Doubly stochastic Markov chains

Ludovick Bouthat

Université Laval

July 14, 2022

Remerciements

Cette recherche est le fruit d'une collaboration avec Pr. Apoorva Khare, Pr. Javad Mashreghi et Pr. Frédéric Morneau-Guérin.

Elle a été menée avec l'aide financière de la bourse à la maîtrise du CRSNG.

NSERC
CRSNG

Processus stochastiques et chaînes de Markov

Definition

Un processus de Markov à temps discret est une séquence $X_1, X_2, X_3, \ldots$ de variables aléatoires à valeurs dans l'espace des états qui respecte la propriété de Markov : *l'information utile pour la prédiction du futur ne dépend pas du passé*.

Processus stochastiques et chaînes de Markov

Definition

Un processus de Markov à temps discret est une séquence $X_1, X_2, X_3, \ldots$ de variables aléatoires à valeurs dans l'espace des états qui respecte la propriété de Markov : *l'information utile pour la prédiction du futur ne dépend pas du passé*.

Example (Le problème du collectionneur de vignettes)

Contexte : Une personne collectionne les cartes des n joueurs d'une équipe sportive, qu'il trouve à l'intérieur de tablettes de chocolat ; à chaque tablette, il a 1 chance sur n d'obtenir la carte $\#n$.

Définition : $X_k :=$ "Avoir k cartes distinctes".

Matrice de transition

Une matrice de transition, ou *matrice stochastique*, est une matrice

$$P = \begin{bmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,j} & \cdots & P_{1,n} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,j} & \cdots & P_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{i,1} & P_{i,2} & \cdots & P_{i,j} & \cdots & P_{i,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{n,1} & P_{n,2} & \cdots & P_{n,j} & \cdots & P_{n,n} \end{bmatrix},$$

où $P_{i,j}$ est la probabilité de passer de l'état i à l'état j d'une chaîne de Markov en une unité de temps (une étape du processus).

Matrice de transition

Une matrice de transition, ou *matrice stochastique*, est une matrice

$$P = \begin{bmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,j} & \cdots & P_{1,n} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,j} & \cdots & P_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{i,1} & P_{i,2} & \cdots & P_{i,j} & \cdots & P_{i,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{n,1} & P_{n,2} & \cdots & P_{n,j} & \cdots & P_{n,n} \end{bmatrix},$$

où $P_{i,j}$ est la probabilité de passer de l'état i à l'état j d'une chaîne de Markov en une unité de temps (une étape du processus).

La probabilité de passer de l'état i à l'état j en k unités de temps (k étapes du processus) est donnée par $(P^k)_{i,j}$, soit le coefficient i, j de la matrice P^k.

Le problème du collectionneur de vignettes

Suite

- $X_1 :=$ "0 coupon" ;
- $X_2 :=$ "1 coupon" ;
- $X_3 :=$ "2 coupons distincts" ;
- $X_4 :=$ "3 coupons distincts".

La matrice de transition associée à ce problème est :

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Le problème du collectionneur de vignettes

Suite

- $X_1 :=$ "0 coupon" ;
- $X_2 :=$ "1 coupon" ;
- $X_3 :=$ "2 coupons distincts" ;
- $X_4 :=$ "3 coupons distincts".

La matrice de transition associée à ce problème est :

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Ainsi, on a par exemple

$$P^2 = \begin{bmatrix} 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & \frac{1}{9} & \frac{2}{3} & \frac{2}{9} \\ 0 & 0 & \frac{4}{9} & \frac{5}{9} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad P^5 = \begin{bmatrix} 0 & \frac{1}{81} & \frac{10}{27} & \frac{50}{81} \\ 0 & \frac{1}{243} & \frac{62}{243} & \frac{20}{27} \\ 0 & 0 & \frac{32}{243} & \frac{211}{243} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Matrices doublement stochastiques

Definition

Une matrice est *doublement stochastique* si et seulement si elle et sa transposée son stochastique. En particulier, une matrice carrée est doublement stochastique si et seulement si ses coefficients sont non-négatif et toutes ses lignes et toutes ses colonnes ont une somme égale à 1.

L'ensemble des matrices doublement stochastiques d'ordre n est dénotée par Ω_n.

Example

$$D = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.4 & 0.2 & 0.4 \\ 0.5 & 0.5 & 0 \end{bmatrix}.$$

Cas particuliers

- Les matrices de permutation de dimension $n \times n$ sont des matrices doublement stochastiques et elles sont les *points extrémaux* de Ω_n.

Cas particuliers

• Les matrices de permutation de dimension $n \times n$ sont des matrices doublement stochastiques et elles sont les *points extrémaux* de Ω_n.

Theorem (Théorème de Birkhoff ; 1946)

Ω_n est l'enveloppe convexe des matrices de permutation de dimension $n \times n$.

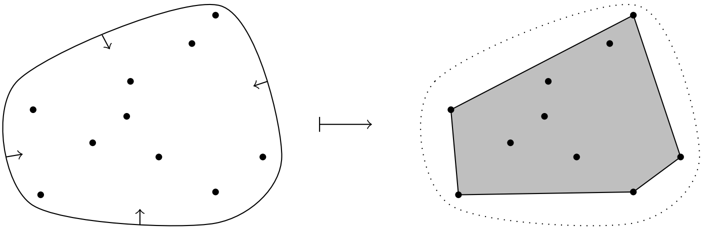

Figure – L'enveloppe convexe d'un nombre fini de points dans le plan

Cas particuliers

- La matrice uniforme $J_n = \frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$ agit comme l'élément neutre de Ω_n. C'est-à-dire que $J_n D = DJ_n = J_n$ pour toute matrice doublement stochastique D.

Cas particuliers

- La matrice uniforme $J_n = \frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$ agit comme l'élément neutre de Ω_n. C'est-à-dire que $J_n D = D J_n = J_n$ pour toute matrice doublement stochastique D.

- Géométriquement, J_n est le *centre* de Ω_n.

Cas particuliers

- La matrice uniforme $J_n = \frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$ agit comme l'élément neutre de Ω_n. C'est-à-dire que $J_n D = D J_n = J_n$ pour toute matrice doublement stochastique D.

- Géométriquement, J_n est le *centre* de Ω_n.

- Ω_n est clos sous la multiplication matricielle classique.

Comportement à long terme

Question : Quel est le comportement à long terme d'une chaîne de Markov formée de matrices doublement stochastiques ?

Comportement à long terme

Question : Quel est le comportement à long terme d'une chaîne de Markov formée de matrices doublement stochastiques ?

Example

$$D = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.4 & 0.2 & 0.4 \\ 0.5 & 0.5 & 0 \end{bmatrix} \rightsquigarrow D^8 = \begin{bmatrix} 0.334636 & 0.334634 & 0.33073 \\ 0.333332 & 0.333335 & 0.333332 \\ 0.332031 & 0.332031 & 0.335938 \end{bmatrix}$$

Comportement à long terme

Question : Quel est le comportement à long terme d'une chaîne de Markov formée de matrices doublement stochastiques ?

Example

$$D = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.4 & 0.2 & 0.4 \\ 0.5 & 0.5 & 0 \end{bmatrix} \rightsquigarrow D^8 = \begin{bmatrix} 0.334636 & 0.334634 & 0.33073 \\ 0.333332 & 0.333335 & 0.333332 \\ 0.332031 & 0.332031 & 0.335938 \end{bmatrix}$$

En fait, on vérifie aisément que $D^k \to J_3$ lorsque $k \to \infty$.

Comportement à long terme

Question : Quel est le comportement à long terme d'une chaîne de Markov formée de matrices doublement stochastiques ?

Example

$$D = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.4 & 0.2 & 0.4 \\ 0.5 & 0.5 & 0 \end{bmatrix} \rightsquigarrow D^8 = \begin{bmatrix} 0.334636 & 0.334634 & 0.33073 \\ 0.333332 & 0.333335 & 0.333332 \\ 0.332031 & 0.332031 & 0.335938 \end{bmatrix}$$

En fait, on vérifie aisément que $D^k \to J_3$ lorsque $k \to \infty$.

Contre-exemple : $I^k = I$ pour tout $k \in \mathbb{N}$.

Résultats initiaux

Theorem

Si $D \in \Omega_n$*, alors* $\lim_{k \to \infty} D^k = J_n$ *si et seulement si* D *a une unique valeur propre de module égal à* 1.

Résultats initiaux

Theorem

Si $D \in \Omega_n$, *alors* $\lim_{k\to\infty} D^k = J_n$ *si et seulement si* D *a une unique valeur propre de module égal à* 1.

Theorem (Š. Schwarz, 1980)

Soit $(A_i) \subseteq \Omega_n$ *une suite de matrices doublement stochastique et soit* $\nu(A) := \min_{i,j} a_{ij}$. *Si* $\sum_{i=1}^{\infty} \nu(A_i) = \infty$, *alors* $\lim_{k\to\infty} A_1 A_2 \cdots A_k = J_n$.

La méthode d'approche

But : Trouver une "bonne" condition suffisante pour que $\|A_1 \cdots A_k - J_n\|$ converge vers 0.

La méthode d'approche

But : Trouver une "bonne" condition suffisante pour que $\|A_1 \cdots A_k - J_n\|$ converge vers 0.

Proposition

$A_1 \cdots A_k - J_n = (A_1 - J_n) \cdots (A_k - J_n)$ *pour tout* $A_1, \ldots, A_k \in \Omega_n$.

La méthode d'approche

But : Trouver une "bonne" condition suffisante pour que $\|A_1 \cdots A_k - J_n\|$ converge vers 0.

Proposition

$A_1 \cdots A_k - J_n = (A_1 - J_n) \cdots (A_k - J_n)$ *pour tout* $A_1, \ldots, A_k \in \Omega_n$.

$$\implies \|A_1 \cdots A_k - J_n\| \leq \|A_1 - J_n\| \cdots \|A_k - J_n\|.$$

La méthode d'approche

But : Trouver une "bonne" condition suffisante pour que $\|A_1 \cdots A_k - J_n\|$ converge vers 0.

Proposition

$A_1 \cdots A_k - J_n = (A_1 - J_n) \cdots (A_k - J_n)$ *pour tout* $A_1, \ldots, A_k \in \Omega_n$.

$$\implies \|A_1 \cdots A_k - J_n\| \;\leq\; \|A_1 - J_n\| \cdots \|A_k - J_n\|.$$

Conclusion : On veut une norme matricielle telle que $\sup_{A \in \Omega_n} \|A - J_n\|$ est *petit*.

Le centre de Tchebychev

Definition

Si $\|\cdot\|$ est une norme matricielle invariante par permutation, alors $\sup_{A\in\Omega_n}\|A - J_n\|$ correspond au *rayon de Tchebychev* de Ω_n et J_n est le *centre de Tchebychev* de Ω_n.

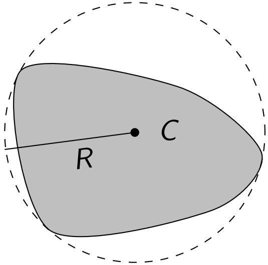

Figure – Le rayon de Tchebychev R et le centre de Tchebychev C d'un ensemble dans $\mathbb{R}^2$.

Un nouveau résultat

Parmis les normes matricielles invariantes par permutation, la norme d'opérateur $\ell^2 \to \ell^2$ définie par

$$\|A\|_{\ell^2 \to \ell^2} := \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$$

minimise la quantité $\sup_{A \in \Omega_n} \|A - J_n\|$.

Un nouveau résultat

Parmis les normes matricielles invariantes par permutation, la norme d'opérateur $\ell^2 \to \ell^2$ définie par

$$\|A\|_{\ell^2 \to \ell^2} := \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$$

minimise la quantité $\sup_{A \in \Omega_n} \|A - J_n\|$. En utilisant cette propriété, on peut montrer le théorème suivant :

Theorem

Soit $A_1, A_2, \ldots$ une suite de matrices doublement stochastiques d'ordre n et soit $\sigma_2(A)$ la seconde plus grande valeur propre de A. Si $\sum_{k=1}^{\infty}(1 - \sigma_2(A_k))$ diverge, alors $\lim_{m \to \infty} A_1 A_2 \cdots A_m = J$.

Une amélioration du résultat de Schwarz

Il est possible de montrer que $n\nu(A) \leq \ 1 - \sigma_2(A)$ pour toute matrice doublement stochastique A. Par conséquent,

$$\sum_{k=1}^{\infty} \nu(A_k) \ = \ \infty \quad \Longrightarrow \quad \sum_{k=1}^{\infty} (1 - \sigma_2(A_k)) \ = \ \infty.$$

Une amélioration du résultat de Schwarz

Il est possible de montrer que $n\nu(A) \leq 1 - \sigma_2(A)$ pour toute matrice doublement stochastique A. Par conséquent,

$$\sum_{k=1}^{\infty} \nu(A_k) = \infty \quad \Longrightarrow \quad \sum_{k=1}^{\infty} (1 - \sigma_2(A_k)) = \infty.$$

Example

Considérons le cas $A_k = A$ pour tout k, où

$$A = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}.$$

On a $A \in \Omega_n$, $\nu(A) = 0$, et les valeurs singulières de A sont $1, 1/2$ et 0 et donc que $1 - \sigma_2(A) = 1/2$.

References

Štefan Schwarz (1980) Infinite products of doubly stochastic matrices. *Acta Math. Univ. Comenian.*, 39 : 131–150.

Ludovick Bouthat, Apoorva Khare, Javad Mashreghi & Frédéric Morneau-Guérin (2021) The p-norm of circulant matrices, *Linear and Multilinear Algebra*, DOI : 10.1080/03081087.2021.1983513

K. R. Sahasranand (2022) The p-norm of circulant matrices via Fourier analysis. arXiv preprint arXiv :2111.11389.