Chaînes de Markov à temps discret

Moutassim Abderrazak
Ph.D. Mathematiques de l'ingénieur

Université de Mundialpolis

Quelques définitions

- Expérience aléatoire (E): tout expérience qui peut être répétée sous les mêmes conditions et dont le résultat ne peut être prédit avec certitude.
  – Exemple: On lance une pièce de monnaie trois fois et on note le nombre de "Pile" obtenues.
- Espace échantillon (S): ensemble de tous les résultats possibles d'une expérience E. Un sous ensemble e de S est appelé événement.
  – Dans notre exemple, $S = \{0, 1, 2, 3\}$, et $e = \{1, 2, 3\}$. Le singleton $\{1\}$ s'appèle événement élémentaire.

Espace de probabilités

- Un espace de probabilités est un triplet $(S, \mathcal{F}, P)$, où S désigne l'espace Échantillons, $\mathcal{F}$ est l'espace des événements, et P est la mesure associée à la probabilité P. Celle-ci est une fonction définie sur l'ensemble d'événements $\mathcal{F}$ et qui associe à chaque événement $e \in \mathcal{F}$ une probabilité $p \in [0, 1]$.
- On définit sur l'ensemble d'événements $\mathcal{F}$ les opérations d'union, d'intersection, d'inclusion et de complément.
- En théorie de mesure, l'ensemble $\mathcal{F}$ est appelé la σ-algèbre satisfaisant à ce qui suit:
  1. $\mathcal{F}$ contient l'espace échantillons S (noté parfois Ω et appelé univers), et l'événement impossible $(\emptyset)$,
  2. si $A \in \mathcal{F}$ alors $A^c \in \mathcal{F}$,
  3. l'union dénombrable d'événement $A_i \in \mathcal{F}, i \in \mathbb{N}$ est aussi un événement de $\mathcal{F}$. On a la même chose pour l'intersection.

Quelques propiétés

On a les axiomes de la probabilité P suivantes:

- $P[A] \in [0,1], \ \forall A \subset S.$.
- $P[S] = 1$ et $P[\emptyset] = 0$
- $P\left[\cup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} P(A_i)$, pour toute suite d'événements incompatibles deux à deux, i.e $A_i \cap A_j = \emptyset$ pour $i \neq j$.

La probabilité P(A) pour qu'un événement $A \subset S$ ait lieu est

$$P(A) = \frac{n(A)}{n},$$

où n est le nombre d'événements élémentaires de S, supposés équiprobables, et $n(A)$ est le nombre d'événements élémentaires dans A.

- $P[A^c] = 1 - P[A] \ \forall A \subset S.$
- $P[A \cup B] = P[A] + P[B] - P[A \cap B],$

Probabilité conditionnelle

- $P[A \mid B]$ désigne la probabilité conditionnelle de A sachant B.
- **Règle de multiplication:** si $P[A]\,P[B] > 0$, alors

$$P[A \cap B] = P[A \mid B]\,P[B] = P[B \mid A]\,P[A].$$

- **Règle de la probabilité totale:** Si $A \subset S$ et $\{B_k\}, k \leq n$ une partition de S, alors

$$P[A] = \sum_{k=1}^{n} P[A \cap B_k] = \sum_{k=1}^{n} P[A \mid B_k]\,P[B_k]$$

- **Règle de Bayes:** Si $P[A] > 0$, alors

$$P[B_k] = \frac{P[A \mid B_k]\,P[B_k]}{\displaystyle\sum_{k=1}^{n} P[A \mid B_k]\,P[B_k]}, \forall k \leq n.$$

- A et B sont des événements indépendants si et seulement si $P[A \mid B] = P[A]$, ou ssi $P[A \cap B] = P[A] \times P[B]$.

Variables aléatoires

- Variables aléatoires (v.a.): toute fonction X qui associe un nombre réel à chaque élément $e \in S$, où S l'espace échantillon associé à E. On désigne par S_X l'ensemble des valeur de X.
- Fonction de répartition: soit l'événement A tel que $P(A) > 0$
$$F_X(x) = P(X \leq x) \; \forall x \in \mathbb{R}, \quad F_X(x \mid A) = \frac{P(\{X \leq x\} \cap A)}{P(A)}.$$
- Variables aléatoires discrètes:
  1. Fonction de masse de probabilité: $p_X(x_k) = P(X = x_k)$, conditionnelle: $p_X(x_k \mid A) = \dfrac{P(\{X = x_k\} \cap A)}{P(A)} \; \forall x_k \in S_X$
  2. Examples: Loi de Bernoulli, binomiale, géométrique, poisson.
- Variables aléatoires continues:
  1. Fonction de densité de probabilité: $f_X(x) = \dfrac{d}{dx} F_X(x)$, conditionnelle: $f_X(x \mid A) = \dfrac{d}{dx} F_X(x \mid A) = \dfrac{f_X(x)}{p(A)}$.
  2. Examples: Loi uniforme, exponentielle, gamma, gaussienne,

Espérance mathématique et variance

- L'espérance mathématique d'une variable aléatoire v.a. X:
  - cas discret: $\mathbb{E}[X] = \sum_{k=1}^{\infty} x_k p_X(x_k)$
  - cas continu: $\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx$,

On obtient l'espérance conditionnelle en remplaçant $p_X(x)$ et $f_X(x)$ par $p_X(x \mid A)$ et $f_X(x \mid A)$, respectivement.

- La variance de X et la variance conditionnelle de $X \mid A$ sont

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2],\ \mathbb{V}(X \mid A) = \mathbb{E}[(X - \mathbb{E}(X \mid A))^2].$$

$$\mathbb{V}(X) = \mathbb{E}[\mathbb{V}(X \mid Y)] + \mathbb{V}[\mathbb{E}(X \mid Y)]$$

- Fonction caractéristique d'une v.a.: $C_X(w) = \mathbb{E}[e^{jwX}]$, $j^2 = -1$.
- Moments d'ordre n: $\mathbb{E}(X^n) = (-j)^n \frac{d^n}{dw^n} C_X(w)_{|w=0}$

Processus stochastiques

- l'ensemble $\{X(t,s), t \in T\}$, où $T \subset \mathbb{R}^+$, et $s \in S$ s'appellent Processus stochastique, en abrégé p.s.

– La fonction $X(t,s)$ est une v.a. pour toute valeur particulière t.

- Remarque:

– Sauf pour l'ergodicité qu'on discutera plus tard, il ne sera pas nécessaire d'écrire l'argument s de la fonction $X(t,s)$, ainsi le processus sera noté $\{X(t), t \in T\}$.

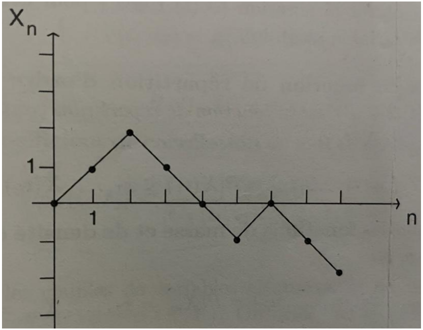
– Si $T \subseteq \{0,1,...\}$, on dit que le processus est à temps discret, et on le note avec $\{X_n, n = 0,1..\}$.

– Si $T \subseteq [0,\infty]$, alors il est à temps continu.

– l'ensemble $S_{X(t)}$ des valeurs que prend $X(t)$ est appelé espace d'états du p.s. $\{X(t), t \in T\}$. Le processus est dit à état discret (respectivement à état continu), lorsque T est dénombrable (resp. non dénombrable).

cas discret et continu

• cas discret: une particule se trouve à l'origine à l'instant $t = 0$. À chaque unité de temps, on lance une pièce de monnaie. Si on a "pile" (respectivement "ace"), la particule se déplace d'une unité vers la droite (resp. gauche). La v.a. X_n désigne la position de la particule au bout de n lancers de la pièce, alors que le p.s. $\{X_n, n = 0, 1..\}$. est une marche aléatoire particulière.



• cas continu: soit le p.s $\{X(t) = Y.t, t \geq 0\}$. où Y une v.a.

Fonctions et propriétés des processus stochastiques

- fonction de répartition d'ordre k du p.s. $\{X(t), t \in T\}$:

$$F(x_1, ..., x_k; t_1 ..., t_k) = P[X(t_1) \leq x_1, ..., X(t_k) \leq x_k]$$

- les fonctions de masse et de densité de probabilité d'ordre k:

$p(x_1, ..., x_k; t_1 ..., t_k) = P[X_{n_1} = x_1, ..., X_{n_k} = x_k]$, et
$f(x_1, ..., x_k; t_1 ..., t_k) = \frac{\partial^k}{\partial x_1 ... \partial x_k} F(x_1, ..., x_k; t_1 ..., t_k)$.

- un p.s. $\{X(t), t \in T\}$ est décrit à l'aide de sa moyenne $\mathbb{E}[X(t)]$, notée par $m_X(t)$ à chaque instant t, et à l'aide de sa fonction d'auto-covariance $C_X(t_1, t_2) = R_X(t_1, t_2) - m_X(t_1) m_X(t_2)$, où $R_X(t_1, t_2) = \mathbb{E}[X(t_1) X(t_2)]$ est la fonction d'autocorrélation.

- un p.s. est à accroissements indépendants si $X(t_4) - X(t_3)$ et $X(t_2) - X(t_1)$ sont indépendants pour $\forall t_1 < t_2 < t_3 < t_4$.

- un p.s. est à accroissements stationnaires si $X(t_2 + s) - X(t_1 + s)$ et $X(t_2) - X(t_1)$ ont la même fonction de répartition $\forall s \geq 0$.

Stationnarité et ergodicité

• Coefficient de corrélation: $\frac{C_X(t_1,t_2)}{\sqrt{[C_x(t_1,t_1)\times C_x(t_2,t_2)]}}$

• un p.s. $\{X_t\}, t \in T$ est stationnaire au sens stict (SSS) si $F(x_1, ..., x_k; t_1..., t_k) = F[x_1, ..., x_k; t_1 + s..., t_k + s], \ \forall s \geq 0,$

et il est stationnaire au sens large (SSL) si $m_X(t) \equiv m(constante)$ et le coefficient d'auto-corrélation satisfait à

$$R_X(t_1, t_2) = R_X(t_2 - t_1) \ \forall t_1, t_2 \in T.$$

• un p.s. $\{X(t), t \in T\}$ est ergodique si toute caractéristique du processus peut être obtenue avec une probabilité 1, à partir d'une seule réalisation $X(t, s)$

• un p.s. $\{X(t), t \in T\}$ pour lequel $m_X(t) = m \forall t \in T$ est ergodique par rapport à la moyenne si

$P\left[\lim_{S\to\infty} < X(t) >_S = m.\right] = 1$, où $< X(t) >_S = \frac{1}{2S}\int_{-S}^{S} X(t,s)\, dt$ la moyenne temporelle du p.s.

Propriété markovienne (Chaîne de Markov)

- Jusqu'à maintenant, on considère les p.s. à états discrets.

Un p.s. à temps discret $\{X_n, n = 0, 1, ...\}$ est une chaîne de Markov si $\forall n \geq 0$

$$P(X_{n+1} = j \mid X_n = i_n, ...X_0 = i_0) = P(X_{n+1} = j \mid X_n = i_n) = p_{i,j}(n).$$

Et si les $p_{i,j}$ ne dépendent pas de n, la chaîne est dite stationnaire.

- Exemple d'une chaîne de Markov non stationnaire:

Si l'on modélise le débit d'une rivière par une chaîne de Markov $\{X_n, n = 1, ...\}$, et que l'on utilise trois qualificatifs pour le débit X_n pendant la n^e journée de l'année: faible (0), moyen(1) et élevé (2), alors cette chaîne n'est pas stationnaire car la probabilité de passer d'un débit à un autre n'est pas la même pendant toute l'année.

Pour la suite, nous considérons que les p.s. sont stationnaires et donc $p_{i,j}(n) = p_{i,j}$.

Matrice P des probabilités de transition en une étape

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \dots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \dots \end{array} \\ \begin{pmatrix} p_{0,0} & p_{0,1} & p_{0,2} & p_{0,3} & \cdots \\ p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} & \cdots \\ p_{2,0} & p_{2,1} & p_{2,2} & p_{2,3} & \cdots \\ p_{3,0} & p_{3,1} & p_{3,2} & p_{3,3} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \end{array}$$

• Remarque: Nous avons indiqué les états possibles de la chaîne de Markov à gauche et au-dessus de la matrice, pour voir les transitions et leurs probabilités. L'état à gauche conditionné est celui dans lequel se trouve le processus à l'instant n, et ce au-dessus est dans lequel se trouvera le processus à l'instant $n+1$.

Comme le processus doit se trouver dans un et un seul état à l'instant $n+1$, $\sum_{j=1}^{\infty} p_{i,j} = 1 \ \forall i$. Si de plus $\sum_{i=1}^{\infty} p_{i,j} = 1$, alors la matrice est dite doublement stochastique.

Propriétés des états

- Accessibilité: on dit que l'état j est accessible à partir de i après n étapes (ou transitions) si $p_{i,j}^{(n)} > 0$. Si de plus, i est accessible à partir de j, on dit que les deux états communiquent. Les états qui communiquent sont dans la même classe.

- Classe fermée: un sous-ensemble C de l'espace des états d'une chaîne de Markov qui satisfait à

$$P[X_{n+1} \in C \mid X_n = i \in C] = 1 \, \forall i \in C.$$

- Chaîne irréductible: si tous les états de la chaîne communiquent, ou s'il existe un chemin dont la probabilité est > 0, qui part d'un état et y retourne en passant par tous autres états au moins une fois.
- État i récurrent: si $f_{i,i} = P\left[\cup_{n=1}^{\infty}\{X_n = i\} \mid X_0 = i\right] = 1$
- État transitoire: si $f_{i,i} < 1$.

Propriétés des états (suite)

• Soit N_i le nombre de fois que l'état i sera visité, étant donné que $X_0 = i$. Alors, l'état i est récurrent si et seulement si $\mathbb{E}[N_i] = \infty$ et aussi si et seulement si $\sum_{i=0}^{\infty} p_{i,i}^{(n)} = \infty$.

• Tous les états d'une chaîne irréductible finie sont récurrents.

$\rightarrow$ la récurrence est une propriété de classe.

• Soit i un état récurrent. Si l'on définit la probabilité de passer de l'état i à l'état j pour la première fois à la n^e transitions par

$$\rho_{i,j}^{(n)} = P\left[X_n = j, X_{n-1} \neq j, ..., X_1 \neq I \mid X_0 = i\right],$$

où $n \geq 1$ et $i, j \geq 0$, alors le nombre moyen de transitions μ_i requises pour que le processus, parti de i, y retourne pour la première fois est $\mu_i := \sum_{n=1}^{\infty} n\rho_{i,i}^{(n)}$. Si $\mu_i < \infty$, alors i est récurrent positif, sinon récurrent nul.

Propriétés des états (suite)

→ On accepte aussi que tout état récurrent d'une chaîne fini est récurrent positif.

• Un état est dit périodique de période d si $p_{i,i}^{(n)} = 0$ n qui n'est pas divisible par d, où d est le plus grand entier qui possède cette propriété.

□ Si $d = 1$ l'état est dit apériodique.

→ La périodicité est une propriété de classe.

• Une chaîne est dite apériodique si tous ses états sont apériodiques

Remarques:

1. Si $p_{i,i}^{(1)} > 0$, alors l'état i est évidemment apériodique.
2. Si $p_{i,i}^{(2)} > 0$ et $p_{i,i}^{(3)} > 0$, alors l'état i est apériodique.
3. Si $d = 4$, alors $d = 2$ vérifie aussi la définition de périodicité, en effet, $p_{i,i}^{(2n+1)} = 0, \forall n > 0$.

Théorème ergodique et distribution stationnaire

- Un état récurrent positif et apériodique est appelé ergodique.

$\rightarrow$ L'ergodicité est aussi une propriété de classe.

- Les chaînes de Markov irréductible finies et apériodiques sont ergodiques.

Théorème: Dans le cas d'une chaîne de Markov irréductible et ergodique, alors la probabilité limite

$$\pi_j := \lim_{n\to\infty} p_{i,j}^{(n)}$$

existe et est indépendante de i. De plus on a:

$$\pi_j = \frac{1}{\mu_j} > 0 \text{ pour tout } j \in \{0, 1, ...\},$$

où μ_j définie plus-haut comme

$$\mu_j := \sum_{n=1}^{\infty} n\rho_{j,j}^{(n)}.$$

Théorème ergodique et distribution stationnaire

- On peut montrer que $\pi = (\pi_j)_{j \geq 0}$ est la solution unique du système suivant:

$$\pi = \pi \mathrm{P} \quad \text{et} \quad \sum_{\mathrm{j}=0}^{\infty} \pi_{\mathrm{j}} = 1$$

Cette probabilité limite est appelée distribution stationnaire de la chaîne de Markov.

Proposition:

Si on a une chaîne irréductible et apériodique, i.e ergodique dont l'espace d'états est $\{0, 1, ..., k\}$, et si de plus elle est doublement stochastique, alors les probabilités limites existent et données par

$$\pi_j = \frac{1}{k+1} \text{ pour } j = 0, 1, \ldots k.$$

Mise en contexte et modélisation

Au XIX^e siècle, en Angleterre, on s'est intéressé à la possibilité que certains noms de famille disparaissent, faute de descendants mâles. Certains chercheurs ont modélisé ce problème mathématiquement à l'aide des processus stochastiques correspondant, parfois appelés processus de branchement.

Définition:
Soit $\{Z_{n,j}, n = 0, 1, ...; j = 1, 2, ...\}$ un ensemble de variables aléatoires i.i.d. dont les valeurs possibles sont des entiers non négatifs. C'est-à-dire que $S_{Z_{n,j}} \subset \{0, 1, ...\}$. Un processus de branchement est une chaîne de Markov $\{X_n, n = 0, 1, ...\}$ définie par

$$X_n = \begin{cases} \sum_{j=1}^{X_{n-1}} Z_{n-1,j} \ si \ X_{n-1} > 0 \\ 0 \quad si \quad X_{n-1} = 0 \end{cases}$$

pour $n = 1, 2, ...$

Remarques

i)- Dans le cas de l'application au problème de la disparition des noms, X_0 représente le nombre de membres de la génération initiale (nombre d'ancêtres de la population). Souvent, on suppose que $X_0 = 1$ de sorte qu'on s'intéresse à une lignée. $Z_{n-1,j}$ désigne le nombre de descendants de la j^e membre de la $(n-1)^e$ génération.

ii)- Soit

$$p_i = P[Z_{n-1,j} = i] \text{ pour tout } n \text{ et } j$$

On suppose que $p_i > 0$ pour tout $i = 0, 1, \ldots$ et $p_0 > 0$.
L'espace des états S_{X_n} est $\{0, 1, \ldots\}$, l'état 0 est absorbant et les autres sont transitoires, car on peut montrer que

$$P[X_n \neq i \ \forall n \in \{1, 2, \ldots\} \mid X_0 = i] \geq p_{i,0} \overset{\text{ind}}{=} p_0^i > 0.$$

Ainsi $f_{i,i} < 1$, et les états $i = 1, 2, \ldots$ sont transitoires.

Remarques (suite)

Supposons que $X_0 = 1$. Soit ν_n le nombre moyen d'individus de la n^e génération pour $n = 1, 2, \ldots$. Notons par $\nu_n = \mathbb{E}[X_n]$, on peut montrer par récurrence que $\nu_n = \nu_1^n$.

Si l'on pose que $\sigma_1^2 := \mathbb{V}[X_1]$, on trouve lorsque $\nu_1 \neq 1$ que

$$\mathbb{V}[X_n] = \sigma_1^2 \nu_1^{n-1} \left(\frac{\nu_1^{n-1} - 1}{\nu^{n_1} - 1} \right).$$

On désire déterminer la probabilité d'extinction éventuelle de la population, soit

$$q_{0,i} = \lim_{n \to \infty} P[X_n = 0 \mid X_0 = i] \stackrel{\text{ind}}{=} q_{0,1}^i.$$

Donc, il suffit de calculer $q_{0,1}$, noté aussi q_0. On peut obtenir $P[X_n = 0 \mid X_0 = 1] = 1 - \nu_1^n$

- Théorème: La probabilité q_0 d'extinction éventuelle de la population est égale à 1 si $\nu_1 < 1$, tandis que $q_0 < 1$ si $\nu_1 > 1$.

Remarques concernant le résultat du théorème

i)- On déduit du théorème qu'une condition nécessaire pour que la probabilité q_0 soit inférieure à 1 est que p_j soit supérieur à 0 pour au moins un $j \geq 2$. En effet, si $p_0 = p > 0$ et $p_1 = 1 - p$, alors on a directement $\mu_1 = 1 - p > 0$.

ii)- Dans le cas où $p_0 = 0$ et $p_1 = 1$, on a $\mu_1 = 1$. Selon le théorème, on devrait avoir $q_0 = 1$. Pourtant, si $p_1 = 1$, il est évident que $X_n = X_0$ pour tout n, et alors on peut écrire que $q_0 = 0$. Cependant, le théorème ne s'applique que lorsque $p_0 > 0$.

Soit F l'évènement défini par $F = \cup_{n=1}^{\infty}\{X_n = 0\}$ de sorte que $q_0 = P\big[F \mid X_0 = 1\big]$. Pour obtenir la valeur de q_0, on peut résoudre l'équation suivante:

$$q_0 = \sum_{j=0}^{\infty} P\big[F \mid X_1 = j\big]p_j = \sum_{j=0}^{\infty} {q_0}^j p_j.$$

On peut montrer que, dans le cas où $\mu_1 > 1$, q_0 est la plus petite solution positive de l'équation ci-dessus.