simulation des chaînes de Markov

Dans cette section, on se donnera pour but de vérifier numériquement le théorème ergodique et le théorème central limite à l'aide de simulations.

3.1 Mesure stationnaire

Définition

Soit X_n une chaîne de Markov irréductible à valeurs dans E, de loi initiale π_0 et de matrice de transition P.

A l'aide de la méthode d'inversion,nous savons simuler des variables aléatoires suivant une loi discrète quand elles sont i.i.d (indépendantes et identiquement distribuées). Cependant, X_n est une chaîne de Markov., donc il n'y a donc pas d'indépendance entre les X_i. On connaît la loi de X_0 (la loi initiale π_0), mais on ne connaît pas la loi de X_n pour $n > 0$. Par contre, on connaît la loi de X_{n+1}) sachant X_n grâce à P. Pour simuler $(X_0, \ldots, X_k)$ avec $k \geq 1$, on procède comme suit :

1. On simule $X_0 = x_0$
2. Pour $n = 1, \ldots, k$, on $X_n = x_n$ avec la valeur de X_{n-1}

La chaîne est supposée irréductible donc récurrente positive, ce qui implique qu'elle admet une unique probabilité invariante que nous notons π. Par définition, elle vérifie $\pi P = \pi$ et $\sum_{k=1}^{m} \pi(k) = 1$. Elle satisfait donc :

$$\pi P = \pi \quad \text{et} \quad \sum_{k=1}^{m} \pi(k) = 1$$

i.e

$$P^T \pi^T - \pi^T = 0 \quad \text{et} \quad (1, \ldots, 1)\pi^T = 1$$

Nous avons là un système à résoudre, et il peut être résolu grâce à la fonction `invariant.measure()` que nous avons codé en Python. Cette fonction calcule la loi stationnaire π d'une chaîne de Markov homogène irréductible dans un espace d'états fini avec une matrice de transition P de la manière suivante :

- Nous stockons la taille m de la matrice P ;
- Nous créons la matrice $(P^T - I_m, e)$ et le vecteur $(0_m, 1)$;
- Nous déterminons π en résolvant le système $(P^T - I_m, e)\pi^T = (0_m, 1)$;
- Nous retournons π.

3.2 Estimateur de matrice de transition

Définition

Soit $\{X_n\}$ une chaîne de Markov à temps discret prenant des valeurs dans un ensemble fini d'états $E = \{1, 2, \ldots, m\}$. La matrice de transition P est une matrice $m \times m$ où l'élément $P(i, j)$ représente la probabilité de transition de l'état i à l'état j en une seule étape.

Un estimateur $\hat{P}$ de la matrice de transition P peut être construit à partir d'une séquence d'observations $\{X_0, X_1, \ldots, X_N\}$. L'estimateur empirique de P est défini comme suit :

$$\hat{P}(i, j) = \frac{N(i, j)}{N(i)}$$

où :

- $N(i, j)$ est le nombre de transitions observées de l'état i à l'état j.
- $N(i)$ est le nombre de fois où l'état i a été observé.

Démonstration : Montrons que $\hat{P}(i, j)$ est un estimateur consistant pour $P(i, j)$. Soit $N(i) = \sum_{k=0}^{N-1} \mathbf{1}_{\{X_k=i\}}$ le nombre de fois où l'état i a été observé, et soit $N(i, j) = \sum_{k=0}^{N-1} \mathbf{1}_{\{X_k=i, X_{k+1}=j\}}$ le nombre de transitions observées de l'état i à l'état j.
L'estimateur empirique de la probabilité de transition de i à j est donné par :

$$\hat{P}(i, j) = \frac{N(i, j)}{N(i)}$$

Montrons que $\hat{P}(i, j)$ est un estimateur consistant, c'est-à-dire que $\hat{P}(i, j) \to P(i, j)$ en probabilité lorsque $N \to \infty$.
Pour cela, remarquons que $N(i, j)$ suit une loi binomiale Binom$(N(i), P(i, j))$. Par la loi des grands nombres, nous avons :

$$\frac{N(i)}{N} \to \pi(i) \quad et \quad \frac{N(i, j)}{N(i)} \to P(i, j)$$

où $\pi(i)$ est la probabilité invariante de l'état i.
Puisque $\frac{N(i,j)}{N(i)} \to P(i, j)$ en probabilité, nous obtenons :

$$\hat{P}(i, j) = \frac{N(i, j)}{N(i)} \to P(i, j) \quad en\ probabilité\ lorsque \quad N \to \infty$$

Ainsi, $\hat{P}(i, j)$ est un estimateur consistant pour $P(i, j)$. j) .

4 Chaînes de Markov cachées

Dans la suite, nous ne démontrerons plus de propositions ou de théorèmes, mais nous donnerons plûtot des propriétés des chaînes de Markov cachées, de la modélisation de la méta-météo et de la prédiction d'erreur.

4.1 Définition et propriétés

Définition

Une *chaîne de Markov cachée* (dit aussi CMC) est un modèle statistique où le système modélisé est supposé être un processus de Markov avec des états cachés. Autrement dit, contrairement aux chaînes de Markov simples où les états sont directement observables, dans les CMC, les états sont cachés et ne peuvent être observés directement. À la place, on observe une séquence de symboles ou d'observations, chacun d'eux étant généré par un des états cachés selon une certaine distribution de probabilité.

Formellement, une CMC est définie par :

- Un ensemble fini d'états cachés $E = \{X_1, X_2, \ldots, X_N\}$.
- Une matrice de transition des états cachés $A = [a_{ij}]$ où a_{ij} est la probabilité de transition de l'état X_i à l'état X_j, soit $a_{ij} = \mathbb{P}(X_{t+1} = X_j \mid X_t = X_i)$.
- Un ensemble fini d'observations $V = \{v_1, v_2, \ldots, v_M\}$.
- Une matrice de distribution d'observations $Q = [b_j(k)]$ où $b_j(k)$ est la probabilité d'observer le symbole v_k à partir de l'état X_j, soit $b_j(k) = \mathbb{P}(O_t = v_k \mid X_t = X_j)$.
- Un vecteur de distribution initiale $\pi = [\pi_i]$ où π_i est la probabilité que l'état initial soit X_i, soit $\pi_i = \mathbb{P}(X_1 = X_i)$.

Quelques propriétés des CMC

- Les transitions entre les états cachés suivent une chaîne de Markov. Ainsi, la probabilité de transition de l'état X_i à l'état X_j ne dépend que de l'état actuel X_i et non des états précédents.:

$$a_{ij} = \mathbb{P}(X_{t+1} = X_j \mid X_t = X_i)$$

- Chaque état caché génère une observation selon une distribution de probabilité spécifique à cet état. La probabilité d'observer le symbole v_k lorsque le système est dans l'état X_j est donnée par :

$$b_j(k) = \mathbb{P}(O_t = v_k \mid X_t = X_j)$$

- La distribution initiale des états cachés est donnée par le vecteur π. La probabilité que le système soit dans l'état S_i au début est :

$$\pi_i = \mathbb{P}(X_1 = X_i)$$

- La probabilité conjointe d'une séquence d'états cachés $S = \{X_1, X_2, \ldots, X_T\}$ et d'une séquence d'observations $O = \{O_1, O_2, \ldots, O_T\}$ est donnée par :

$$\mathbb{P}(S, O) = \pi_{X_1} b_{X_1}(O_1) \prod_{t=2}^{T} a_{X_{t-1}, X_t} b_{X_t}(O_t)$$

- Algorithme de Viterbi L'algorithme de Viterbi est utilisé pour trouver la séquence d'états cachés la plus probable donnée une séquence d'observations. Il utilise une approche dynamique pour maximiser la probabilité conjointe des observations et des états cachés.

- Probabilité d'une séquence d'observations La probabilité d'une séquence d'observations $O = \{O_1, O_2, \ldots, O_T\}$ donnée les paramètres du modèle $\lambda = (A, B, \pi)$ est calculée en utilisant **l'algorithme forward** (cf prédiction d'erreurs) :

$$\mathbb{P}(O \mid \lambda) = \sum_{S} \mathbb{P}(E, O \mid \lambda)$$

4.2 Modélisation de la méta-météo

Définition-proposition

La modélisation de la *méta-météo* utilise une chaîne de Markov cachée (CMC) pour représenter l'évolution stochastique des conditions météorologiques. Dans cette approche, les états cachés représentent différents régimes météorologiques, tandis que les observations correspondent à des variables météorologiques mesurables telles que la température, les précipitations et le vent.

Quelques définitions et propositions

- Soit X_t l'état caché à l'instant t, représentant le régime météorologique (par exemple, "beau temps", "pluie", "neige"). Soit O_t l'observation à l'instant t, représentant les mesures météorologiques observées (par exemple, la température, la précipitation, etc.).

- La matrice de transition $P = (a_{ij})$ décrit les probabilités de passage d'un régime météorologique à un autre. Pour des états X_i et X_j,

$$a_{ij} = \mathbb{P}(X_{t+1} = X_j \mid X_t = X_i)$$

- Le vecteur de distribution initiale $\pi = (\pi_1, ..., \pi_n)$ représente les probabilités initiales des régimes météorologiques. Pour un état initial X_i,

$$\pi_i = \mathbb{P}(X_1 = X_i)$$

- Chaque régime météorologique génère des observations météorologiques selon une distribution spécifique à ce régime. La probabilité d'observer une mesure v_k lorsque le régime est X_j est donnée par la matrice de distribution d'observations Q.

$$b_j(k) = \mathbb{P}(O_t = v_k \mid X_t = X_j)$$

4.3 Prédiction d'erreurs

Définition

La *prédiction d'erreur* dans le contexte des chaînes de Markov cachées (CMC) se réfère à la capacité du modèle à prédire des observations futures et à mesurer l'écart entre les observations prédites et les observations réelles. Cette erreur de prédiction est très importante pour évaluer la performance du modèle et pour l'améliorer.

Quelques propriétés

- **L'erreur de prédiction** à un instant t est définie comme la différence entre l'observation réelle O_t et l'observation prédite $\hat{O}_t$ basée sur le modèle CMC.

$$\text{Erreur de Prédiction} = O_t - \hat{O}_t$$

- **Erreur Quadratique Moyenne (EQM)** : La moyenne des carrés des erreurs de prédiction, définie par :

$$\text{EQM} = \frac{1}{N}\sum_{t=1}^{N}(O_t - \hat{O}_t)^2$$

- **Erreur Absolue Moyenne (EAM)** : La moyenne des valeurs absolues des erreurs de prédiction, définie par :

$$\text{EAM} = \frac{1}{N}\sum_{t=1}^{N}|O_t - \hat{O}_t|$$

- L'algorithme forward est utilisé pour calculer la probabilité des observations futures données les observations passées. Il permet de faire des prédictions basées sur le modèle CMC.
- L'analyse de l'erreur de prédiction permet d'évaluer la précision et la fiabilité du modèle CMC. Un modèle avec une faible erreur de prédiction est généralement considéré comme performant.
- L'étude des erreurs de prédiction peut révéler des faiblesses dans le modèle, telles que des états cachés mal définis ou des probabilités de transition incorrectes. En analysant ces erreurs, on peut ajuster et améliorer le modèle CMC pour obtenir de meilleures prédictions.

Applications dans divers domaines

La prédiction d'erreur est cruciale dans de nombreuses applications pratiques, telles que :

- **Reconnaissance de la Parole** : Améliorer la précision des systèmes de reconnaissance vocale en réduisant l'erreur de prédiction des phonèmes.
- **Bioinformatique** : Prédire les séquences d'ADN avec une meilleure précision en analysant les erreurs de prédiction.

- **Finance** : Prédire les tendances des marchés financiers et réduire les risques en améliorant les modèles CMC utilisés.
- **Météorologie** : Améliorer les prévisions météorologiques en réduisant les erreurs de prédiction des modèles de méta-météo.

Détection et Correction des Anomalies

L'analyse des erreurs de prédiction peut également être utilisée pour détecter des anomalies dans les données. Par exemple, des erreurs de prédiction importantes peuvent indiquer des événements rares ou des changements dans le processus sous-jacent, permettant ainsi de prendre des mesures correctives.

Conclusion

Ce mémoire a exploré en profondeur les chaînes de Markov cachées (CMC) et leurs diverses applications à travers l'exemple de la marche aléatoire du voilier, en commençant par une introduction aux concepts fondamentaux et en progressant vers des applications pratiques.

Nous avons commencé par définir les chaînes de Markov simples, établissant les bases nécessaires pour comprendre les chaînes de Markov cachées.

Nous avons ensuite détaillé les définitions et les propriétés des chaînes de Markov homogènes, abordant des concepts clés tels que la récurrence et l'irréductibilité, ainsi que la probabilité invariante.

Nous avons également présenté deux théorèmes fondamentaux dans l'étude des chaines de Markov et de leur convergence :

- Le théorème central limite, qui décrit la convergence en distribution des sommes normalisées de variables aléatoires.
- Le théorème de convergence ergodique, qui traite de la convergence des distributions des états vers une distribution invariante.

Ces théorèmes ont ensuite été appliqués aux chaînes de Markov pour illustrer leur utilité pratique dans la mise en place de simulations pour des temps donnés très grands.

Nous avons par la suite discuté de la mesure stationnaire et de l'estimation de la matrice de transition, deux aspects essentiels pour la simulation et l'analyse des chaînes de Markov.

S'en est suivi l'introduction des chaînes de Markov cachées, définissant leurs propriétés et leur structure. Ensuite, nous avons exploré la modélisation de la méta-météo d'une marche aléatoire de voilier comme exemple concret d'application des CMC. Cela nous a également permis de mettre en lumière la difficulté liée à la détermination d'une matrice de transition exacte pour la méta-météo, raison pour laquelle nous avions introduit la notion d'estimateur de matrice.

La section sur la prédiction d'erreurs a mis en lumière l'importance de l'évaluation et de l'amélioration des modèles CMC en analysant les erreurs de prédiction.

Enfin, nous avons mentionné d'autres exemples de chaînes de Markov cachées, comme les séquences ADN, démontrant l'étendue des applications possibles des CMC.

Les chaînes de Markov cachées représentent donc un outil puissant pour modéliser des systèmes stochastiques complexes. Leur capacité à gérer des états cachés les rend particulièrement utiles dans de nombreux domaines, de la reconnaissance de la parole à la bioinformatique. Les théorèmes de convergence, les techniques de simulation et les méthodes d'estimation sont autant d'outils essentiels pour exploiter pleinement le potentiel des CMC.

References

[1] Guilbaut, A., Belotti, F. (2018). *Chaînes de Markov cachées*. Lien : https://math.univ-lille1.fr/~calgaro/TER_2018/wa_files/guilbaut_belotti.pdf

[2] Rösler, F., Bleistein, B., Koo, C. *Théorème de convergence ergodique*. Lien : https://www.math.ens.psl.eu/shared-files/9934/?Rosler_Bleistein_Koo.pdf

[3] Lepetit, G. *Théorème Central Limite*. Lien : https://agreg-maths.fr/uploads/versions/846/TCL.pdf

[4] ISAE-Supaero. *Notes de cours*. Lien : https://pagespro.isae-supaero.fr/IMG/pdf/notes_de_cours.pdf

[5] Barbé, P., Ledoux, M. (2007). *Probabilité*. EDP Sciences.

[6] Brémaud, P. (2009). *Initiation aux probabilités et aux chaînes de Markov*. Berlin: Heidelberg: Springer. (2e édition entièrement révisée)