

B. R. G. M.
Département Géostatistique

Jean Serra C-13 B

15-11-65

COURS de PROBABILITES

(2ème Partie)

CHAPITRES V, VI, VII, VIII, IX et X

LES PROCESSUS STOCHASTIQUES

G. MATHERON

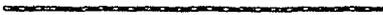
Octobre 1965

LES PROCESSUS STOCHASTIQUES

Table des Matières Pages

<u>CHAPITRE V.- LOIS INDEFINIMENT DIVISIBLES ET PROCESSUS A ACCROISSEMENTS INDEPENDANTS.</u>	
I.- Les lois indéfiniment divisibles	1
Théorème fondamental	5
II.- Processus à accroissements indépendants	10
Processus de Wiener-Levy	16
Processus de Poisson	18
 <u>CHAPITRE VI.- CHAINES DE MARKOV HOMOGENES A TEMPS DISCRET</u>	
I.- Généralités	23
II.- Classification des états	28
Etats essentiels et non essentiels	29
Etats transitoires et persistants	32
III.- Propriétés ergodiques des chaînes irréductibles	37
IV.- Etude des états transitoires	43
Calcul de la probabilité d'absorption	44
 <u>CHAPITRE VII.- CHAINES DE MARKOV HOMOGENES A TEMPS CONTINU</u>	
I.- Définitions et hypothèses	48
II.- Les équations de Kolmogorov	52
Première équation de Kolmogorov	54
Deuxième équation de Kolmogorov	55
Résolution des équations de Kolmogorov	56
III.- Les processus de renouvellement	59

<u>CHAPITRE VIII.- LES PROCESSUS MARKOVIENS CONTINUS</u>	62
I.- Définition	62
Equation de Markov	63
Hypothèses de continuité	64
II.- Les équations de Kolmogorov	66
Première équation de Kolmogorov	66
Théorème de Feller	68
Deuxième équation de Kolmogorov	69
<u>CHAPITRE IX.- LES PROCESSUS MARKOVIENS COMPLETEMENT DISCONTINUS</u>	74
I.- Généralités	74
II.- Les équations de Feller	75
III.- Loi de probabilité du nombre de sauts	77
<u>CHAPITRE X.- LES FONCTIONS ALEATOIRES STATIONNAIRES</u>	79
I.- Généralités	79
II.- Processus stationnaires d'ordre 2	81
Moments de la loi temporelle	82
La covariance $K(h)$	84
III.- Dérivation stochastique	85
IV.- Intégration en moyenne quadratique	89
Convolution stochastique	93
V.- Analyse harmonique d'un processus stationnaire	95



CHAPITRE V

LOIS INDEFINIMENT DIVISIBLES ET PROCESSUS A ACCROISSEMENTS INDEPENDANTS

Les lois indéfiniment divisibles se situent, en quelque sorte, à la charnière entre l'outil probabiliste classique, tel qu'il a été présenté dans la première partie de ce cours, et la théorie des processus stochastiques, qui constitue la partie la plus vivante, aujourd'hui encore en pleine évolution, et aussi la plus féconde en applications futures, du calcul des probabilités.

Jusque vers les années 30, les recherches en probabilités, avaient comme objectif principal de déterminer sous quelles conditions (aussi générales que possibles) des sommes de variables indépendantes ou faiblement corrélées convergent en loi vers une loi de Gauss ou une autre loi limite. Le théorème fondamental de P. LEVY, qui donne la forme générale de toutes les lois indéfiniment divisibles à variance finie, et que nous établirons ci-dessous a permis à Kolmogorov, parmi d'autres, d'énoncer sous forme à peu près définitive des conditions nécessaires et suffisantes de convergence vers une loi limite, gaussienne ou non. Malgré le très grand intérêt de ces résultats, nous n'insisterons pas sur cet aspect des lois indéfiniment divisibles. Le centre d'intérêt, en effet, aussi bien pour les recherches théoriques que pour les applications, s'est déplacé en direction des processus stochastiques et des fonctions aléatoires. Le même théorème de P. LEVY, qui marquait la fin d'une période de l'histoire des probabilités, ouvre la voie à la période suivante, en tant qu'il permet de caractériser complètement la classe très importante des processus à accroissements indépendants. Ce deuxième aspect des lois indéfiniment divisibles sera seul abordé ici, et constituera une transition naturelle entre les deux parties de ce cours (1).

I.- LES LOIS INDEFINIMENT DIVISIBLES.

On sait que toutes les sommes de variables indépendantes ne convergent pas nécessairement en loi vers une variable normale. Par exemple, dans certaines conditions définies, la loi binomiale converge vers la loi de Poisson. Mais alors on

(1) - La théorie des lois indéfiniment divisibles et de la convergence en loi des sommes de variables aléatoires est traitée dans Gnedenko " théorie des probabilités", Moscou 1961 - Traduction française en cours d'impression chez Dunod, Paris -.

peut se demander, inversement, si n'importe quelle loi peut jouer ce rôle de loi limite. En fait, il n'en est rien. Moyennant des hypothèses assez générales, on montre que toute loi limite, obtenue à partir d'une somme de variables indépendantes, est une sorte de mélange de lois de Poisson et de lois normales, cette sorte de mélange portant le nom de loi indéfiniment divisible. De là provient l'importance des lois de Gauss et de Poisson, comme prototypes de la classe beaucoup plus vaste de ces lois limites. Par ailleurs, ces lois indéfiniment divisibles jouent, nous le verrons, un rôle fondamental dans la théorie des processus stochastiques à accroissements indépendants, dont les deux types fondamentaux seront l'un à loi gaussienne (processus de Wiener-Levy), l'autre à loi de Poisson (processus poissoniens).

Définition : Loi indéfiniment divisible. Une fonction de répartition $F(x)$ est dite indéfiniment divisible si, quel que soit l'entier n positif, il existe une fonction de répartition $F_n(x)$ telle que la somme

$$X = X_1 + X_2 + \dots + X_n$$

de n variables indépendantes de même loi $F_n(x)$ obéisse à la loi $F(x)$.

D'après les propriétés élémentaires des fonctions caractéristiques, on voit qu'une loi $F(x)$ est indéfiniment divisible si, et seulement si, quel que soit n , sa fonction caractéristique $\bar{\Phi}(u)$ peut se mettre sous la forme

$$\bar{\Phi}(u) = \left[\bar{\Phi}_n(u) \right]^n$$

$\bar{\Phi}_n(u)$ étant une fonction caractéristique. Énonçons :

Théorème 1 - Pour qu'une loi $F(x)$ de fonction caractéristique $\bar{\Phi}(u)$ soit indéfiniment divisible, il faut et il suffit que, quel que soit l'entier $n > 0$, $[\bar{\Phi}(u)]^{1/n}$ soit une fonction caractéristique.

Exemples : la loi normale $\bar{\Phi}(u) = e^{i u m - \frac{\sigma^2}{2} u^2}$, la loi de Poisson $\bar{\Phi}(u) = e^{\theta(e^{iu} - 1)}$,

la loi Gamma $\bar{\Phi}(u) = \frac{1}{(1 - i \frac{u}{b})^\alpha}$ sont des lois indéfiniment divisibles : de plus, on remarque que, dans chacun de ces trois exemples, $[\bar{\Phi}(u)]^{1/n}$ est du même type

que $\Phi(u)$ avec des paramètres différents.

En ce qui concerne la loi de Poisson, il sera utile d'en élargir un peu la définition. Etant données deux constantes a et b quelconques, une variable X susceptible de prendre les valeurs discrètes $a + n + b$ (n entier positif) avec les probabilités :

$$P(X = a + n + b) = \frac{\theta^n}{n!} e^{-\theta}$$

sera dite variable poissonienne (au sens large). Sa fonction caractéristique est :

$$\Phi(u) = e^{\theta(e^{iau} - 1) + i b u}$$

Elle est manifestement du type indéfiniment divisible.

Propriété 1 - La fonction caractéristique d'une loi indéfiniment divisible ne s'annule jamais.

En effet, $\Phi(u)$ étant continue, quel que soit le nombre β ($0 < \beta < 1$) on peut trouver $\alpha > 0$ tel que :

$$|u| \leq \alpha \implies |\Phi(u)| \geq \beta > 0$$

Plaçons-nous en un point u fixé avec $|u| \leq \alpha$. $\Phi(u)$ n'est pas nul, et par suite :

$$\lim_{n \rightarrow \infty} \Phi_n(u) = \lim_{n \rightarrow \infty} [\Phi(u)]^{1/n} = 1$$

Alors, on peut trouver un entier N (qui dépendra de u) tel que :

$$n \geq N \implies \begin{aligned} 1 - \varepsilon &\leq \Re \Phi_n(u) \leq 1 \\ 1 - \varepsilon &\leq |\Phi_n(u)| \leq 1 \end{aligned}$$

($\varepsilon > 0$ arbitrairement petit). Utilisons l'inégalité (13) du ch. II :

$$\left| \Phi_n(2u) - \Phi_n(u) \right| \leq \sqrt{2 [1 - \Re \Phi_n(u)]} \leq \sqrt{2} \varepsilon$$

On en tire

$$|\Phi_n(2u)| \geq |\Phi_n(u)| - \sqrt{2} \varepsilon \geq 1 - \varepsilon - \sqrt{2} \varepsilon$$

Comme ε est arbitraire, $|\Phi_n(2u)|$, pour n assez grand est strictement positif, donc aussi $|\Phi(2u)| = |\Phi_n(2u)|^n > 0$

Comme $|\Phi(2u)|$ est continue, et qu'une fonction continue sur un intervalle fini atteint sa borne inférieure, on a aussi :

$$\inf_{|u| \leq \alpha} |\Phi(2u)| = \gamma > 0$$

D'où l'implication :

$$|u| \leq 2\alpha \implies |\Phi(u)| \geq \gamma > 0$$

Il suffit alors d'itérer le raisonnement précédent pour voir que $\Phi(u)$ ne peut s'annuler pour aucune valeur finie de u .

Propriété 2.- Si X et Y sont deux variables indépendantes, à lois indéfiniment divisibles leur somme $X + Y$ a une loi indéfiniment divisible.

En effet, si Φ et Ψ sont les fonctions caractéristiques de X et Y , on a

$$\Phi = [\Phi_n]^n$$

$$\Psi = [\Psi_n]^n$$

Φ_n et Ψ_n étant, quel que soit n , des fonctions caractéristiques. On a donc aussi :

$$\Phi \Psi = [\Phi_n \Psi_n]^n$$

et, comme $\Phi_n \Psi_n$ est une fonction caractéristique, la propriété est établie.

Propriété 3.- Toute limite (au sens de la convergence en loi) de lois indéfiniment divisibles est indéfiniment divisible.

En effet, soient $\Phi_k(u)$ les fonctions caractéristiques d'une suite de lois indéfiniment divisibles, avec, pour tout n :

$$\Phi_k = [\Phi_{k,n}]^n$$

$\overline{\Phi}_{k,n}$ étant une fonction caractéristique. Par hypothèse, les $\overline{\Phi}_k(u)$ pour $k \rightarrow \infty$ convergent vers la fonction continue $\overline{\Phi}(u)$. Alors, les $\overline{\Phi}_{k,n}(u)$ convergent vers la fonction $\overline{\Phi}_n = (\overline{\Phi})^{\frac{1}{n}}$, et $\overline{\Phi}_n$ est continue, donc est une fonction caractéristique.

Comme on a

$$\overline{\Phi} = (\overline{\Phi}_n)^n$$

$\overline{\Phi}$ est bien la fonction caractéristique d'une loi indéfiniment divisible.

Théorème Fondamental. - Pour qu'une loi $F(x)$ possédant une variance finie σ^2 et une espérance mathématique m soit indéfiniment divisible, il faut et il suffit que sa caractéristique seconde se mette sous la forme :

$$(1) \quad \log \overline{\Phi}(u) = i m u + \sigma^2 \int \frac{e^{iux} - 1 - iux}{x^2} d G(x)$$

où $G(x)$ est une fonction de répartition.

a/ La condition est suffisante.

En effet, l'intégrale de Stieltjes qui figure au deuxième membre de (1) est somme de trois termes :

$$(2) \quad \left\{ \begin{array}{l} \log \overline{\Phi}_0 = -\sigma^2 \left[G(+0) - G(-0) \right] \frac{u^2}{2} \\ \log \overline{\Phi}_1 = \sigma^2 \int_{+0}^{+\infty} \frac{e^{iux} - 1 - iux}{x^2} d G(x) \\ \log \overline{\Phi}_2 = \sigma^2 \int_{-\infty}^{-0} \frac{e^{iux} - 1 - iux}{x^2} d G(x) \end{array} \right.$$

$\log \overline{\Phi}_0$ est la caractéristique seconde d'une variable normale, donc d'une variable X_0 à loi indéfiniment divisible.

$\log \overline{\Phi}_1$ est limite, pour $\varepsilon \rightarrow 0$, de

$$\sigma^2 \int_{\varepsilon}^{\frac{1}{\varepsilon}} \frac{e^{iux} - 1 - iux}{x^2} d G(x)$$

Cette dernière intégrale est elle-même limite de somme du type

$$\sigma^2 \sum_k \frac{e^{iu \xi_k} - 1 - iu \xi_k}{\xi_k^2} \left[G(x_{k+1}) - G(x_k) \right]$$

($\varepsilon \leq x_k \leq \xi_k \leq x_{k+1} \leq \frac{1}{\varepsilon}$). Chacun des termes d'une telle somme est la caractéristique seconde d'une loi poissonnienne (indéfiniment divisible). Cette somme, caractéristique seconde d'une somme de variables poissonniennes, représente elle-même une loi indéfiniment divisible d'après la propriété 2. Alors, la propriété 3 montre que l'intégrale étendue de ε à $\frac{1}{\varepsilon}$ représente aussi la caractéristique seconde d'une loi indéfiniment divisible. Toujours d'après la propriété 3, il en est de même de $\log \overline{\Phi}_1$, comme on le voit en faisant tendre ε vers 0.

Le même raisonnement s'applique à $\log \overline{\Phi}_2$. Finalement, la somme $\log \overline{\Phi}_0 + \log \overline{\Phi}_1 + \log \overline{\Phi}_2$, caractéristique seconde de la somme de trois variables à lois indéfiniment divisibles, est bien la caractéristique seconde d'une loi indéfiniment divisible.

b/ La condition est suffisante.

Soient F une loi indéfiniment divisible et $\overline{\Phi}$ sa fonction caractéristique. On a, pour tout n ,

$$\log \overline{\Phi} = n \log \overline{\Phi}_n$$

où $\overline{\Phi}_n$ est une fonction caractéristique. Comme $\overline{\Phi}(u) \neq 0$, on voit que $\overline{\Phi}_n(u)$ tend vers 1, pour $n \rightarrow \infty$, en tout point u . Alors, on peut écrire, pour n assez grand :

$$\log \overline{\Phi}(u) = n \log \left[1 - (1 - \overline{\Phi}_n) \right] = -n \left[1 - \overline{\Phi}_n(u) \right] - n \theta (1 - \overline{\Phi}_n)^2$$

$$(0 \leq |\theta| \leq 1)$$

et, à la limite

$$\log \overline{\Phi}(u) = \lim_{n \rightarrow \infty} n \left[\overline{\Phi}_n(u) - 1 \right]$$

$$= \lim_{n \rightarrow \infty} n \int (e^{iux} - 1) dF_n(x)$$

$F_n(x)$ étant la fonction de répartition associée à $\overline{\Phi}_n(u)$. Par ailleurs, l'espérance

mathématique m est égale à :

$$m = \int x \, dF(x) = n \int x \, dF_n(x)$$

Par suite

$$\log \bar{\Phi} = ium + \lim_{n \rightarrow \infty} n \int (e^{iux} - 1 - iux) \, dF_n(x)$$

On fait ensuite le changement de variables :

$$G_n(x) = \frac{n}{\sigma^2 + \frac{m^2}{n}} \int_{-\infty}^{x-\sigma} y^2 \, dF_n(y)$$

ce qui est légitime, puisque F_n admet un moment d'ordre 2. $G_n(x)$ est une fonction non décroissante, continue à gauche, et on a $G_n(-\infty) = 0$ et $G_n(+\infty) = 1$: C'est une fonction de répartition. Avec ce changement de variable, on obtient :

$$\log \bar{\Phi}(u) = ium + \lim_{n \rightarrow \infty} \left(\sigma^2 + \frac{m^2}{n} \right) \int \frac{e^{iux} - 1 - iux}{x^2} \, dG_n(x)$$

Afin de passer à la limite sous le signe d'intégration, montrons que les G_n convergent en loi vers une fonction de répartition $G(x)$. Pour cela, considérons la transformée de Fourier de G_n :

$$\int e^{iux} \, dG_n(x) = \frac{n}{\sigma^2 + \frac{m^2}{n}} \int x^2 e^{iux} \, dF_n(x) = - \frac{n}{\sigma^2 + \frac{m^2}{n}} \bar{\Phi}_n''(u)$$

cherchons la limite de $n \bar{\Phi}_n''(u)$. De :

$$\log \bar{\Phi}(u) = n \log \bar{\Phi}_n(u)$$

on tire (puisque $\bar{\Phi}$ et $\bar{\Phi}_n$ ne s'annulent jamais)

$$\frac{n \bar{\Phi}'_n}{\bar{\Phi}_n} = \frac{d}{du} \log \bar{\Phi}$$

En dérivant une deuxième fois :

$$\frac{n \bar{\Phi}''_n}{(\bar{\Phi}_n)^2} = n \frac{(\bar{\Phi}'_n)^2}{(\bar{\Phi}_n)^2} = \frac{d^2}{du^2} \log \bar{\Phi}$$

Comme $\frac{n \Phi'_n}{\Phi_n} = \frac{d}{du} \log \Phi$, le deuxième terme du premier membre tend vers 0 pour $n \rightarrow \infty$, et $(\Phi_n)^2 \rightarrow 1$, de sorte que l'on a :

$$\lim_{n \rightarrow \infty} n \Phi_n''(u) = \frac{d^2}{du^2} \log \Phi(u)$$

Cette limite existe, et c'est une fonction continue (puisque le moment d'ordre 2 existe). Alors, la transformée de Fourier des G_n admettant une limite continue, les G_n convergent en loi, vers une fonction de répartition $G(x)$.

Comme $\left| \frac{e^{iux} - 1 - iux}{x^2} \right| \leq \frac{u^2}{2}$, le deuxième théorème de Helly donne :

$$\lim_{n \rightarrow \infty} \int \frac{e^{iux} - 1 - iux}{x^2} dG_n(x) = \int \frac{e^{iux} - 1 - iux}{x^2} dG(x)$$

Il en résulte immédiatement :

$$\log \Phi(u) = ium + \sigma^2 \int \frac{e^{iux} - 1 - iux}{x^2} dG(x)$$

Remarque

La signification profonde du théorème fondamental est mise en évidence par la décomposition (2). Toute variable X à loi indéfiniment divisible se met sous la forme d'une somme

$$X = X_0 + X_1 + X_2$$

de trois variables indépendantes, dont la première X_0 est gaussienne, la deuxième X_1 limite d'une somme de variables poissonniennes à valeurs positives, la troisième X_2 limite d'une somme de variables poissonniennes à valeurs négatives. Ainsi la loi normale et la loi de Poisson sont les deux prototypes à partir desquels sont constituées toutes les lois indéfiniment divisibles à variance finie.

Avec $G(x) = \Theta(x)$

$$\Theta(x) = \begin{cases} 0 & \text{pour } x \leq 0 \\ 1 & \text{pour } x > 0 \end{cases}$$

la formule (1) représente la loi normale de moyenne m et de variance σ^2 .

Avec $G(x) = \theta(x - a)$, il vient :

$$\log \bar{\Phi}(u) = i m u + \frac{\sigma^2}{a^2} (e^{iua} - 1 - iua) = iu(m - \frac{\sigma^2}{a}) + \frac{\sigma^2}{a^2} (e^{iua} - 1)$$

On reconnaît la loi de Poisson définie par :

$$P(X = m - \frac{\sigma^2}{a} + ak) = \frac{1}{k!} \left(\frac{\sigma^2}{a}\right)^k e^{-\frac{\sigma^2}{a^2}}$$

... / ...

II.- PROCESSUS STOCHASTIQUES A ACCROISSEMENTS INDEPENDANTS ET STATIONNAIRES

Comme application des lois indéfiniment divisibles, nous allons étudier une classe particulière de processus stochastique. La définition générale d'un processus stochastique $X(t)$ a été donnée au chapitre 2, ainsi que celle de la loi temporelle associée, c'est-à-dire l'ensemble des fonctions de répartition :

$$F(x_1, \dots, x_n; t_1, \dots, t_n) = P(X(t_1) < x_1, X(t_2) < x_2, \dots, X(t_n) < x_n)$$

pour tous les entiers n et tous les instants (t_1, t_2, \dots, t_n) .

Nous dirons qu'un processus stochastique est à accroissements indépendants si, quels que soient les intervalles, en nombre fini, ne se recouvrant pas $(t_1, t'_1), (t_2, t'_2) \dots (t_k, t'_k)$ les variables aléatoires $X(t'_i) - X(t_i)$ sont mutuellement indépendantes.

De tels processus seront toujours étudiés à partir de leurs accroissements. En effet, si l'on a :

$$t_0 < t_1 < \dots < t_n$$

la variable $X(t_n) - X(t_0)$ se met sous la forme

$$(3) \quad X(t_n) - X(t_0) = \sum_{k=1}^n \left[X(t_k) - X(t_{k-1}) \right]$$

d'une somme de n variables indépendantes. Une origine t_0 étant choisie arbitrairement, $X(t)$, à une constante près $X(t_0)$, pourra toujours se mettre sous la forme (3).

En réalité, la formule (3) devrait se mettre sous la forme :

$$X(t) = X(t_0) = \int_{t_0}^t dX(tu)$$

(sous réserve d'avoir défini la notion d'intégrale de Stieltjes stochastique). C'est

en fait la mesure aléatoire définie par $dX(t)$ qui représente l'élément stochastique significatif. Pour éviter d'avoir à introduire une telle mesure aléatoire, on la remplace par son intégrale étendue de t_0 à t , qui est simplement une fonction aléatoire (de la même manière, l'étude d'une probabilité P , qui est une mesure se ramène à celle de sa fonction de répartition $F(x)$ qui est l'intégrale de P étendue de $-\infty$ à x).

Un processus stochastique $X(t)$ à accroissements indépendants sera dit, de plus, à accroissements stationnaires si la loi de probabilité de la variable $X(t + t_0) - X(t_0)$ ne dépend que de t et non de l'origine arbitraire t_0 . On dit parfois aussi, dans ce cas, que le processus est homogène dans le temps. Les lois qui régissent la variabilité locale de $X(t)$ sont, en effet, indépendantes de l'instant particulier t considéré.

Un exemple typique de processus à accroissements indépendants et stationnaires est donné par le mouvement brownien d'une particule dans un milieu homogène. Les coordonnées $X(t)$, $Y(t)$... de la particule vérifient bien, en effet, les définitions précédentes.

Comme l'étude d'un tel processus à une constante près se ramène à celle de ses accroissements, nous pouvons prendre une origine $t = 0$ arbitraire et remplacer $X(t)$ par $X(t) - X(0)$, ce qui revient à supposer $X(0) = 0$. Alors la loi temporelle est entièrement définie par la donnée de la fonction de répartition :

$$F(x, t) = P(X(t) < x)$$

En effet, si t_1, \dots, t_n sont des instants quelconques, avec par exemple

$$0 < t_1 < \dots < t_n$$

(si certains des t_i étaient négatifs, il y aurait quelques complications d'écriture, mais rien de réellement différent), on pose :

$$Y_k = X(t_k) - X(t_{k-1})$$

Les n variables Y_k sont indépendantes. Leur fonction de répartition :

$$F(y_1, t_1) \cdot F(y_2, t_2 - t_1) \dots F(y_n, t_n - t_{n-1})$$

permet, par changement de variable, de déterminer celle des n variables $X(t_k)$

Théorème 1.- La fonction de répartition $F(x, t)$ d'un processus à accroissements indépendants et stationnaires est indéfiniment divisible.

En effet, considérons les instants successifs :

$$0 = t_0 < \frac{t}{n} < \frac{2t}{n} < \dots < t$$

On a

$$X(t) = \sum_{k=1}^n \left[X\left(\frac{k}{n} t\right) - X\left(\frac{k-1}{n} t\right) \right]$$

Les variables $X\left(\frac{k}{n} t\right) - X\left(\frac{k-1}{n} t\right)$ sont, par définition, indépendantes et ont la même loi de probabilité $F(x, \frac{t}{n})$. Il en résulte bien que la loi $F(x, t)$ est indéfiniment divisible.

Limitons nous maintenant au cas des processus à accroissements indépendants et stationnaires possédant une variance finie, qui est évidemment une fonction de t , $\sigma^2(t)$. Un tel processus possède également une espérance mathématique $m(t)$. Naturellement, $\sigma^2(t)$ est la variance non seulement de $X(t)$ ($= X(t) - X(0)$, d'après nos conventions), mais de tout accroissement $X(t + t_0) - X(t_0)$, quel que soit t_0 . Le théorème 1 entraîne alors les conséquences suivantes :

Conséquence 1.- La variance $\sigma^2(t)$ est proportionnelle à t :

$$(4) \quad \sigma^2(t) = a t \quad (a > 0, t \geq 0)$$

En effet, les accroissements étant indépendants, on a immédiatement :

$$(5) \quad \sigma^2(t_1 + t_2) = \sigma^2(t_1) + \sigma^2(t_2)$$

On sait que les seules fonctions continues vérifiant cette propriété sont les fonctions linéaires de la forme $a t$. La propriété $\sigma^2(t) = a t$ sera donc établie si l'on montre que $\sigma^2(t)$ est continue. D'après (5), il suffit de montrer que $\sigma^2(t)$ est continue en $t = 0$.

Soit $\alpha > 0$. Du fait que les accroissements sont indépendants, ou de la relation (5), on déduit

$$\sigma^2(\alpha) = n \sigma^2\left(\frac{\alpha}{n}\right)$$

soit $\sigma^2\left(\frac{\alpha}{n}\right) = \frac{\sigma^2(\alpha)}{n}$. D'autre part, toujours d'après (5), $\sigma^2(t)$ est une fonction croissante de t . Donc :

$$0 \leq t \leq \frac{\alpha}{n} \implies \sigma^2(t) \leq \frac{1}{n} \sigma^2(\alpha)$$

Par suite $\sigma^2(t)$ est continue en $t = 0$.

Conséquence 2.- Si l'espérance mathématique $m(t)$ est bornée sur un intervalle fini $0 \leq t \leq t_0$, alors $m(t)$ est proportionnelle à t :

$$(6) \quad m(t) = b t$$

On a, comme ci-dessus $m(t_1 + t_2) = m(t_1) + m(t_2)$, et il suffit de montrer que $m(t)$ est continue en $t = 0$. Par hypothèse, il existe $B > 0$ tel que :

$$0 \leq t \leq t_0 \implies |m(t)| \leq B$$

Comme $m\left(\frac{t}{n}\right) = \frac{1}{n} m(t)$, pour tout entier n , on en déduit :

$$0 \leq t < \frac{t_0}{n} \implies |m(t)| \leq \frac{B}{n}$$

ce qui établit la continuité, et la relation (6).

L'hypothèse que $m(t)$ est bornée sur un intervalle fini, toujours vérifiée dans les applications, sera adoptée dans tout ce qui suit, de sorte que (6) sera supposée valable.

Conséquence 3.- Lorsque h tend vers 0, $X(t+h)$ converge en probabilité, donc aussi en loi, vers $X(t)$. Autrement dit, $X(t)$ est continu en probabilité.

La variable $Y(h) = X(t+h) - X(t)$ possède l'espérance mathématique $b h$ et la variance $a h$. L'inégalité de Tchebychev donne :

$$P\left(\left|Y(h) - b h\right| \geq \varepsilon\right) \leq \frac{a h}{\varepsilon^2}$$

Comme on a toujours $\left|Y\right| - \left|b h\right| \leq \left|Y - b h\right|$, on a aussi

$$P\left(\left|Y(h)\right| \geq \varepsilon + \left|b h\right|\right) \leq \frac{a h}{\varepsilon^2}$$

Avec $h \leq \inf\left(\frac{\varepsilon}{|b|}, \frac{\varepsilon^2}{a} \alpha\right)$, on a donc aussi :

$$P\left(\left|Y(h)\right| \geq 2 \varepsilon\right) \leq \alpha$$

d'où résulte la continuité en probabilité, et par suite aussi la continuité en loi.

Conséquence 4.- La fonction caractéristique $\underline{\Phi}(u, t)$ de la loi $F(x, t)$ est continue par rapport à t .

Comme on a

$$\underline{\Phi}(u, t + h) = \underline{\Phi}(u, t) \underline{\Phi}(u, h)$$

et $\underline{\Phi}(u, 0) = 1$, il suffit de montrer que $\underline{\Phi}(u, t)$ est continue en $t = 0$. D'après la conséquence 3, $X(t)$ converge en loi vers la variable presque certainement nulle (de loi $\theta(x) = 0$ pour $x \leq 0$ et $\theta(x) = 1$ pour $x > 0$). Par suite $\underline{\Phi}(u, t)$ converge, pour $t \rightarrow 0$, vers la fonction caractéristique de la loi $\theta(x)$, c'est-à-dire vers :

$$\int e^{i u x} d \theta(x) = 1 = \underline{\Phi}(u, 0)$$

Donc $\underline{\Phi}(u, t)$ est continue en $t = 0$.

Conséquence 5. - La fonction caractéristique $\bar{\Phi}(u, t)$ est de la forme :

$$\bar{\Phi}(u, t) = \left[\bar{\Phi}(u, 1) \right]^t \quad (t \geq 0)$$

où $\bar{\Phi}(u) = \bar{\Phi}(u, 1)$ est une fonction caractéristique.

En effet, les accroissements étant indépendants, on a pour tout entier n :

$$\bar{\Phi}(u, n t) = \left[\bar{\Phi}(u, t) \right]^n$$

et, pour tout rationnel $\frac{p}{q} > 0$:

$$\bar{\Phi}(u, \frac{p}{q} t) = \left[\bar{\Phi}(u, p t) \right]^{\frac{1}{q}} = \left[\bar{\Phi}(u, t) \right]^{\frac{p}{q}}$$

Comme $\bar{\Phi}(u, t)$ est une fonction continue de t , on en déduit, pour tout $\alpha > 0$ (rationnel ou non)

$$\bar{\Phi}(u, \alpha t) = \left[\bar{\Phi}(u, t) \right]^\alpha$$

et, en particulier $\bar{\Phi}(u, \alpha) = \left[\bar{\Phi}(u, 1) \right]^\alpha$

Conséquence 6. - La caractéristique seconde $\log \bar{\Phi}(u, t)$ est de la forme :

$$(7) \quad \log \bar{\Phi}(u, t) = i u b t + a t \int \frac{e^{iux} - 1 - iux}{x^2} d G(x)$$

où $G(x)$ est une fonction de répartition.

Cela résulte immédiatement de la conséquence 5. Comme $\bar{\Phi}(u, 1)$ est caractéristique de la loi indéfiniment divisible $F(x, 1)$, on a, d'après le théorème fondamental :

$$\log \bar{\Phi}(u, 1) = i u b + a \int \frac{e^{iux} - 1 - iux}{x^2} d G(x)$$

et (7) en résulte immédiatement.

Cette formule (7) caractérise complètement les processus à accroissements indépendants et stationnaires, variance finie et espérance mathématique bornée sur tout intervalle fini.

Toute loi indéfiniment divisible est une superposition de lois de Gauss et de Poisson. Il en résulte que les processus du type (7) sont eux-aussi combinaison de

processus de deux types simples : l'un, gaussien est le processus de Wiener-Lévy, l'autre, poissonien, est le processus de Poisson.

Processus de Wiener-Lévy - Si nous prenons, dans (15), $G(x) = \theta(x)$ ($= 0$ pour $x \leq 0$ et 1 pour $x > 0$), il vient :

$$(8) \quad \log \Phi(u, t) = i b u t - \frac{a t}{2} u^2$$

et $X(t)$ est une variable normale, de moyenne $b t$ et de variance $a t$. En remplaçant $X(t)$ par $\frac{1}{\sqrt{a}} [X(t) - b t]$ on se ramène au cas type :

$b = 0$, $a = 1$ qui correspond au processus de Wiener-Lévy défini par :

$$\log \Phi(u, t) = - \frac{t}{2} u^2$$

Parmi tous les processus à accroissements indépendants et stationnaires, le processus Gaussien défini par (8) est le seul à présenter la propriété de continuité uniforme en probabilité. L'énoncé précis est le suivant :

Théorème 2. - Pour qu'un processus à accroissements indépendants et stationnaires et à variance finie soit soumis à la loi de Gauss, il faut et il suffit que, pour tout $\varepsilon > 0$ et tout t on ait :

$$(9) \quad \lim_{n \rightarrow \infty} P \left[\sup_{1 \leq k \leq n} \left| X\left(\frac{k}{n}t\right) - X\left(\frac{k-1}{n}t\right) \right| > \varepsilon \right] = 0$$

a/- Remarquons, en premier lieu, que l'on a :

$$P \left[\sup_{1 \leq k \leq n} \left| X\left(\frac{k}{n}t\right) - X\left(\frac{k-1}{n}t\right) \right| \leq \varepsilon \right] \\ = \left[F_n(\varepsilon + 0) - F_n(-\varepsilon) \right]^n = \left[1 - (1 - F_n(\varepsilon + 0) + F_n(-\varepsilon)) \right]^n$$

Cette expression tend vers 1 si, et seulement si :

$$(10) \quad \lim_{n \rightarrow \infty} n \left[1 - F_n(\varepsilon + 0) + F_n(-\varepsilon) \right] = 0$$

On a écrit $F_n(x)$ au lieu de $F(x, \frac{t}{n})$. Ainsi les conditions (9) et (10) sont équivalentes.

b/- Transformons la formule (7) à l'aide des deux fonctions auxiliaires

$$(11) \quad \left\{ \begin{array}{l} H(u) = \int_{-\infty}^u \frac{1}{x^2} dG(x) \quad (u < 0) \\ K(u) = - \int_u^{\infty} \frac{1}{x^2} dG(x) \quad (u > 0) \end{array} \right.$$

On a

$$\log \Phi(u, t) = i u b t - \frac{at}{2} u^2 [G(+0) - G(-0)] \\ + at \int_{-\infty}^{-0} (e^{iux} - 1 - iux) dH(u) + at \int_{+0}^{+\infty} (e^{iux} - 1 - iux) dK(u)$$

ainsi, le processus sera Gaussien si, et seulement si,

$$(12) \quad \left\{ \begin{array}{l} H(u) \equiv 0 \quad (u < 0) \\ K(u) \equiv 0 \quad (u > 0) \end{array} \right.$$

c/- Pour achever la démonstration, il faut montrer que les conditions (10) et (11) sont équivalentes. On sait que $G(x)$ est limite en loi de :

$$G_n(x) = n \int_{-\infty}^x y^2 dF_n(y)$$

Posons :

$$\left\{ \begin{array}{l} H_n(x) = \int_{-\infty}^x \frac{1}{y^2} dG_n(y) = n F_n(x) \quad (x < 0) \\ K_n(x) = - \int_x^{\infty} \frac{1}{y^2} dG_n(y) = -n [1 - F_n(x)] \quad (x > 0) \end{array} \right.$$

Le 2ème théorème de Helly donne, (puisque $\frac{1}{y^2}$ est borné sur $(-\infty, x)$)

$$\lim_{n \rightarrow \infty} H_n(x) = \lim_{n \rightarrow \infty} n F_n(x) = H(x) \quad (x < 0)$$

$$\lim_{n \rightarrow \infty} K_n(x) = - \lim_{n \rightarrow \infty} n [1 - F_n(x)] = K(x) \quad (x > 0)$$

Alors les conditions (12) sont équivalentes à :

$$\lim_{n \rightarrow \infty} n F_n(-\varepsilon) = 0$$

$$\lim_{n \rightarrow \infty} n [1 - F_n(\varepsilon)] = 0$$

pour tout $\varepsilon > 0$, et sont donc également équivalentes aux conditions (10).

Processus Poissonien

Les processus poissonniens, au contraire, vont être caractérisés par leur discontinuité. Prenons, dans (7),

$$G(x) = \theta(x-1) = \begin{cases} 0 & \text{pour } x \leq 1 \\ 1 & \text{pour } x > 1 \end{cases}$$

On obtient la loi poissonnienne :

$$\log \Phi(u, t) = i u (b-a)t + at(e^{iu} - 1)$$

Quitte à remplacer $X(t)$ par $X(t) - (b-a)t$, on peut supposer $a = b$.

Il reste alors

$$\log \Phi(u, t) = a t (e^{iu} - 1)$$

c'est la loi de Poisson de paramètre at . On a :

$$(13) \quad P(X(t) = n) = \frac{(at)^n}{n!} e^{-at}$$

$X(t)$ est une fonction non décroissante, complètement discontinue, variant par sauts positifs d'amplitude $+1$.

En particulier, la probabilité pour qu'il ne se produise aucun saut dans l'intervalle (t_0, t_1) est donnée par :

$$P(X(t_1) - X(t_0) = 0) = e^{-a(t_1 - t_0)}$$

Soit, avec $t_0 = 0$ et $X(t_0) = 0$:

$$(14) \quad P(X(t) = 0) = e^{-at}$$

Cette loi exponentielle est très caractéristique des phénomènes exempts de vieillissement, comme l'émission radioactive. En effet, pour un intervalle de temps Δt très petit on a :

$$(15) \quad \begin{cases} P(X(\Delta t) = 0) = 1 - a \Delta t + o(\Delta t) \\ P(X(\Delta t) = 1) = a \Delta t + o(\Delta t) \\ P(X(\Delta t) > 1) = o(\Delta t) \end{cases}$$

$o(\Delta t)$ désignant des quantités quelconques tendant vers 0 plus vite que Δt (infinitement petits d'ordre supérieur à 1).

Réciproquement. Réciproquement, le processus poissonnien ainsi obtenu est le seul processus $X(t)$ variant par sauts discontinus $+1$ à des instants aléatoires ($X(t)$ représentera, par exemple, le nombre d'apparitions d'un phénomène donné entre les instants 0 et t : nombre de particules radioactives émises, nombre d'appels téléphoniques enregistrés à un standard, etc ...), et vérifiant les conditions suivantes :

1 - les accroissements sont stationnaires, c'est-à-dire : $X(t+t_0) - X(t_0)$ obéit à la même loi que $X(t) - X(0)$, quel que soit t_0 .

2 - Les accroissements sont indépendants : si les intervalles (t_1, t_1^0) ne se recouvrent pas, les $X(t_1^0) - X(t_1)$ sont indépendants. En particulier, pour $t \geq t_1$, l'accroissement $X(t) - X(t_1)$ est indépendant de toutes les valeurs $X(t^0)$ prises par X à des instants t^0 antérieurs à t_1 .

3 - Pour Δt petit, on a les relations (15).

En effet, prenons $X(0) = 0$ et posons

$$P(X(t) = n) = P_n(t)$$

Calculons $P_0(t)$. L'évènement " $X(t) = 0$ " est le produit des deux évènements " $X(\Delta t) = 0$ " et " $X(t) - X(\Delta t) = 0$ ", qui sont indépendants, d'après 2, et ont

comme probabilités $P_0(\Delta t)$ et $P_0(t - \Delta t)$ d'après 1. D'où, compte tenu de (23) :

$$P_0(t) = \left[1 - a\Delta t + o(\Delta t) \right] P_0(t - \Delta t)$$

et

$$P_0(t) - P_0(t - \Delta t) = -a\Delta t P_0(t - \Delta t) + o(\Delta t)$$

Lorsque Δt tend vers 0, le deuxième membre tend vers 0, et par suite :

$$\lim_{t \rightarrow 0} P_0(t - \Delta t) = P_0(t)$$

On a alors :

$$\frac{P_0(t) - P_0(t - \Delta t)}{\Delta t} = -a P_0(t - \Delta t) + \frac{o(\Delta t)}{\Delta t}$$

Par hypothèse, $\frac{o(\Delta t)}{\Delta t} \rightarrow 0$ pour $\Delta t \rightarrow 0$, de sorte que le deuxième membre admet la limite $-a P_0(t)$. Donc $P_0(t)$ est dérivable à gauche et admet la dérivée $-a P_0(t)$. De la même façon, on montrera que l'on a

$$P_0(t + \Delta t) = \left[1 - a\Delta t + o(\Delta t) \right] P_0(t)$$

et on en déduira que $P_0(t)$ est dérivable à droite. Finalement $P_0(t)$ est dérivable, et

$$\frac{d}{dt} P_0(t) = -a P_0(t)$$

D'où $P_0(t) = C e^{-at}$ et, puisque $P_0(0) = 1$:

$$P_0(t) = e^{-at}$$

Calculons maintenant $P_n(t)$. L'évènement " $X(t) = n$ " est somme des trois évènements incompatibles suivants :

$$" X(\Delta t) = 0 \text{ et } X(t) - X(\Delta t) = n "$$

$$" X(\Delta t) = 1 \text{ et } X(t) - X(\Delta t) = n - 1 "$$

$$" X(\Delta t) = k, \quad k > 1, \text{ et } X(t) - X(\Delta t) = n - k "$$

On en déduit comme ci-dessus :

$$\begin{aligned} P_n(t) &= P_0(\Delta t) P_n(t - \Delta t) + P_1(\Delta t) P_{n-1}(t - \Delta t) + o(\Delta t) \\ &= P_n(t - \Delta t) + a \Delta t \left[P_{n-1}(t - \Delta t) - P_n(t - \Delta t) \right] + o(\Delta t) \end{aligned}$$

et de même, pour l'évènement " $X(t + \Delta t) = n$ "

$$P_n(t + \Delta t) = P_n(t) + a \Delta t \left[P_{n-1}(t) - P_n(t) \right] + o(\Delta t)$$

La continuité de $P_{n-1}(t)$ étant admise, comme hypothèse de récurrence vraie pour $P_0(t)$, on en déduit celle de $P_n(t)$ et, après division par Δt et passage à la limite, on voit que $P_n(t)$ est dérivable et admet comme dérivée

$$(16) \quad \frac{d P_n(t)}{d t} = a \left[P_{n-1}(t) - P_n(t) \right]$$

Soit alors

$$U(s, t) = \sum_{n=0}^{\infty} P_n(t) s^n$$

la fonction génératrice des $P_n(t)$. Multipliant (16) par s^n et sommant de $n=1$ à l'infini, il vient :

$$\frac{\partial}{\partial t} U(s, t) - \frac{d P_0(t)}{d t} = a s U(s, t) - a [U - P_0]$$

soit, compte tenu de $P_0(t) = e^{-at}$

$$\frac{\partial U}{\partial t} = -a(1-s)U$$

La solution générale est de la forme :

$$U = c(s) e^{-a(1-s)t}$$

Mais on ^(a)la condition aux limites $U(s, 0) = 1$, d'où $c(s) = 1$ et, finalement :

$$U(s, t) = e^{-a(1-s)t} = e^{-at + ast}$$

On reconnaît la fonction génératrice de la loi de Poisson de paramètre at ,
soit :

$$P_n(t) = \frac{(at)^n}{n!} e^{-at}$$

Application - Prenant toujours $X(0) = 0$, cherchons, à titre d'exercice, la probabilité d'avoir $X(t) = k$ sachant que $X(t_1) = n$ ($0 < t < t_1$ et $k \leq n$). On écrit la formule des probabilités conditionnelles :

$$\begin{aligned} P(X(t) = k \mid X(t_1) = n) &= \frac{P(X(t) = k \text{ et } X(t_1) = n)}{P(X(t_1) = n)} \\ &= \frac{P(X(t) = k \text{ et } X(t_1) - X(t) = n - k)}{P(X(t_1) = n)} \\ &= \frac{e^{-at} \frac{(at)^k}{k!} e^{-a(t_1-t)} \frac{[a(t_1-t)]^{n-k}}{(n-k)!}}{e^{-at_1} \frac{(at_1)^n}{n!}} \\ &= C_n^k \left(\frac{t}{t_1} \right)^k \left(1 - \frac{t}{t_1} \right)^{n-k} \end{aligned}$$

On obtient, comme loi conditionnelle, la loi binomiale de paramètres n et $p = \frac{t}{t_1}$. Ce résultat signifie que chacun des n sauts qui se sont produits dans l'intervalle $(0, t)$ sont répartis, indépendamment les uns des autres et avec une densité de probabilité uniforme sur l'intervalle $(0, t)$.

C H A P I T R E V I

CHAINES DE MARKOV HOMOGENES (à temps discret)

I.- GENERALITES

Dans un schéma d'alternative répétée, supposons que chaque épreuve soit susceptible d'amener l'un ou l'autre des résultats (ou évènements élémentaires) e_1, e_2, \dots , en nombre fini ou infini dénombrable. Si les épreuves sont indépendantes les unes des autres, on retombe sur le problème classique de Bernoulli. Dans ce qui suit, nous nous placerons dans le cas où ces épreuves ne sont pas indépendantes, mais où le résultat de chacune d'elle est influencé uniquement par le résultat de celle qui la précède immédiatement. Autrement dit, si nous introduisons les variables aléatoires X_n définies par :

$$X_n = i \quad \text{si la } n^{\text{ième}} \text{ épreuve amène le résultat } e_i$$

nous admettons que toutes les probabilités conditionnelles :

$$P\left(X_n = i_n \mid X_{n_1} = i_1, X_{n_2} = i_2, \dots, X_{n_k} = i_k, X_{n-1} = i_{n-1}\right)$$

sont égales entre elles, quel que soit $k < n-1$ et quelles que soient les suites d'indices $n_1 < n_2 < \dots < n_k (< n-1)$ et i_1, i_2, \dots, i_k . En particulier, toutes ces probabilités conditionnelles sont égales à $P(X_n = i_n \mid X_{n-1} = i_{n-1})$. Nous poserons, dans la suite,

$$(1) \quad P_{ij}(n) = P\left(X_n = j \mid X_{n-1} = i\right)$$

Dans les applications physiques, les évènements élémentaires e_1, e_2, \dots sont souvent interprétés comme les différents états qu'un système physique est susceptible de prendre. Si l'on admet que le système ne peut changer d'états qu'aux temps $t = 1, 2, \dots$, on énoncera souvent l'évènement " $X_n = j$ " en disant : "le système est dans l'état j au temps $t = n$ ". Avec cette terminologie, la probabilité conditionnelle $P_{ij}(n)$ s'appelle probabilité de transition (ou de passage) de l'état i à l'état j au temps $t = n$. C'est, par définition, la probabilité pour que le

ystème soit dans l'état j au temps $t = n$, sachant qu'il était dans l'état i au temps $t = n - 1$.

Lorsque cette probabilité est indépendante des états antérieurs, c'est-à-dire, plus précisément, si l'on a :

$$P_{ij}^{(n)} = P\left(X_n = j \mid X_{n_1} = i_1, \dots, X_{n_k} = i_k, X_{n-1} = i\right)$$

pour tous les systèmes d'indices $n_1 < n_2 < \dots < n_k < n-1$, et i_1, i_2, \dots, i_k on dit que l'on a une chaîne de MARKOV.

Remarque

La donnée des $P_{ij}^{(n)}$ ne suffit pas pour définir complètement la chaîne de MARKOV. Il faut, de plus, se donner les probabilités initiales :

$$p_k = P(X_0 = k)$$

c'est-à-dire les probabilités a priori p_k pour que le système soit dans l'état e_k au temps $t = 0$.

D'un point de vue purement axiomatique, désignons par E l'ensemble des événements élémentaires (e_1, e_2) , ou états du système, et par \mathcal{A} la σ -algèbre sur E définie comme l'ensemble de tous les sous-ensembles de E , qui est la σ -algèbre engendrée par les $\{e_i\}$. Alors, la σ -algèbre $\mathcal{A}^{(\omega)}$ sur $E^{(\omega)}$ est engendrée par les événements simples de la forme :

$$(2) \quad A = e_{i_1} \times e_{i_2} \times \dots \times e_{i_k} \times E_{k+1} \times E_{k+2}$$

On peut montrer qu'il existe une probabilité P , et une seule, sur $(E^{(\omega)}, \mathcal{A}^{(\omega)})$ telle que l'on ait, pour tout événement simple de la forme 2 :

$$P(A) = p_{i_1} P_{i_1 i_2}^{(1)} P_{i_2 i_3}^{(2)} \dots P_{i_{k-1} i_k}^{(k)}$$

Ainsi, la donnée de $P_{ij}^{(n)}$ et des p_k suffit à définir parfaitement

l'espace probabilisé $(E^{\omega}, \mathcal{A}^{\omega}, P)$ qui représente l'évolution du système physique.

En fait, nous le verrons, les probabilités initiales ne jouent qu'un rôle accessoire, et les propriétés essentielles des chaînes de Markov sont liées à la matrice de transition $P_i^j(n)$.

Définitions : chaîne de Markov homogène. Une chaîne de Markov est dite homogène si $P_{ij}(n)$ ne dépend pas de n :

$$P_{ij}(n) = P_{ij}$$

Pour une chaîne homogène (on dit aussi stationnaire) la matrice de transition ne dépend pas du temps. Dans ce chapitre, nous nous limiterons au cas des chaînes homogènes.

Chaînes finies et infinies

Une chaîne de Markov est dite finie s'il n'y a qu'un nombre fini d'états possibles e_1, e_2, \dots, e_n . Elle est dite infinie s'il y a une infinité (dénombrable) d'états possibles. Les chaînes finies sont, naturellement, beaucoup plus faciles à étudier. Néanmoins, un certain nombre de propriétés fondamentales sont communes aux chaînes finies et infinies.

Propriété 1 - La matrice de transition P_{ij} d'une chaîne homogène est une matrice stochastique.

Cela signifie que l'on a :

$$(3) \quad \left\{ \begin{array}{l} P_{ij} \geq 0 \\ \sum_{j=1}^{\infty} P_{ij} = 1 \end{array} \right. , \quad P_{ij} \leq 1$$

Les relations (3) sont une conséquence immédiate des axiomes des probabilités (il est nécessaire de faire intervenir l'axiome d'additivité complète dans le cas d'une chaîne infinie).

Transitions d'ordre supérieur

Désignons par $P_{ij}^{(n)}$ la probabilité pour que le système soit dans l'état j au temps $t = n$, sachant qu'il était dans l'état i au temps $t = 0$. En particulier, on a $P_{ij}^{(1)} = P_{ij}$. On dira que $P_{ij}^{(n)}$ est la probabilité de transition d'ordre n de l'état i à l'état j .

Soit m un entier tel que $0 < m < n$. L'évènement " $X_n = j$ " est somme pour tous les indices k des évènements incompatibles " $X_m = k$ et $X_n = j$ ". On en déduit que l'on a (quel que soit i) :

$$(4) \quad P_{ij}^{(n)} = \sum_{k=1}^{\infty} P_{ik}^{(m)} P_{kj}^{(n-m)}$$

Cette équation porte le nom d'équation de Markov. En notations matricielles, si l'on désigne par $P^{(n)}$ la matrice des $P_{ij}^{(n)}$, elle s'écrit :

$$P^{(n)} = P^{(m)} P^{(n-m)}$$

En particulier, si P désigne la matrice des transitions d'ordre 1, on a :

$$P^{(2)} = P P = P^2$$

et, par une récurrence immédiate :

$$(5) \quad P^{(n)} = P^n$$

Enonçons :

Propriété 2 - La matrice de transition d'ordre n est la puissance $n^{\text{ième}}$ de la matrice de transition du premier ordre.

On notera que la propriété 2 s'applique aux chaînes infinies. On déduit immédiatement des relations (3) que la série qui figure au deuxième membre de (4) est nécessairement convergente. En effet, on a :

$$P_{ij}^{(n+1)} = \sum_{k=1}^{\infty} P_{ik} P_{kj}^{(n)}$$

Pour $n=1$, la matrice P^1 est stochastique. Supposons que P^n soit stochastique, et montrons que P^{n+1} est alors nécessairement stochastique.

De $0 \leq P_{kj}^{(n)} \leq 1$, on déduit :

$$0 \leq P_{ij}^{(n+1)} \leq \sum_{k=1}^{\infty} P_{ik} = 1$$

Sommons en j :

$$\begin{aligned} \sum_{j=1}^{\infty} P_{ij}^{(n+1)} &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} P_{ik} P_{kj}^{(n)} \\ &= \sum_{k=1}^{\infty} P_{ik} \sum_{j=1}^{\infty} P_{kj}^{(n)} = \sum_{k=1}^{\infty} P_{ik} = 1 \end{aligned}$$

S'agissant de séries à termes positifs, il est permis, comme on sait, d'invertir l'ordre des sommations. Les deux dernières égalités découlent immédiatement de l'hypothèse de récurrence. Énonçons :

Propriété 3 - La matrice de transition d'ordre n est stochastique.

Ergodicité - Connaissant les probabilités de transition d'ordre n et les probabilités a priori initiales p_k , on obtient immédiatement les probabilités a priori au temps $t = n$. Désignons par $p_k^{(n)}$ la probabilité pour que le système soit dans l'état k au temps $t = n$. On a :

$$(6) \quad p_k^{(n)} = \sum_{i=1}^{\infty} p_i P_{ik}^{(n)}$$

L'un des problèmes les plus importants qui se posent à propos des chaînes de Markov homogène est celui de l'examen du comportement des $p_k^{(n)}$ lorsque n augmente indéfiniment. Si les limites

$$(7) \quad u_k = \lim_{n \rightarrow \infty} p_k^{(n)}$$

existent et ne dépendent pas des probabilités initiales p_k , on dit qu'il y a ergodicité. Si une chaîne est ergodique, l'influence des conditions initiales devient de plus en plus faible et tend vers 0 lorsque le temps n augmente indéfiniment, et le système est, à la limite, soumis à une loi de probabilité qui ne dépend plus du temps: on dit parfois qu'un tel système est stable (on prendra garde que ce sont les probabilités qui sont stables, et non pas les états eux-mêmes du système).

Comme on a, dans tous les cas

$$p_j^{(n+1)} = \sum_{i=1}^{\infty} p_i^{(n)} P_{ij}$$

on voit que, si la chaîne est ergodique, les limites (7) seront solutions du système d'équation :

$$(8) \quad u_j = \sum_{i=1}^{\infty} u_i P_{ij}$$

S'il y a ergodicité, les limites u_j sont indépendantes, par définition, des probabilités initiales. Avec $p_i = 1$ et $p_k = 0 (k \neq i)$, les équations (6) et (7) donnent, quel que soit i :

$$u_j = \lim_{n \rightarrow \infty} P_{ij}^{(n)}$$

On voit qu'il y aura ergodicité si, et seulement si, la matrice de transition P^n tend, pour n infini, vers une matrice limite dont toutes les lignes sont identiques.

II.- CLASSIFICATION DES ETATS

Nous dirons qu'un état e_j peut être atteint à partir d'un état e_i , ou qu'il existe un chemin de e_i à e_j , s'il existe un entier n tel que :

$$P_{ij}^{(n)} > 0$$

c'est-à-dire si une transition de i à j n'est pas presque impossible.

Un état e_i sera dit non essentiel si l'on peut trouver un état e_j tel qu'il existe un chemin allant de e_i à e_j et qu'il n'existe aucun chemin retournant de e_j à e_i .

Un état e_i sera dit essentiel, au contraire, si, quel que soit l'état e_j pouvant être atteint à partir de e_i , il existe un chemin de retour de e_j à e_i .

A partir d'un état essentiel, on ne peut atteindre que des états essentiels.

En effet, soit e_i un état essentiel, et e_j un état quelconque pouvant être atteint à partir de e_i . Montrons que e_j est essentiel. Soit e_k un état atteint à partir de e_j . Il existe un chemin joignant e_i et e_k , puisque l'on peut aller de e_i à e_j puis de e_j à e_k avec des probabilités non nulles. Comme e_i est essentiel par hypothèse, il existe un chemin retournant de e_k à e_i . Par suite, il existe aussi un chemin de e_k à e_j , puisque l'on peut aller de e_k à e_i puis de e_i à e_j avec des probabilités non nulles : e_j est donc essentiel.

Convenons de dire que deux états essentiels e_i et e_j sont en correspondance s'il existe un chemin joignant e_i à e_j (et par suite aussi un chemin joignant e_j à e_i). La relation de correspondance est reflexive : e_i étant essentiel, il y a un chemin menant de e_i à e_i . Elle est symétrique, puisque s'il existe un chemin de e_i à e_j il existe aussi un chemin de retour de e_j à e_i . Elle est transitive : S'il existe un chemin de e_i à e_j et un chemin de e_j à e_k , on peut atteindre e_k à partir de e_i . Ainsi la relation de correspondance est une équivalence. Il existe donc une partition de l'ensemble des états essentiels en classes de correspondance. Par définition, la classe à laquelle appartient un état essentiel e_i est constituée par l'ensemble de tous les états (nécessairement essentiels) qui peuvent être atteints à partir de e_i . Ou encore : deux états essentiels e_i et e_j peuvent être reliés par un chemin si, et seulement si, ils appartiennent à la même classe. Ou enfin : si, à un instant quelconque, le système a pris un état essentiel e_i , il est presque certain qu'il ne prendra dans la suite que des états appartenant à la classe de e_i .

Soit C la classe d'un état essentiel e_i . Désignons par $P_{jk}(C)$ la sous-matrice de P_{ij} obtenue en ne conservant que les indices j et k désignant des états e_j et e_k de la classe C . La matrice des $P_{jk}(C)$ est stochastique. En effet, on a d'abord :

$$0 \leq P_{jk} \leq 1$$

D'autre part, si e_ℓ n'appartient pas à C , il n'existe aucun chemin joignant un état e_j de C à e_ℓ , et par suite $P_{j\ell} = 0$. Donc on a :

$$\sum_{k \in C} P_{jk} = \sum_{\ell=1}^{\infty} P_{j\ell} = 1$$

On voit qu'une fois atteint l'état essentiel e_i , la chaîne initiale définie par les P_{ij} peut être remplacée par la chaîne plus simple définie par les $P_{ij}(C)$, qui ne comporte que des états essentiels et une seule classe de correspondance : une telle chaîne est dite irréductible.

Définition : Une chaîne est irréductible si elle ne comporte que des états essentiels et une seule classe de correspondance, autrement dit si tout état peut être atteint à partir de tout autre état.

D'après ce qui précède, l'étude d'une classe d'états essentiels se ramène à celle d'une chaîne irréductible.

Périodicité d'une chaîne irréductible (ou d'une classe d'états essentiels).

Soit e_i un état d'une chaîne irréductible, et N_i l'ensemble des entiers n tels que $P_{ii}^{(n)}$ ne soit pas nul. N_i n'est pas vide, puisque e_i est un état essentiel. Si n appartient à N_i , il en est de même de tous les multiples de n . Si n et m appartiennent à N_i , il en est de même de $n + m$. Désignons par d_i le plus grand commun diviseur de tous les nombres appartenant à N_i . On dira que d_i est la période de l'état e_i , puisque $P_{ii}^{(n)}$ ne peut être différent de 0 que si n est un multiple de d_i . Soit alors e_j un état distinct de e_i , et d_j le plus grand commun diviseur de N_j . On a $d_i = d_j$. En effet, e_i et e_j étant des états essentiels, on peut trouver des entiers n et m tels que :

$$(9) \quad P_{ij}^{(n)} > 0 \quad P_{ji}^{(m)} > 0$$

Alors $n + m$ appartient à la fois à N_i et à N_j , car

$$P_{ii}^{(n+m)} \geq P_{ij}^{(n)} P_{ji}^{(m)}$$

$$P_{jj}^{(n+m)} \geq P_{ji}^{(m)} P_{ij}^{(n)}$$

De même, si n_i est un entier quelconque de N_i , on a :

$$P_{jj}^{(n_i+n+m)} \geq P_{ji}^{(m)} P_{ii}^{(n_i)} P_{ij}^{(n)} > 0$$

et $n_i + n + m$ appartient à N_j . Ainsi d_j , qui divise $n+m$ et $n+m+n_i$, divise n_i . Divisant tous les nombres de N_i , d_j divise leur plus grand commun diviseur d_i . Mais, de la même manière, on voit que d_i divise d_j . Par suite :

$$d_i = d_j$$

Ainsi tous les états d'une chaîne irréductible, ou tous les états essentiels d'une même classe dans une chaîne quelconque, ont la même période d .

L'analyse peut être poussée plus loin. En effet, si n et m sont des entiers vérifiant les inégalités (9), $n + m$ est multiple de d , et n est congru à $-m$ modulo d . Tous les entiers n tels que $P_{ij}^{(n)} > 0$ sont donc congrus à un même entier α_j ($0 \leq \alpha_j < d$). Ainsi, à tout état e_j correspond un entier α_j modulo d . Si e_j et e_k correspondent à un même entier α ($0 \leq \alpha < d$), alors $P_{jk}^{(n)}$ n'est différent de 0 que si n est un multiple de d . En effet, tout entier m tel que $P_{ij}^{(m)} > 0$ est congru à α . De $P_{ik}^{(m+n)} \geq P_{ij}^{(m)} P_{jk}^{(n)} > 0$ on déduit que $m + n$ est également congru à α , et par suite n est congru à 0. Appelons S_α l'ensemble des états tels que e_j, e_k correspondant au même entier α modulo d . Pour $\alpha = 0, 1, 2, \dots, d-1$, les S_α constituent une partition de l'ensemble des états. Deux états e_j et e_k appartiennent à une même classe S_α si, et seulement si, $P_{jk}^{(n)}$ ne diffère de 0 que lorsque n est multiple de d . Si e_j et e_k appartiennent à deux classes distinctes S_α et S_β , $P_{jk}^{(n)}$ ne diffère de 0 que lorsque n est congru à $\beta - \alpha$ modulo d .

Si le système est au temps $t = 0$ dans un état e_j de la classe S_α , il passera alternativement par des états des classes $S_{\alpha+1}, S_{\alpha+2}, \dots, S_{\alpha+d} = S_\alpha$ aux temps $t = 1, 2, \dots, d$. Au temps $t = kd + \gamma$ il sera nécessairement dans un état de la classe $S_{\alpha+\gamma}$. Si l'on ne considère que les instants $t = kd$ multiples de la période, le système sera toujours dans la même classe S_α . De cette manière, l'étude d'une chaîne périodique, de période $d > 1$, peut se ramener à celle d'une chaîne de période $d = 1$, dite apériodique. La périodicité introduit, dans l'énoncé

des théorèmes généraux, des complications qui n'ont rien d'essentiel, et nous nous limiterons, en général, au cas des chaînes apériodiques.

Etats transitoires, Etats persistants.

Le système étant, au temps $t = 0$, dans l'état e_i peut retourner, au temps $t = n$, au même état e_i , avec la probabilité $P_{ii}^{(n)}$. On dira que ce retour en e_i au temps n est un premier retour en e_i si les états $e_{i_1}, e_{i_2}, \dots, e_{i_{n-1}}$ pris par le système aux temps $t = 1, 2, \dots, n-1$ sont tous distincts de e_i , c'est-à-dire si l'on a :

$$X_1 \neq i, \quad X_2 \neq i, \quad \dots \quad X_{n-1} \neq i, \quad X_n = i$$

La proposition : " le premier retour en e_i a lieu pour $t = n$ " définit un événement (conditionnel) et possède une probabilité conditionnelle (relative à l'hypothèse que le système se trouve en e_i au temps $t = 0$) que nous noterons $R_n(i)$

$$R_n(i) = P(X_1 \neq i, \dots, X_{n-1} \neq i, X_n = i \mid X_0 = i)$$

et que nous appellerons probabilité de premier retour en e_i au temps n .

L'évènement conditionnel " le système repasse par l'état e_i " est la somme logique des évènements conditionnels " le premier retour en e_i a lieu en $t = n$ " manifestement incompatibles. La somme

$$(10) \quad R_i = \sum_{n=1}^{\infty} R_n(i)$$

représente donc la probabilité (conditionnelle) que le système repasse au moins une fois par l'état e_i . Nous l'appellerons - probabilité de retour en e_i .

Un état e_i sera dit persistant si $R_i = 1$ c'est-à-dire s'il est presque certain que le système repasse par l'état e_i , (sachant qu'il était dans l'état e_i au temps $t = 0$).

Un état e_i sera dit transitoire si $R_i < 1$, c'est-à-dire s'il y a une probabilité non nulle pour que le système ne repasse jamais par l'état e_i .

On notera qu'un état non essentiel est nécessairement transitoire, puisque, par définition, il y a une probabilité non nulle pour que le système passe par un état à partir duquel tout retour est presque impossible. La réciproque est fautive, du moins pour les chaînes infinies : un état essentiel n'est pas nécessairement persistant.

Définition : temps de retour moyen. On appelle temps de retour moyen d'un état persistant l'espérance mathématique (si elle existe)

$$\mu_i = \sum_{n=1}^{\infty} n R_n(i)$$

de l'instant de premier retour.

Un état persistant e_i dont le temps de retour moyen μ_i est fini est appelé état positif. Un état persistant dont le temps de retour moyen est infini est appelé état nul.

Théorème 1 : Pour qu'un état e_i soit transitoire (resp. persistant) il faut et il suffit que la série $\sum_{n=1}^{\infty} P_{ii}^{(n)}$ soit convergente (resp. divergente).

Pour qu'un état e_j soit un état nul il faut et il suffit que la série $\sum_{n=1}^{\infty} P_{jj}^{(n)}$ diverge et que l'on ait $\lim_{n \rightarrow \infty} P_{jj}^{(n)} = 0$.

a) D'après le critère de Borel-Cantelli, si la série $\sum_{n=1}^{\infty} P_{ii}^{(n)}$ converge, l'évènement conditionnel A : "le système ne repasse qu'un nombre fini de fois par l'état e_i " est presque certain. Or A est somme des évènements incompatibles A_k : "le système repasse k fois exactement par l'état e_i ". On a

$$P(A_k) = (R_i)^k (1 - R_i)$$

$$P(A) = \sum_{k=0}^{\infty} P(A_k)$$

Si R_i était égal à 1, tous les $P(A_k)$ seraient nuls, et on aurait $P(A) = 0$ contrairement à l'hypothèse. Donc on a $R_i < 1$, et l'état e_i est transitoire.

b) Inversement, supposons e_i transitoire ($R_i < 1$). Introduisons les fonctions génératrices :

$$G_i(s) = \sum_{n=0}^{\infty} P_{ii}^{(n)} s^n$$

$$H_i(s) = \sum_{n=0}^{\infty} R_n(i) s^n$$

(on pose, conventionnellement, $P_{ii}^{(0)} = 1$ et $R_0(i) = 0$).

Un retour au temps n peut être un premier retour, ou bien avoir été précédé d'un premier retour en un temps k ($0 < k < n$). On en déduit :

$$P_{ii}^{(n)} = R_n(i) + \sum_{k=1}^{n-1} R_k(i) P_{ii}^{(n-k)} = \sum_{k=1}^n R_k(i) P_{ii}^{(n-k)}$$

Multiplions les deux membres par s^n et sommons de $n=1$ à l'infini. A gauche, il vient $G_i(s) - 1$. A droite, on obtient :

$$\sum_{n=1}^{\infty} \sum_{k=1}^n R_k(i) P_{ii}^{(n-k)} s^n = \sum_{k=1}^{\infty} R_k(i) s^k \sum_{n=k}^{\infty} P_{ii}^{(n-k)} s^{n-k}$$

D'où :

$$G_i(s) - 1 = H_i(s) G_i(s)$$

et :

$$(11) \quad G_i(s) = \frac{1}{1 - H_i(s)}$$

Si e_i est transitoire, on a $R_i = H_i(1) < 1$, et $G_i(1) = \frac{1}{1 - R_i}$ prend une valeur finie. La fonction $G_i(s)$ est, non décroissante et continue pour $|s| \leq 1$.

Pour tout N on a :

$$\sum_{n=0}^N P_{ii}^{(n)} \leq \lim_{s \rightarrow 1} G_i(s) = G_i(1) = \frac{1}{1 - R_i}$$

Par suite la série des $P_{ii}^{(n)}$ est convergente : sa somme est du reste égale à $\frac{1}{1 - R_i}$, à cause de la continuité de $G_i(s)$ en $s = 1$.

c) La démonstration de l'énoncé relatif aux états nuls est un peu plus délicate. Elle découle du théorème suivant, que nous donnerons sans démonstration et qui nous servira plus tard pour établir les propriétés ergodiques des chaînes infinies.

Théorème 2 - Soit e_i un état persistant et apériodique d'une chaîne markovienne.

Si e_i est un état positif et μ_i son temps de retour moyen, $P_{ii}^{(n)}$ tend vers une limite pour $n \rightarrow \infty$, et on a :

$$\lim_{n \rightarrow \infty} P_{ii}^{(n)} = \frac{1}{\mu_i}$$

si e_i est un état nul, on a :

$$\lim_{n \rightarrow \infty} P_{ii}^{(n)} = 0$$

Considérons, maintenant, une chaîne de Markov irréductible, et apériodique. Tout état est un état essentiel (par définition). Soient e_i et e_j deux états distincts, N et M deux entiers tels que :

$$P_{ij}^{(N)} > 0 \quad P_{ji}^{(M)} > 0$$

Alors, pour tout entier n , on a :

$$1 \geq P_{ii}^{(n+N+M)} \geq P_{ij}^{(N)} P_{jj}^{(n)} P_{ji}^{(M)}$$

$$1 \geq P_{jj}^{(n+N+M)} \geq P_{ji}^{(M)} P_{ii}^{(n)} P_{ij}^{(N)}$$

Ces inégalités montrent que $P_{ii}^{(n)}$ et $P_{jj}^{(n)}$ se comportent de la même

manière lorsque n tend vers l'infini. Si $P_{ii}^{(n)}$ a une limite différente de 0, $P_{jj}^{(n)}$ a aussi une limite différente de 0. Si $P_{ii}^{(n)}$ tend vers 0, il en est de même de $P_{jj}^{(n)}$. Si la série $\sum P_{ii}^{(n)}$ converge, la série $\sum P_{jj}^{(n)}$ est aussi convergente. Les théorèmes 1 et 2 permettent donc d'énoncer :

Théorème 3 - Tous les états d'une chaîne irréductible aperiodique sont de même nature. Ils sont tous transitoires, ou bien tous persistants et nuls, ou bien tous persistants et positifs.

Le même résultat s'applique, dans le cas d'une chaîne quelconque, aux états essentiels appartenant à une même classe de correspondance. D'autre part, il s'applique également aux chaînes périodiques : il suffit, pour le voir, de faire varier le temps par multiples entiers de la période d et d'appliquer le résultat précédent à chacune des classes S_{α} , qui correspondent à autant de chaînes aperiodiques. En ce qui concerne les chaînes finies, on a le résultat suivant :

Théorème 4 - Dans une chaîne finie, tout état essentiel est persistant et positif. Il n'existe aucun état nul, et il est impossible que tous les états soient transitoires.

Montrons d'abord que les états ne peuvent pas être tous transitoires. En effet, si e_i est transitoire, l'évènement " au delà d'un temps fini, le système ne passe plus par l'état e_i " est presque certain. Si un nombre fini d'états $e_{i_1} \dots e_{i_k}$ sont transitoires, l'évènement " au delà d'un temps fini, le système ne passe plus par aucun des évènements $e_{i_1} \dots e_{i_k}$ " est également presque certain. Par suite, il n'est pas possible que tous les e_i soient transitoires.

D'autre part, d'après le théorème 3, dans une chaîne irréductible tous les états sont de même nature : si la chaîne est finie, ils sont donc tous persistants. Dans une chaîne quelconque, un état essentiel appartient à une classe de correspondance qui constitue une chaîne irréductible. Donc tout état essentiel est persistant.

Enfin si e_i était un état nul, on aurait

$$\lim_{n \rightarrow \infty} P_{ii}^{(n)} = 0$$

Si N est le nombre total des états de la chaîne, on a :

$$P_{ii}^{(n+m)} = \sum_{k=1}^N P_{ik}^{(n)} P_{ki}^{(m)}$$

Par suite

$$(12) \quad \lim_{n \rightarrow \infty} \sum_{k=1}^N P_{ik}^{(n)} = 0$$

pour tout indice k tel qu'il existe un chemin de k à i . Donc, dans le cas d'une chaîne irréductible, ceci aurait lieu pour tous les indices k . Or on a nécessairement pour tout i :

$$\sum_{k=1}^N P_{ik}^{(n)} = 1$$

et par suite (12) est impossible : e_i ne peut pas être un état nul. Si e_i est un état essentiel d'une chaîne quelconque, la conclusion subsiste évidemment. Comme tout état persistant est essentiel, il ne peut donc exister aucun état nul.

III.- PROPRIÉTÉS ERGODIQUES DES CHAINES IRREDUCTIBLES.

Proposons nous, maintenant, de chercher sous quelles conditions une chaîne irréductible possède la propriété d'ergodicité, c'est-à-dire sous quelles conditions les probabilités $p_k^{(n)}$ définies en (6) tendent vers des limites u_k indépendantes des probabilités initiales. Ce problème est étroitement lié au comportement, pour n infini, de la matrice de transition $P_{ij}^{(n)}$. Pour simplifier, nous nous limiterons au cas des chaînes apériodiques. Le résultat essentiel est le suivant :

Théorème 5 - Soit une chaîne de Markov irréductible et apériodique. La probabilité de transition $P_{ij}^{(n)}$ a toujours une limite u_j indépendante de i . Si les états sont soit tous transitoires, soit tous persistants et nuls, ces limites u_j sont toutes nulles. Si les états sont tous persistants positifs u_j est égal à l'inverse $\frac{1}{\mu_j}$

du temps de retour moyen de l'état e_j . Les u_j vérifient, dans ce cas, les relations :

$$(13) \quad \left\{ \begin{array}{l} u_j = \sum_{i=1}^{\infty} u_i P_{ij} \\ \sum_{j=1}^{\infty} u_j = 1 \end{array} \right.$$

et, quelles que soient les probabilités initiales p_k , les probabilités au temps n vérifient :

$$(14) \quad \lim_{n \rightarrow \infty} p_k^{(n)} = u_k = \frac{1}{\mu_k}$$

a) Supposons d'abord que les états soient ou bien tous transitoires, ou bien tous persistants ou nuls. D'après les théorèmes 1 et 2, on a dans chacun de ces deux cas :

$$\lim_{n \rightarrow \infty} P_{ii}^{(n)} = 0$$

Il reste à montrer que, si e_i et e_j sont deux états distincts, on a aussi :

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)} = 0$$

Or, la chaîne étant irréductible, on peut trouver deux entiers N et M avec :

$$P_{jj}^{(N)} > 0 \quad P_{ji}^{(M)} > 0$$

Dans l'inégalité :

$$P_{ij}^{(n)} P_{jj}^{(N)} P_{ji}^{(M)} \leq P_{ii}^{(n+N+M)}$$

faisons tendre n vers l'infini : le deuxième membre tend vers 0, et par suite aussi le premier, ce qui établit la première partie du théorème.

b) Nous supposons, maintenant, que les états sont tous persistants positifs. Le théorème 2 montre que l'on a, pour tout i :

$$\lim_{n \rightarrow \infty} P_{ii}^{(n)} = \frac{1}{\mu_i} = u_i$$

Montrons que, pour $i \neq j$, $P_{ij}^{(n)}$ possède également une limite.

Etant supposé que le système est dans l'état e_i à l'instant initial, l'évènement conditionnel B_n : "le système passe pour la première fois par l'état e_j au temps $t = n$ " possède une probabilité définie que nous noterons b_n .

$$b_n = P(B_n)$$

On a d'ailleurs $b_0 = 0$. Comme les B_n sont incompatibles, on a :

$$(15) \quad \sum_{n=1}^{\infty} b_n \leq 1$$

de sorte que, quel que soit $\varepsilon > 0$, on peut trouver un entier N avec :

$$(16) \quad \sum_{n=N+1}^{\infty} b_n \leq \varepsilon$$

Par ailleurs, l'évènement A_n : "le système passe par l'état e_j au temps n " est somme des évènements incompatibles B_n et $B_k A_{n-k}$, d'où :

$$P_{ij}^{(n)} = b_n + b_{n-1} P_{jj}^{(1)} + b_{n-2} P_{jj}^{(2)} + \dots + b_1 P_{jj}^{(n-1)}$$

Pour n supérieur à N , et compte tenu des inégalités (15) et (16), on en déduit :

$$b_1 P_{jj}^{(n-1)} + \dots + b_N P_{jj}^{(n-N)} \leq P_{ij}^{(n)} \leq b_1 P_{jj}^{(n-1)} + \dots + b_N P_{jj}^{(n-N)} + \varepsilon$$

Passant à la limite en n , et puisque $P_{jj}^{(n)} \rightarrow u_j$, on obtient :

$$\frac{b_1 + \dots + b_N}{\mu_j} \leq \lim_{n \rightarrow \infty} P_{ij}^{(n)} \leq \overline{\lim}_{n \rightarrow \infty} P_{ij}^{(n)} \leq \frac{b_1 + \dots + b_N}{\mu_j} + \varepsilon$$

Passons à la limite en N , en posant :

$$b = \sum_{n=1}^{\infty} b_n$$

$$\text{Il vient : } \frac{b}{\mu_j} \leq \underline{\lim} P_{ij}^{(n)} \leq \overline{\lim} P_{ij}^{(n)} \leq \frac{b + \varepsilon}{\mu_j}$$

Comme ε est arbitrairement petit, on en déduit :

$$\underline{\lim} P_{ij}^{(n)} = \overline{\lim} P_{ij}^{(n)} = \lim P_{ij}^{(n)} = \frac{b}{\mu_j}$$

Ainsi $P_{ij}^{(n)}$ possède une limite $\frac{b}{\mu_j}$, où b représente la probabilité pour

que, partant de l'état e_i , le système passe au moins une fois par l'état e_j .

- c) Montrons que l'on a nécessairement $b = 1$. En effet, supposons $1 - b > 0$. Par définition, l'état e_j étant persistant, l'évènement conditionnel "le système repasse par e_j " est presque certain lorsque l'état initial est e_j . D'autre part, la chaîne étant irréductible, on peut trouver m avec $P_{ji}^{(m)} > 0$. L'état initial étant e_j , la probabilité pour que le système ne repasse jamais par e_j est supérieure ou égale à $(1 - b) P_{ji}^{(m)} > 0$, d'où la contradiction. Par suite, nécessairement, $b = 1$ et :

$$(17) \quad \lim_{n \rightarrow \infty} P_{ij}^{(n)} = \frac{1}{\mu_j} = u_j$$

- d) Montrons que les u_j vérifient les relations (13). Comme, pour tout n , et tout N :

$$\sum_{k=1}^N P_{ik}^{(n)} \leq 1$$

On a aussi, en passant à la limite d'abord en n , puis en N

$$(18) \quad \sum_{k=1}^{\infty} u_k \leq 1$$

Mais, de :

$$P_{ik}^{(n+1)} \geq \sum_{j=1}^N P_{ij}^{(n)} P_{jk}$$

on déduit aussi :

$$u_k \geq \sum_{j=1}^N u_j P_{jk}$$

et par suite :

$$(19) \quad u_k \geq \sum_{j=1}^{\infty} u_j P_{jk}$$

Si l'on somme en k les deux membres de (19), on obtient $\sum u_k$ des deux côtés : par suite l'inégalité stricte est impossible, et on a :

$$(20) \quad u_k = \sum_{j=1}^{\infty} u_j P_{jk}$$

Sous forme matricielle, (20) s'écrit $u = uP$, u est vecteur propre de P pour la valeur propre 1. Par une récurrence immédiate, on en déduit $u = uP^n$, soit

$$u_k = \sum_{j=1}^{\infty} u_j P_{jk}^{(n)}$$

D'après (18), on peut trouver N tel que

$$\sum_{k=N+1}^{\infty} u_k \leq \varepsilon$$

Comme $P_{jk}^{(n)} \leq 1$, on a donc

$$\sum_{j=1}^N u_j P_{jk}^{(n)} \leq u_k \leq \sum_{j=1}^N u_j P_{jk}^{(n)} + \varepsilon$$

Passant à la limite en n :

$$u_k \sum_{j=1}^N u_j \leq u_k \leq u_k \sum_{j=1}^N u_j + \varepsilon$$

Par suite aussi

$$u_k \sum_{j=1}^{\infty} u_j \leq u_k \leq u_k \sum_{j=1}^{\infty} u_j + \varepsilon$$

Comme les u_k ne sont pas nuls et que ε est arbitraire, on en déduit :

$$\sum_{j=1}^{\infty} u_j = 1$$

Par suite les relations (13) sont vérifiées, et les u_j constituent un système de probabilité.

e) Si les p_k sont les probabilités initiales, on a :

$$p_k^{(n)} = \sum_{j=1}^{\infty} p_j P_{jk}^{(n)}$$

Passant, comme ci-dessus, à la limite en n il vient :

$$\lim_{n \rightarrow \infty} p_k^{(n)} = u_k \sum_{j=1}^{\infty} p_j = u_k$$

Convenons de dire qu'un système p_k de probabilités initiales est

stationnaire si l'on a :

$$(23) \quad p_k = \sum_{i=1}^{\infty} p_i P_{ik}$$

On a immédiatement le résultat suivant :

Corollaire 1 - S'il existe un système p_k de probabilités stationnaires, ce système est unique, les états de la chaîne irréductible sont nécessairement persistants positifs, et on a $p_k = u_k = \lim_{n \rightarrow \infty} P_{ik}^{(n)}$

En effet, de (23) on déduit immédiatement :

$$p_k = \sum_{i=1}^{\infty} p_i P_{ik}^{(n)}$$

Si les états sont nuls, ou transitoires, $\lim_{n \rightarrow \infty} P_{ik}^{(n)} = 0$ d'après le théorème 5 et par suite $p_k = 0$: les p_k ne constituent pas un système de probabilités. Si les états sont persistants positifs, on a, pour $n \rightarrow \infty$, $p_k = u_k \sum_{i=1}^{\infty} p_i = u_k$ d'où résulte l'unicité du système de probabilités stationnaires.

Dans le cas d'une chaîne finie irréductible et aperiodique, l'énoncé se simplifie, puisque les états sont tous persistants positifs. En particulier, si une chaîne finie est telle que, pour un entier s , tous les $P_{ij}^{(s)}$ soient différents de zéro, elle est nécessairement irréductible et aperiodique. D'où :

Corollaire 2 - Si une chaîne d'ordre k finie est telle que, pour un entier s , aucune des probabilités de transition $P_{ij}^{(s)}$ ne soit nulle, cette chaîne est ergodique.

On a :

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)} = u_j = \frac{1}{\mu_j}$$

$$\sum_{j=1}^k u_j = 1$$

$$u_j = \sum_{i=1}^k u_i P_{ij}$$

IV.- ETUDE DES ETATS TRANSITOIRES.

Soit, maintenant, e_i un état d'une chaîne quelconque. Si e_i est un état essentiel, il appartient à une classe d'états essentiels qui constitue une chaîne irréductible, et le théorème 5 permet d'étudier le comportement de $P_{ij}^{(n)}$ pour $n \rightarrow \infty$. Il reste à examiner le cas où e_i est non essentiel, donc aussi transitoire.

Soit donc e_i un état non essentiel et transitoire, et soit e_j un autre état. Désignons par b_{ij} la probabilité conditionnelle de l'évènement : "le système passera au moins une fois par l'état e_j " sachant que l'état initial était e_i . Si $P_{jj}^{(n)}$ a une limite u_j , il suffit de répéter mot pour mot la partie b) de la démonstration du théorème 5 pour voir que $P_{ij}^{(n)}$ a aussi une limite, et que l'on a :

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)} = u_j b_{ij}$$

Énonçons :

Théorème 6 - Dans une chaîne quelconque, si e_j est un état transitoire ou persistant et nul, $P_{ij}^{(n)}$ tend vers 0 quel que soit e_i . Si e_i est transitoire et e_j persistant, positif et apériodique, de temps de retour moyen μ_j , on a :

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)} = \frac{b_{ij}}{\mu_j}$$

b_{ij} représentant la probabilité conditionnelle pour que le système, partant de l'état e_i , passe au moins une fois par l'état e_j .

Si e_j est persistant, il est aussi essentiel. Soit e_k un autre état essentiel appartenant à la même classe C que e_j . On a, nécessairement :

$$b_{ij} = b_{ik}$$

En effet, si le système est passé une fois par e_j il est presque certain qu'il passera ensuite au moins une fois par chacun des états de la classe de e_j (ce point a fait l'objet de la partie c) de la démonstration du théorème 5). Par suite, on a $b_{ij} \leq b_{ik}$. De la même manière, on a aussi $b_{ik} \leq b_{ij}$, d'où l'égalité. Posons :

$$a_i = b_{ij} = b_{ik}$$

a_i représente la probabilité pour que le système passe par l'un quelconque des états de la classe C à laquelle appartient e_j . Si le système entre une fois dans la classe C , il est presque certain qu'il n'en sortira plus jamais : a_i représente donc aussi la probabilité d'absorption du système dans la classe C .

Calcul de la probabilité d'absorption. Proposons nous de calculer la probabilité a_i pour que le système, partant de l'état initial non essentiel e_i , soit absorbé par une classe C donnée d'états essentiels, et, corrélativement, la probabilité h_i pour que le système ne soit absorbé par aucune classe C , c'est-à-dire pour que le système ne prenne jamais que des états transitoires non essentiels.

Désignons par A_n l'évènement conditionnel : " le système entre pour la première fois dans un état de la classe C au temps $t = n$ " et par $a_i^{(n)}$ sa probabilité. On a :

$$(24) \quad a_i = \sum_{n=1}^{\infty} a_i^{(n)}$$

Soit T l'ensemble des états non essentiels. Pour que l'absorption ait lieu pour la première fois au temps $t = n$, il faut qu'au temps $t = n-1$ le système ait pris un état non essentiel quelconque de T . On en déduit :

$$(25) \quad a_i^{(n+1)} = \sum_{k \in T} P_{ik} a_k^{(n)}$$

Par ailleurs, pour $n=1$, on a manifestement :

$$(26) \quad a_i^{(1)} = \sum_{j \in C} P_{ij}$$

Les équations (25) et (26) permettent de calculer, par récurrence, tous les $a_i^{(n)}$, et l'équation (24) donne la probabilité d'absorption a_i .

Sommant (25) de $n=1$ à l'infini, on voit aussi que les a_i sont solution du système d'équations :

$$(27) \quad a_i - \sum_{k \in T} P_{ik} a_k = a_i^{(1)}$$

Inversement, on peut se demander si le système (27) possède une solution unique, et permet ainsi de caractériser complètement les probabilités d'absorption a_i . Ce problème est lié étroitement au suivant : partant d'un état non essentiel e_i quelconque, est-il presque certain que le système sera finalement absorbé (dans une classe quelconque), autrement dit a-t-on $h_i = 0$?

En effet, désignons par $h_i^{(n)}$ la probabilité pour que le système (partant de l'état e_i) soit dans un état non essentiel au temps $t = n$. On a immédiatement :

$$(28) \quad \left\{ \begin{array}{l} h_i^{(1)} = \sum_{k \in T} P_{ik} \\ h_i^{(n+1)} = \sum_{k \in T} P_{ik} h_k^{(n)} \end{array} \right.$$

On en déduit : $h_i^{(1)} \leq 1$, et, par récurrence, $h_i^{(n+1)} \leq h_i^{(n)}$. Suite décroissante, les $h_i^{(n)}$ possèdent une limite, qui n'est autre que la probabilité h_i pour que le système ne soit jamais absorbé :

$$\lim_{n \rightarrow \infty} h_i^{(n)} = h_i$$

Passant à la limite $n \rightarrow \infty$ dans (28), on voit que h_i est solution du système :

$$(29) \quad h_i = \sum_{k \in T} P_{ik} h_k$$

Plus précisément, les h_i peuvent être caractérisés comme la solution maximale du système (29) vérifiant $|h_i| \leq 1$. En effet, soient x_i des quantités vérifiant :

$$(30) \quad x_i = \sum_{k \in T} P_{ik} x_k \quad |x_i| \leq 1$$

On a manifestement

$$|x_i| \leq \sum_{k \in T} P_{ik} = h_i^{(1)}$$

et, par récurrence, les équations (28) donnent $|x_i| \leq h_i^{(n)}$ quel que soit n , par suite aussi

$$|x_i| \leq h_i$$

Les h_i sont donc bien la solution maximale de (30) pour $|x_i| \leq 1$.

Par suite, si tous les h_i sont nuls, le système (30) n'admet pas d'autres solutions bornées que la solution zéro. Dans ce cas, le système (27) possède une solution unique et caractérise complètement les probabilités d'absorption a_i . Au contraire, si l'un au moins des h_i est non nul, le système (27) admet d'autres solutions que les a_i . Ainsi :

Théorème 7 - Les probabilités h_i pour que le système ne sorte jamais de l'ensemble T des états non essentiels sont la solution maximale du système (30). Les probabilités d'absorption a_i sont solution du système (27). Pour que la solution du système (27) soit unique, et caractérise ainsi complètement les a_i , il faut et il suffit que les h_i soient toutes nulles.

Dans une chaîne finie, cette condition $h_i = 0$ est toujours vérifiée. En effet, on a toujours

$$h_i^{(n)} = \sum_{k \in T} P_{ik}^{(n)}$$

Mais il n'y a qu'un nombre fini d'états non essentiels e_k dans T, et, pour chacun d'eux, on a $P_{ik}^{(n)} \rightarrow 0$. Donc :

$$h_i = \lim_{n \rightarrow \infty} h_i^{(n)} = \lim_{n \rightarrow \infty} \sum_{k \in T} P_{ik}^{(n)} = 0$$

D'où :

Corollaire : Dans une chaîne finie, il est presque impossible que le système ne sorte jamais de l'ensemble T : l'absorption finale est presque certaine, et les probabilités d'absorption a_i constituent la solution unique du système (27).

CHAPITRE VII

CHAINES DE MARKOV HOMOGENES A TEMPS CONTINU

I.- DEFINITION ET HYPOTHESE

Les chaînes de Markov étudiées au chapitre précédent décrivent l'évolution d'un système susceptible de changer d'états aux instants (discrets) $t = 1, 2, \dots$ donnés d'avance, et non aléatoires. Nous dirons que ce sont des chaînes à temps discret. Très souvent, on a affaire à des systèmes physiques dont les changements d'état, au contraire, peuvent se produire à des instants quelconques, inconnus à l'avance, c'est-à-dire à des instants t_1, t_2, \dots aléatoires. L'évolution d'un tel système dans le temps est représentée par le processus stochastique $X(t)$ défini par :

$$X(t) = i \quad \text{si le système est dans l'état } e_i \text{ au temps } t.$$

L'ensemble des états possibles e_i est supposé dénombrable (ou fini), de sorte que les valeurs possibles de $X(t)$ sont les entiers positifs $1, 2, \dots$. $X(t)$ est une fonction (aléatoire) complètement discontinue, ou fonction de sauts.

Le processus (complètement discontinu) $X(t)$ est dit markovien si, quels que soient les instants :

$$t_1 < t_2 < \dots < t_k < t_0 < t$$

la probabilité (conditionnelle) d'avoir $X(t) = i$, sachant que l'on avait $X(t_0) = i_0$ n'est pas affectée par les valeurs prises par $X(t)$ aux instants antérieurs t_1, t_2, \dots, t_k , autrement dit, si l'on a :

$$\begin{aligned} P(X(t) = i \mid X(t_0) = i_0, X(t_1) = i_1, \dots, X(t_k) = i_k) \\ = P(X(t) = i \mid X(t_0) = i_0) \end{aligned}$$

En langage bref : le processus est markovien si à chaque instant, l'évolution future du système ne dépend que de son état présent, et non de son histoire antérieure. On dit aussi que ce processus constitue une chaîne de Markov à temps continu.

Posons (pour $\mathcal{T} \leq t$)

$$P_{ij}(\mathcal{T}, t) = P(X(t) = j \mid X(\mathcal{T}) = i)$$

$P_{ij}(\mathcal{T}, t)$ est la probabilité de transition de l'état i (au temps \mathcal{T}) à l'état j (au temps t). La matrice $P(\mathcal{T}, t)$ des $P_{ij}(\mathcal{T}, t)$, ou matrice de transition de la chaîne de Markov, vérifie les relations :

$$\left\{ \begin{array}{l} 0 \leq P_{ij}(\mathcal{T}, t) \leq 1 \\ \sum_{j=1}^{\infty} P_{ij}(\mathcal{T}, t) = 1 \end{array} \right.$$

et l'équation de Markov :

$$P_{ij}(\mathcal{T}, t_1 + t_2) = \sum_k P_{ik}(\mathcal{T}, t_1) P_{kj}(t_1, t_1 + t_2)$$

Tout comme la relation (4) du chapitre VI, cette équation exprime que la transition de e_i (au temps \mathcal{T}) à e_j (au temps $t_1 + t_2$) implique le passage par un état e_k ($k = 1, 2 \dots$) quelconque au temps intermédiaire t_1 .

Enfin si $P_{ij}(\mathcal{T}, t)$ ne dépend que de la différence $t - \mathcal{T}$, c'est-à-dire si l'on a :

$$P_{ij}(\mathcal{T} + h, t + h) = P_{ij}(\mathcal{T}, t) \quad \forall h$$

on dit que la chaîne de Markov à temps continu est homogène. Nous nous limiterons ici à l'étude du cas homogène. Au lieu de $P_{ij}(\mathcal{T}, t)$, nous écrirons $P_{ij}(t - \mathcal{T})$. La matrice de transition se met sous la forme $P_{ij}(t)$, et vérifie les conditions :

$$(1) \quad \left\{ \begin{array}{l} 0 \leq P_{ij}(t) \leq 1 \\ \sum_{j=1}^{\infty} P_{ij}(t) = 1 \end{array} \right.$$

$$(2) \quad P_{ij}(t + t') = \sum_{k=1}^{\infty} P_{ik}(t) P_{kj}(t')$$

Remarque

La définition donnée ci-dessus n'a rien de rigoureux, mais suffit dans les applications. Du point de vue axiomatique, on devrait prendre comme espace d'événements E l'ensemble des fonctions $X(t)$ à valeurs entières positives, et définir sur E une σ -algèbre et une probabilité. La nécessité d'une telle définition apparaît lorsqu'on se pose la question suivante : les propositions " le système reste dans l'état i de $t = t_1$ à $t = t_2$ " ou " le système ne passe par l'état i à aucun instant compris entre t_1 et t_2 " définissent-elles des événements ? Si " $X(t) = i$ " par exemple, est un événement, la deuxième proposition apparaît comme le complément de la somme logique :

$$\sum_{t \in [t_1, t_2]} "X(t) = i"$$

d'une infinité non dénombrable d'événements, et il n'est nullement évident qu'elle constitue elle-même un événement. Seule la définition axiomatique rigoureuse permet de lever de telles difficultés.

Pour éviter ce genre de complications, nous ferons des hypothèses assez restrictives (qui seront d'ailleurs presque toujours vérifiées dans les applications).

Supposons que le système soit dans l'état e_i à l'instant $t = 0$. Nous admettrons que les propositions :

A = " le système reste dans l'état e_i de $t = 0$ à $t = t_1$ "

$$= " X(t) = i \quad \forall t, \quad 0 \leq t \leq t_1 "$$

et

B = " le système ne subit qu'un changement d'état dans l'intervalle $0 \leq t \leq t_1$,
et se trouve en e_j au temps $t = t_1$ "

$$= " \exists t_0, \quad 0 < t_0 \leq t_1 \quad \text{tel que : } X(t) = i \quad \text{pour } 0 \leq t < t_0$$

et $X(t) = j \quad \text{pour } t_0 \leq t \leq t_1 "$

définissent des évènements (conditionnels), et nous désignerons leurs probabilités par :

$$P_{ii}^{(0)}(t_1) = P(A)$$

$$P_{ij}^{(1)}(t_1) = P(B) \quad (i \neq j)$$

Plus généralement, $P_{ij}^{(n)}(t)$ représentera la probabilité de passer, pendant une durée t , de l'état e_i à l'état e_j en effectuant exactement n changements d'état.

Posons :

$$\overline{P}_{ij}(t) = \sum_{n=0}^{\infty} P_{ij}^{(n)}(t)$$

$\overline{P}_{ij}(t)$ représente la probabilité pour que le système passe de e_i à e_j en effectuant un nombre fini (quelconque) de changements d'état.

On a nécessairement :

$$\left\{ \begin{array}{l} \overline{P}_{ij}(t) \leq P_{ij}(t) \\ \sum_j \overline{P}_{ij}(t) \leq 1 \end{array} \right.$$

Mais, sans hypothèse supplémentaire, on peut très bien avoir des inégalités strictes, parce que rien n'indique a priori que l'évènement : "le système subit une infinité de changements d'états pendant le temps fini t " soit presque impossible.

Si on a l'inégalité stricte

$$\sum_j \overline{P}_{ij}(t) < 1$$

Il y a une probabilité non nulle $1 - \sum_j \overline{P}_{ij}(t)$ pour que $X(t)$ possède une infinité de points de discontinuités t_1, t_2, \dots sur l'intervalle fini $(0, t)$. Mais, dans ce cas, t_1, t_2, \dots possèdent un point d'accumulation t_0 et, en t_0 , $X(t)$ présente une discontinuité non isolée, de nature plus complexe qu'un simple saut. C'est pour éviter (presque certainement) cette éventualité que des hypothèses supplémentaires sont obligatoires.

En fait, nous n'insisterons pas sur ce genre de difficultés. Nous admettrons sans démonstration les propriétés suivantes, qui sont effectivement presque toujours vérifiées dans les applications :

1.- Pour tout intervalle de temps fini t , il est presque certain que le système ne subit qu'un nombre fini de changements d'état de 0 à t :

2.- Classification des états : Pour tout nombre $\alpha > 0$, les $P_{ij}(n\alpha)$ constituent la matrice de transition d'ordre n de la chaîne à temps discret définie par $P_{ij}(\alpha)$. Pour tout $\alpha > 0$, la chaîne à temps continu $P_{ij}(t)$ et la chaîne à temps discret $P_{ij}(n\alpha)$ ont mêmes états essentiels et mêmes états non essentiels. De plus, les états essentiels sont toujours apériodiques. Si le système est en $t = 0$ dans un état E_i appartenant à une classe C d'états essentiels, il est presque certain qu'il ne sortira de C à aucun instant ultérieur.

3.- Limites ergodiques. La chaîne à temps continu $P_{ij}(t)$ possède, pour $t \rightarrow \infty$, les mêmes limites ergodiques que la chaîne à temps discret $P_{ij}(n\alpha)$. De même, la probabilité d'absorption dans une classe C d'états essentiels est la même pour les deux chaînes.

Ainsi, les questions relatives à la classification des états ou au comportement ergodique de la chaîne $P_{ij}(t)$ pourront se traiter par les méthodes du chapitre précédent.

Par contre, du fait que le temps est traité comme un paramètre continu, on peut se demander dans quelle mesure la matrice des $P_{ij}(t)$ est continue et dérivable. Ce genre de problème, qui n'avait évidemment pas d'équivalent dans le cas des chaînes à temps discret, va nous conduire aux célèbres équations de Kolmogorov, grâce auxquelles les chaînes à temps continu sont souvent plus faciles à traiter que leurs homologues à temps discret.

II.- LES EQUATIONS DE KOLMOGOROV.

Nous partirons de l'équation de Markov

$$(3) \quad P_{ij}(t_1 + t_2) = \sum_k P_{ik}(t_1) P_{kj}(t_2)$$

ou, sous forme matricielle :

$$P(t_1 + t_2) = P(t_1) P(t_2)$$

Admettant provisoirement qu'une telle opération soit légitime, dérivons en t_1 et faisons $t_1 = 0$. Il vient formellement

$$(K_1) \quad P'(t) = P'(0) P(t)$$

De même, dérivons en t_2 et faisons $t_2 = 0$: il vient

$$(K_2) \quad P'(t) = P(t) P'(0)$$

Si l'on connaît la dérivée $P'(0)$ de la matrice de transition, l'une quelconque de ces équations différentielles doit donc permettre d'en déduire $P(t)$. Ce sont les équations de Kolmogorov. La première s'obtient en faisant varier l'instant initial dans l'équation de Markov, la deuxième en faisant varier l'instant final.

Pour établir rigoureusement ces deux équations, nous ferons les trois hypothèses suivantes :

H_1 - Lorsque $\Delta t \rightarrow 0$, on a les limites suivantes :

$$(4) \quad \left\{ \begin{array}{l} \Delta t \xrightarrow{\lim} 0 \frac{1 - P_{ii}(\Delta t)}{\Delta t} = \lambda_i \\ \Delta t \xrightarrow{\lim} 0 \frac{P_{ik}(\Delta t)}{\Delta t} = \mu_{ik} \end{array} \right.$$

H_2 - Ces limites vérifient la relation :

$$(5) \quad \sum_{k \neq i} \mu_{ik} = \lambda_i$$

H_3 - La deuxième limite (4) est uniforme en i .

Remarque - Dans le cas d'une chaîne finie, la relation (5) découle immédiatement de (4)

et de $\sum_j P_{ij}(t) = 1$ et l'hypothèse H_3 est automatiquement vérifiée, de sorte que la seule hypothèse requise est H_1 . Par contre, lorsqu'il y a une infinité dénombrable d'états, H_2 et H_3 sont des hypothèses supplémentaires qui ne découlent pas de H_1 .

En fait, la première équation de Kolmogorov n'exige que les deux premières hypothèses H_1 et H_2 . L'hypothèse d'uniformité H_3 n'est indispensable que pour

la deuxième équation de Kolmogorov.

Première équation de Kolmogorov. (K_1)

Posons $t_1 = \Delta t$ dans l'équation de Markov écrite en (3). Il vient :

$$(6) \quad P_{ij}(t + \Delta t) = P_{ii}(\Delta t) P_{ij}(t) + \sum_{k \neq i} P_{ik}(\Delta t) P_{kj}(t)$$

La somme qui figure au 2ème membre de (6) est majorée par

$$\sum_{k \neq i} P_{ik}(\Delta t) = 1 - P_{ii}(\Delta t)$$

Mais, d'après l'hypothèse H_1 , $1 - P_{ii}(\Delta t)$ tend vers 0 lorsque $\Delta t \rightarrow 0$.
L'équation (6) montre ainsi que l'on a :

$$\lim_{\Delta t \rightarrow 0} P_{ij}(t + \Delta t) = P_{ij}(t)$$

autrement dit que $P_{ij}(t)$ est continue à droite. On montre de même que $P_{ij}(t)$ est continue à gauche (il suffit de remplacer dans (6) t par $t - \Delta t$). Par suite $P_{ij}(t)$ est continue.

Compte tenu de l'hypothèse (4), posons

$$(7) \quad P_{ii}(\Delta t) = 1 - \lambda_i \Delta t - o_i(\Delta t)$$

les $o_i(\Delta t)$ étant des infiniment petits d'ordre supérieur à 1, et portons dans (6).
Il vient :

$$(8) \quad \frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} = -(\lambda_i + \frac{o_i(\Delta t)}{\Delta t}) P_{ij}(t) + \sum_{k \neq i} \frac{P_{ik}(\Delta t)}{\Delta t} P_{kj}(t)$$

Il faut montrer que le passage à la limite sous le signe somme est légitime.

Soit N un entier supérieur à i . On a :

$$(9) \quad \left\{ \begin{aligned} 0 &\leq \sum_{k=N+1}^{\infty} \frac{P_{ik}(\Delta t)}{\Delta t} P_{kj}(t) \\ &\leq \frac{1}{\Delta t} \sum_{k=N+1}^{\infty} P_{ik}(\Delta t) = \frac{1}{\Delta t} \left[1 - \sum_{k=1}^N P_{ik}(\Delta t) \right] \end{aligned} \right.$$

Cette dernière expression ne comporte qu'un nombre fini de termes. Lorsque Δt tend vers 0, elle admet donc la limite :

$$L_N = \lambda_i - \sum_{\substack{k=1 \\ (k \neq i)}}^{-N} \mu_{ik}$$

D'après (5), on peut prendre N assez grand pour que cette limite L_N soit inférieure à ε

$$L_N < \varepsilon$$

Prenant Δt suffisamment petit, on déduira de (9)

$$0 \leq \sum_{k=N+1}^{\infty} \frac{P_{ik}(\Delta t)}{\Delta t} P_{kj}(t) \leq L_N + \varepsilon \leq 2\varepsilon$$

Il en résulte bien que le passage à la limite sous le signe somme, dans (8), est légitime :

$$\Delta t \rightarrow 0 \quad \lim \sum_{k \neq i} \frac{P_{ik}(\Delta t)}{\Delta t} P_{kj}(t) = \sum_{k \neq i} \mu_{ik} P_{kj}(t)$$

Le second membre de (8) admettant une limite pour $\Delta t \rightarrow 0$, il en est de même du premier membre : $P_{ij}(t)$ est donc dérivable à droite. On montre de la même manière qu'elle est dérivable à gauche. Elle est donc dérivable, et sa dérivée vérifie la première équation de Kolmogorov.

$$(K_1) \quad \frac{d P_{ij}(t)}{dt} = -\lambda_i P_{ij}(t) + \sum_{k \neq i} \mu_{ik} P_{kj}(t)$$

On notera que la démonstration n'a pas utilisé l'hypothèse H_3 d'uniformité. L'équation (K_1) possède un plus haut degré de généralité que (K_2) .

Deuxième équation de Kolmogorov

Posons $t_2 = t + \Delta t$ dans l'équation de Markov (3). Il vient :

$$(10) \quad P_{ij}(t + \Delta t) = P_{ij}(t) P_{jj}(\Delta t) + \sum_{k \neq j} P_{ik}(t) P_{kj}(\Delta t)$$

D'après l'hypothèse H_3 d'uniformité, on peut, pour tout $\varepsilon > 0$, trouver $\alpha > 0$ tel que l'inégalité $\Delta t \leq \alpha$ entraîne pour tous les indices k :

$$\left\{ \begin{array}{l} 1 - (\lambda_j + \varepsilon) \Delta t \leq P_{jj}(\Delta t) \leq 1 - (\lambda_j - \varepsilon) \Delta t \\ (\mu_{kj} - \varepsilon) \Delta t \leq P_{kj}(\Delta t) \leq (\mu_{kj} + \varepsilon) \Delta t \end{array} \right.$$

Portons ces inégalités dans (10), en tenant compte de $\sum_k P_{ik} = 1$.
On obtient :

$$\begin{aligned} P_{ij}(t) - \varepsilon \Delta t - \lambda_j \Delta t P_{ij}(t) + \Delta t \sum_{k \neq j} P_{ik} \mu_{kj} &\leq P_{ij}(t + \Delta t) \\ &\leq P_{ij}(t) + \varepsilon \Delta t - \lambda_j \Delta t P_{ij}(t) + \Delta t \sum_{k \neq j} P_{ik} \mu_{kj} \end{aligned}$$

Ce qui peut aussi bien s'écrire :

$$\left| \frac{P_{ij}(t + \Delta t) - P_{ij}(t)}{\Delta t} + \lambda_j P_{ij}(t) - \sum_{k \neq j} P_{ik}(t) \mu_{kj} \right| \leq \varepsilon$$

Par suite $P_{ij}(t)$ est dérivable, et sa dérivée vérifie la deuxième équation de Kolmogorov :

$$\frac{d P_{ij}(t)}{dt} = - P_{ij}(t) \lambda_j + \sum_{k \neq j} P_{ik}(t) \mu_{kj}$$

Résolution des équations de Kolmogorov.

Lorsque $P_{ij}(t)$ vérifie les hypothèses H_1, H_2 , et la condition supplémentaire H_3 d'uniformité, elle constitue une solution commune des deux équations de Kolmogorov. Dans les applications, c'est le problème inverse qui présente le plus grand intérêt : étant donnés les nombres positifs λ_i et μ_{ij} vérifiant

$\lambda_i = \sum_{j \neq i} \mu_{ij}$, les équations de Kolmogorov K_1 et K_2 ont-elles une solution commune $P_{ij}(t)$ unique vérifiant :

$$(11) \quad \left. \begin{aligned} 0 &\leq P_{ij}(t) \leq 1 \\ \sum_j P_{ij}(t) &= 1 \\ \sum_k P_{ik}(t_1) P_{kj}(t_2) &= P_{ij}(t_1 + t_2) \end{aligned} \right\}$$

En fait, on démontre que K_1 et K_2 possèdent toujours une solution $\overline{P_{ij}(t)}$ commune vérifiant la relation de Markov et $0 \leq \overline{P_{ij}(t)} \leq 1$. Mais, λ_i et μ_{ij} étant quelconques, on a seulement :

$$\sum_j \overline{P_{ij}(t)} \leq 1$$

Pour que l'égalité ait lieu, les λ_i et μ_{ij} doivent vérifier une condition supplémentaire (toujours remplie en pratique), et c'est seulement dans ce cas que les équations de Kolmogorov permettent de résoudre complètement le problème.

Donnons simplement quelques indications sur ce résultat. Soit $P_{ij}^{(n)}(t)$ la probabilité de transition de e_i (en $t=0$) à e_j (en t) avec exactement n changements d'état dans l'intervalle $(0, t)$. On doit avoir :

$$P_{ij}^{(0)}(t) = \delta_{ij} e^{-\lambda_i t} \quad (\delta_{ij} = 1 \text{ si } i = j \text{ et } 0 \text{ si } i \neq j)$$

et, par récurrence :

$$(12) \quad \left\{ \begin{aligned} P_{ij}^{(n+1)}(t) &= \sum_{k \neq i} \int_0^t \mu_{ik} e^{-\lambda_i(t-s)} P_{kj}^{(n)}(s) ds \\ &= \sum_{k \neq j} \int_0^t P_{ik}^{(n)}(s) \mu_{kj} e^{-\lambda_j(t-s)} ds \end{aligned} \right.$$

Ces relations permettent le calcul effectif des $\overline{P_{ij}^{(n)}}(t)$. Posons :

$$\overline{P_{ij}^{(n)}}(t) = \sum_{n=0}^{\infty} P_{ij}^{(n)}(t)$$

On vérifie que les $\overline{P_{ij}^{(n)}}(t)$ constituent une solution commune des deux équations de Kolmogorov.

Par ailleurs, comme $P_{ij}(t)$ représente la probabilité de transition de e_i à e_j par un nombre fini de changements d'état, on doit s'attendre à avoir :

$$\overline{P_{ij}^{(n)}}(t) \leq P_{ij}(t)$$

Effectivement, on peut démontrer que tout $P_{ij}(t)$ vérifiant la première équation K_1 [qui ne suppose aucune hypothèse d'uniformité, et doit par conséquent être vraie pour tout $P_{ij}(t)$, s'il en existe, vérifiant H_1, H_2] et les conditions (11) est supérieur ou égal à $\overline{P_{ij}^{(n)}}(t)$.

Donc, si on a :

$$(13) \quad \sum_j \overline{P_{ij}^{(n)}}(t) = 1$$

le seul $P_{ij}(t)$ possible est $\overline{P_{ij}^{(n)}}(t)$ lui-même, solution de K_1 et K_2 , vérifiant les hypothèses H_1, H_2 et les conditions (11). Le problème admet, dans ce cas, une solution unique : on sait que la condition (13) exprime, en fait, qu'il est presque impossible que le système subisse, en un temps fini, une infinité de changements d'états.

Par contre, lorsque (13) n'est pas vérifiée, la possibilité d'une infinité de transitions dans un intervalle fini donne lieu à des anomalies de fuite. On peut montrer qu'il existe alors une infinité de $P_{ij}(t)$ vérifiant K_1 et les conditions (14), mais aucune d'elles ne vérifie K_2 .

IV. - PROCESSUS DE RENOUVELLEMENT

La classe la plus simple de chaînes de Markov homogène à temps continu, et la plus utilisée dans les applications, est celle des processus de renouvellement pour lesquels on a :

$$\mu_{ik} = 0 \quad \text{pour } k \neq i+1 \quad \text{et } k \neq i-1$$

Interprétons l'indice i comme le nombre d'individus (théorie des files d'attente) ou de particules présents dans un dispositif à l'instant t . Pendant l'intervalle $(t, t + \Delta t)$ ce nombre peut augmenter d'une unité (avec une probabilité équivalente à $\mu_i \Delta t$), ou diminuer d'une unité (avec une probabilité équivalente à $\nu_i \Delta t$) soit rester stationnaire (avec une probabilité $1 - \lambda_i \Delta t = 1 - (\mu_i + \nu_i) \Delta t$, en négligeant les infiniments petits d'ordre supérieur).

Les équations de Kolmogorov s'écrivent dans ce cas :

$$(K_1) \quad \frac{d P_{ij}(t)}{d t} = - \lambda_i P_{ij} + \mu_i P_{i+1,j} + \nu_i P_{i-1,j}$$

$$(K_2) \quad \frac{d P_{ij}(t)}{d t} = - \lambda_j P_{ij} + \mu_{i,j-1} P_{i,j-1} + \nu_{i,j+1} P_{i,j+1}$$

$$(\text{avec } \lambda_i = \mu_i + \nu_i)$$

Exemple : File d'attente

Le nombre de personnes se présentant devant un guichet est supposé obéir à un processus poissonnien : on admet qu'il y a une probabilité $1 - \lambda \Delta t + o(\Delta t)$ pour qu'aucune arrivée ne se produise pendant la durée Δt , et $\lambda \Delta t + o(\Delta t)$ pour qu'une arrivée exactement se produise.

De même les temps de service (les sorties) sont supposées poissonniennes. Lorsque la file n'est pas vide, c'est-à-dire lorsqu'une personne au moins est présente devant le guichet, il y a une probabilité $\mu \Delta t$ pour que le service de

cette personne ait pris fin pendant la durée Δt (et $1 - \mu \Delta t$ pour qu'il n'ait pas pris fin, en négligeant les infiniment petits d'ordre supérieur).

On a ici :

$$(K_1) \left\{ \begin{array}{l} \frac{d P_{ij}}{dt} = -(\lambda + \mu) P_{ij} + \lambda P_{i+1,j} + \mu P_{i-1,j} \quad (i \neq 0) \\ \frac{d P_{0j}}{dt} = -\lambda P_{0j} + \lambda P_{1j} \end{array} \right.$$

$$(K_2) \left\{ \begin{array}{l} \frac{d P_{ij}}{dt} = -(\lambda + \mu) P_{ij} + P_{i,j-1} \lambda + P_{i,j+1} \mu \quad (j \neq 0) \\ \frac{d P_{i0}}{dt} = -\lambda P_{i0} + \mu P_{i1} \end{array} \right.$$

On peut montrer que ces équations ont une solution commune vérifiant les conditions (14). En fait, en général, on s'intéresse surtout à la limite, si elle existe, de $P_{ij}(t)$ pour $t \rightarrow \infty$, c'est-à-dire au régime stationnaire. Tous les états étant ici essentiels, on doit s'attendre à avoir :

$$\lim_{t \rightarrow \infty} P_{ij}(t) = u_j$$

les u_j étant tous nuls, ou bien étant tous différents de 0 et constituant un système de probabilités stationnaires.

En fait, si $\lambda \gg \mu$, on est dans le cas $u_j = 0$: il n'y a pas de régime stationnaire, et, quel que soit j , $P_{ij}(t) \rightarrow 0$, ce qui signifie que le nombre de personnes en attente augmente indéfiniment.

Si $\lambda < \mu$, au contraire, le régime stationnaire existe.

On retrouve immédiatement ce résultat en remarquant que les limites u_j doivent vérifier les équations obtenues en annulant le deuxième membre de (K_2) :

$$\left\{ \begin{array}{l} u_1 = \frac{\lambda}{\mu} u_0 \\ \mu u_{j+1} - (\lambda + \mu) u_j + \lambda u_{j-1} = 0 \end{array} \right.$$

La solution générale de cette équations aux différences est :

$$u_j = A \left(\frac{\lambda}{\mu} \right)^j + B$$

Si $\lambda \geq \mu$, $A = B = 0$ (car $\sum_j u_j$ est nul ou égal à 1). Si $\lambda < \mu$, la solution est de la forme :

$$u_j = A \left(\frac{\lambda}{\mu} \right)^j$$

avec

$$u_1 = A \frac{\lambda}{\mu} = u_0 \frac{\lambda}{\mu}$$

Soit

$$u_j = u_0 \left(\frac{\lambda}{\mu} \right)^j$$

La condition :

$$\sum_{j=0}^{\infty} u_j = u_0 \left[1 + \sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu} \right)^j \right] = u_0 \frac{\mu}{\mu - \lambda} = 1$$

donne :

$$\left\{ \begin{array}{l} u_0 = \frac{\mu - \lambda}{\mu} \\ u_j = \frac{\mu - \lambda}{\mu} \left(\frac{\lambda}{\mu} \right)^j \end{array} \right.$$

C H A P I T R E V I I I

PROCESSUS MARKOVIENS CONTINUS

I. - DEFINITION

Soit $X(t)$ un processus stochastique (c'est-à-dire, si l'on veut, une variable aléatoire à une infinité non dénombrable de composantes X_t , $t \in \mathbb{R}$, chaque composante X_t représentant la valeur prise par $X(t)$ au temps t). Au chapitre VII, nous nous sommes limités au cas où $X(t)$ ne pouvait prendre que des valeurs entières. Nous admettons, maintenant, que $X(t)$ peut prendre n'importe quelle valeur (entière ou non). Le point de vue axiomatique, qui nécessite la définition d'une σ -algèbre sur l'espace des fonctions réelles, seul rigoureux, est cependant trop complexe pour être envisagé ici. Nous nous contenterons d'étudier les propriétés de $X(t)$ liées à sa loi temporelle. Si k est un entier, t_1, t_2, \dots, t_k des instants quelconques, l'ensemble $X(t_1), X(t_2) \dots X(t_k)$ des valeurs prises par $X(t)$ en chacun de ces instants constitue une variable aléatoire vectorielle à k composantes, qui possède une fonction de répartition :

$$F(x_1, x_2, \dots, x_k; t_1, t_2, \dots, t_k) = P(X(t_1) < x_1, X(t_2) < x_2, \dots, X(t_k) < x_k)$$

Par définition, on appelle loi temporelle du processus $X(t)$ l'ensemble de toutes ces lois F pour tous les entiers k et tous les systèmes d'instants t_1, t_2, \dots, t_k .

Au chapitre précédent, par ailleurs, nous avons caractérisé les chaînes de Markov par la propriété suivante : A chaque instant t , l'évolution future du système ne dépend que de son état présent $X(t)$ et non pas de son histoire antérieure à t . La même définition peut être retenue pour les processus $X(t)$ à valeurs quelconques. Plus précisément, étant donnés des instants

$$t_1 < t_2 < \dots < t_k < t_0 < t$$

nous dirons que le processus $X(t)$ est markovien si les probabilités conditionnelles (dont nous admettrons l'existence)

$$P(X(t) < x \mid X(t_0) = x_0, X(t_1) = x_1, \dots, X(t_k) = x_k)$$

sont égales à $P(X(t) < x \mid X(t_0) = x_0)$, quels que soient les instants $t_1 \dots t_k$ (antérieurs à t_0) et les nombres x_1, \dots, x_k .

Posons (pour $\mathcal{T} \leq t$)

$$P(X(t) < x \mid X(\mathcal{T}) = y) = F(y, \mathcal{T}; x, t)$$

$F(y, \mathcal{T}; x, t)$ est la fonction de répartition (conditionnelle) de la variable aléatoire $X(t)$, lorsque l'on sait que $X(\mathcal{T}) = y$. Elle représente la probabilité de la transition de l'état $X(\mathcal{T}) = y$ au temps \mathcal{T} à un état $X(t)$ inférieur à x au temps t , et constitue ainsi la généralisation naturelle de la matrice de transition $P_{ij}(\mathcal{T}, t)$ des chaînes de Markov à temps continu. Nous l'appellerons, pour abrégé, fonction de transition.

Lorsque la fonction de transition $F(y, \mathcal{T}; x, t)$ ne dépend pas séparément des instants t et \mathcal{T} , mais seulement de leur différence $t - \mathcal{T}$, on dit que le processus markovien $X(t)$ est homogène, ou stationnaire. Toutefois, et contrairement à ce que nous avons fait au chapitre VII, nous ne nous limiterons pas systématiquement à l'étude du cas homogène, et nous considérerons le cas général des fonctions de transition non stationnaires.

S'il existe une fonction $f(y, \mathcal{T}; x, t)$ telle que :

$$F(y, \mathcal{T}; x, t) = \int_{-\infty}^x f(y, \mathcal{T}; z, t) dz$$

c'est-à-dire si F est (en x) une loi absolument continue, f sera appelée densité associée à la fonction de transition F .

Equation de Markov - Exactement comme dans le cas des chaînes de Markov, on remarque que la transition d'un état $X(\mathcal{T}) = y$ à un état $X(t) < x$ implique nécessairement le passage par un état quelconque $X(s) = z$ au temps s intermédiaire ($\mathcal{T} < s < t$), et on en déduit l'équation de Markov :

$$(1) \quad F(y, \mathcal{T}; x, t) = \int F(z, s; x, t) d_x F(y, \mathcal{T}; z, s)$$

Dans le cas où il existe une densité f , l'équation de Markov s'écrit sous la forme :

$$(2) \quad f(y, \tau; x, t) = \int f(z, s; x, t) f(y, \tau; z, s) dz$$

Hypothèses de Continuité.

Lorsque $t = \tau$, on doit avoir (avec une probabilité unité) $X(t) = X(\tau)$. Par suite :

$$F(y, t; x, t) = \theta(x - y) = \begin{cases} 0 & \text{si } x \leq y \\ 1 & \text{si } x > y \end{cases}$$

Une première hypothèse de continuité, toute naturelle, consiste à supposer que la fonction de transition converge (en loi) vers $\theta(x - y)$ lorsque $t \rightarrow \tau$ ou lorsque $\tau \rightarrow t$, soit (pour $x \neq y$) :

$$(3) \quad \lim_{\Delta t \rightarrow 0} F(y, t - \Delta t; x, t) = \lim_{\Delta t \rightarrow 0} F(y, t; x, t + \Delta t) = \theta(x - y)$$

En fait, pour construire la théorie des processus markoviens continus, et notamment, pour établir les équations fondamentales de Kolmogorov, nous aurons besoin d'hypothèses de continuité beaucoup plus strictes que (3). Ces hypothèses sont les suivantes :

A - Quel que soit $\varepsilon > 0$, on a la limite :

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{|y-x| \geq \varepsilon} d_x F(y, t - \Delta t; t, x) = 0$$

B - Les dérivées partielles en y (par rapport à l'état initial)

$$\frac{\partial}{\partial y} F(y, \tau; x, t) \quad \text{et} \quad \frac{\partial^2}{\partial y^2} F(y, \tau; x, t)$$

existent et sont continues quels que soient y, x, τ et $t > \tau$.

C - Quel que soit $\varepsilon > 0$, il existe deux fonctions $a(x, t)$ et $b(x, t)$, telles que l'on ait les limites suivantes :

$$\left\{ \begin{array}{l} \Delta t \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{|y-x| < \varepsilon} (x-y) d_x F(y, t - \Delta t, x, t) = a(y, t) \\ \Delta t \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{|x-y| < \varepsilon} (x-y)^2 d_x F(y, t - \Delta t, x, t) = b(y, t) \end{array} \right.$$

et ces limites sont uniformes relativement à y .

Remarques : 1/ Ces trois hypothèses sont très sévères. Il existe des processus markoviens très usuels qui ne les vérifient pas. En fait, on peut, à bon droit, se poser la question inverse : peut-il exister une fonction de transition vérifiant à la fois la condition (1) et les trois hypothèses ci-dessus ? En fait, la réponse est positive : on peut montrer que le processus de Wiener-Levy, défini au chapitre V, et qui possède la densité de transition :

$$f(y, \tau, x, t) = \frac{1}{\sqrt{2\pi C(t-\tau)}} \exp \left[-\frac{1}{2} \frac{(x-y)^2}{C(t-\tau)} \right]$$

répond à la question (avec $a = 0$ et $b = C = \text{constante}$).

Il faut remarquer que, parmi tous les processus à accroissements indépendants et stationnaires, les processus du type de Wiener-Levy (à loi gaussienne) sont les seuls à répondre à la question.

Lorsque le processus $X(t)$ représente une coordonnée d'une particule animée d'un mouvement brownien, les limites $a(y, t)$ et $b(y, t)$ possèdent une interprétation physique simple : $a(y, t)$ est la vitesse probable de la particule (placée en y au temps t) et $b(y, t)$ représente son énergie cinétique (en valeur probable également). Pour cette raison, $a(y, t)$ est parfois appelée dérivée du processus.

2/ Dans la condition C, on pourrait penser que les limites $a(x, t)$ et $b(x, t)$ dépendent du choix de $\varepsilon > 0$. En fait, il n'en est rien, comme on peut le voir immédiatement, compte tenu de la condition A.

II.- LES EQUATIONS DE KOLMOGOROV.

Dans l'équation de Markov (1) nous pouvons (comme au chapitre précédent) faire varier l'état initial (en remplaçant τ par $\tau - \Delta\tau$ et en passant à la limite $\Delta\tau \rightarrow 0$), soit faire varier l'état final (en remplaçant au contraire t par $t + \Delta t$ et passant à la limite $\Delta t \rightarrow 0$). On obtient ainsi, respectivement, la première et la deuxième équations de Kolmogorov.

Première équation de Kolmogorov.

Soit $F(y, \tau; x, t)$ la fonction de transition d'un processus markovien vérifiant les conditions A, B et C. L'équation de Markov (1) permet d'écrire :

$$F(y, \tau - \Delta\tau; x, t) = \int F(z, \tau; x, t) d_z F(y, \tau - \Delta\tau; z, \tau)$$

Par ailleurs, comme $\int d_z F(y, \tau - \Delta\tau; z, \tau) = 1$, on a aussi :

$$F(y, \tau; x, t) = \int F(y, \tau; x, t) d_z F(y, \tau - \Delta\tau; z, \tau)$$

Formons la différence, et divisons par $\Delta\tau$. Il vient :

$$\frac{F(y, \tau - \Delta\tau; x, t) - F(y, \tau; x, t)}{\Delta\tau} = \frac{1}{\Delta\tau} \int \left[F(z, \tau; x, t) - F(y, \tau; x, t) \right] d_z F(y, \tau - \Delta\tau; z, \tau)$$

L'hypothèse B permet d'écrire le développement limité :

$$F(z, \tau; x, t) - F(y, \tau; x, t) = (z - y) \frac{\partial F(y, \tau; x, t)}{\partial y} + \frac{1}{2} (z - y)^2 \frac{\partial^2 F(y, \tau; x, t)}{\partial y^2} + o \left[(z - y)^2 \right]$$

où $o \left[(z - y)^2 \right]$ est d'ordre supérieur à 2. On en déduit :

$$\begin{aligned}
& \frac{F(y, \tau - \Delta\tau; x, t) - F(y, \tau; x, t)}{\Delta\tau} = \\
& = \frac{1}{\Delta\tau} \int_{|z-y| \geq \varepsilon} [F(z, \tau; x, t) - F(y, \tau; x, t)] d_z F(y, \tau - \Delta\tau; z, \tau) \\
& \quad + \frac{1}{\Delta\tau} \int_{|z-y| < \varepsilon} [F(z, \tau; x, t) - F(y, \tau; x, t)] d_z F(y, \tau - \Delta\tau; z, \tau) = \\
& = \frac{1}{\Delta\tau} \int_{|z-y| \geq \varepsilon} [F(z, \tau; x, t) - F(y, \tau; x, t)] d_z F(y, \tau - \Delta\tau; z, \tau) \\
& \quad + \frac{\partial F(y, \tau; x, t)}{\partial y} \frac{1}{\Delta\tau} \int_{|z-y| < \varepsilon} (z - y) d_z F(y, \tau - \Delta\tau; z, \tau) \\
& \quad + \frac{1}{2} \frac{\partial^2 F(y, \tau; x, t)}{\partial y^2} \frac{1}{\Delta\tau} \int_{|z-y| < \varepsilon} [(z-y)^2 + o((z-y)^2)] d_z F(y, \tau - \Delta\tau; z, \tau)
\end{aligned}$$

Dans cette dernière expression, faisons tendre $\Delta\tau$ vers 0. L'hypothèse A nous garantit que le premier terme tend vers 0. D'après C, le deuxième terme tend vers $a(y, \tau) \frac{\partial}{\partial y} F(y, \tau; x, t)$. Le troisième terme comprend deux parts. L'une, obtenue en intégrant $(z-y)^2$ possède, d'après C la limite $\frac{1}{2} b(y, \tau) \frac{\partial^2}{\partial y^2} F(y, \tau; x, t)$. La deuxième part est :

$$R = \frac{1}{2} \frac{\partial^2 F(y, \tau; x, t)}{\partial y^2} \frac{1}{\Delta\tau} \int_{|z-y| < \varepsilon} [(z-y)^2] d_z F(y, \tau - \Delta\tau; z, \tau)$$

Elle peut être rendue aussi petite que l'on veut en prenant ε assez petit. Plus précisément, donnons-nous $\alpha > 0$ arbitrairement petit. On peut trouver ε tel que :

$$|z - y| < \varepsilon \implies |o[(z-y)^2]| \leq \alpha (z-y)^2$$

et par suite :

$$|R| \leq \alpha \frac{1}{2} \left| \frac{\partial^2 F(y, \tau; x, t)}{\partial y^2} \right| \frac{1}{\Delta \tau} \int_{|z-y| < \varepsilon} (z-y)^2 dz F(y, \tau - \Delta \tau; z, \tau)$$

Passons à la limite $\Delta \tau \rightarrow 0$: de C, nous déduisons :

$$\overline{\lim}_{\Delta \tau \rightarrow 0} |R| \leq \alpha \frac{1}{2} \left| \frac{\partial^2 F}{\partial y^2} \right| b(y, t)$$

Par ailleurs, le premier membre de (4) ne dépend pas de ε , et les trois premières limites ne dépendent pas non plus de ε . Désignons par \bar{M} et \underline{M} les limites supérieure et inférieure de ce premier membre. On a :

$$a \frac{\partial F}{\partial y} + \frac{1}{2} b \frac{\partial^2 F}{\partial y^2} - \frac{1}{2} \alpha \left| b \frac{\partial^2 F}{\partial y^2} \right| \leq \underline{M} \leq \bar{M} \leq a \frac{\partial F}{\partial y} + \frac{1}{2} b \frac{\partial^2 F}{\partial y^2} + \frac{1}{2} \alpha \left| b \frac{\partial^2 F}{\partial y^2} \right|$$

Comme ceci est vrai, quel que soit α , on a $\underline{M} = \bar{M}$: autrement dit $F(y, \tau; x, t)$ est dérivable en τ , et cette dérivée est donnée par l'équation suivante, dite première équation de Kolmogorov :

$$(K_1) \quad \frac{\partial}{\partial \tau} F(y, \tau; x, t) = - a(y, \tau) \frac{\partial}{\partial y} F(y, \tau; x, t) - \frac{1}{2} b(y, \tau) \frac{\partial^2}{\partial y^2} F(y, \tau; x, t)$$

Si la fonction de transition $F(y, \tau; x, t)$ admet une densité $f(y, \tau; x, t)$, on voit immédiatement en dérivant (en x) l'équation précédente que cette densité vérifie également la première équation de Kolmogorov :

$$(K'_1) \quad \frac{\partial}{\partial \tau} f(y, \tau; x, t) = - a(y, \tau) \frac{\partial}{\partial y} f(y, \tau; x, t) - \frac{1}{2} b(y, \tau) \frac{\partial^2}{\partial y^2} f(y, \tau; x, t)$$

Théorème de Feller. Moyennant des hypothèses convenables de continuité et de dérivabilité sur $a(x, t)$ et $b(x, t)$, il est possible de démontrer que l'équation (K_1) admet une solution, et une seule $F(y, \tau; x, t)$ qui soit (en x) une fonction de répartition et vérifie la condition (A). La donnée des fonctions $a(x, t)$ et $b(x, t)$ suffit ainsi à déterminer complètement la loi du processus markovien.

Avec des hypothèses un peu plus strictes encore sur a et b , on peut démontrer, de plus, que la fonction de transition F ainsi déterminée admet une densité $f(y, \mathcal{U}, x, t)$ et que cette densité f vérifie nécessairement (en x et t) la deuxième équation de Kolmogorov que nous allons maintenant établir.

Deuxième équation de Kolmogorov

Il se produit ici une circonstance très analogue à celle que l'on a pu observer au chap. VII. La deuxième équation de Kolmogorov - celle que l'on obtient en faisant varier l'état final - exige, pour être établie, des hypothèses plus strictes que la première équation.

En plus des hypothèses A, B et C, nous admettrons ici les hypothèses supplémentaires suivantes :

D/ $a(x, t)$, $b(x, t)$ et $\frac{\partial}{\partial x} b(x, t)$ sont bornés.

E/ Il existe une densité $f(y, \mathcal{U}, x, t) = \frac{\partial}{\partial x} F(y, \mathcal{U}, x, t)$ (pour $t > \mathcal{U}$), et les dérivées suivantes existent et sont continues :

$$\frac{\partial}{\partial t} f(y, \mathcal{U}, x, t) ; \quad \frac{\partial}{\partial x} \left[a(x, t) f(y, \mathcal{U}, x, t) \right] ; \quad \frac{\partial^2}{\partial x^2} \left[b(x, t) f(y, \mathcal{U}, x, t) \right]$$

F/ $f(y, \mathcal{U}, x, t) \rightarrow 0$ pour $y \rightarrow \pm \infty$

G/ Quel que soit ε , on a uniformément en y :

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{|x-y| > \varepsilon} f(y, t, x, t + \Delta t) = 0$$

Considérons l'opérateur L défini par :

$$L(u) = \frac{\partial u}{\partial t} + a(z, t) \frac{\partial u}{\partial z} + \frac{1}{2} b(z, t) \frac{\partial^2 u}{\partial z^2}$$

où $u = u(z, t)$ est une fonction dérivable. L'équation (K_1) s'écrit $L(u) = 0$.
L'opérateur adjoint de L est l'opérateur L^* défini par :

$$L^*(v) = -\frac{\partial v}{\partial t} - \frac{\partial}{\partial z} (av) + \frac{1}{2} \frac{\partial^2}{\partial z^2} (bv)$$

Avec les conditions D , et en supposant que u , v et $\frac{\partial v}{\partial z}$ tendent vers 0 pour $z \rightarrow \pm \infty$, on établit facilement l'identité suivante :

$$(5) \quad \int_{t_1}^{t_2} dt \int_{-\infty}^{+\infty} [v L(u) - u L^*(v)] dz = \int_{-\infty}^{+\infty} [v(t_2, z) u(t_2, z) - v(t_1, z) u(t_1, z)] dz$$

Appliquons cette identité aux fonctions u et v définies comme suit :

$$\begin{cases} u(z, t) = f(z, t; x, t_0) \\ v(z, t) = f(y, \tau; z, t) \end{cases}$$

qui vérifient (en z et t) les conditions requises, et prenons :

$$\tau < t_1 < t_2 < t_0$$

L'équation de Markov montre que le deuxième membre est identiquement nul, et l'équation (K_1) donne $L(u) = 0$. Dérivant (5) en t_2 , et remplaçant t_2 par t , on obtient (pour $\tau < t < t_0$)

$$(6) \quad \int_{-\infty}^{+\infty} f(z, t; x, t_0) L^* [f(y, \tau; z, t)] dz \equiv 0$$

Il est entendu que l'opérateur L^* agit sur les variables z et t .
De (6), nous allons déduire que $L^* f(y, \tau; z, t)$ est identiquement nul, et ce résultat constituera la deuxième équation (K_2) de Kolmogorov.

Soit $R(x)$ une fonction uniformément continue et intégrable de $-\infty$ à $+\infty$.

Posons :

$$(7) \quad R(z, t, t_0) = \int_{-\infty}^{+\infty} f(z, t; x, t_0) R(x) dx$$

De la condition G, on conclut facilement que l'on a, uniformément en z :

$$(8) \quad \lim_{t_0 \rightarrow t} R(z, t, t_0) = R(z)$$

Or, multipliant (6) par $R(x)$ et intégrant en x , on obtient d'après (7) :

$$\int_{-\infty}^{+\infty} R(z, t, t_0) L^* f(y, \tau; z, t) dz \equiv 0$$

Lorsque $t_0 \rightarrow t$, il est légitime de passer à la limite sous le signe d'intégration, car $L^* f$ est intégrable et la limite (8) est uniforme. Il vient ainsi :

$$(9) \quad \int_{-\infty}^{+\infty} R(z) L^* f(y, \tau; z, t) dz \equiv 0$$

Or ceci a lieu quelle que soit la fonction $R(z)$, intégrable et uniformément continue. Il en résulte que $L^* f = 0$. En effet, d'après les hypothèses, $L^* f$ est une fonction continue de z . Si elle est différente de 0 - par exemple positive - en un point z_0 , on peut trouver un petit intervalle $(z_0 - \alpha, z_0 + \alpha)$ où elle reste strictement positive. Il suffit de prendre pour $R(z)$ une fonction non négative identiquement nulle à l'extérieur de $(z_0 - \alpha, z_0 + \alpha)$ et de somme $\int R(z) dz = 1$, pour obtenir :

$$\int R(z) L^*(f) dz > 0$$

résultat qui contredit (9). Donc on a nécessairement $L^*(f) = 0$, et ceci démontre la deuxième équation de Kolmogorov, qui s'écrit explicitement :

$$(K_2) \quad \frac{\partial}{\partial t} f(y, \tau; x, t) = - \frac{\partial}{\partial x} [a(x, t) f(y, \tau; x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [b(x, t) f(y, \tau; x, t)]$$

Exemple :

Si $a(x, t) = a(t)$ et $b(x, t) = b(t)$ ne dépendent que du temps t (et non de x), la densité $f(y, \tau; x, t)$ vérifie les équations :

$$(10) \quad \left\{ \begin{array}{l} \frac{\partial f}{\partial \tau} = - a(\tau) \frac{\partial f}{\partial y} - \frac{1}{2} b(\tau) \frac{\partial^2 f}{\partial y^2} \\ \frac{\partial f}{\partial t} = - a(t) \frac{\partial f}{\partial x} + \frac{1}{2} b(t) \frac{\partial^2 f}{\partial x^2} \end{array} \right.$$

Avec le changement de variables :

$$x' = x - \int_0^t a(u) du$$

$$y' = y - \int_0^{\tau} a(u) du$$

$$t' = \int_0^t b(u) du$$

$$\tau' = \int_0^{\tau} b(u) du$$

Ces équations deviennent (en posant $g(y', \tau'; x', t') = f(y, \tau; x, t)$:

$$\left\{ \begin{array}{l} \frac{\partial g}{\partial \tau'} = - \frac{1}{2} \frac{\partial^2 g}{\partial y'^2} \\ \frac{\partial g}{\partial t'} = \frac{1}{2} \frac{\partial^2 g}{\partial x'^2} \end{array} \right.$$

dont la solution (unique, d'après le théorème de Feller) est le processus de Wiener-Levy :

$$g(y', \tau'; x', t') = \frac{1}{\sqrt{2\pi(t' - \tau')}} \exp \left[-\frac{1}{2} \frac{(x' - y')^2}{(t' - \tau')} \right]$$

La solution de (10) s'en déduit : c'est le processus Gaussien défini par :

$$f(y, \tau, x, t) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{(x-y-\mu)^2}{2 \sigma^2} \right]$$

$$\sigma^2 = \int_{\tau}^t b(u) du$$

$$\mu = \int_{\tau}^t a(u) du$$

C H A P I T R E I X

PROCESSUS MARKOVIENS COMPLETEMENT DISCONTINUS

I.- GENERALITES

Les processus markoviens continus étudiés au chapitre précédent constituent un cas limite d'extrême continuité. A l'opposé, les processus dont il va être maintenant question - et où l'on reconnaîtra une généralisation des chaînes de Markov à temps continu - vont être caractérisés par leur extrême discontinuité. En termes intuitifs, un processus markovien $X(t)$ sera dit complètement discontinu si toute réalisation de $X(t)$ est (presque certainement) une fonction de saut, c'est-à-dire une fonction constante partout, sauf sur un ensemble dénombrable de points où elle subit des discontinuités finies. Nous ne chercherons pas à donner une définition rigoureuse (axiomatique), et nous nous contenterons des considérations à peu près intuitives suivantes :

Plaçons-nous (conditionnellement) dans l'hypothèse où $X(t) = x$. S'il existe une fonction $p(t, x)$ telle qu'il y ait une probabilité

$$1 - p(t, x) \Delta t + o(\Delta t)$$

(avec $\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} o(\Delta t) = 0$) pour que $X(t)$ reste constant et égal à x sur tout

l'intervalle $(t, t + \Delta t)$, une probabilité $p(t, x) \Delta t + o'(\Delta t)$ pour que $X(t)$ subisse exactement un saut dans l'intervalle $(t, t + \Delta t)$, et par conséquent une probabilité $o''(\Delta t)$, infiniment petite d'ordre supérieur à 1 pour que $X(t)$ subisse plus d'un saut, nous dirons que le processus Markovien $X(t)$ est complètement discontinu.

On désignera par $P(t, y, x)$ la fonction de répartition (conditionnelle) de $X(t)$ dans l'hypothèse où un saut vient juste de se produire au temps t , la valeur immédiatement antérieure de X étant $X(t-0) = y$.

Si $F(y, \tau, x, t)$ désigne la fonction de transition du processus $X(t)$, on a la relation suivante :

$$(1) \quad F(y, t, x, t + \Delta t) = \left[1 - p(t, y) \Delta t \right] \theta(x-y) + \Delta t p(t, y) P(t, y, x) + o(\Delta t)$$

($\theta(x-y) = 0$ pour $x \leq y$ et 1 pour $x > y$)

Nous adopterons les hypothèses suivantes :

A - La fonction $p(t, x)$ est bornée.

B - Les fonctions $p(t, y)$ et $P(t, y, x)$ sont continues en t et y .

Moyennant ces hypothèses, nous allons montrer que la fonction de transition F vérifie deux équations, ou équations de Feller, qui ont exactement la même signification et jouent le même rôle que les équations de Kolmogorov dans le cas des processus continus.

II. - LES EQUATIONS DE FELLER

Pour établir la première équation de Feller, on fait varier l'instant initial dans l'équation de Markov, soit :

$$F(y, \tau, x, t) = \int F(z, \tau + \Delta\tau, x, t) d_z F(y, \tau, z, \tau + \Delta\tau)$$

Exprimons $F(y, \tau, z, \tau + \Delta\tau)$ à l'aide de (1). Il vient :

$$\begin{aligned} F(y, \tau, x, t) &= \int F(z, \tau + \Delta\tau, x, t) d_z \left[1 - p(\tau, y) \Delta\tau + o(\Delta\tau) \right] \theta(z-y) \\ &\quad + \int F(z, \tau + \Delta\tau, x, t) d_z \left[p(\tau, y) \Delta\tau + o(\Delta\tau) \right] P(\tau, y, z) \end{aligned}$$

Mais, par définition de la mesure de Dirac θ , on a :

$$\int F(z, \tau + \Delta\tau, x, t) d_z \theta(z-y) = F(y, \tau + \Delta\tau, x, t)$$

D'où :

$$\begin{aligned} F(y, \tau, x, t) &= \left[1 - p(\tau, y) \Delta\tau \right] F(y, \tau + \Delta\tau, x, t) \\ &\quad + \Delta\tau p(\tau, y) \int F(z, \tau + \Delta\tau, x, t) d_z P(\tau, y, z) + o(\Delta\tau) \end{aligned}$$

et

$$\begin{aligned} \frac{F(y, \tau + \Delta\tau, x, t) - F(y, \tau, x, t)}{\Delta\tau} &= p(\tau, y) F(y, \tau + \Delta\tau, x, t) \\ &\quad - p(\tau, y) \int F(z, \tau + \Delta\tau, x, t) d_z P(\tau, y, z) + \frac{1}{\Delta\tau} o(\Delta\tau) \end{aligned}$$

Passant à la limite $\Delta\tau \rightarrow 0$, on obtient la première équation de Feller :

$$(F_1) \quad \frac{\partial}{\partial \tau} F(y, \tau; x, t) = p(\tau, y) \left[F(y, \tau; x, t) - \int F(z, \tau; x, t) d_z P(\tau, y; z) \right]$$

Remarque

L'équation (F_1) peut s'obtenir directement, sans passer par l'équation de Markov en remarquant que le passage de $X(\tau - \Delta\tau) = y$ à $X(t) < x$ est possible de deux manières seulement : ou bien X reste égal à y dans tout l'intervalle $(\tau - \Delta\tau; \tau)$, ou bien X subit une discontinuité dans cet intervalle. D'où :

$$F(y, \tau - \Delta\tau; x, t) = \left[1 - \Delta\tau p(\tau, y) \right] F(y, \tau; x, t) + \Delta\tau p(\tau, y) \int F(z, \tau; x, t) d_z P(\tau, y; z) + o(\Delta\tau)$$

D'où (F_1) par un passage à la limite immédiate.

De la même manière, faisons varier l'état final : le passage de $X(\tau) = y$ à $X(t + \Delta t) < x$ implique le passage par un état intermédiaire $X(t) = z$, avec deux possibilités : ou bien $X(t)$ reste constant (et égal à z) dans tout l'intervalle $(t, t + \Delta t)$, ou bien il y subit un saut. D'où :

$$F(y, \tau; x, t + \Delta t) = \int_{-\infty}^x \left[1 - \Delta t p(t, z) \right] d_z F(y, \tau; z, t) + \Delta t \int p(t, z) P(t, z; x) d_z F(y, \tau; z, t) + o(\Delta t)$$

Passant à la limite comme ci-dessus, on en déduit la deuxième équation de Feller :

$$(F_2) \quad \frac{\partial}{\partial t} F(y, \tau; x, t) = - \int_{-\infty}^x p(t, z) d_z F(y, \tau; z, t) + \int p(t, z) P(t, z; x) d_z F(y, \tau; z, t)$$

Inversement, et moyennant des hypothèses convenables de régularité pour $p(t, x)$ et $P(t, y; x)$, on peut démontrer que le système (F_1) (F_2) possède une solution unique $F(y, \tau; x, t)$ vérifiant les conditions requises pour être la fonction de transition d'un processus markovien.

III.- LOI DE PROBABILITE DU NOMBRE DE SAUTS

Dans les applications, on s'intéresse souvent au nombre de sauts subis par $X(t)$ dans un intervalle donné (\mathcal{T}, t) . Désignons par $p_n(\mathcal{T}, y; t)$ la probabilité conditionnelle pour qu'il y ait exactement n sauts dans l'intervalle (\mathcal{T}, t) , sachant que l'on avait $X(\mathcal{T}) = y$.

Calcul de $p_0(\mathcal{T}, y; t)$. L'évènement "il n'y a aucun saut de \mathcal{T} à $t + \Delta t$ " est identique à: "il n'y a aucun saut de \mathcal{T} à t et aucun saut de t à $t + \Delta t$ ".

D'où l'on tire :

$$p_0(\mathcal{T}, y; t + \Delta t) = p_0(\mathcal{T}, y; t) \left[1 - p(y, t) \Delta t \right] + o(\Delta t)$$

Faisant le passage à la limite habituel, on obtient :

$$\frac{\partial}{\partial t} p_0(\mathcal{T}, y; t) = - p(y, t) p_0(\mathcal{T}, y; t)$$

D'où l'on tire :

$$p_0(\mathcal{T}, y; t) = C \exp \left[- \int_{\mathcal{T}}^t p(y, u) du \right]$$

Compte tenu de $p_0(\mathcal{T}, y; \mathcal{T}) = 1$, on a $C = 1$ et :

$$(2) \quad p_0(\mathcal{T}, y; t) = e^{- \int_{\mathcal{T}}^t p(y, u) du}$$

Calcul de $p_n(\mathcal{T}, y; t)$. Les p_n peuvent ensuite se calculer par récurrence à partir du p_0 donné par la formule (2). En effet, l'évènement : "le système subit $n + 1$ changements de \mathcal{T} à t " est somme d'évènements définis comme suit :

- De \mathcal{T} à s ($\mathcal{T} < s < t$) aucun changement ne se produit :

$$\text{probabilité } p_0(\mathcal{T}, y; s)$$

- Dans l'intervalle $(s, s + \Delta s)$ le système subit un changement qui amène une valeur comprise entre z et $z + \Delta z$. Probabilité :

$$\left[\Delta s p(s, y) + o(\Delta s) \right] \left[P(s, y; z + \Delta z) - P(z, y; z) \right]$$

- Enfin de $(s + \Delta s)$ à t le système subit n changements, avec une probabilité $p_n(s + \Delta s, z, t)$

Un tel évènement possède la probabilité :

$$p_0(\mathcal{C}, y; s) p(s, y) \Delta_z P(s, y; z) p_n(s, z; t) \Delta s + o(\Delta s)$$

D'où, en sommant en s et en z :

$$(3) \quad p_{n+1}(\mathcal{C}, y; t) = \int_{\mathcal{C}}^t ds \int p_0(\mathcal{C}, y; s) p(s, y) p_n(s, z; t) d_z P(s, y; z)$$

Dans tous les cas, on aura :

$$\overline{P(\mathcal{C}, y; t)} = \sum_{n=0}^{\infty} p_n(\mathcal{C}, y; t) \leq 1$$

$\overline{P(\mathcal{C}, y; t)}$ représente la probabilité pour qu'il n'y ait qu'un nombre fini (quelconque) de sauts de \mathcal{C} à t . Dans la plupart des applications, on aura l'égalité $\overline{P} = 1$, et il sera presque certain que le processus ne subit qu'un nombre fini de sauts sur un intervalle de temps fini. Le cas où l'on a l'inégalité stricte $\overline{P} < 1$ permet, avec une probabilité non nulle, l'apparition de discontinuités de $X(t)$ de nature plus complexe qu'un simple saut. Cette circonstance a été déjà rencontrée à propos des chaînes de Markov.

C H A P I T R E X

LES FONCTIONS ALEATOIRES STATIONNAIRES

I.- GENERALITES

Dans ce chapitre, nous allons nous intéresser aux fonctions aléatoires $Z(x)$, où x désigne le point courant de l'espace R^n à n dimensions. Dans le cas particulier $n = 1$, et en remplaçant x par t (interprété comme un temps) $Z(t)$ est un processus stochastique. Mais, en général $Z(t)$ ne sera pas markovien : son évolution pour $t > t_0$ dépendra, en général, non seulement de $Z(t_0)$ mais également de toutes les $Z(\mathcal{U})$ pour $\mathcal{U} < t_0$. Dans le cas général où $n \neq 1$, la notion de caractère markovien pour une fonction aléatoire $Z(x)$ n'est même plus définissable (du moins de manière simple).

En termes intuitifs, une fonction aléatoire $Z(x)$ peut être regardée comme une variable aléatoire à une infinité de composantes Z_x ($x \in R^n$), infinité d'ailleurs non dénombrable. La composante Z_x n'est autre que la valeur prise par la fonction aléatoire au point x .

Du point de vue de l'axiomatique, ce passage à une infinité de composantes soulève des problèmes assez difficiles. On doit prendre comme espace E des événements élémentaires l'ensemble des fonctions numériques réelles définies sur R^n . Un événement élémentaire e est ainsi identifié à une fonction particulière que nous pouvons noter : $f_e(x)$. Supposons qu'il nous ait été possible de définir une σ -algèbre \mathcal{A} sur l'espace E des fonctions numériques, et une probabilité P sur l'espace probabilisable (E, \mathcal{A}) . La fonction aléatoire Z est alors définie comme l'application $Z(e) = f_e(x)$, qui est l'application identique de E .

D'un point de vue beaucoup plus intuitif, on peut imaginer que l'indice e ($e \in E$) de chaque fonction $f_e(x)$ est inscrit sur un carton : l'ensemble (infini non dénombrable !) de ces cartons est placé dans un chapeau, et on procède à un tirage au sort selon la loi de probabilité P défini sur (E, \mathcal{A}) . L'épreuve ainsi définie consiste donc à tirer au sort (selon la loi P) une fonction particulière $f_e(x)$ parmi l'ensemble de toutes les fonctions possibles. Le résultat d'un tel tirage, qui est une fonction particulière $f_e(x)$ (e fixé) s'appelle une réalisation de la fonction aléatoire.

Soient x_1, x_2, \dots, x_k k points définis de R^n et z_1, \dots, z_k k nombres réels. Parmi l'ensemble de toutes les fonctions $f_e(x)$ de E considérons le sous-ensemble A constitué par les fonctions $Z_e(x)$ vérifiant les inégalités :

$$Z_e(x_1) < z_1, \dots, Z_e(x_k) < z_k$$

La σ -algèbre \mathcal{A} est toujours choisie de manière à ce que A soit un événement. Soit $P(A)$ sa probabilité, qui dépend à la fois du choix des points d'appui x_1, \dots, x_k et des nombres z_1, \dots, z_k . $P(A)$ est, par définition, la probabilité pour que la fonction aléatoire $Z(x)$ prenne, aux points x_1, \dots, x_k , des valeurs numériques inférieures respectivement à z_1, \dots, z_k . Posons :

$$P(A) \equiv P[Z(x_1) < z_1, \dots, Z(x_k) < z_k] = F(z_1, \dots, z_k; x_1, \dots, x_k)$$

On voit que F est la fonction de répartition de la variable aléatoire à k composantes (Z_1, \dots, Z_k) définie par :

$$\begin{aligned} Z_1 &= Z(x_1) \\ &\dots \dots \dots \\ Z_k &= Z(x_k) \end{aligned}$$

L'ensemble de toutes les lois F , pour tous les systèmes possibles de points d'appui x_1, \dots, x_k , constitue la loi spatiale. Si le nombre n des dimensions est égal à 1, et si $x = t$ est interprété comme un temps, l'ensemble des lois F s'appelle loi temporelle du processus stochastique $Z(t)$.

On peut encore présenter cette notion de la manière suivante (équivalente) : Si x_1 est un point défini de R_n , supposé fixé, $Z_e(x_1)$ est une application de E dans R qui, à tout e de E , fait correspondre la valeur numérique $Z_e(x_1)$ que prend au point x_1 la fonction particulière $Z_e(x)$ associée à e . On s'arrange toujours pour qu'une telle application soit mesurable. Alors $Z_e(x_1)$, application mesurable de E dans R est une variable aléatoire ordinaire $Z(x_1)$.

De même, avec k points d'appui x_1, \dots, x_k , les $Z_e(x_1) \dots Z_e(x_k)$ définissent une application (mesurable) de E dans R^k , c'est-à-dire une variable aléatoire vectorielle à k composantes.

Ainsi, à tout groupe de k points d'appui x_1, \dots, x_k , nous associons une variable vectorielle $Z(x_1), \dots, Z(x_k)$ dont les k composantes ne sont autres que les k valeurs (aléatoires) prises par la fonction aléatoire $Z(x)$ en ces k points d'appui. A cette variable vectorielle est associée sa fonction de répartition $F(z_1, \dots, z_k; x_1, \dots, x_k)$, et l'ensemble de toutes les lois F pour tous les points d'appui possibles constitue la loi spatiale de la fonction aléatoire $Z(x)$.

La loi spatiale permet de résoudre tous les problèmes où n'interviennent qu'un nombre fini quelconque de points d'appui. En général, les axiomes des σ -algèbres permettent également, par passage à la limite, de traiter les cas où intervient une infinité dénombrable de points d'appui. Par contre, des difficultés très réelles surgissent dès que les points d'appui sont en infinité non dénombrable. Par exemple, pour un processus stochastique $Z(t)$, on peut s'intéresser à la probabilité:

$$P(Z(t) < a, \quad \forall t, \quad t_0 \leq t \leq t_1)$$

pour que $Z(t)$ reste inférieur à une constante a pendant tout l'intervalle de temps (t_0, t_1) . La loi temporelle ne permet pas, à elle seule, de calculer une telle probabilité. Il est nécessaire de faire intervenir des hypothèses supplémentaires concernant la probabilité P définie sur l'espace fonctionnel E et sa σ -algèbre (par exemple, on pourra supposer que P est telle que $Z(t)$ soit presque sûrement une fonction continue).

Nous n'insisterons pas sur ce genre de difficultés, et nous nous contenterons d'étudier les propriétés qui ne dépendent que de la loi spatiale. Pour abrégier les notations, nous nous limiterons au cas $n = 1$, c'est à dire à l'étude des processus stochastiques et de leurs lois temporelles. Les résultats obtenus (à l'exception de ceux qui concernent les processus markoviens étudiés dans les chapitres précédents) s'étendent sans difficulté aux fonctions aléatoires et à leurs lois spatiales, c'est-à-dire au cas $n > 1$.

II.- PROCESSUS STATIONNAIRES D'ORDRE 2

Soit $X(t)$ un processus stochastique. Soient t_1, \dots, t_k des instants, et

$$F(x_1, \dots, x_k; t_1, \dots, t_k) = P(X(t_1) < x_1, \dots, X(t_k) < x_k)$$

la fonction de répartition des variables aléatoires $X(t_1), \dots, X(t_k)$. Il est évident que l'ordre dans lequel on a rangé les instants d'appui ne joue aucun rôle ici. Autrement dit, pour toute permutation (i_1, \dots, i_k) de $(1, \dots, k)$, on doit avoir :

$$(1) \quad F(x_{i_1}, \dots, x_{i_k}; t_{i_1}, \dots, t_{i_k}) = F(x_1, \dots, x_k; t_1, \dots, t_k)$$

La condition (1) est une condition de symétrie que la loi temporelle doit nécessairement respecter.

Moments de la loi temporelle.

Il n'y a aucune difficulté à définir le moment $M_{\alpha_1 \dots \alpha_k}(t_1, \dots, t_k)$ d'ordre $(\alpha_1, \dots, \alpha_k)$, dépendant des instants t_1, \dots, t_k , comme l'espérance mathématique, si elle existe, du produit $[X(t_1)]^{\alpha_1} [X(t_2)]^{\alpha_2} \dots M_{\alpha_1 \dots \alpha_k}(t_1, t_2, \dots, t_k)$ ne dépend que de la fonction de répartition $F(x_1, \dots, x_k; t_1, \dots, t_k)$, c'est-à-dire uniquement de la loi temporelle.

En particulier, soit $m(t)$ le moment d'ordre 1 (s'il existe) :

$$M_1(t) = m(t) = E[X(t)] = \int x d_x F(x; t)$$

C'est une fonction de la seule variable t . Comme on peut toujours remplacer $X(t)$ par $X(t) - m(t)$, nous la supposons dans la suite identiquement nulle :

$$(2) \quad m(t) \equiv 0$$

De même, le moment d'ordre $(1, 1)$

$$M_{1,1}(t_1, t_2) = E[X(t_1) X(t_2)]$$

ne dépend que des deux instants t_1 et t_2 . Comme $m(t)$ est supposée nulle, ce moment est égal à la covariance de $X(t_1)$ et $X(t_2)$. Nous écrirons $K(t_1, t_2)$ au lieu de $M_{1,1}(t_1, t_2)$

$$(3) \quad K(t_1, t_2) = E[X(t_1) X(t_2)]$$

En particulier, prenant $t_1 = t_2 = t$, on obtient la variance $\sigma^2(t)$ de $X(t)$:

$$(4) \quad \sigma^2(t) = K(t, t)$$

Processus Stationnaire.

Un processus stochastique $X(t)$ est dit stationnaire lorsque sa loi temporelle est homogène dans le temps, c'est-à-dire lorsque l'on a :

$$(5) \quad F(x_1, \dots, x_k; t_1, \dots, t_k) = F(x_1, \dots, x_k; t_1 + h, \dots, t_k + h)$$

quels que soient les instants t_1, \dots, t_k et la translation h .

Les moments d'un processus stationnaire possèdent la même propriété d'homogénéité dans le temps. En particulier :

$$\left\{ \begin{array}{l} m(t) = m(t + h) \\ K(t_1, t_2) = K(t_1 + h, t_2 + h) \end{array} \right.$$

Il en résulte que $m(t)$ est une constante, que nous supposons d'ailleurs nulle conformément à (2), et que la covariance $K(t_1, t_2)$ ne dépend que de la différence $t_1 - t_2$, et non de t_1 et t_2 séparément (pour le voir, il suffit de prendre $h = -t_2$ dans la relation ci-dessus) :

$$(6) \quad \left\{ \begin{array}{l} m(t) = 0 \\ K(t_1, t_2) = K(t_1 - t_2) \end{array} \right.$$

Inversement, si un processus stochastique vérifie les relations (6) on dit qu'il est stationnaire au sens large ou stationnaire d'ordre deux.

On notera qu'un processus stationnaire au sens large peut très bien ne pas être stationnaire au sens strict, car (5) n'est nullement une conséquence de (6). Dans beaucoup d'applications, cependant, on n'utilise que des propriétés liées aux moments des deux premiers ordres, de sorte qu'il n'y a pas lieu de distinguer entre processus stationnaires sensu lato ou sensu stricto. En particulier, les propriétés de la covariance restent les mêmes.

La covariance $K(h)$.

Soit $X(t)$ un processus stationnaire. D'après les relations (6), la fonction

$$(7) \quad K(h) = E [X(t) X(t + h)]$$

ne dépend pas de t , et représente la covariance (stationnaire) des valeurs prises par $X(t)$ en deux instants quelconques distants de h .

Propriété 1 - La covariance $K(h)$ est une fonction de type positif.

En effet, soient t_1, \dots, t_k k instants et $\lambda_1, \dots, \lambda_k$ k nombres complexes.

Considérons la variable aléatoire

$$Y = \left| \sum_{i=1}^k \lambda_i X(t_i) \right|^2 = \sum_{i,j} \lambda_i \bar{\lambda}_j X(t_i) X(t_j)$$

Elle ne prend que des valeurs positives ou nulles. Donc on a :

$$E(Y) \geq 0$$

Or, d'après (7)

$$E(Y) = \sum_{i,j} \lambda_i \bar{\lambda}_j K(t_j - t_i)$$

Par suite $K(h)$ est de type positif (chapitre III, paragraphe 4).

Conséquence. Les relations (12) du chap. III, donnent :

$$(8) \quad \begin{cases} K(0) \geq 0 \\ K(-h) = K(h) \\ |K(h)| \leq K(0) \end{cases}$$

Les relations (8) sont du reste évidentes: $K(0)$ est la variance de $X(t)$ et $K(h)$ la covariance de $X(t)$ et $X(t+h)$, donc aussi de $X(t-h)$ et $X(t)$ puisque le processus est stationnaire.

Propriété 2 - Pour qu'un processus stationnaire (de variance finie) $X(t)$ soit continu en moyenne quadratique, il faut et il suffit que sa covariance $K(h)$ soit continue en $h = 0$.

Par définition (ch. IV, parag. 3) $X(t)$ est continu en moyenne quadratique si l'on a :

$$\lim_{h \rightarrow 0} E \left[(X(t+h) - X(t))^2 \right] = 0$$

Or, on a :

$$E \left[(X(t+h) - X(t))^2 \right] = 2 [K(0) - K(h)]$$

d'où la propriété 2.

Remarque : On sait (ch. III, parag. 4, propriété 2) que $K(h)$ est continue en tout point h si elle est continue en $h = 0$: ainsi la continuité en moyenne quadratique de $X(t)$ est équivalente à la continuité ordinaire de sa covariance $K(h)$.

Propriété 3 - Pour qu'une fonction $K(h)$ soit la covariance d'un processus continu en moyenne quadratique, et de variance finie, il faut et il suffit qu'il existe une fonction de répartition $G(x)$ telle que l'on ait

$$(9) \quad K(h) = K(0) \int \cos hx \, dG(x)$$

La condition est nécessaire : $K(h)$, covariance d'un processus continu en moyenne quadratique, est continue en $h = 0$ (propriété 2) et de type positif (propriété 1). Donc, d'après le théorème de Bochner (ch. III, paragr. IV), il existe une fonction de répartition $G(x)$ telle que :

$$K(h) = K(0) \int e^{ihx} \, dG(x)$$

Comme $K(h)$ est réelle, la répartition définie par $G(x)$ est symétrique [on a $G(x) = 1 - G(-x + 0)$] et (9) en résulte.

Réciproquement, si $K(h)$ est de la forme (9), elle est continue, et le théorème de Bochner montre qu'elle est de type positif. On démontre qu'il existe alors un processus stationnaire à loi temporelle gaussienne admettant $K(h)$ comme covariance (donc continu en moyenne quadratique d'après la propriété 2).

Ainsi la classe des fonctions de covariance des processus continus en moyenne quadratique et de variance finie s'identifie à la classe des fonctions de type (9) et aussi à la classe des fonctions de type positif réelles et continues.

III.- DERIVATION STOCHASTIQUE.

Bien que les notions d'intégration et de dérivation en moyenne quadratique soient surtout utiles dans le cas des processus stationnaires, il est préférable de les introduire dans leur généralité. Dans ce qui suit $X(t)$ désignera un processus non stationnaire de variance finie, de moyenne $m(t) \equiv 0$ et de covariance $K(t_1, t_2)$. Les énoncés se particulariseront sans difficulté au cas où $X(t)$ est stationnaire (sensu stricto ou sensu lato) en prenant $K(t_1, t_2) = K(t_1 - t_2)$.

Définition

On dit qu'un processus $X'(t)$ est la dérivée en moyenne quadratique du processus $X(t)$ si $\frac{X(t+h) - X(t)}{h}$ converge en moyenne quadratique vers $X'(t)$ lorsque h tend vers 0.

c'est-à-dire si l'on a :

$$(10) \quad \lim_{h \rightarrow 0} E \left\{ \left[\frac{X(t+h) - X(t)}{h} - X'(t) \right]^2 \right\} = 0$$

Pour trouver une condition nécessaire et suffisante de dérivabilité, nous utiliserons le critère de Cauchy (ch.IV, parag.3) sous la forme suivante :

Lemme - Soit $X_\lambda(t)$ un processus admettant une variance finie et dépendant d'un paramètre réel λ . Pour que $X_\lambda(t)$ converge en moyenne quadratique vers un processus $X(t)$ lorsque $\lambda \rightarrow 0$, il faut et il suffit que pour tout t la limite :

$$\lim_{\substack{\lambda \rightarrow 0 \\ \lambda' \rightarrow 0}} E [X_\lambda(t) X_{\lambda'}(t)]$$

existe et soit la même quelle que soit la manière dont λ et λ' tendent vers 0.

La covariance $K(t, t')$ de la limite $X(t)$ est alors donnée par :

$$(11) \quad K(t, t') = \lim_{\lambda \rightarrow 0} E [X_\lambda(t) X_\lambda(t')]$$

a/ D'après le critère de Cauchy, $X_\lambda(t)$ converge en moyenne quadratique si, et seulement si :

$$(12) \quad \lim_{\substack{\lambda \rightarrow 0 \\ \lambda' \rightarrow 0}} E \left[\left[X_\lambda(t) - X_{\lambda'}(t) \right]^2 \right] \\ = \lim_{\substack{\lambda \rightarrow 0 \\ \lambda' \rightarrow 0}} \left\{ E [X_\lambda(t)^2] + E [X_{\lambda'}(t)^2] - 2 E [X_\lambda(t) X_{\lambda'}(t)] \right\} = 0$$

On voit immédiatement que la condition de l'énoncé entraîne l'égalité ci-dessus, donc aussi la convergence en moyenne quadratique.

b/ Inversement, supposons que $X_\lambda(t)$ converge en moyenne quadratique vers $X(t)$. L'inégalité de Schwartz peut s'écrire

$$E \left[\left[X_\lambda(t) - X(t) \right]^2 \right] = E [X_\lambda(t)^2] + E [X(t)^2] - 2 E [X(t) X_\lambda(t)] \\ \geq \left| \sqrt{E [X_\lambda(t)^2]} - \sqrt{E [X(t)^2]} \right|^2$$

et donne

$$\lim_{\lambda \rightarrow 0} E \left[X_{\lambda}(t)^2 \right] = E \left[X(t)^2 \right]$$

Le critère (12) de Cauchy donne alors :

$$(13) \quad \lim_{\substack{\lambda \rightarrow 0 \\ \lambda' \rightarrow 0}} E \left[X_{\lambda}(t) X_{\lambda'}(t) \right] = E \left[X(t)^2 \right]$$

et la condition de l'énoncé est vérifiée.

c/ La relation (13) s'identifie avec (11) lorsque l'on prend $t = t'$.

Pour $t \neq t'$, la relation (11) se déduit facilement de l'inégalité de Schwartz. On a :

$$E \left[\left(X_{\lambda}(t) - X(t) + X_{\lambda}(t') - X(t') \right)^2 \right] \leq \left\{ \sqrt{E \left[X_{\lambda}(t) - X(t) \right]^2} + \sqrt{E \left[X_{\lambda}(t') - X(t') \right]^2} \right\}^2$$

d'où l'on déduit que $X_{\lambda}(t) + X_{\lambda}(t')$ converge en moyenne quadratique vers $X(t) + X(t')$, ce qui entraîne, d'après (13) :

$$\begin{aligned} \lim_{\lambda \rightarrow 0} \left\{ E \left[X_{\lambda}(t)^2 \right] + E \left[X_{\lambda}(t')^2 \right] + 2 E \left[X_{\lambda}(t) X_{\lambda}(t') \right] \right\} \\ = E \left[X(t)^2 \right] + E \left[X(t')^2 \right] + 2 E \left[X(t) X(t') \right] \end{aligned}$$

Comme

$$\lim_{\lambda \rightarrow 0} E \left[X_{\lambda}(t)^2 \right] = E \left[X(t)^2 \right]$$

$$\lim_{\lambda \rightarrow 0} E \left[X_{\lambda}(t')^2 \right] = E \left[X(t')^2 \right]$$

Il vient bien

$$\lim_{\lambda \rightarrow 0} E \left[X_{\lambda}(t) X_{\lambda}(t') \right] = E \left[X(t) X(t') \right]$$

c'est-à-dire (11) :

Théorème 1 - Pour qu'un processus de variance finie $X(t)$ soit dérivable en moyenne quadratique, il faut et il suffit qu'en tout t la limite :

$$\lim_{\substack{h \rightarrow 0 \\ \ell \rightarrow 0}} \frac{K(t+h, t+\ell) - K(t+h, t) - K(t, t+\ell) + K(t, t)}{h \ell} = \left[\frac{\partial^2 K(t_1, t_2)}{\partial t_1 \partial t_2} \right]_{t_1 = t_2 = t}$$

existe quelle que soit la manière dont h et ℓ tendent vers 0. Dans ces conditions, la dérivée seconde $\frac{\partial^2 K(t_1, t_2)}{\partial t_1 \partial t_2}$ existe aussi pour $t_1 \neq t_2$ et coïncide avec la covariance de la dérivée de $X(t)$.

C'est une simple application du Lemme : $X(t)$ est dérivable si, et seulement si

$$\lim_{\substack{h \rightarrow 0 \\ \ell \rightarrow 0}} E \left[\left(\frac{X(t+h) - X(t)}{h} \right) \left(\frac{X(t+\ell) - X(t)}{\ell} \right) \right]$$

$$= \lim_{\substack{h \rightarrow 0 \\ \ell \rightarrow 0}} \frac{K(t+h, t+\ell) - K(t+h, t) - K(t, t+\ell) + K(t, t)}{h \ell}$$

existe. Cette limite, qui est la dérivée seconde $\frac{\partial^2 K(t_1, t_2)}{\partial t_1 \partial t_2}$ en $t_1 = t_2 = t$ est alors égale, d'après (13), à $E [X'(t)^2]$. La covariance de $X'(t)$ est alors donnée par (11)

$$\begin{aligned} \lim_{h \rightarrow 0} E \left[\frac{X(t+h) - X(t)}{h} \cdot \frac{X(t'+h) - X(t')}{h} \right] \\ = \lim_{h \rightarrow 0} \frac{K(t+h, t'+h) - K(t, t'+h) - K(t+h, t') + K(t, t')}{h^2} \\ = \frac{\partial^2 K(t, t')}{\partial t \partial t'} \end{aligned}$$

On notera que l'existence de la dérivée seconde de $K(t_1, t_2)$ sur la bissectrice $t_1 = t_2 = t$ entraîne son existence en tout point (t_1, t_2) .

Corollaire - Pour qu'un processus stationnaire au sens large soit dérivable en moyenne quadratique il faut et il suffit que sa covariance $K(h)$ soit deux fois dérivable en $h = 0$. Le processus dérivé est alors stationnaire au sens large et admet comme covariance stationnaire la dérivée seconde de $K(h)$ changée de signe, soit $-K''(h)$.

Il suffit d'appliquer le théorème précédent avec $K(t_1, t_2) = K(t_1 - t_2)$.

Généralisation - Les résultats précédents s'étendent d'eux mêmes aux dérivées successives d'un processus $X(t)$. Ainsi, pour que la dérivée $X^{(k)}(t)$ d'un processus stationnaire (au sens large) existe, il faut et il suffit que $K(h)$ soit $2k$ fois dérivable en $h = 0$. Elle est alors, comme on sait (ch. III parag. 4, propriété 4) $2k$ fois dérivable en tout point. La dérivée $X^{(k)}(t)$ est également stationnaire (au sens large) et admet la covariance :

$$(-1)^k \frac{d^{2k} K(h)}{d h^{2k}}$$

IV. - INTEGRATION EN MOYENNE QUADRATIQUE.

Soit $X(t)$ un processus (non stationnaire) de variance finie et de covariance $K(t_1, t_2)$. Soit aussi $p(t)$ une fonction numérique (réelle) non aléatoire. Nous cherchons à donner un sens à l'intégrale

$$I = \int_a^b X(t) p(t) dt$$

Une telle intégrale, qui dépend de la fonction aléatoire $X(t)$, ne peut définir qu'une variable aléatoire : la variable aléatoire I (lorsqu'elle aura reçu une définition précise) sera appelée intégrale stochastique.

Pour définir I , on procédera exactement comme dans le cas d'une intégrale ordinaire. Divisons l'intervalle (a, b) par des points :

$$a = t_0 < t_1 < \dots < t_n = b$$

et formons la somme :

$$(14) \quad I_n = \sum_{i=1}^n X(t_i) p(t_i) (t_i - t_{i-1})$$

I_n est une variable aléatoire (ordinaire) parfaitement définie.

Faisons tendre n vers l'infini, de telle manière que les intervalles (t_{i-1}, t_i) convergent uniformément vers 0 :

$$\sup_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0$$

Si il existe une variable aléatoire I telle que I_n converge en moyenne quadratique vers I , on posera (conventionnellement) :

$$I = \int_a^b X(t) p(t) dt$$

et on dira que I est l'intégrale (en moyenne quadratique) de $p(t) X(t)$ sur le segment (a, b) .

Si $a = -\infty$, ou $b = +\infty$, on définira l'intégrale comme la limite en moyenne quadratique (si elle existe) des intégrales correspondantes pour $a \rightarrow -\infty$ ou $b \rightarrow +\infty$.

Théorème 2 - Soit $X(t)$ un processus de variance finie, de covariance $K(t_1, t_2)$.

Pour que l'intégrale en moyenne quadratique

$$(15) \quad I = \int_a^b X(t) p(t) dt$$

existe, il faut et il suffit que l'intégrale double ordinaire :

$$(16) \quad S^2 = \int_a^b \int_a^b K(t, t') p(t) p(t') dt dt'$$

existe : dans ce cas, on a de plus

$$(17) \quad S^2 = E [I^2]$$

Considérons les sommes de Riemann définies en (14). D'après le critère de Cauchy, l'intégrale I existe si, et seulement si :

$$\lim_{\substack{n \rightarrow \infty \\ m \rightarrow \infty}} E \left[(I_n - I_m)^2 \right] = 0$$

Le lemme du paragraphe précédent montre (moyennant des changements de notations évidents) qu'il en est ainsi si, et seulement si la limite

$$\lim_{\substack{n \rightarrow \infty \\ m \rightarrow \infty}} E(I_n I_m) = A$$

existe et que l'on a alors $A = E(I^2)$.

Soient t_i et s_j les points de subdivisions utilisés dans la définition de I_n et I_m . On a :

$$I_n I_m = \sum_{i=1}^n \sum_{j=1}^m X(t_i) X(s_j) p(t_i) p(s_j) \Delta t_i \Delta s_j$$

et

$$E(I_n I_m) = \sum_{i=1}^n \sum_{j=1}^m K(t_i, s_j) p(t_i) p(s_j) \Delta t_i \Delta s_j$$

Par suite la limite A existe si, et seulement si, l'intégrale (16) existe et on a alors $A = S^2$, donc $S^2 = E(I^2)$.

Par passage à la limite, on voit facilement que le théorème subsiste si a , ou b (ou les deux) deviennent infinis.

Corollaire 1 - Si $X(t)$ est un processus stationnaire au sens large, de covariance $K(h)$, l'intégrale (15) existe si, et seulement si, l'intégrale :

$$(18) \quad S^2 = \int_a^b \int_a^b K(t - t') p(t) p(t') dt dt'$$

existe. Dans ce cas, on a $S^2 = E(I^2)$.

Corollaire 2 - Si l'intégrale (à limite variable)

$$(19) \quad Y(t) = \int_a^t X(s) p(s) ds$$

existe, la dérivée $Y'(t)$ du processus $Y(t)$ ainsi défini existe aussi et coïncide (presque certainement) avec $X(t)$

$$Y'(t) = X(t)$$

En effet, puisque $Y(t)$ existe, l'intégrale

$$S^2(t) = E \left[Y(t)^2 \right] = \int_a^t \int_a^t K(s, s') p(s) p(s') ds ds'$$

existe aussi, et $Y(t)$ a une variance finie, donc aussi une covariance $Q(t, t')$ définie par :

$$Q(t, t') = E \left[Y(t) Y(t') \right]$$

Appliquons la relation (11) du lemme du paragraphe précédent. On a :

$$Q(t, t') = \lim_{n \rightarrow \infty} E \left(Y_n(t) Y_n(t') \right)$$

où $Y_n(t)$ et $Y_n(t')$ désignent des sommes de Riemann approchant $Y(t)$ et $Y(t')$.

Soit, par exemple, $t < t' = t_n$ et $t_k = t$ un point de subdivision coïncidant avec t :

$$E \left(Y_n(t) Y_n(t') \right) = \sum_{i=1}^k \sum_{j=1}^n K(t_i, t_j) p(t_i) p(t_j) \Delta t_i \Delta t_j$$

d'où l'on déduit :

$$(20) \quad Q(t, t') = \int_a^t ds \int_a^t K(s, s') p(s) p(s') ds'$$

Comme Q est dérivable en t et t' , la dérivée $Y'(t)$ existe en moyenne quadratique.

Enfin, par des passages à la limite très analogues, on montre que l'on a :

$$E \left[\left(Y'(t) - X(t) \right)^2 \right] = 0$$

d'où résulte la dernière affirmation.

Remarque 1 - Si l'espérance mathématique de $X(t)$ n'est pas identiquement nulle, on voit facilement que l'on a :

$$E \left[\int_a^b X(t) p(t) dt \right] = \int_a^b E \left[X(t) \right] p(t) dt$$

sous réserve que l'intégrale stochastique existe en moyenne quadratique. De même, d'ailleurs, pour la dérivation

$$E \left[\frac{d}{dt} X(t) \right] = \frac{d}{dt} E \left[X(t) \right]$$

Ainsi le symbole E (espérance mathématique) peut être permuté avec les symboles d'intégration et de dérivation (sous réserve que l'intégrale ou la dérivée stochastiques existent en moyenne quadratique).

Remarque 2 - Supposons que, dans l'intégrale (19) à limite variable, $X(t)$ soit stationnaire au sens large. Il n'en résulte pas du tout que $Y(t)$ soit également stationnaire; car la covariance $Q(t, t')$ définie en (20) n'est nullement de la forme $Q(t-t')$, lorsque $K(s, s')$ est de la forme $K(s-s')$. Cela tient à ce que la borne inférieure d'intégration a reste fixe lorsque t varie. Pour obtenir une intégrale stationnaire, on doit faire varier simultanément les deux bornes. Ceci nous conduit à définir une opération de convolution stochastique, très utilisée notamment en électronique.

Si $f_1(t)$ et $f_2(t)$ sont deux fonctions (ordinaires) on définit leur produit de convolution :

$$g = f_1 * f_2$$

par l'intégrale :

$$g(t) = \int_{-\infty}^{+\infty} f_1(t-s) f_2(s) ds = \int_{-\infty}^{+\infty} f_1(s) f_2(t-s) ds$$

Dans le cas particulier (très important en physique) où

$$f_1(t) = f_2(t) = 0 \quad \text{pour } t < 0$$

la convolution prend la forme :

$$\left\{ \begin{array}{l} g(t) = \int_0^t f_1(t-s) f_2(s) \quad (t \geq 0) \\ g(t) \equiv 0 \quad (t < 0) \end{array} \right.$$

Convolution Stochastique - Soit $X(t)$ un processus stationnaire au sens large et $p(t)$ une fonction (dite souvent fonction de pondération). Si l'intégrale stochastique :

$$(21) \quad Y(t) = \int_{-\infty}^{+\infty} X(t-s) p(s) ds = \int_{-\infty}^{+\infty} X(s) p(t-s) ds$$

existe en moyenne quadratique, on pose

$$Y = p * X$$

et on dit que $Y(t)$ est le convolué de $X(t)$ par $p(t)$.

Les intégrales à limite fixes (a,b) peuvent toujours s'obtenir comme cas particulier de (21). Il suffit de prendre $t = 0$, et d'annuler $p(t)$ en dehors d'un intervalle convenablement choisi.

Soit $K(h)$ la covariance stationnaire de $X(t)$. La convolution (21) est définie si l'intégrale.

$$E \left[Y(t)^2 \right] = \iint K(s - s') p(s) p(s') ds ds'$$

est définie. La covariance de $Y(t)$ existe aussi, dans ce cas, et se présente sous la forme :

$$Q(t, t') = \iint K(t - t' - s + s') p(s) p(s') ds ds'$$

Elle est de la forme $Q(t-t')$. Ainsi la convolution $Y(t)$ définie en (21) conserve le caractère stationnaire au sens large. La covariance stationnaire $Q(h)$ est donnée par :

$$(22) \quad Q(h) = \iint K(h - s + s') p(s) p(s') ds ds'$$

Soit $\overset{v}{p}(s) = p(-s)$ la transposée de $p(s)$ et

$$P = p * \overset{v}{p}$$

ou, explicitement

$$P(u) = \int p(v) p(u+v) dv$$

le produit de convolution de p par sa transposée. Posons $s' = s + u$ et calculons (22) en variables s et u : il vient :

$$Q(h) = \int K(h + u) du \int p(s) p(s + u) ds = \int K(h + u) P(u) du$$

Comme $K(h) = K(-h)$ et $Q(h) = Q(-h)$ (en fait, on a également $P(u) = P(-u)$), la covariance $Q(h)$ de la convoluée se met sous la forme d'un produit de convolution :

$$(23) \quad Q = K * P = K * p * \overset{v}{p}$$

Cas particulier - Dans beaucoup d'applications, $Y(t)$ ne doit dépendre que des valeurs prises par le processus antérieurement à l'instant présent t . La fonction de pondération qui figure dans (21) doit alors vérifier :

$$p(t) \equiv 0 \quad \text{pour } t < 0$$

et les formules précédentes s'écrivent :

$$\left\{ \begin{array}{l} Y(t) = \int_0^{+\infty} X(t-s) p(s) ds \\ P(u) = \int_0^{+\infty} p(v) p(u+v) dv \quad (u \geq 0) \\ P(-u) = P(u) \\ Q(h) = K * P = \int_{-\infty}^{+\infty} K(h+u) P(u) du \end{array} \right.$$

Mentionnons encore le théorème suivant, dit parfois théorème ergodique, dont l'importance est grande en pratique. Il sert, en effet, de fondement à l'inférence statistique en permettant l'estimation de la valeur probable $E[X(t)]$ d'un processus stochastique par la moyenne temporelle (calculée sur une seule réalisation de $X(t)$).

Théorème ergodique - Si $X(t)$ est un processus stationnaire (au sens strict) continu en moyenne quadratique et de variance finie, l'intégrale stochastique

$$\frac{1}{T} \int_0^T X(t) dt$$

converge presque sûrement vers $E[X(t)]$ lorsque T tend vers l'infini.

V.- ANALYSE HARMONIQUE D'UN PROCESSUS STATIONNAIRE (AU SENS LARGE).

Avant d'énoncer le théorème général donnant la décomposition harmonique d'un processus stationnaire, examinons quelques exemples simples qui nous prépareront à en comprendre la signification et la très grande importance, notamment, pour les applications physiques.

Exemple 1 : Processus monochromatique. Soit Y et Z deux variables aléatoires ordinaires d'espérances mathématiques nulles, possédant une même variance finie σ_1^2 et indépendantes (il suffit, en réalité, de supposer que Y et Z ont une corrélation nul-

le). Soit λ un nombre réel constant. Posons :

$$X(t) = Y \cos \lambda t + Z \sin \lambda t$$

Le processus $X(t)$ ainsi défini est caractérisé par la propriété suivante : toute réalisation de $X(t)$ est une fonction sinusoidale de période (non aléatoire) $\frac{2\pi}{\lambda}$. Seules la phase et l'amplitude de $X(t)$ apparaissent comme aléatoires, sa longueur d'onde est déterminée.

Calculons la covariance :

$$\begin{aligned} E[X(t+h) X(t)] &= E \left\{ Y^2 \left[\cos \lambda(t+h) \cos \lambda t \right] + Y Z \left[\cos(\lambda t+h) \sin \lambda t \right. \right. \\ &\quad \left. \left. + \sin(\lambda t+h) \cos \lambda t \right] + Z^2 \left[\sin \lambda(t+h) \sin \lambda t \right] \right\} \\ &= \sigma^2 \left[\cos \lambda(t+h) \cos \lambda t + \sin \lambda(t+h) \sin \lambda t \right] = \sigma^2 \cos \lambda h \end{aligned}$$

Ainsi $X(t)$ est stationnaire au sens large et possède la covariance :

$$K(h) = \sigma^2 \cos \lambda h$$

Appliquons la propriété 3 du paragraphe 2 : on trouve immédiatement :

$$K(h) = \sigma^2 \int e^{ih\gamma} dG(\gamma)$$

avec

$$G(\gamma) \begin{cases} 0 & \text{si } \gamma \leq -\lambda \\ \frac{1}{2} & \text{si } -\lambda < \gamma \leq \lambda \\ 1 & \text{si } \gamma > \lambda \end{cases}$$

$G(\gamma)$ est la fonction de répartition associée à la distribution discrète comportant une masse $\frac{1}{2}$ aux points $\gamma = \lambda$ et $\gamma = -\lambda$. En physique, $K(h)$, liée au carré de l'amplitude, possède toujours une signification énergétique. Le résultat obtenu signifie que l'énergie du phénomène $X(t)$ est localisable sur la fréquence $\gamma = \lambda$ (les fréquences $-\lambda$ et $+\lambda$ ne sont pas réellement distinctes). Ainsi un phénomène stationnaire monochromatique possède une énergie monochromatique.

La fonction $G(\gamma)$ est souvent appelée fonction de répartition spectrale de l'énergie. Elle donne, en effet, dans le cas d'un processus stationnaire la décompo-

sition de l'énergie selon les bandes de fréquence. Ici $X(t)$ est caractérisé par son spectre monochromatique.

Il faut bien voir que la possibilité d'une telle décomposition spectrale est liée à l'indépendance (absence de corrélation) entre les composantes aléatoires Y et Z . Supposons, en effet, que Y et Z aient une covariance non nulle

$$\beta = E(Y Z) \neq 0$$

On obtient alors :

$$E \left[X(t+h) X(t) \right] = \sigma^2 \cos \lambda h + \beta \sin \lambda(2t + h)$$

$X(t)$ n'est plus stationnaire au sens large, puisque la covariance dépend de t . La variance (énergie) s'écrit :

$$E \left(X(t)^2 \right) = \sigma^2 + \beta \sin 2 \lambda t$$

Elle est fluctuante. L'analyse harmonique de la covariance donnerait ici, outre la composante σ^2 localisée sur la fréquence fixe $\gamma = \lambda$, une composante imaginaire fluctuante (sur la même fréquence).

Il y a là une circonstance qui est de la même nature que le principe d'incertitude de Heisenberg en mécanique quantique. L'énergie n'est parfaitement déterminée (indépendante du temps) que lorsque le phénomène est stationnaire (de position totalement indéterminée dans le temps).

Exemple 2 : Soient maintenant Y_k et Z_k deux suites de variables indépendantes et d'espérances mathématiques nulles. Posons :

$$E(Y_i Y_j) = E(Z_i Z_j) = 0 \quad (i \neq j)$$

$$E(Y_i Z_j) = 0 \quad (\forall i \text{ et } j)$$

$$E(Y_k^2) = E(Z_k^2) = \sigma_k^2$$

On suppose la série $\sum_{k=1}^{\infty} \sigma_k^2 = \sigma^2$ convergente. Soit aussi λ_k une suite de nombres. Le processus :

$$X(t) = \sum_{k=1}^{\infty} (Y_k \cos \lambda_k t + Z_k \sin \lambda_k t)$$

existe (la série converge en moyenne quadratique) et possède la variance finie σ^2 . On vérifiera facilement que $X(t)$ est stationnaire au sens large et possède la covariance :

$$K(h) = \sum_{k=1}^{\infty} \sigma_k^2 \cos \lambda_k h$$

La décomposition spectrale de $K(h)$ fait apparaître un spectre discret où chacune des fréquences λ_k est porteuse de l'énergie σ_k^2 . Ici encore cette affectation de quantités d'énergie déterminées à chacune des fréquences cesse d'être possible, lorsque les Y_k et les Z_k ne sont plus indépendantes, en même temps que $X(t)$ cesse d'être stationnaire.

Il n'y a aucune raison pour qu'une telle propriété soit particulière aux spectres discrets. Le théorème fondamental que nous énoncerons dans un instant montre qu'elle est générale. Il nous faut auparavant introduire deux notions : intégrales de Stieltjes stochastiques, et processus à accroissements indépendants, qui sont de simples généralisations de notions déjà rencontrées.

Intégrales stochastiques de Stieltjes.

Soit $X(t)$ un processus de variance finie et de covariance $K(t_1, t_2)$. L'intégrale stochastique de Stieltjes :

$$(24) \quad I = \int_a^b p(t) dX(t)$$

se définit comme la limite en moyenne quadratique (lorsqu'elle existe) des sommes finies :

$$I_n = \sum_{i=1}^n p(t_i) [X(t_i) - X(t_{i-1})]$$

On montre, comme pour l'intégrale stochastique de Riemann, que I existe si et seulement si l'intégrale double de Stieltjes :

$$S^2 = \int_a^b \int_a^b p(t) p(t') d^2 K(t, t')$$

existe, et on a dans ce cas :

$$S^2 = \mathbb{E}(I^2)$$

Processus à accroissements indépendants. Il s'agit d'une généralisation évidente des processus à accroissements indépendants et stationnaires étudiés au chapitre V. $X(t)$ est dit à accroissements indépendants si, quels que soient les intervalles (t_i, t'_i) ne se recouvrant pas, les $X(t'_i) - X(t_i)$ sont mutuellement indépendants. Si l'on suppose simplement que les $X(t'_i) - X(t_i)$ ont des corrélations nulles :

$$(t_i, t'_i) \cap (t_j, t'_j) = \emptyset \Rightarrow E \left[X(t'_i) - X(t_i) \right] \left[X(t'_j) - X(t_j) \right] = 0$$

on dira que le processus est à accroissements sans corrélation.

Nous supposons que $X(t) - X(0)$ a une variance finie $\sigma^2(t)$ et une espérance mathématique nulle :

$$\begin{cases} E \left[X(t) - X(0) \right] = 0 \\ E \left[\left(X(t) - X(0) \right)^2 \right] = \sigma^2(t) \end{cases}$$

On établit que $\sigma^2(t)$ est une fonction croissante (mais non nécessairement bornée ni continue) et que l'on a pour $0 < t_0 < t$

$$E \left[\left(X(t) - X(t_0) \right)^2 \right] = \sigma^2(t) - \sigma^2(t_0)$$

Désignons par $K(t, t')$ la covariance de $X(t) - X(0)$ et $X(t') - X(0)$. L'intégrale stochastique :

$$X(t) - X(0) = \int_0^t dX(t)$$

existe toujours (il suffit de considérer les sommes finies I_n pour le voir), et on a :

$$\sigma^2(t) = \int_0^t \int_0^t d^2 K(t, t')$$

Calculons $K(t, t')$ avec, par exemple, $t < t'$:

$$\begin{aligned} K(t, t') &= E \left[\left(X(t) - X(0) \right) \left(X(t') - X(0) \right) \right] \\ &= E \left[\left(X(t) - X(0) \right)^2 \right] + E \left[\left(X(t) - X(0) \right) \left(X(t') - X(t) \right) \right] \\ &= \sigma^2(t) \end{aligned}$$

(Comme t est borne commune de $(0, t)$ et (t, t') ce résultat n'est vrai que si $\sigma^2(t)$ est continu en t , mais nous n'avons pas en vue un raisonnement rigoureux).

Une intégrale stochastique du type (24) admet de même la variance :

$$\begin{aligned}
 S^2 \int_a^b \int_a^b p(t) p(t') d^2 K(t, t') \\
 &= \lim_{n \rightarrow \infty} \sum_{i, j} p(t_i) p(t_j) \mathbb{E} \left[X(t_i) - X(t_{i-1}) \right] \left[X(t_j) - X(t_{j-1}) \right] \\
 &= \lim_{n \rightarrow \infty} \sum_i [p(t_i)]^2 \left[\sigma^2(t_i) - \sigma^2(t_{i-1}) \right] \\
 &= \int_a^b [p(t)]^2 d\sigma^2(t)
 \end{aligned}$$

Introduisant la fonction de répartition Θ de la mesure de Dirac. ($\Theta(t) = 0$ pour $-t \leq 0$ et $\Theta(t) = 1$ pour $t > 0$), le résultat obtenu montre que l'on peut écrire :

$$(25) \quad d^2 K(t, t') = d\Theta(t'-t) d\sigma^2(t)$$

et cette relation caractérise parfaitement les processus à accroissements sans corrélation.

Nous sommes maintenant en mesure d'énoncer le théorème fondamental de la décomposition spectrale.

Théorème fondamental - Tout processus stationnaire au sens large $X(t)$, de variance finie et de covariance $K(h)$ peut être mis sous la forme :

$$(26) \quad X(t) = \int_0^\infty \cos \lambda t d\xi_1(\lambda) + \int_0^\infty \sin \lambda t d\xi_2(\lambda)$$

où $\xi_1(\lambda)$ et $\xi_2(\lambda)$ sont deux processus à accroissements sans corrélation tels que, de plus, quels que soient $\lambda_1, \lambda_1', \lambda_2$ et λ_2' :

$$\mathbb{E} \left[\left(\xi_1(\lambda_1') - \xi_1(\lambda_1) \right) \left(\xi_2(\lambda_2') - \xi_2(\lambda_2) \right) \right] = 0$$

$\xi_1(\lambda)$ et $\xi_2(\lambda)$ ont la même variance $\sigma^2(\lambda)$, bornée et vérifiant :

$$(27) \quad K(h) = \int_0^\infty \cos \lambda h d\sigma^2(\lambda)$$

et sont donnés par les relations :

$$(28) \quad \left\{ \begin{aligned} \xi_1(\lambda) &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{\sin \lambda t}{t} X(t) dt \\ \xi_2(\lambda) &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{1 - \cos \lambda t}{t} X(t) dt \end{aligned} \right.$$

Ainsi la transformation de Fourier échange les processus stationnaires au sens large et les processus à accroissements sans corrélation. A chaque bande de fréquence $(\lambda, \lambda + \Delta \lambda)$ de la décomposition (26) est affectée une énergie bien déterminée $\Delta \sigma^2(\lambda) = \sigma^2(\lambda + \Delta \lambda) - \sigma^2(\lambda)$ mise en évidence par le spectre de $K(h)$ qui apparaît en (27).

Nous ne démontrerons pas le théorème fondamental. Nous nous contenterons de la vérification suivante : soit $\xi_1(\lambda)$ et $\xi_2(\lambda)$ deux processus à accroissements sans corrélation et sans corrélation mutuelle, possédant la même variance bornée $\sigma^2(\lambda)$. Alors la relation (26) définit bien un processus stationnaire au sens large, dont la covariance est donnée par (27).

En effet, en premier lieu, $X(t)$ existe, car l'intégrale :

$$E [X(t)^2] = \int_0^{\infty} (\cos^2 \lambda t + \sin^2 \lambda t) d\sigma^2(\lambda) = \int_0^{\infty} d\sigma^2(\lambda)$$

existe puisque $\sigma^2(\lambda)$ est bornée.

En deuxième lieu, calculons la covariance de $X(t)$:

$$\begin{aligned} E [X(t+h)X(t)] &= \int_0^{\infty} \int_0^{\infty} \cos \lambda(t+h) \cos \lambda' t \, d\theta(\lambda' - \lambda) d\sigma^2(\lambda) \\ &\quad + \int_0^{\infty} \int_0^{\infty} \sin \lambda(t+h) \sin \lambda' t \, d\theta(\lambda' - \lambda) d\sigma^2(\lambda) \\ &= \int_0^{\infty} \cos \lambda(t+h) \cos \lambda t \, d\sigma^2(\lambda) + \int_0^{\infty} \sin \lambda(t+h) \sin \lambda t \, d\sigma^2(\lambda) \\ &= \int_0^{\infty} \cos \lambda h \, d\sigma^2(\lambda) \end{aligned}$$

Elle ne dépend pas de t . $X(t)$ est bien stationnaire au sens large, et sa covariance est bien donnée par (27).