Master I

Option : Probabilité et Statistiques

Département de Mathématiques, Faculté Des Sciences, UMMTO

CHAÎNES DE MARKOV DISCRETES

Y. Berkoun

1 Généralités sur les processus stochastiques

Soient $(\Omega,\mathcal{A},P)$ et $(E,\mathcal{F})$ respectivement un espace de probabilité et un espace probabilisable. Notons $\mathbb{T}$ un espace des temps pouvant être discret ($\mathbb{N}$ ou $\mathbb{Z}$) ou continu ($\mathbb{R}$).

Definition 1. *Un processus stochastique X est une application définie par*

$$\begin{aligned} X : \Omega \times \mathbb{T} &\longmapsto E \\ (w,t) &\longmapsto X(w,t) = X_t(w) \end{aligned}$$

Dans toute la suite, on notera $X = (X_t)_t$. E est dit espace des états pouvant être discret ou continu.
Si $\mathbb{T}$ et E sont discrets (respectivement continus), le processus X est dit processus à temps discret (continu) et à espace d'états discret (continu).
Ex: X peut représenter la valeur d'un indice boursier au fil du temps, l'évolution d'une population, etc.

Definition 2. *Pour ω fixé, l'application*

$$\begin{aligned} X : \mathbb{T} &\longmapsto E \\ t &\longmapsto X(w,t) = X_t(w) \end{aligned}$$

est dite trajectoire du processus associée l'aléa w.
Remarques 1
1- Pour t fixé et w variable, $X(w,t)$ est une variable aléatoire.
2- Le processus est dit du second ordre si $E(X_t^2) < \infty$, $\forall t$

Definition 3. *Bruit blanc*
Une suite $(\epsilon_t)_t$ de variables aléatoires identiquement distribuées est dite bruit blanc (au sens faible) si
a) $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2 < \infty$ $\forall t$
b) $Cov(\epsilon_t,\epsilon_s) = 0$ $\forall t \neq s$
Remarque 2
On parlera de bruit blanc au sens fort si seulement la condition précédente a est vérifiée.

1.1 Stationnarité

Definition 4. *Stationnarité au sens faible*
Un processus $(X_t)_t$ est dit stationnaire au sens faible si
- $E(X_t) = constante < \infty$, $\forall t$
- $Cov(X_t,X_s)$ ne dépend que de la différence $|t-s|$

Definition 5. *Stationnarité au sens fort*
Un processus $(X_t)_t$ est dit stationnaire au sens fort ou strict si

$$\mathcal{L}(X_{t_1+h},X_{t_2+h},\dots X_{t_n+h}) = \mathcal{L}(X_{t_1},X_{t_2},\dots X_{t_n})$$

$$\forall t_1 < t_2 < \ldots < t_n, t_i \in \mathbb{T}\ \forall i,\ et\ \forall h \in \mathbb{Z}$$

Où $\mathcal{L}$ désigne la loi de probabilité.
Ce qui signifie que $\mathcal{L}(X_{t+h}) = \mathcal{L}(X_t)$, $\forall t$, $\forall h$, i.e., l'invariance de la loi par translation.

Exemple
Soit $(X_t)_t$ un processus stationnaire au sens faible et notons $(Y_t)_t$ le processus défini par $Y_t = \sum_{i=1}^{\infty} a_i X_{t-i}$ où $(a_i)_i$ est une suite réelle vérifiant $\sum_{i\geq 1} |a_i| < \infty$. On a
- $E(Y_t) = m \sum_{i\geq 1} a_i < \infty$, $m = E(X_t)$
- $Cov(Y_t,Y_s) = \sum_{i\geq 1} \sum_{j\geq 1} a_i a_j Cov(X_t,X_s)$
Ce qui montre la stationnarité au sens faible de $(Y_t)_t$.

Remarques importantes 3
- Il est facile de montrer que la stationnarité au sens fort implique la sationnarité au sens faible (à condition que le moment d'ordre deux existe). Si $(\varepsilon_t)_t$ est une suite de variables aléatoires identiquement distribuées de même loi de Cauchy standard, alors $(\varepsilon_t)_t$ est strictement stationnaire, mais pas faiblement stationnaire du fait de la non-existence du moment d'ordre deux.
- L'inverse est généralement faux. Considérons une suite $(X_t)_t$ de variables aléatoires i.i.d de loi $N(0,1)$ et posons

$$Y_t = \begin{cases} X_t & \text{si t est pair} \\ \frac{1}{\sqrt{2}}(X_t^2 - 1) & \text{si t est impair} \end{cases}$$

$E(Y_t) = 0\ \forall t$, $Cov(Y_t,Y_s) = 0, \forall t,s$
Ce qui donne la stationnarité au sens faible de $(Y_t)_t$
$P(Y_t \leq 0) = P(X_t \leq 0) = \frac{1}{2}$ si t est pair
$P(Y_t \leq 0) = P(|X_t| \leq 1) \neq \frac{1}{2}$ si t est impair
Ce qui implique la non-stationnarité au sens strict de $(Y_t)_t$.

Definition 6. *Processus à accroissements indépendants*
Un processus $(X_t)_t$ est dit à accroissements indépendants si
$\forall t_1 < t_2 < \ldots < t_k, \forall k$, les variables aléatoires X_{t_1}, $X_{t_2} - X_{t_1}$, $\ldots$, $X_{t_k} - X_{t_{k-1}}$ sont indépendantes.
Exemple Un bruit blanc est un processus à accroissements indépendants.

Definition 7. *Processus à accroissements stationnaires*
Un processus $(X_t)_t$ est dit à accroissements stationnaires si
$\forall t \in \mathbb{T}$, $\forall h > 0$, la $\mathcal{L}(X_{t+h} - X_{t+h})$ ne dépend pas de t.

Definition 8. ***Fonctions d'autocovariance et d'autocorrélation***
Soit $(X_t)_t$ un processus faiblement stationnaire. La fonction d'autocovariance $\gamma_X(.)$ est la fonction définie par

$$\begin{aligned} \gamma_X : \mathbb{Z} &\longmapsto \mathbb{R} \\ h &\longmapsto \gamma_X(h) = Cov(X_t,X_{t+h}) \end{aligned}$$

la fonction d'autocorrélation notée $\varrho_X(.)$ est définie par

$$\varrho_X : \mathbb{Z} \longmapsto [-1,1]$$
$$h \longmapsto \varrho_X(h) = \frac{Cov(X_t, X_{t+h})}{\sqrt{V(X_t)}V(X_{t+h})}$$

Proposition 1. *La fonction d'autocorrélation ϱ_X posséde les propriétès suivantes*
-$\varrho_X(0) = 1$
- $|\varrho_X(h)| \leq 1$
- $\varrho_X(h) = \varrho_X(-h)$

Definition 9. *Lois fini-dimensionnelles*
On appelle lois fini- dimensionnelles d'un processus $(X_t)_t$, l'ensemble des lois

$$\{\mathcal{L}(X_{t_1}, X_{t_2}, \ldots, X_{t_k}),\, t_1, \ldots, t_k \in \mathbb{T},\, k \in \mathbb{N}^*\}$$

Proposition 2. *(Kolmogorov) Loi d'un processus*
La loi $\mathbb{P}_X$ d'un processus est entiérement déterminée par ses lois fini-dimensionnelles.

Exercices

Exercice1
Dans la suite, $(\varepsilon_t)_t$ désignera un bruit blanc de variance σ^2.
a- Etudier la stationnarité du processus défini par $X_t = a + b\varepsilon_t + \varepsilon_{t-1}$, où $(a,b,c) \in \mathbb{R}^3$
b-A quelle condition le processus $X_t = at + b + \varepsilon_t$ est-il stationnaire?

Exercice 2
Soit X une variable aléatoire non-symétrique telle que $E(X) = 0$, $V(X) = 1$. On pose $Y_t = (-1)^t X$, $t \in \mathcal{Z}$
Montrer que $(Y_t)_t$ est stationnaire au sens faible mais pas au sens fort.

Exercice 3
Soient X and Y deus processus faiblement stationnaires indépendants de fonction d'autocovariance ϱ_X et ϱ_Y. Montrer que le processus $X + Y$ est faiblement stationnaire de fonction d 'autocovariance $\varrho_{X+Y} = \varrho_X + \varrho_Y$.

Exercice 4
Soient A et B deux variables aléatoires de lois respectives $N(m_a,\sigma_a)$ et $N(m_b,\sigma_b)$ telles que $Cov(A,B) = \delta$. Notons $(Y_t)_t$ le processus défini par

$$Y_t = \begin{cases} A + \varepsilon_t & \text{si t est pair} \\ B + \varepsilon_t & \text{si t est impair} \end{cases}$$

où $(\varepsilon_t)_t$ est un bruit blanc de variance σ^2 indépendant de A et B.
a) $(Y_t)_t$ est-il sationnaire? est-il du second ordre?
b) Etudier la stationnarité au sens fort?

Exercice 5
Soit $(Y_t)_t$ un processus défini par $Y_t = \sum_{i=0}^{t}(-1)^i\varepsilon_i$, où $(\varepsilon_i)_i$ est un bruit blance de variance σ^2.
- $(Y_t)_t$ est-l du second ordre?
- Calculer $Cov(Y_t,Y_{t+h})$ pour $h \in \mathbb{Z}$. On distinguera le cas $h > 0$ et $h < 0$.
- $(Y_t)_t$ est-il stationnaire?

Exercice 6
On considère le processus défini par
$X_t = \varepsilon_t - \theta\varepsilon_{t-1}$, $t \in \mathbb{Z}$ où $(\varepsilon)_t$ est un bruit blanc et $\theta \in]-1,+1[$.
a) Montrer que $(X_t)_t$ est stationnaire et calculer sa fonction d'auto-covariance.

2 Chaînes de Markov

Les chaînes de Markov ont été introduites par Andreï Andreïevitch Markov (1856-1922) qui était un éléve de Tchebychev. En lisant le roman *Eugène Onéguine* du russe Alexandre Pouchkine, Il examine les 20 000 premières lettres du texte et les considère comme 20 000 expériences aléatoires dont le résultat est consonne ou voyelle.Il montre alors, que l'occurrence d'une voyelle immédiatement après une consonne est un processus de Markov. Autrement dit, si on tire une lettre au hasard, c'est une voyelle ou une consonne selon des probabilités données par les statistiques (comme lorsqu'on joue à pile ou face), mais le statut d'une lettre dépend du statut de la précédente (événements non indépendants).

2.1 Chaînes de Markov discrètes

Considérons un processus stochastique $(X_t)_t$ à temps discret $\mathbb{T}$ et à espace d'états discret E.

Definition 10. *$(X_t)_t$ est dite chaîne de Markov (cm) si*

$$P(X_{n+1} = j/X_0 = i_0, \ldots ,X_{n-1} = i_{n-1},X_n = i) = P(X_{n+1} = j/X_n = i) = P_{ij}(n)$$

$$i_0,i_1, \ldots ,i_{n-1},i,j \in E,n \in \mathbb{T}$$

Ce qui signifie que l'évoultion future du processus ne dépend que de son passé le plus récent, i.e., que le présent contient toute l'information passée du processus. En d'autres termes, le processus a tendance à oublier son passé; c'est ce qu'on appelle un processus sans mémoire. $P_{ij}(n)$ est la probabilité que le système démarrant de l'état initial i_0 atteigne l'état j au bout de $n + 1$ transitions.

Exemple

Soit $(X_n)_n$ une suite positive de variables aléatoires indépendantes et équidistribuées de loi P donnée par

$$P(X_n = k) = q_k, \sum_{k\geq 0} q_k = 1$$

Soit $Y_n = \sum_{i=0}^{n} X_i$, alors $(Y_n)_n$ est une chaîne de Markov.
Remarquons dabord que $Y_{n+1} = Y_n + X_{n+1}$.

$$P(Y_{n+1} = j/Y_0 = i_0, \ldots ,Y_{n-1} = i_{n-1},Y_n = i) = P(X_{n+1} = j-i/X_0 = i_0,\ X_1 = i_1-i_0 \ldots,\ X_n = i-i_{n-1}$$

Du fait que les variables aléatoires X_n sont indépendantes, le dernier terme est égal à $P(X_{n+1} = j - i) = P(Y_{n+1} = j/Y_n = i)$
$(Y_n)_n$ est donc une chaîne de Markov.

Pour l'instant, nous allons nous intéresser au cas des chaînes de Markov à valeurs dans un ensemble fini E, qui sont homogènes dans le temps. Leur loi est entièrement d'écrite par une matrice carrée (la matrice de transition de la chaîne), et leur étude se ramène alors essentiellement à des problèmes d'algèbre linéaire.

Definition 11. *La chaîne de Markov $(X_n)_n$ est dite Homogène (dans le temps) ou à transitions stationnaires si*

$$P(X_{n+1} = j/X_n = i) = P(X_n = j/X_{n-1} = i) \dots, = P(X_1 = j/X_0 = i) = P_{ij}$$

Ce qui signifie que la probabilité de passer de l'état i l'état j ne dépend pas du temps. P_{ij} est la probabilité de passage de l'état i à j en une étape.

Remarque 1
• $P_{ij} \geq 0, \forall i,j$ et $\sum_{j\in E} P_{ij} = 1 \; \forall i$
Du fait que X_1 est discrète, $\Omega = \bigcup_{j\in E}(X_1 = j)$.

$$P(\Omega/X_0 = i) = 1 = P(\bigcup_{j\in E}(X_1 = j/X_0 = i) = \sum_{j\in E} P(X_1 = j/X_0 = i) = \sum_{j\in E} P_{ij}$$

Definition 12. *Matrice de Transition*
Soit $(X_n)_n$ une chaîne de Markov homogène à espace d'états E fini.On appelle matrice de transition P associée à $(X_n)_n$ la matrice $P = (P_{ij})_{i,j\in E}$.

Si $E = \{1,2,\dots,n\}$, la matrice P s'écrit

$$P = \begin{pmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{n1} & \cdots & P_{nn} \end{pmatrix}$$

Definition 13. *Matrice Stochastique*
Une matrice $A = (a_{ij})_{i,j}$ est dite stochastique si
• $a_{ij} \geq 0, \forall i,j$
•• $\sum_j a_{ij} = 1, \forall i$

Proposition 3. *1-Si A et B sont deux matrices stochastiques, alors le produit $A \times B$ est aussi une matrice stochastique.*
2- Toute matrice stochastique admet la valeur $\lambda = 1$ comme valeur propre. C'est la seule laquelle on peut associer une distribution de probabilité.
3- Toute valeur propre λ est de module inférieur ou égal à 1.

Démonstration

1- Soient $A = (a_{ik})_{i,k}$, $1 \leq i \leq n$, $1 \leq k \leq p$ et $B = (b_{kj})_{k,j}$, $1 \leq k \leq p$, $1 \leq j \leq q$ deux matrices stochastiques. Notons $C = A \times B = (c_{ij})_{i,j}$

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}, \forall i,j$$

$$\sum_{j=1}^{q} c_{ij} = \sum_{j=1}^{q}\sum_{k=1}^{p} a_{ik} b_{kj} = \sum_{k=1}^{p} a_{ik} \sum_{j=1}^{q} b_{kj}$$

Du fait que A et B sont stochastiques $\sum_{j=1}^{q} b_{kj} = \sum_{k=1}^{p} a_{ik} = 1$.
Ce qui donne $\sum_{j=1}^{q} c_{ij} = 1$, $\forall i$.
2- la preuve est simple et elle est laissée au lecteur.
3- Soit λ une valeur propre d'une matrice stochastique P et notons π le vecteur propre associé.
On a $\lambda\pi_i = \sum_j P_{ij}\pi_j \ \forall i$.Ce qui implique que $|\lambda||\pi_i| \leq \sum_j P_{ij}|\pi_j| \ \forall i$
Notons $|\pi_k| = \max_i\{|\pi_i|\}$ alors $|\lambda||\pi_k| = \sum_j P_{ij}|\pi_k|$. Ce qui donne $|\lambda| \leq \sum_j P_{ij} = 1$ d'ou le résultat.

Exemple Soit P la matrice donnée par

$$P = \begin{pmatrix} 1/3 & 0 & 2/3 \\ 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 \end{pmatrix}$$

Les valeurs propres sont données par le calcul du déterminant

$$det(P-\lambda I) = \begin{vmatrix} \frac{1}{3}-\lambda & 0 & \frac{2}{3} \\ \frac{1}{4} & \frac{1}{4}-\lambda & \frac{1}{2} \\ 0 & 0 & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda & 0 & \frac{2}{3} \\ 1-\lambda & \frac{1}{4}-\lambda & \frac{1}{2} \\ 1-\lambda & 0 & 1-\lambda \end{vmatrix} = (1-\lambda)\begin{vmatrix} 1 & 0 & \frac{2}{3} \\ 1 & \frac{1}{4}-\lambda & \frac{1}{2} \\ 1 & 0 & 1-\lambda \end{vmatrix} =$$

$$(1-\lambda)\begin{vmatrix} 1 & 0 & \frac{2}{3} \\ 0 & \frac{1}{4}-\lambda & -\frac{1}{6} \\ 0 & 0 & \frac{1}{3}-\lambda \end{vmatrix} = (1-\lambda)(\tfrac{1}{4}-\lambda)(\tfrac{1}{3}-\lambda).$$

Les valeurs propres sont $\lambda_1 = 1$, $\lambda_2 = \frac{1}{4}$, $\lambda_3 = \frac{1}{3}$
le vecteur propre associé à la valeur propre λ_1 est $\pi_1 = (0,0,1)$. On voit bien que les assertions de la proposition précédente sont vérifiées.

Remarques 2
1) Une matrice de transition P est une matrice stochastique et P admet la valeur $\lambda = 1$ comme valeur propre.
2) Il existe un vecteur propre à gauche π associé à la valeur propre $\lambda = 1$ donné par $\pi P = \pi$ qui définit une distribution de probabilité.
si $\pi = (\pi_1,\pi_2,\ldots,\pi_n)$, Pour déterminer π il suffit de résoudre l'équation $\pi P = \pi$ ce qui est équivalent à

$$\pi_i = \sum_{j=1}^{n} P_{ij}\pi_j \ \forall i,\ \pi_i \geq 0,\ \sum_{j=1}^{n} \pi_j = 1$$

3) La valeur propre $\lambda = 1$ est la seule valeur propre à laquelle on peut associer une distribution de probabilité(vecteur propre associé).

Definition 14. *Graphe de transitions*
Soit $(X_n)_n$ une chaîne de Markov homogène à espace d'états E fini de matrice de transition $P = (P_{ij})_{i,j\in E}$. On associe à P un graphe $G = \{(i,j),P_{i,j}\}$ dit graphe de transitions.

Le graphe d'une chaîne de Markov à temps discret est défini comme un graphe orienté dans lequel les noeuds représentent les états du processus et les arcs correspondant aux transitions, i.e., les probabilités de transition strictement positives.

Remarque 3
Si $(X_n)_n$ est une chaîne de Markov et A est un événement engendré par les variables $(X_1,X_2,\ldots,X_n)$, alors

$$(P(X_{n+1}=j/\ A,X_n=i)=P(X_{n+1}=j/X_n=i)$$

Definition 15. *Probabilité de transition en n-étapes*
On appelle probabilité de transition en n-étapes, la probabilité de passer de l'état i à l'état j en n étapes, probabilité notée $P_{ij}^{(n)}=P(X_n=j/X_0=i)$ *avec* $P_{ij}^0=\delta_{ij}=\begin{cases}1 & si\ i=j\\ 0 & sinon.\end{cases}$

La matrice de transitions en n- étapes est la matrice.
Il est intéressant de voir s'il y a un lien entre les matrices P et P_n. Le résultat suivant nous apporte une réponse.

Théorème 1. *Soit* $P_n=(P_{ij}^{(n)})_{i,j}$ *la matrice de transitions en n-étapes associée à une chaîne de Markov homogène finie de matrice de transition* P*, alors* $P_n=P^n$*.*
Preuve
- Si $n=1$ évident du fait que $P_1=P^1=P$
- Si $n=2$,

$$P_{ij}^{(2)}=P(X_2=j/X_0=i)=P(X_2=j,\bigcup_{k\in E}(X_1=k)/X_0=i)=\sum_{k\in E}P(X_2=j,(X_1=k)/X_0=i)$$

$$=\sum_{k\in E}P(X_1=k/X_0=i)P(X_2=j/X_0=i,X_1=k)=\sum_{k\in E}P(X_1=k/X_0=i)P(X_2=j/X_1=k)$$

$=\sum_{k\in E}P_{ik}P_{kj}=PP=P^2$

Supposons que la propriétè est vraie à l'ordre n, montrons qu'elle est vraie à l'ordre $n+1$.

$$\begin{aligned}P_{ij}^{(n+1}&=P(X_{n+1}=j/X_0=i)=P(X_{n+1}=j,\bigcup_{k\in E}(X_n=k)/X_0=i)\\&=\sum_{k\in E}P(X_{n+1}=k/X_n=k,X_0=i)P(X_n=k/X_0=i)\\&=\sum_{k\in E}P(X_{n+1}=k/X_n=k)P(X_n=k/X_0=i)=\sum_{k\in E}P_{ik}^{(n)}P_{kj}\end{aligned}$$

Ce qui donne que $P_n=P^nP=P^{n+1}$, d'où le résultat.

Théorème 2. *Equation de Kolmogorov-Chapman*

$$\forall m,n\in\mathbb{N}^\star,\ P_{ij}^{(n+m)}=\sum_{k\in E}P_{ik}^{(m)}P_{kj}^{(n)},\ \forall i,j\in E$$

Ce qui signifie que $P_{n+m}=P^nP^m=P^{n+m}$

Preuve

$$P_{ij}^{(n+m)} = P(X_{n+m} = j/X_0 = i) = \sum_{k\in E} P(X_{n+m} = j, X_n = k/X_0 = i)$$
$$= \sum_{k\in E} P(X_{n+m} = j/X_n = k)P(X_n = k/X_0 = i) = \sum_{k\in E} P_{ik}^{(n)} P_{kj}^{(m)}$$

Ce qui donne $P_{n+m} = P^n P^m = P^{n+m}$

Dans tout ce qui précède, on a utiliser que des probabilités conditionnelles. Si nous voulons calculer la probabilité d'un état i à l'instant n, il est nécessaire de connaître la loi initiale processus.

Loi d'une chaîne de Markov

Notons $\Pi_k(n) = P(X_n = k)$, $n \geq 0$, $k \geq 1$.
Nous pouvons écrire la distributionde X_n sous forme d'un vecteur ligne

$$\Pi(n) = (\Pi_1(n), \Pi_2(n), \dots,) \text{ avec} \sum_{j\geq 1} \Pi_j(n) = 1$$

Pour déterminer $\Pi(n)$, il est nécessaire de connaître la loi initiale du processus, soit sa distribution
$\Pi(0) = (\Pi_1(0), \Pi_0(n), \dots,)$ avec $\Pi_k(0) = P(X_0 = k)$

$$\Pi_k(n) = P(X_n = k) = P(X_n = k, \bigcup_{j\in E}(X_0 = j)) = \sum_{j\in E} P(X_n = k, X_0 = j)$$
$$= P(X_0 = j)P(X_n = k/X_0 = j) = \sum_{j\in E} \Pi_j(0)P_{jk}^{(n)}$$

Ce qui donne $P(X_n = k) = \sum_{j\in E} \Pi_j(0)P_{jk}^{(n)}$ et donc $\Pi(n) = \Pi(0)P^n$, $\forall n \geq 1$.

Ce qui signifie que la loi de X_n est entièrement déterminée par sa loi initiale et sa matrice de transition.

Remarque 4

$\Pi(n+1) = \Pi(0)P^{n+1} = \Pi(0)P^n P = \Pi(n)P$, $\forall n \geq 1$