Chapitre 3 : Processus stochastiques, exemple des chaînes de Markov

3.1 Processus stochastiques : définitions

Définition (3.1.1)

On appelle processus aléatoire (ou stochastique) une famille $X = (X_\theta)_{\theta\in\Theta}$ de variables aléatoires X_θ à valeurs dans $(E, \mathcal{E})$ et définies sur un même espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ (c'est à dire que $X_\theta : \Omega \to E$ est une application mesurable pour tout $\theta \in \Theta$), indexée par $\theta \in \Theta$ où Θ est un ensemble.
Si Θ est fini ou dénombrable (ex : $\Theta = \mathbb{N}, \mathbb{Z}, \ldots$) on parle de processus à temps discret.
Si Θ a la puissance du continu (ex : $\Theta = \mathbb{R}_+, \mathbb{R}, \mathbb{R}^2, \ldots$) on parle de processus à temps continu.
L'espace E est appelé l'espace d'état du processus X (on dit aussi que le processus X est à valeurs dans E).

Remarque 3.1.1 : [au tableau]

Grenoble INP Ensimag

Définition (3.1.2)

Un processus $X = (X_\theta)_{\theta\in\Theta}$ *est dit stationnaire si pour tout* $n \in \mathbb{N}^*$, *pour tous* $\theta_1, \ldots, \theta_n \in \Theta$, *et pour tout* $h \in \Theta$ *avec* $\theta_i + h \in \Theta$, $\forall 1 \leq i \leq n$, *les vecteurs* $(X_{\theta_1+h}, \ldots, X_{\theta_n+h})'$ *et* $(X_{\theta_1}, \ldots, X_{\theta_n})'$ *ont même loi.*

Définition (3.1.3)

Un processus $X = (X_\theta)_{\theta\in\Theta}$ *est à accroissements stationnaires si la loi de* $X_{t+\theta} - X_t$ *ne dépend pas de* t, *pour tous* $\theta, t \in \Theta$ *t.q.* $t + \theta \in \Theta$.

Dorénavant on suppose qu'il existe une relation d'ordre sur Θ : l'indice $\theta \in \Theta$ représente le temps (typiquement on a $\Theta = \mathbb{N}$ ou $\Theta = \mathbb{R}_+$).

Définition (3.1.4)

Un processus $X = (X_\theta)_{\theta\in\Theta}$ *est dit à accroissements indépendants si pour tous* $\theta_1 < \ldots < \theta_k$ *dans* Θ *les v.a.* $X_{\theta_2} - X_{\theta_1}, \ldots, X_{\theta_k} - X_{\theta_{k-1}}$ *sont mutuellement indépendantes.*

Définition (3.1.5)

Une filtration $(\mathcal{F}_\theta)_{\theta\in\Theta}$ *sur un espace de probabilités* $(\Omega, \mathcal{F}, \mathbb{P})$ *c'est un ensemble croissant de sous-tribus* $\mathcal{F}_\theta$ *de* $\mathcal{F}$*, i.e. pour tous* $s \leq t$ *dans* Θ*, on a* $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$*.*

Définition (3.1.6)

On dit qu'un processus $X = (X_\theta)_{\theta\in\Theta}$ *est adapté à une filtration* $(\mathcal{F}_\theta)_{\theta\in\Theta}$ *si pour tout* $\theta \in \Theta$ *la v.a.* X_θ *est* $\mathcal{F}_\theta$*-mesurable.*

Exemple 3.1.1 : (filtration naturelle). Soit $X = (X_\theta)_{\theta\in\Theta}$ un processus aléatoire. Définissons pour tout $\theta \in \Theta$ la tribu $\mathcal{F}^X_\theta$ par $\mathcal{F}^X_\theta = \sigma(X_s : s \leq \theta)$, où $\sigma(X_s : s \leq \theta)$ est la plus petite tribu qui rend mesurable tous les X_s pour $s \leq \theta$ (cette tribu existe car on peut la définir comme étant la plus petite qui contient $\mathcal{C}_\theta = \{X_s^{-1}(A),\ A \in \mathcal{E},\ s \leq \theta\}$).

☛ $(\mathcal{F}^X_\theta)_{\theta\in\Theta}$ est une filtration et X est adapté à $(\mathcal{F}^X_\theta)$ [au tableau]

Définition (3.1.7)

Soit $(\mathcal{F}_\theta)_{\theta\in\Theta}$ une filtration et $X=(X_\theta)_{\theta\in\Theta}$ un processus qui lui est adapté. On dit que X est $(\mathcal{F}_\theta)$-Markov si pour tous $s<t$ dans Θ, et tout Γ dans $\mathcal{E}$ on a

$$\mathbb{P}(X_t\in\Gamma\,|\,\mathcal{F}_s)=\mathbb{P}(X_t\in\Gamma\,|\,X_s).$$

Exercice 3.1.1 : Montrer que si $X=(X_\theta)_{\theta\in\Theta}$ est $(\mathcal{F}_\theta)$-Markov pour une certaine filtration $(\mathcal{F}_\theta)_{\theta\in\Theta}$, alors il est $(\mathcal{F}^X_\theta)$-Markov.

Compte tenu de l'exercice 3.1.1 un processus de Markov X vérifie, pour tous $s<t$ dans Θ, et tout Γ dans $\mathcal{E}$,

$$\mathbb{P}(X_t\in\Gamma\,|\,\mathcal{F}^X_s)=\mathbb{P}(X_t\in\Gamma\,|\,X_s).$$

Cela signifie que, pour un processus de Markov X et $s<t$, la loi de X_t connaissant le comportement du processus jusqu'à l'instant s compris, ne dépend que de sa position à l'instant s : on oublie tout ce qui s'est passé juste avant l'instant s.

Remarque 3.1.2 : Dans le contexte où E est fini ou dénombrable le caractère $(\mathcal{F}_\theta^X)$-Markov de X peut s'écrire de façon équivalente : pour tout $n \in \mathbb{N}$, tous $0 \leq t_1 \leq \ldots \leq t_n \leq t_{n+1}$, et tous $x_1, \ldots, x_n, x_{n+1} \in E$ on a

$$\mathbb{P}(X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n, \ldots, X_{t_1} = x_1) = \mathbb{P}(X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n)$$

dès que $\mathbb{P}(X_{t_n} = x_n, \ldots, X_{t_1} = x_1) > 0$.

Définition (3.1.8)

Un processus de Markov $X = (X_\theta)_{\theta \in \Theta}$ est dit de Markov homogène si pour tous $s < t$ dans Θ, et tout $\Gamma \in \mathcal{E}$ on a

$$\mathbb{P}(X_t \in \Gamma \,|\, X_s) = g_\Gamma(t - s, X_s)$$

où la fonction $g_\Gamma(t - s, \cdot)$ dépend de l'écart $t - s$ mais pas de s.
Dans ce cas on a pour tous $s < t$ dans Θ et tout $x \in E$,

$$\mathbb{P}(X_t \in \Gamma \,|\, X_s = x) = g_\Gamma(t - s, x) = \mathbb{P}(X_{t-s} \in \Gamma \,|\, X_0 = x).$$

Remarque 3.1.4 : A nouveau dans le cas où E est discret la définition de Markov homogène admet une variante : pour tous $s < t$ dans Θ et tous x, y dans E

$$\mathbb{P}(X_t = y | X_s = x) = p_{t-s}(x, y)$$

où $p_{t-s}(x, y)$ dépend de x, de y et de l'écart $t - s$ mais pas de l'instant s.

... ☛...

3.2 Chaînes de Markov

Une chaîne de Markov c'est un processus de Markov homogène à valeurs dans E au plus dénombrable et indexé par $n \in \mathbb{N}$ (i.e. $X = (X_n)_{n \in \mathbb{N}}$; on notera aussi souvent $X = (X_n)_{n \geq 0}$, le n indiquant la nature discrète du temps).

Proposition (3.2.1)

Pour une chaîne de Markov $X = (X_n)_{n \geq 0}$ la propriété de Markov homogène s'écrit de façon équivalente : pour tout $n \in \mathbb{N}^*$, et tous $x_0, x_1, \ldots, x_n, x_{n+1} \in E$

$$\begin{aligned} \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_1 = x_1, X_0 = x_0) &= \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n) \\ &= \mathbb{P}(X_1 = x_{n+1} | X_0 = x_n) \end{aligned}$$

dès que $\mathbb{P}(X_n = x_n, \ldots, X_1 = x_1, X_0 = x_0) > 0$.

Exemple 3.2.1 : Soit $(\xi_i)_{i \geq 1}$ une suite i.i.d. de v.a. à valeurs dans $\mathbb{Z}$. Le processus $(S_n)_{n \geq 0}$ défini par

$$S_0 = 0 \quad \text{et} \quad S_n = \sum_{i=1}^{n} \xi_i \;\; \forall n \geq 1$$

est une chaîne de Markov (cf Exercice 2 de la feuille de TD 4). C'est une marche aléatoire sur $\mathbb{Z}$.

3.2.1 Notion de matrice de transition

Ainsi le comportement d'une chaîne de Markov va être déterminé par la donnée

- de la loi de sa position initiale X_0
- de ses probabilités de transition, i.e. les quantités du type $\mathbb{P}(X_1 = y | X_0 = x)$.

Cela nous amène à la notion de matrice de transition (ou matrice stochastique) d'une chaîne de Markov : soit $X = (X_n)_{n \geq 0}$ une chaîne de Markov à valeurs dans E, on pose pour tous $x, y \in E$

$$Q(x, y) = \mathbb{P}(X_1 = y | X_0 = x).$$

On appelle $Q = \{Q(x, y)\}_{(x,y) \in E^2}$ la matrice de transition de la chaîne X.

Notons qu'on a $\sum_{y \in E} Q(x, y) = 1$ pour tout $x \in E$ (les lignes d'une matrice de transition somment toujours à 1).

Proposition (3.2.2)

Soit $X = (X_n)_{n\geq 0}$ *une chaîne de Markov de matrice de transition* Q*. Alors pour tous* $n \geq 1$ *et* $x, y \in E$ *on a*

$$\mathbb{P}(X_n = y \,|\, X_0 = x) = Q^n(x, y)$$

(ici Q^n *désigne simplement le produit matriciel constitué par* Q *multipliée* $n-1$ *fois par elle-même).*
En particulier on a les équations de Chapman-Kolmogorov : pour tous $x, y \in E$ *et tous* $m, n \in \mathbb{N}$*,*

$$\mathbb{P}(X_{m+n} = y \,|\, X_0 = x) = \sum_{z\in E} \mathbb{P}(X_m = z | X_0 = x)\mathbb{P}(X_n = y | X_0 = z)$$

ou, écrit matriciellement,

$$\forall x, y \in E, \;\; \forall m, n \in \mathbb{N}, \quad Q^{m+n}(x, y) = \sum_{z\in E} Q^m(x, z) Q^n(z, y).$$

3.2.2 Phénomène d'équilibre

Donc $Q^n(x, y)$: probabilité que la chaîne soit à l'état y à l'instant n, sachant qu'elle est partie de x.

On verra en TD (Exercice 1 de la Feuille 5) que, pour $|E| = k < \infty$ et sous certaines conditions supplémentaires, on a le phénomène de convergence

$$Q^n \xrightarrow[n\to\infty]{} \begin{pmatrix} p_1 & \dots & p_k \\ \vdots & & \vdots \\ p_1 & \dots & p_k \end{pmatrix}$$

où $\sum_{i=1}^k p_i = 1$.

Noter que les lignes de la matrice limite sont toutes les mêmes !

Cela signifie que, en temps long, le système est distribué selon μ définie par $\mu(i) = p_i$, $1 \leq i \leq k$, et ce, quelle que soit sa position de départ.
Le mesure μ apparait comme la "mesure d'équilibre" ou "distribution à l'équilibre" de la chaîne (il s'agit de théorie des mesures invariantes et de théorie ergodique...).