Stationnarité des processus

Bernard Delyon

1 Définition. Généralités

On ne considérera dans la suite que des processus indexés par $\mathbb{N}$ ou $\mathbb{Z}$.

1 - Definition

Un processus $(X_i)_{i\geqslant 1}$ est stationnaire si pour tout $p \geq 1$, $(X_1, \ldots X_p)$ a même loi que $(X_2, \ldots X_{p+1})$.
Un processus $(X_i)_{i\in\mathbb{Z}}$ est stationnaire si pour tout $p \geq 1$, $(X_{-p}, \ldots X_p)$ a même loi que $(X_{-p+1}, \ldots X_{p+1})$.

Cette définition est un peut minimale. Considérons le cas des processus indexés par $\mathbb{N}$. Il faut bien voir que la définition implique que $(X_k, \ldots X_p) \sim (X_{k+l}, \ldots X_{p+l})$ pour tous $k \leqslant p$, $l > 0$. En effet, par transitivité, il suffit de le montrer pour $l = 1$. Notons que comme $(X_1, \ldots X_p) \sim (X_2, \ldots X_{p+1})$, on obtient en particulier que $(X_k, \ldots X_{p-1}) \sim (X_{k+1}, \ldots X_p)$, ce qui montre bien le résultat pour $l = 1$.

Mesures sur $\mathbb{R}^{\mathbb{N}}$. Se donner une suite infinie de variables aléatoires indexées par $\mathbb{N}$, c'est se donner une distribution sur $\mathbb{R}^{\mathbb{N}}$ muni de la tribu $\mathscr{B}_\infty$ engendrée par la famille $\mathscr{C}$ des ensembles définis par un nombre fini de coordonnées, c-à-d de la forme $\{(x_1, \ldots x_n) \in A\}$, $A \in \mathscr{B}(\mathbb{R}^n)$ (tribu de Borel). La loi d'un processus est donc par définition caractérisée par ses distributions finidimensionnelles ; la stationnarité se résume donc en d'autres termes à :

$$(X_i)_{i\geqslant 1} \sim (X_{i+1})_{i\geqslant 1} \quad : \quad \textit{Un décalage temporel n'affecte pas la distribution.}$$

Rappelons au passage un résultat qui sera utilisé dans la suite :

2 - Théorème

Pour toute probabilité P sur $(\mathbb{R}^{\mathbb{N}}, \mathscr{B}_\infty)$, l'ensemble des fonctions ne dépendant que d'un nombre fini de coordonnées est dense dans $L_1(P)$.
On peut même se restreindre aux fonction C^∞ à support compact, ou aux fonctions étagées.

Démonstration: Montrons d'abord la densité des fonctions étagées basées sur une algèbre génératrice $\mathscr{C}$; ce dernier point se vérifie simplement en observant d'abord que les ensembles de $\mathscr{B}_\infty$ dont l'indicateur peut être approché dans L_1 par une suite d'indicateurs d'ensembles de $\mathscr{C}$ forment une tribu, et donc forment tous les ensembles de $\mathscr{B}_\infty$. le dernier point vient de la densité des fonctions C^∞ à support compact dans $L_1(\mathbb{R}^d, P)$ pour tout probabilité P. ∎

3 - Proposition

Si $(X_i)_{i\geqslant 1}$ est stationnaire et φ une application mesurable, alors $Y_k = \varphi(X_k, X_{k+1}, \ldots)$ aussi. De même si $(X_i)_{i\in\mathbb{Z}}$ est stationnaire, alors $Y_k = \psi(\ldots, X_{k-1}, X_k, X_{k+1}, \ldots)$ également.

Démonstration: La famille des ensembles $A \subset \mathbb{R}^{\mathbb{N}}$ tels que la suite $Y_k = 1_{(X_k, X_{k+1}, \dots) \in A}$ soit stationnaire constitue une tribu (élémentaire). Comme elle contient la famille $\mathscr{C}$ définie plus haut, elle contient toute la tribu. La propriété reste donc vraie si φ est étagée, puis s'étend à toute φ mesurable par approximation par des fonctions étagées (en tronquant φ à $[-n, n]$ et en arrondissant au plus proche multiple de $1/n$). On procède de même avec ψ. ∎

Processus autoregressif d'ordre 1. Soit $(X_n)_{n>0}$ une suite i.i.d. de v.a. intégrables, α de valeur absolue < 1 et

$$Y_n = \sum_{j=0}^{\infty} \alpha^j X_{n-j} = \alpha Y_{n-1} + X_n$$

Alors Y est stationnaire.

Le modèle autorégressif à moyenne mobile (ARMA). Il est donné par la formule suivante :

$$Y_n = \sum_{k=1}^{p} a_k Y_{n-k} + \varepsilon_n + \sum_{k=1}^{q} b_k \varepsilon_{n-k}, \tag{1}$$

où les ε_n sont des $\mathcal{N}(0, \sigma^2)$ indépendantes. Les paramètres sont les a_k, b_k et σ. On va voir que par un bon choix des conditions initiales, on peut rendre ce processus stationnaire.

Si l'on note

$$Z_n = \begin{pmatrix} Y_n \\ \vdots \\ Y_{n-p+1} \end{pmatrix}, \ \eta_n = \begin{pmatrix} \varepsilon_n \\ \vdots \\ \varepsilon_{n-q} \end{pmatrix}, \ A = \begin{pmatrix} a_1 & a_2 & \dots & a_p \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix}, \ B = \begin{pmatrix} 1 & b_1 & \dots & b_q \\ 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 0 \end{pmatrix}$$

on a

$$Z_n = AZ_{n-1} + B\eta_n$$

ce qui permet de faire certains calculs de façon analogue au cas $p = 1, q = 0$, en particulier de représenter la loi stationnaire par

$$Z_n = B\eta_n + AB\eta_{n-1} + A^2 B\eta_{n-2} + \dots,$$

et également de voir que $\widetilde{Z}_n = (Z_n, \varepsilon_n, \dots \varepsilon_{n-q+1})$ admet la représentation markovienne $\widetilde{Z}_n = \widetilde{A}\widetilde{Z}_{n-1} + \widetilde{B}\varepsilon_n$.

L'opérateur de décalage. Soit $X = (X_i)_{i \geqslant 1}$ un processus et une variable aléatoire Y X-mesurable, soit $Y = f(X)$, par exemple

$$f(X) = X_3 + 2\cos(X_7)$$

L'opérateur de décalage T définit TY comme la valeur obtenue en décalant X :

$$Tf(X) = X_4 + 2\cos(X_8).$$

La stationnarité n'est autre que de dire TY a la même loi que Y ; on dit que T préserve la mesure. Notons que comme T est une isométrie de L_1, l'avoir défini seulement sur les fonctions ne dépendant que d'un nombre fini des x_i permet de l'étendre de manière unique à tout L_1 car ces dernières sont denses (théorème 2). Dans les trois exemples

$$Y = \overline{\lim}\, X_i$$

$$Y = 1_{X_i \text{ prend la valeur 1 infiniment souvent}}$$

$$Y = \overline{\lim}\, \tfrac{1}{n} \sum_{i=1}^{n} X_i.$$

on a $TY = Y$.

Exemple : Fractions continues. Soit la suite X_n obtenue comme les termes du développement en fraction continue d'un nombre réel ξ sur $\Omega = [0,1]$ tiré aléatoirement selon la mesure $\frac{1}{\log 2}\frac{dx}{1+x}$:

$$\xi = \cfrac{1}{X_1 + \cfrac{1}{X_2 + \cdots}}.$$

Pour vérifier la stationnarité, il suffit de vérifier que ξ a même loi que

$$T\xi = \cfrac{1}{X_2 + \cfrac{1}{X_3 + \cdots}} = \xi^{-1} - [\xi^{-1}]$$

ce qui est laissé en exercice.

Exemple : Invariance de la mesure de Liouville pour un flot hamiltonien. Soit un corps en mouvement dont on note x la position et v la vitesse ; par exemple un pendule, x est l'angle et v la vitesse angulaire. On suppose qu'il est soumis à un potentiel $V(x)$ ($\cos(x)$ pour le pendule) ; l'énergie associée est $W(x,v) = \frac{1}{2}m\|v\|^2 + V(x)$, et en posant $y = (x,v)$, l'équation du mouvement est

$$m\dot{y}_t = m\frac{d}{dt}\begin{pmatrix} x_t \\ v_t \end{pmatrix} = \begin{pmatrix} mv_t \\ -\nabla V(x_t) \end{pmatrix}. \tag{2}$$

Noter que $W(y_t) = W(y_0)$. Soit $a \in \mathbb{R}$ tel que l'ensemble $E = \{y : W(y) \le a\}$ est compact (l'existence de a fait partie des hypothèses). Notons $y_t(y)$ la solution partant de $y_0 = y$. On montre que pour tout t la transformation $T_t : y_0 \mapsto y_t$ préserve la mesure de Lebesgue sur E, ce qui s'écrit ici :

$$\frac{d}{dt}\int_E f(y_t(y))\,dy = 0, \qquad \text{soit} \qquad \frac{d}{dt}\int_E f(x_t(x,v), v_t(x,v))\,dxdv = 0. \tag{3}$$

Considérons la suite y_n où y_0 est tiré uniformément sur E. Comme, par (3), $E[f(y_n)] = E[f(y_1)]$, on a stationnarité car les y_i étant fonction déterministe de y_1, toute fonction $\varphi(y_1 \ldots y_p)$ est en fait une fonction de y_1 seul. Une interprétation un peu différente est que si l'on part d'un point y_0 et que l'on considère un voisinage (une petite boule) autour, l'équation différentielle fait évoluer ce voisinage en le déformant mais en conservant son volume.

La démonstration de (3) se fait pour f régulière à support compact dans E, en faisant le calcul pour $t = 0$ dans un premier temps (faire une intégration par parties ; la clé est que la divergence du champ, membre de droite de (2), est nulle), puis en utilisant que pour $t_0 \neq 0$ on peut écrire en posant $g(y) = f(y_{t_0}(y))$

$$\frac{d}{dt}\int_E f(y_t(y))\,dy_{|t=t_0} = \frac{d}{ds}\int_E g(y_s(y))\,dy_{|s=0}$$

et le membre de droite est nul en utilisant (3) avec g au lieu de f.

2 Ergodicité

4 - Definition

Un suite stationaire X_i est dite ergodique si les seules variables Y telles que $Y = TY$ sont presque sûrement constantes.

Comme la stationnarité, c'est une propriété de la mesure P sur $\mathscr{B}_\infty$. On montre qu'il suffit de le vérifier pour les indicateurs d'ensembles. Si $T1_A = 1_A$ on dit que A est invariant, ce qui revient à $T^{-1}A = A$, et l'ergodicité signifie que tout ensemble invariant est de probabilité 0 ou 1. L'ergodicité est parfois difficile à vérifier car les variables Y à prendre en considération dépendent a priori de toute la suite. Toutefois une suite i.i.d. est stationnaire ergodique (conséquence du Corollaire 6 plus bas) et l'on verra au Corollaire 7 que les fonctions de ces suites le sont aussi (p. ex. le processus autorégressif présenté plus haut) ce qui fait déjà une très grande famille d'exemples.

Venons-en à l'un des théorèmes les plus importants de la théorie des probabilités :

5 - Théorème (Théorème ergodique de Birkoff)

Soit X_k une suite stationnaire ergodique. Pour toute fonction f mesurable telle que

$$E[|f(X_1)|] < \infty$$

on a

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \longrightarrow E[f(X_1)]$$

où la convergence a lieu presque surement et dans L_1.

La démonstration de ce théorème est reportée en appendice. Comme pour tout p la suite $\{(X_{k+1}, \dots X_{k+p})\}_{k>0}$ est stationnaire ergodique (corollaire 7 vérification élémentaire), pour toute fonction f mesurable telle que $E[|f(X_1, \dots X_p)|] < \infty$ on a

$$\frac{1}{n}\sum_{k=1}^{n} f(X_{k+1}, \dots X_{k+p}) \xrightarrow{L_1} E[f(X_1, \dots X_p)].$$

Réciproquement on a

6 - Corollaire

Soit X_k une suite stationnaire à valeurs dans $\mathbb{R}^d$. Si pour toute fonction $f \in C^\infty(\mathbb{R}^{pd})$ à support compact on a

$$\frac{1}{n}\sum_{k=1}^{n} f(X_{k+1}, \dots X_{k+p}) \xrightarrow{L_1} E[f(X_1, \dots X_p)]$$

alors pour toutes v.a. $Y = \varphi(X_1, X_2, \dots)$ intégrable on a

$$\frac{1}{n}\sum_{k=1}^{n} T^k Y \xrightarrow{L_1} E[Y] \tag{4}$$

et en particulier la suite est ergodique.

Démonstration: Comme T est une isométrie de $L_1(P)$, les applications linéaires $Y \mapsto \frac{1}{n}\sum_{k=1}^{n} T^k Y - E[Y]$ sont toutes de norme $\leqslant 2$. Par conséquent l'ensemble des v.a. Y $\mathscr{B}_\infty$-mesurables qui satisfont (4) est un fermé de $L_1(P)$, qui contient les $f \in C^\infty(\mathbb{R}^{pd})$ à support compact. En vertu du théorème 2, il contient $L_1(P)$. ∎

Les suites i.i.d. sont donc stationnaires ergodiques.

EXEMPLE. Soit λ irrationnel, ξ tiré uniformément sur $[0,1]$ et $X_n = \xi + n\lambda \pmod 1$. L'invariance par translation de la mesure de Lebesgue fait que la suite est stationnaire. Toute fonction de $X_1, \dots X_n$ est une fonction de ξ. Soit f une fonction bornée de ξ, alors elle appartient à $L_2([0,1])$ et admet un développement en série de Fourier :

$$f(x) = \sum_k c_n e^{2i\pi kx}, \quad c_k = \int_0^1 f(x)e^{-2i\pi kx}dx.$$

Comme $e^{2i\pi kT\xi} = e^{2i\pi k(\xi+\lambda)}$, le développement de Tf est

$$Tf(\xi) = \sum_k c_k e^{2i\pi k\lambda} e^{2i\pi n\xi}.$$

L'identité presque sûre entre f et Tf ne peut avoir lieu que si les coefficients de Fourrier coïncident, ce qui ne peut se produire que si $c_n = 0$ pour $n \neq 0$, c.-à-d. si f est constante. La transformation est ergodique.

Pour les fractions continues, on a également ergodicité mais c'est beaucoup plus difficile à montrer [2].

7 - COROLLAIRE

Si $(X_n)_{n\geqslant 1}$ est stationnaire ergodique, alors toute suite de la forme $Y_n = \varphi(X_n, X_{n+1}, \dots)$, $n \geqslant 1$, l'est encore.
De même si $(X_n)_{n\in\mathbb{Z}}$ est ergodique, alors $Y_n = \psi(\dots, X_{n-1}, X_n, X_{n+1}, \dots)$ également.

Démonstration: Il suffit d'appliquer le corollaire précédent. ∎

On peut ainsi fabriquer de nombreux processus stationnaires ergodiques à partir de suites de v.a.i.i.d. comme on l'a fait déjà pour les processus autorégressifs.

CHAÎNES DE MARKOV. Soit X_n est une chaîne de Markov à nombre fini d'états indécomposable (non nécessairement apériodique), c-à-d que la valeur propre 1 de sa matrice de transition est simple, ou encore qu'il n'existe aucune partition de l'ensemble d'états en deux ensembles non vides stables ($E_1 \rightsquigarrow E_1, E_2 \rightsquigarrow E_2$). Alors, il est classique que X_n admet une unique mesure invariante π et que l'on a convergence des moyennes

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow{L_1} \pi(f)$$

pour toute fonction f bornée et toute mesure initiale. Mais pour tout p, $(X_{k+1}, \dots X_{k+p})$ est encore une chaîne de Markov indécomposable (à vérifier !). Il s'ensuit que partant de la mesure invariante, on a bien un processus stationnaire ergodique. La proposition 3 et le corollaire 7 permettent d'en fabriquer ensuite bien d'autres.

3 Mélange fort

Il existe une propriété qui est plus forte que l'ergodicité et qui peut être plus simple à vérifier, c'est le mélange fort. Cette appellation est peu adéquate car il existe des propriétés de mélange plus spécifiques et plus fortes que le mélange fort.

8 - DEFINITION

On a mélange fort si pour toutes fonctions réelles mesurables bornées f et g sur $\mathbb{R}^p$ on a

$$E[f(X_1, ..X_p)g(X_{1+n}, \dots X_{p+n})] \longrightarrow E[f(X_1, ..X_p)]E[g(X_1, ..X_p)].$$

On montre par densité que ceci implique que pour deux v.a. X-mesurables Y et Z, on a

$$E[YT^nZ] \longrightarrow E[Y]E[Z]. \tag{5}$$

Il suffit de vérifier cette propriété pour f et g prises dans une famille totale de L_2.

Une chaîne de Markov non-apériodique n'est pas fortement mélangeante (exercice). La transformation $x \mapsto x + \lambda \pmod 1$ non-plus.

9 - Proposition

Le mélange fort implique l'ergodicité.

Démonstration: Choisir A invariant, appliquer (5) avec $Y = Z = 1_A$, et en déduire que $P(A) = 0$ ou 1. ∎

4 Exercices

Exercice 1. Soit $(X_n)_{n\geqslant 0}$ une suite i.i.d. $\mathcal{N}(0,1)$ et $(Y_n)_{n\geqslant 0}$ une suite i.i.d. $\mathcal{B}(1,p)$ indépendante de $(X_n)_{n\geqslant 0}$. Soit C une variable $\mathcal{B}(1,q)$, c.-à-d. $P(C=1) = 1 - P(C=0) = q$, indépendante des deux suites précédentes. Soit Z_n la suite construite ainsi

$$(Z_1, Z_2, ...) = \begin{cases} (X_1, Y_2, X_3, Y_4, ...) & \text{si } C = 0 \\ (Y_1, X_2, Y_3, X_4, ...) & \text{si } C = 1 \end{cases}$$

Cette suite induit une mesure sur $\mathbb{R}^{\mathbb{N}}$. On notera T la transformation $(TZ)_n = Z_{n+1}$.

(1) A quelle condition sur q la suite (Z_n) est-elle stationnaire ? Démontrer.

Dans toute la suite on suppose cette condition satisfaite.

(2) Démontrer que (Z_n) est ergodique. On utilisera le corollaire 6 : Considérer une fonction $f(Z_1, Z_2)$, et calculer la limite de $\frac{1}{n}\sum_{k=1}^n T^k f$ en remarquant que la suite $(X_1, Y_2), (X_3, Y_4), ...$ est stationnaire ergodique puisque i.i.d. et en décomposant la somme en plusieurs termes. Etendre aux fonctions de la forme $f(Z_1, Z_2, ..Z_k)$.

Exercice 2. Soit α un nombre réel, φ une variable aléatoire uniforme sur $[0, 2\pi]$ et la suite

$$X_n = \cos(2\pi n\alpha + \varphi).$$

1. Démontrer que X_n n'est pas ergodique si α est rationnel.
2. La suite X_n est-elle stationnaire ? (On démontrera)
3. On suppose α irrationnel
   (a) Soit $k \in \mathbb{Z}$. Quelle est la limite de $\frac{1}{n}\sum_{i=1}^n e^{2i\pi kn\alpha}$ lorsque n tend vers $+\infty$?
   (b) Soit f une fonction continue sur 2π-périodique. Donner la limite de $\frac{1}{n}\sum_{i=1}^n f(n\alpha)$.
   Indication : On commencera par le cas où f est un polynôme trigonométrique ($f(x) = \sum_{|k|\leqslant K} c_k e^{2i\pi kx}$).
   On rappelle que l'ensemble des polynômes trigonométrique est dense dans l'ensemble des fonctions continues 2π-périodiques avec la norme $\|.\|_\infty$.
   (c) La suite X_n est-elle ergodique ?

A Démonstration du théorème 5

Posons

$$\begin{aligned}
Y_n &= f(X_n) \\
S_n &= Y_1 + Y_2 + \dots Y_n
\end{aligned}$$

Commençons par un lemme :

10 - Lemme

Sous les hypothèses du théorème l'ensemble

$$A = \{\omega : \inf_n S_n = -\infty\}$$

satisfait

$$E[1_A Y_1] \leqslant 0.$$

En particulier si $E[Y_1] > 0$ alors $P(A) = 0$.

Démonstration: La dernière remarque vient de l'évidente contradiction si $P(A) = 1$.

Les fonctions $\varphi_p(\omega) = \inf(S_1, \dots S_p)$ satisfont

$$\begin{aligned}
\varphi_p =& \inf(S_1, \dots S_p) \\
=& Y_1 + \inf(0, TS_1, \dots TS_{p-1}) \\
=& Y_1 + \inf(0, T\varphi_{p-1}) \\
\geqslant& Y_1 + \inf(0, T\varphi_p) \\
=& Y_1 + T\varphi_p^-.
\end{aligned}$$

D'où,

$$\begin{aligned}
E[1_A Y_1] \leqslant& E[1_A \varphi_p] - E[1_A T\varphi_p^-] \\
=& E[1_A \varphi_p] - E[1_A \varphi_p^-] \quad \text{car } A \text{ est invariant} \\
=& E[1_A \varphi_p^+]
\end{aligned}$$

qui tend vers 0 lorsque p tend vers l'infini par définition de A (la suite φ_p^+ est décroissante bornée par Y_1^+). ∎

Poursuivons la démonstration du théorème. Soit $\varepsilon > 0$; posons

$$Z_n = Y_n - E[Y_1] + \varepsilon = f(X_n) - E[f(X_n)] + \varepsilon.$$

et appliquons le lemme à cette suite. Comme $E[Z_1] > 0$, $P(A) = 0$, et ceci implique en particulier que

$$\underline{\lim}_n \frac{Z_1 + \dots Z_n}{n} \geqslant 0$$

presque sûrement et donc

$$\underline{\lim}_n \frac{Y_1 + \dots Y_n}{n} \geqslant E[Y_1] - \varepsilon.$$

Comme ε est > 0 arbitraire, il s'ensuit que

$$\underline{\lim}_n \frac{Y_1 + \ldots Y_n}{n} \geqslant E[Y_1].$$

En appliquant ce même résultat à $-Y$ il vient

$$-\overline{\lim}_n \frac{Y_1 + \ldots Y_n}{n} \geqslant -E[Y_1].$$

Par conséquent $\frac{S_n}{n}$ converge vers $E[Y_1]$.

Pour la convergence dans $L_1(P)$, supposons, quitte à translater Y, que $E[Y_1] = 0$; on procède par troncature[1] :

$$\begin{aligned} n^{-1}E[|S_n|] \leqslant & n^{-1}E[|\sum Y_k 1_{|Y_k| \leqslant M}|] + n^{-1}E[|\sum Y_k 1_{|Y_k| > M}|] \\ \leqslant & E[n^{-1}|\sum Y_k 1_{|Y_k| \leqslant M}|] + E[|Y_1| 1_{|Y_1| > M}] \end{aligned}$$

par conséquent, en appliquant le théorème à $Y_k = Y_k 1_{|Y_k| \leqslant M}$, on obtient en vertu du convergence dominée

$$\overline{\lim}_n n^{-1}E[|S_n|] \leqslant |E[Y_1 1_{|Y_1| \leqslant M}]| + E[|Y_1| 1_{|Y_1| > M}] = |E[Y_1 1_{|Y_1| > M}]| + E[|Y_1| 1_{|Y_1| > M}]$$

qui tend vers 0 quand M tend vers l'infini. ∎

Références

[1] L. Breiman, *Probability*, Addison-Wesley, 1968.

[2] P. Billingsley *Ergodic theory and information*, Wiley, 1965.

[3] R. Durrett, *Probability theory and examples*, Duxbury, 1996.

[4] F. Merlevède, M. Peligrad, S. Utev, Recent advances in invariance principles for stationary sequences, *Probab. Surv.* 3 (2006), 1–36

1. On peut procéder autrement en remarquant que la suite S_n/n est uniformément intégrable, car appartenant à l'enveloppe convexe de $\{Y_1, Y_2, \ldots\}$ qui est une famille uniformément intégrable.