PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 11, Fasc. 1 (1990), p. 79-95

ENTROPIE DES PROCESSUS LINÉAIRES

PAR

ABDELKADER MOKKADEM (PARIS)

Abstract. Let $Y = (Y_n)_{n \in Z}$ be a real stationary process with the spectral representation

$$Y_n = \int_{-\pi}^{\pi} e^{in\lambda} dZ_Y,$$

and $X = (X_n)_{n \in Z}$ be a process defined by

$$X_n = \int_{-\pi}^{\pi} e^{in\lambda} \varphi(\lambda) dZ_Y.$$

We prove that the entropy of X satisfies the relation

$$H(X) \geqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Log} |\varphi(\lambda)| d\lambda + H(Y).$$

Some cases where equality holds are obtained. We also give some applications.

1. INTRODUCTION

Soit X une variable aléatoire à valeurs dans R^k, admettant une densité de probabilité $f(x)$ telle que $f(x) \operatorname{Log} f(x)$ soit intégrable; *l'entropie* de X est définie par

$$h(X) = -\int_{R^k} f(x) \operatorname{Log} f(x) dx \tag{1}$$

(cette définition sera étendue dans la section 2 à toute variable aléatoire admettant un moment absolu d'ordre $s > 0$ fini).

Pour un processus réel, stationnaire en loi, $X = (X_n)_{n \in Z}$, l'entropie de X est définie par

$$H(X) = \lim_n \frac{1}{n+1} h(X_0, \ldots, X_n). \tag{2}$$

Considérons maintenant un processus stationnaire en loi, réel et harmonisable (voir section 2) $Y = (Y_n)_{n \in \mathbf{Z}}$ et $X = (X_n)_{n \in \mathbf{Z}}$ filtré de Y par un filtre linéaire réel

$$X_n = \int_{-\pi}^{\pi} e^{in\lambda} \varphi(\lambda) \, dZ_Y(\lambda), \tag{3}$$

où dZ_Y désigne la mesure spectrale de Y, $\varphi(\lambda)$ est intégrable par rapport à dZ_Y et vérifie

$$\varphi(\lambda) = \overline{\varphi(-\lambda)}, \tag{4}$$

cette dernière condition assurant que X est réel.

On établit dans cet article le résultat suivant qui généralise des résultats obtenus par Kanter [9] dans le cas où les Y_n sont indépendants et par Shepp, Slepian et Wyner [15] dans le cas où φ est un polynôme trigonométrique:

$$H(X) \geqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Log} |\varphi(\lambda)| \, d\lambda + H(Y). \tag{5}$$

L'égalité dans (5) est obtenue dans différents cas, en particulier quand Y est régulier ou symétrique α-stable, ce qui permet d'écrire l'entropie de Y en fonction de son spectre et de l'entropie de son innovation. Un tel résultat a été établi par Shannon [14] dans le cas gaussien en faisant directement le calcul de $h(X_0, \ldots, X_n)$ et en passant à la limite. Notre méthode n'utilise pas ce calcul explicite; elle permet en particulier d'avoir une nouvelle interprétation du premier théorème de Szëgo,

$$\lim_n |\det Q_{n+1}|^{1/(n+1)} = \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Log} f(\lambda) \, d\lambda\right\}, \tag{6}$$

où Q_{n+1} est la matrice des covariances de $(X_0, \ldots, X_n)$ et f la densité spectrale de $X = (X_n)_{n \in \mathbf{Z}}$.

Le plan de l'article est le suivant: la section 2 est constituée de préliminaires qu'on utilise dans la suite; on établit notre résultat principal dans la section 3; des applications et une extension au cas des processus n'ayant pas de représentation spectrale sont données dans la section 4; on discute en particulier dans cette section du principe du maximum d'entropie introduit par Burg [2] et [3].

Tous les processus que nous considérerons seront supposés centrés.

2. PRÉLIMINAIRES ET NOTATIONS

Nous introduisons d'abord quelques propriétés de l'information et de l'entropie de variables aléatoires.

Définition 2.1. Soient P_1 et P_2 deux mesures de probabilité définies sur un même espace mesurable $(\Omega, \mathscr{A})$. On définit *l'information de Kullback* de P_1 et P_2 par

(7) $$K(P_1, P_2) = \sup \sum_i P_1(E_i) \operatorname{Log} \frac{P_1(E_i)}{P_2(E_i)},$$

où le supremum est pris sur toutes les partitions finies (E_i) de Ω.

$K(P_1, P_2)$ est à valeurs dans $[0, +\infty]$; elle est infinie si P_2 ne domine pas P_1. Quand P_1 et P_2 admettent des densités de probabilité f_1 et f_2, on a

(8) $$K(P_1, P_2) = \int f_1 \operatorname{Log} f_1/f_2.$$

Définition 2.2. Soit (X, Y) un couple de variables aléatoires de loi de probabilité P_{XY}. Notons P_X (resp. P_Y) la loi de probabilité de X (resp. Y). On définit *l'information mutuelle* de X et Y par

(9) $$I(X, Y) = K(P_{XY}, P_X \otimes P_Y).$$

Les propriétés suivantes sont établies dans Pinsker [10]:

si P_{1n} et P_{2n} convergent respectivement vers P_1 et P_2, alors

(10) $$\liminf K(P_{1n}, P_{2n}) \geqslant K(P_1, P_2);$$

(11) $$I(X, Y) = I(Y, X);$$

(12) $I(X, Y) = 0$ si et seulement si X et Y sont indépendantes;

(13) $I(X, X) = \infty$ si X est absolument continue;

(14) $I(X, Y) \geqslant I(f(X), Y)$ pour toute application mesurable f;

si $X = (X_1, X_2, \ldots)$, alors

(15) $$I(X, Y) = \lim_n I((X_1, \ldots, X_n), Y).$$

Nous passons maintenant à la définition de l'entropie d'une variable aléatoire. Pour une variable aléatoire X dans $\boldsymbol{R}^k$, admettant une densité de probabilité f et un moment absolu d'ordre $s > 0$ fini, l'intégrale

(16) $$h(X) = -\int f(x) \operatorname{Log} f(x)\, dx$$

est bien définie et à valeurs dans $[-\infty, +\infty[$. Pour éviter d'avoir à vérifier à chaque étape l'existence d'une densité, nous poserons

(17) $h(X) = -\infty$ si X n'est pas absolument continue.

Pour justifier (17) et rester dans le cadre du début de la présente section, on pose la

Définition 2.3. Si Q_u est une mesure de probabilité sur $\boldsymbol{R}^k$ de densité u et si la variable aléatoire X à valeurs dans $\boldsymbol{R}^k$ a pour loi de probabilité P et vérifie

(18) $$E_P\{\operatorname{Log} u(X)\} > -\infty,$$

on pose

(19) $$h(X) = -K(P, Q) - E_P\{\operatorname{Log} u(X)\}.$$

On a ainsi défini l'entropie pour toute variable aléatoire à valeurs dans $\boldsymbol{R}^k$, satisfaisant à une condition du type (18). Cette entropie est à valeurs dans $[-\infty, +\infty[$, elle vaut $-\infty$ quand la variable aléatoire n'est pas absolument continue et est égale à l'entropie usuelle (16) sinon. La condition (18) est vérifiée par les variables aléatoires admettant un moment absolu d'ordre $s > 0$ en prenant

(20) $$u(x) = Ce^{-|x|^s},$$

mais aussi, par exemple, par les variables aléatoires X vérifiant $E\operatorname{Log}(1+|X|^2) < \infty$ en prenant

(21) $$u(x) = \frac{C}{1+|x|^2}.$$

Les propriétés usuelles de l'entropie sont naturellement encore vérifiées. En particulier on a (Ash [1]):

(22) $$h(X) = h(X+a) \text{ pour toute constante } a;$$

si A est une application linéaire de $\boldsymbol{R}^k$ dans $\boldsymbol{R}^k$, alors

(23) $$h(AX) = \operatorname{Log}|\det A| + h(A);$$

(24) $$h(X, Y) \leqslant h(X) + h(Y)$$

avec égalité si X et Y sont indépendantes;

(25) si la matrice des covariances de X est fixée, le maximum de $h(X)$ est atteint quand X est gaussienne.

Nous définissons maintenant l'entropie conditionnelle de X sachant Y où Y est à valeurs dans un espace quelconque et X est à valeurs dans $\boldsymbol{R}^k$ et admet une entropie au sens de (19) par

(26) $$h(X|Y) = h(X) - I(X, Y).$$

Les propriétés suivantes proviennent alors de celles de l'information et de l'entropie:

(27) $$h(X, X') = h(X') + h(X|X');$$

(28) $$h(X|g(X')) \geqslant h(X|X'), \text{ si } g \text{ est mesurable};$$

(29) $$h(X + g(X')|X') = h(X|X'), \text{ si } g \text{ est mesurable}.$$

Notons que (28) implique

(30) $$h(X|X', g(X')) = h(X|X').$$

Nous aurons besoin dans la suite des propositions suivantes dont la démonstration est immédiate.

PROPOSITION 2.1. *Soient X et Y deux variables aléatoires indépendantes et à valeurs dans $\boldsymbol{R}^k$. Si X et $X+Y$ admettent une entropie au sens de* (19), *alors* $h(X+Y) \geqslant h(X)$.

PROPOSITION 2.2. *Soit (X_n) une suite de variables aléatoires à valeurs dans $\boldsymbol{R}^k$ convergente en loi vers X et u une densité de probabilité sur $\boldsymbol{R}^k$. Si les variables* $\operatorname{Log} u(X_n)$ *sont uniformément intégrables, alors*

$$\limsup_n h(X_n) \leqslant h(X).$$

Preuve. C'est une conséquence de (10) et (19).

Nous définissons maintenant l'entropie d'un processus stationnaire.

Définition 2.4. Soit $X = (X_n)_{n\in\boldsymbol{Z}}$ un processus réel stationnaire en loi et tel que X_0 satisfait une condition (18). *L'entropie* de X est définie par

$$H(X) = \lim_n \frac{1}{n+1} h(X_0, \ldots, X_n). \tag{31}$$

Notons que par stationnarité on a aussi, pour tout k,

$$H(X) = \lim_n \frac{1}{n+1} h(X_k, \ldots, X_{k+n}). \tag{32}$$

On peut vérifier (voir Kanter [9]) que la suite $(1/n+1)h(X_0, \ldots, X_n)$ est décroissante et que

$$H(X) = \lim_n h(X_0 | X_{-1}, \ldots, X_{-n}) = h(X_0 | X_{-\infty}^{-n}), \tag{33}$$

où $X_{-\infty}^{-n} = (X_{-1}, \ldots, X_{-n}, \ldots)$.

Nous allons maintenant définir des classes de processus que nous utiliserons dans la suite. Pour les propriétés des variables aléatoires symétriques α-stable ($S\alpha S$) on se refère à Cambanis [4] où est définie en particulier la covariation de deux variables $S\alpha S$ X et Y, que nous noterons également $[X, Y]_\alpha$. La norme en covariation sera notée $\|X\|_\alpha = [X, X]_\alpha^{1/\alpha}$. Dans toute la suite on supposera que $1 < \alpha \leqslant 2$.

Un processus réel ou complexe $X = (X_t)_{t\in T}$ est dit $S\alpha S$ si pour tout $t_1, \ldots, t_k$ les variables $X_{t_1}, \ldots, X_{t_k}$ sont conjointement $S\alpha S$; pour $\alpha = 2$, X est alors gaussien.

On notera $\mathfrak{S}_\alpha$ la classe des processus réels, $X = (X_n)_{n\in\boldsymbol{Z}}$ strictement stationnaires, $S\alpha S$ et qui admettent une représentation

$$X_n = \int_{-\pi}^{\pi} e^{in\lambda} d\xi(\lambda), \tag{34}$$

où $\xi(\lambda)$ est processus $S\alpha S$ à accroissements indépendants; $\mu\{[a, b]\} = \|\xi(b)-\xi(a)\|_\alpha^\alpha$ définit une mesure positive sur $T = [-\pi, \pi]$ (voir [4], [8], [13]). On dira que μ est la *mesure spectrale* de X et que $d\xi$ est sa *mesure spectrale aléatoire*; on les notera respectivement μ_X et dZ_X.

On considère aussi la classe $\mathfrak{C}$ des processus réels strictement stationnaires et de carré intégrable (notons que $\mathfrak{S}_2$ est la classe des processus gaussiens et est contenue dans $\mathfrak{C}$).

La théorie spectrale des processus dans $\mathfrak{C}$ est largement développée [11]; Cambanis et Soltani [6] développent une théorie analogue dans $\mathfrak{S}_\alpha$; ils obtiennent en particulier

2.5. Décomposition de Wold. Tout $X \in \mathfrak{S}_\alpha$, s'écrit $X = X_1 + X_2$, où X_1 et X_2 sont indépendants; X_1 est la partie régulière de X et X_2 sa partie singulière; X est dit *régulier* (resp. *singulier*) si $X_2 = 0$ (resp. $X_1 = 0$).

2.6. Moyenne mobile. Si X est régulier de densité spectrale $f(\lambda)$, il existe une fonction extérieure h dans l'espace de Hardy H^α et un processus $S\alpha S$, $W = (W_n)_{n\in \mathbf{Z}}$, dont la mesure spectrale μ_W est la mesure de Lebesgue, tels que $|h(\lambda)|^\alpha = f(\lambda)$ et

$$X_n = \int_{-\pi}^{\pi} e^{in\lambda}\, \bar{h}(\lambda)\, dZ_W.$$

Si $\alpha \neq 2$, Cambanis et Soltani [6] montrent que les variables aléatoires $(W_n)_{n\in \mathbf{Z}}$ ne sont pas indépendantes; mais les covariations $[W_n, W_{n'}]_\alpha$ sont nulles pour $n \neq n'$. W est appelé *innovation* de X.

3. ENTROPIE ET FILTRE LINÉAIRE

L'objet de cette section est d'évaluer l'entropie d'un processus obtenu par un filtre linéaire; nous commençons par calculer l'entropie d'un processus singulier dans $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$.

PROPOSITION 3.1. *Si Y est singulier, alors $H(Y) = -\infty$.*

Preuve. Puisque Y est singulier, on a $Y_0 = T(Y_{-\infty}^{-1})$, où T est linéaire. Donc $I(Y_0, Y_{-\infty}^{-1}) \geq I(Y_0, T(Y_{-\infty}^{-1})) \geq I(Y_0, Y_0)$, mais $H(Y) = h(Y_0) - I(Y_0, Y_{-\infty}^{-1})$, d'où $H(Y) \leq h(Y_0) - I(Y_0, Y_0)$. Si Y_0 est absolument continue on conclut grâce à (13). Si Y_0 n'est pas absolument continue, on sait qu'alors $h(Y_0) = -\infty$ et on conclut encore.

Nous comparons maintenant l'entropie de Y avec l'entropie de sa partie régulière.

PROPOSITION 3.2. *Si Y est non singulier et Y_1 (resp. Y_2) est sa partie régulière (resp. singulière), alors $H(Y_1) \geq H(Y)$. De plus, on a l'égalité si Y_1 et Y_2 sont indépendants.*

Preuve. On pose $Y_1 = (Y_{1,n})_{n\in \mathbf{Z}}$ et $Y_2 = (Y_{2,n})_{n\in \mathbf{Z}}$ on a en utilisant (28):

$$H(Y_1) = h(Y_{1,n} | Y_{1,-\infty}^{n-1}) \geq h(Y_n - Y_{2,n} | Y_{1,-\infty}^{n-1}, Y_{-\infty}^{n-1}).$$

Comme Y_{2n} est fonction de $Y^{n-1}_{-\infty}$, on obtient, par (29), que $H(Y_1) \geqslant h(Y_n | Y^{n-1}_{1,-\infty}, Y^{n-1}_{-\infty})$; de même, $Y^{n-1}_{1,-\infty}$ est fonction de $Y^{n-1}_{-\infty}$ et donc, par (28), $H(Y_1) \geqslant h(Y_n | Y^{n-1}_{-\infty})$, ce qui termine la première partie.

Si Y_1 et Y_2 sont indépendants, on a (d'après la proposition 2.1) $H(Y) \geqslant H(Y_1)$, on conclut donc la démonstration.

Pour évaluer l'entropie d'un processus X obtenu avec un filtre linéaire appliqué à un processus Y de $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$, on a besoin de deux résultats préliminaires. Le premier est dû à Shepp, Slepian et Wyner [15].

LEMME 3.1 [15]. *Soit X et Y deux processus réels stationnaires en loi admettant un moment absolu d'ordre $s > 0$ fini et tels que*

$$X_n = \sum_{j=0}^{k} a_j Y_{n+m-j},$$

où les a_j sont réels et m est fixé. Posons

$$\varphi(\lambda) = \sum_{j=0}^{k} a_j e^{ij\lambda}.$$

On a alors

$$H(X) \geqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Log} |\varphi(\lambda)| \, d\lambda + H(Y).$$

Le deuxième résultat est un lemme d'approximation.

LEMME 3.2. *Soit μ la mesure spectrale d'un processus réel non singulier de $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$. Soit $\varphi \in L^p(\mu, T)$ avec $p \geqslant 1$. On suppose que*

$$\varphi(\lambda) = \overline{\varphi(-\lambda)} \quad \textit{pour tout } \lambda \in [-\pi, \pi] \tag{35}$$

et

$$\int_{-\pi}^{\pi} \operatorname{Log} |\varphi(\lambda)| \, d\lambda > -\infty. \tag{36}$$

Alors il existe une suite φ_n de polynômes trigonométriques à coefficients réels tels que

$$\lim \varphi_n = \varphi \quad \textit{dans } L^p(\mu, T) \tag{37}$$

et

$$\lim \int_{-\pi}^{\pi} \operatorname{Log} |\varphi_n(\lambda)| \, d\lambda = \int_{-\pi}^{\pi} \operatorname{Log} |\varphi(\lambda)| \, d\lambda. \tag{38}$$

P r e u v e. Soit $\mu = \mu_1 + \mu_2$ la décomposition de Lebesgue de μ. Puisque le processus de mesure spectrale μ est non singulier, μ_1 est équivalente à la mesure de Lebesgue.

Faisons d'abord l'approximation dans $L^p(\mu_1, T)$. Soit $\delta > 0$. Il existe alors $\varepsilon > 0$, $\varepsilon < \delta$, tel que pour tout borélien B vérifiant $\mu_1(B) < \varepsilon$ on ait

$$(39) \qquad \int_B |\varphi|^p \, d\mu_1 < \delta, \qquad \int_B \big|\mathrm{Log}\,|\varphi(\lambda)|\big| \, d\lambda < \delta, \qquad \int_B |\varphi| \, d\mu_1 < \delta, \qquad \int_B d\lambda < \delta;$$

cela provient du fait que les mesures $|\varphi|^p \, d\mu_1$, $\big|\mathrm{Log}\,|\varphi|\big| \, d\lambda$, $|\varphi| \, d\mu_1$ et $d\lambda$ sont dominées par $d\mu_1$.

Puisque $\int \mathrm{Log}\,|\varphi| \, d\lambda > -\infty$, la mesure de Lebesgue de l'ensemble N des zéros de φ est nulle.

On se place dans le complémentaire de N; par le théorème de Lusin [12], on peut trouver un compact K_1 qu'on peut choisir symétrique par rapport à $\lambda = 0$ et une fonction continue ψ_1 tels que

$$(40) \qquad \mu_1(T - K_1) < \varepsilon$$

et

$$(41) \qquad \psi_1 = \varphi \quad \text{sur } K_1.$$

ψ_1 étant sans zéros sur K_1, soient

$$(42) \qquad \gamma = \sup_{K_1} \big(\big|\mathrm{Log}\,|\psi_1|\big|\big) \quad \text{et} \quad \beta = \max\{1, \sup_{K_1} |\psi_1|\}.$$

Choisissons maintenant $\varepsilon_2 > 0$ et K_2 un compact dans $T - K_1$, symétrique par rapport à $\lambda = 0$, tels que

$$(43) \qquad \varepsilon_2 < \delta/\beta^p, \quad \varepsilon_2 < \delta/\gamma$$

et

$$(44) \qquad \mu_1(T - (K_1 \cup K_2)) < \varepsilon_2.$$

Définissons sur $K = K_1 \cup K_2$ la fonction

$$(45) \qquad \psi_2 = \begin{cases} \psi_1 & \text{sur } K_1, \\ 1 & \text{sur } K_2. \end{cases}$$

On prolonge ψ_2 de la manière suivante. Si $I =]\lambda_1, \lambda_2[$ est un intervalle dans le complémentaire de K tel que λ_1 et λ_2 soient dans K, on pose, pour tout λ dans I,

$$(46) \qquad \psi_2(\lambda) = \left[\frac{\lambda - \lambda_1}{\lambda_2 - \lambda_1} |\psi_2(\lambda_2)| + \frac{\lambda_2 - \lambda}{\lambda_2 - \lambda_1} |\psi_2(\lambda_1)|\right] \times$$
$$\times \exp\left\{i\left[\frac{\lambda - \lambda_1}{\lambda_2 - \lambda_1} \mathrm{Arg}\,\psi_2(\lambda_2) + \frac{\lambda_2 - \lambda}{\lambda_2 - \lambda_1} \mathrm{Arg}\,\psi_2(\lambda_1)\right]\right\}.$$

On obtient ainsi une fonction continue qui vérifie

$$(47) \qquad \psi_2(\lambda) = \overline{\psi_2(-\lambda)}, \quad |\psi_2| < \beta, \quad \big|\mathrm{Log}\,|\psi_2|\big| < \gamma$$

(ψ_2 n'a pas de zéro sur T).

Il est alors facile de vérifier que

$$\int_{-\pi}^{\pi} |\varphi-\psi_2|^p \, d\mu_1 \leqslant 8\delta \tag{48}$$

et

$$\Big|\int_{-\pi}^{\pi} \operatorname{Log}|\varphi| \, d\lambda - \int_{-\pi}^{\pi} \operatorname{Log}|\psi_2| \, d\lambda\Big| \leqslant 3\delta. \tag{49}$$

Passons maintenant à $L^p(\mu, T)$. On peut prendre un support de μ_2: S symétrique par rapport à $\lambda = 0$. Soit Q un polynôme à coefficients réels tel que

$$\int_{-\pi}^{\pi} |\varphi-Q|^p \, d\mu_2 < \delta. \tag{50}$$

On peut prendre Q sans zéro sur T en perturbant faiblement ses racines et en gardant des coefficients réels. Notons: $\psi_3 = \psi_2$ sur $T-S$ et $\psi_3 = Q$ sur S. Il est clair que

$$\int_{-\pi}^{\pi} |\varphi-\psi_3|^p \, d\mu \leqslant 9\delta \tag{51}$$

et

$$\Big|\int_{-\pi}^{\pi} \operatorname{Log}|\psi_3| \, d\lambda - \int_{-\pi}^{\pi} \operatorname{Log}|\varphi| \, d\lambda\Big| \leqslant 3\delta. \tag{52}$$

ψ_3 n'a pas de zéros sur T; on pose

$$\beta = \sup_T |\psi_3| \quad \text{et} \quad \gamma = \sup_T \big|\operatorname{Log}|\psi_3|\big|. \tag{53}$$

Soient ε' et ε'' tels que

$$\varepsilon' < \delta/\lambda, \qquad \varepsilon'' < \delta/4\beta^p \tag{54}$$

et

$$\mu(B) < \varepsilon'' \Rightarrow \int_B d\lambda < \varepsilon'. \tag{55}$$

Par le théorème de Lusin, il existe un compact K (qu'on prendra symétrique) et une fonction continue ψ' tels que

$$\mu(T-K) < \varepsilon'' \tag{56}$$

et

$$\psi' = \psi_3 \text{ sur } K. \tag{57}$$

On prolonge ψ' comme plus haut en une fonction continue ψ qui vérifie

$$|\psi| \leqslant \beta \quad \text{et} \quad \big|\operatorname{Log}|\psi|\big| \leqslant \gamma. \tag{58}$$

Il ne reste plus qu'à approcher ψ uniformément par un polynôme trigonométrique à coefficients réels pour conclure la démonstration.

Remarque 3.1. Le lemme est encore vrai pour la norme sup; si φ est une fonction continue sur T, on peut l'approcher uniformément par des polynômes trigonométriques à coefficients réels qui vérifient (38).

Remarque 3.2. La condition $\varphi(\lambda) = \overline{\varphi(-\lambda)}$ signifie que le filtre est réel; sans cette condition le lemme est encore vrai mais l'approximation est faite avec des polynômes à coefficients complexes.

On établit maintenant notre résultat principal.

THÉORÈME 3.1. *Soient X et Y deux processus dans $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$ tels que*

$$X_n = \int_{-\pi}^{\pi} e^{in\lambda} \varphi(\lambda)\, dZ_Y \quad \text{avec} \quad \varphi \in L^\alpha(\mu_Y, T).$$

Alors:

(i)
$$H(X) \geqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Log} |\varphi(\lambda)|\, d\lambda + H(Y).$$

(ii) *On a l'égalité dans* (i) *dans les cas suivants*: (a) $\varphi(\lambda) \neq 0$ μ_Y*-presque partout*; (b) X *est singulier*; (c) Y *est régulier*; (d) $Y \in \mathfrak{S}_\alpha$.

Preuve. (i) Si Y est singulier, X l'est aussi et (i) est vérifiée d'après la proposition 3.1 et on a en fait trivialement l'égalité dans (i).

Supposons que Y est non singulier; on a alors deux possibilités. Si X est non singulier, on a

$$\int_{-\pi}^{\pi} \operatorname{Log} |\varphi(\lambda)|\, d\lambda > -\infty. \tag{59}$$

D'après le lemme 3.2 on peut trouver une suite de polynômes trigonométriques à coefficients réels φ_k, tels que

(60) $\lim \varphi_k = \varphi$ dans $L^\alpha(\mu_Y, T)$ (avec $\alpha = 2$ si X et Y sont dans $\mathfrak{C}$)

et

$$\lim_k \int_{-\pi}^{\pi} \operatorname{Log} |\varphi_k(\lambda)|\, d\lambda = \int_{-\pi}^{\pi} \operatorname{Log} |\varphi(\lambda)|\, d\lambda. \tag{61}$$

Soit

$$X_n(k) = \int_{-\pi}^{\pi} e^{in\lambda} \varphi_k(\lambda)\, dZ_Y. \tag{62}$$

On a

$$\|X_n(k) - X_n\|_\alpha^\alpha = \int |\varphi_k(\lambda) - \varphi(\lambda)|^\alpha\, d\mu_Y = \delta_k \tag{63}$$

et

$$\lim \delta_k = 0. \tag{64}$$

On peut extraire une sous-suite telle que

(65) $$\sum \delta_k < \infty$$

de sorte que le processus $X(k)$ converge en loi vers X; et donc

(66) $$\liminf I\left(X_0(k),\, X_{-\infty}^{-1}(k)\right) \geqslant I\left(X_0,\, X_{-\infty}^{-1}\right),$$

mais

(67) $$H\left(X(k)\right) = h\left(X_0(k)\right) - I\left(X_0(k),\, X_{-\infty}^{-1}(k)\right)$$

et d'après la proposition 2.2

(68) $$\limsup_k h\left(X_0(k)\right) \leqslant h(X_0).$$

Il vient, par conséquent,

(69) $$\limsup_k H\left(X(k)\right) \leqslant H(X);$$

on conclut alors avec les lemmes 3.1 et 3.2.

Si maintenant X est singulier, on a

(70) $$\int_{-\pi}^{\pi} \operatorname{Log}|\varphi(\lambda)|\, d\lambda = -\infty$$

et

(71) $$-\infty = H(X) = -\infty + H(Y).$$

(ii) (a) Si $\varphi(\lambda) \neq 0$ μ_Y-presque partout, alors $\psi(\lambda) = 1/\varphi(\lambda)$ est dans $L^\alpha(\mu_X)$ et on a

(72) $$Y_n = \int_{-\pi}^{\pi} e^{in\lambda} \psi(\lambda)\, dZ_X.$$

Par conséquent, d'aprés (i),

(73) $$H(Y) \geqslant -\frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Log}|\varphi(\lambda)|\, d\lambda + H(X),$$

on conclut donc:

(b) a été démontré dans (i).

(c) Si Y est régulier, ou bien X est singulier et on a l'égalité d'après (b), ou bien X est non singulier et alors $\varphi(\lambda) \neq 0$ presque partout pour la mesure de Lebesgue. Comme μ_Y est absolument continue, il s'ensuit que $\varphi(\lambda) \neq 0$ μ_Y-presque partout, ce qui nous ramène à (a).

(d) Si Y est singulier, on a évidemment l'égalité. Si Y est non singulier, soit $Y = Y_1 + Y_2$ sa décomposition de Wold. Y_1 et Y_2 étant indépendants on a, d'après la proposition 3.2,

(74) $$H(Y) = H(Y_1).$$

D'autre part, il est clair que si X_1 est la partie régulière de X, on a

$$X_{1n} = \int_{-\pi}^{\pi} e^{in\lambda} \varphi(\lambda)\, dZ_{Y_1} \tag{75}$$

et donc

$$H(X_1) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Log} |\varphi(\lambda)|\, d\lambda + H(Y_1), \tag{76}$$

or

$$H(X_1) \geqslant H(X) \tag{77}$$

d'après la proposition 3.2. Les formules (74), (76) et (77) impliquent alors

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Log} |\varphi(\lambda)|\, d\lambda + H(Y) \geqslant H(X), \tag{78}$$

ce qui achève la démonstration.

Remarque 3.3. Dans [15], Shepp, Slepian et Wyner affirment qu'on a l'égalité dans (i) quand Y est ergodique; leur démonstration est malheureusement inexacte. Elle consiste principalement à montrer que le filtre est inversible, ce qui ne peut être le cas si la mesure μ_Y charge les zéros de φ (le seul point qui n'est jamais chargé par μ_Y quand Y est ergodique est $\lambda = 0$). Leur principal argument consiste à montrer que si $\xi \in C$, $|\xi| = 1$, alors la suite

$$c_k = \frac{1}{k} \sum_{j=1}^{k} \xi^{-j} Y_{n+j}$$

de variables aléatoires converge vers une constante presque sûrement. Cela est inexact, si $\xi \neq 1$. Un contrexemple simple est le suivant.

Supposons $\xi = e^{i\lambda}$ avec $\lambda \neq 0$ et prenons $Y_n = e^{in\lambda} V$, (Y_n) ergodique.

Dans [11] la loi de V est complètement caractérisée: $|V|$ est constant p.s. et la loi de $\operatorname{Arg} V$ est uniforme sur le sous-groupe fermé de T engendré par $e^{in\lambda}$. Comme $\lambda \neq 0$, ce sous-groupe est toujours de cardinal strictement plus grand que 1 (il est fini si $\lambda/2\pi$ est rationnel et égal à T sinon). V n'est donc pas constant p.s. Si on regarde maintenant c_k, on obtient

$$c_k = \frac{1}{k} \sum_{j=1}^{k} e^{-i\lambda j} e^{i(n+j)\lambda} V = e^{in\lambda} V$$

et donc $\lim c_k = e^{in\lambda} V$ ce qui n'est ni constant ni indépendant de n.

4. APPLICATIONS

4.1. Entropie d'un processus régulier de $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$. Soit X un processus régulier de densité spectrale f, soit W son innovation (μ_W est la mesure de Lebesgue). On a alors

(79) $$X_n = \int_{-\pi}^{\pi} e^{in\lambda} h(\lambda)\, dZ_W,$$

où

(80) $$|h(\lambda)|^\alpha = f(\lambda) \quad \text{avec } \alpha = 2 \text{ si } X \in \mathfrak{C}.$$

On obtient

(81) $$H(X) = \frac{1}{2\alpha\pi} \int_{-\pi}^{\pi} \operatorname{Log} f(\lambda)\, d\lambda + H(W).$$

Dans $\mathfrak{C}$, $H(W)$ est maximal si W est gaussien et on obtient

(82) $$H(X) \leqslant \frac{1}{4\pi} \int_{-\pi}^{\pi} \operatorname{Log} f(\lambda)\, d\lambda + \frac{1}{2} \operatorname{Log} 2\pi e$$

avec égalité si X est gaussien.

Enfin on a aussi le

COROLLAIRE 4.1. *Soit X un processus régulier de $\mathfrak{C}$ et W son innovation. Alors:*

Si $H(W) > -\infty$, les lois de probabilité marginales de dimension finie de X sont absolument continues.

Si $H(X) > -\infty$, les lois de probabilité marginales de dimension finie de W sont absolument continues.

4.2. Entropie d'un processus non singulier de $\mathfrak{S}_\alpha$. Si $X \in \mathfrak{S}_\alpha$ et X est non singulier, soit X_1 sa partie régulière et f sa densité spectrale. Comme $H(X_1) = H(X)$, on a

(83) $$H(X) = \frac{1}{2\alpha\pi} \int_{-\pi}^{\pi} \operatorname{Log} f(\lambda)\, d\lambda + H(W),$$

où W est l'innovation de X_1. En particulier, si $\alpha = 2$, X est gaussien et $H(W) = \frac{1}{2} \operatorname{Log} 2\pi e$. Il faut remarquer alors que, dans ce cas,

(84) $$\frac{1}{n+1} h(X_0, \ldots, X_n) = \frac{1}{2} \operatorname{Log} 2\pi e\, |\det Q_{n+1}|^{1/(n+1)},$$

où Q_{n+1} est la matrice des covariances de $(X_0, \ldots, X_n)$. Comme

(85) $$H(X) = \lim_n \frac{1}{n+1} h(X_0, \ldots, X_n),$$

on en déduit que

(86) $$\lim_n |\det Q_{n+1}|^{1/(n+1)} = \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Log} f(\lambda)\, d\lambda\right\},$$

ce qui est le premier théorème de Szëgo.

4.3. Comparaison de l'erreur de prédiction linéaire et non linéaire. Soit Y dans $\mathfrak{S}_\alpha$ ou $\mathfrak{C}$ et X tel que

$$X_n = \int_{-\pi}^{\pi} e^{in\lambda}\varphi(\lambda)\,dZ_Y. \tag{87}$$

Soit $\hat{X}_n = f(X_{-\infty}^{n-1})$ un prédicteur de X_n et $p > 0$. On peut montrer alors que [15]

$$\|X_n - \hat{X}\|_p \geqslant C(p)\,e^{H(X)}, \tag{88}$$

où $C(p)$ est une constante qui ne dépend que de p.

Si on note $\varepsilon_L(X)$ et $\varepsilon_L(Y)$ les erreurs de prédiction linéaires à un pas respectivement de X et Y on voit alors que

$$\|\varepsilon_L(X)\|_\alpha = \left[\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\operatorname{Log}|\varphi(\lambda)|\,d\lambda\right\}\right]\|\varepsilon_L(Y)\|_\alpha \tag{89}$$

(pour $\alpha = 2$ voir Rozanov [11] et pour $1 < \alpha \leqslant 2$ Cambanis et Soltani [6]), ce qui donne, en utilisant le théorème 3.1 et (88),

$$\|X_n - \hat{X}\|_p \geqslant \{C(p)\,\|\varepsilon_L(Y)\|_\alpha^{-1}\,e^{H(Y)}\}\,\|\varepsilon_L(X)\|_\alpha. \tag{90}$$

4.4. Processus d'entropie maximale pour une densité spectrale $f(\lambda)$ fixée. D'après la formule (81) il s'agit de trouver l'innovation W pour laquelle $H(W)$ est maximum. Cela est possible dans $\mathfrak{C}$; la seule contrainte sur W est $EW_n^2 = 1$. Comme $H(W) \leqslant h(W_n)$ avec égalité si et seulement si les (W_n) sont indépendants, il suffit donc de prendre des W_n indépendants tels que $h(W_n)$ soit maximum sous la contrainte $EW_n^2 = 1$; cela est réalisé d'après (25) pour W_n gaussienne. Le processus d'entropie maximale est donc le processus gaussien de densité spectrale $f(\lambda)$. En conséquence, pour tout X dans $\mathfrak{C}$ de densité spectrale $f(\lambda)$ on a

$$H(X) \leqslant \frac{1}{4\pi}\int_{-\pi}^{\pi}\operatorname{Log} f(\lambda)\,d\lambda + \frac{1}{2}\operatorname{Log} 2\pi e. \tag{91}$$

Dans $\mathfrak{S}_\alpha$, $\alpha \neq 2$, la situation est différente, car les W_n ne peuvent pas être indépendants. Le problème qui reste donc posé est celui de trouver le processus stationnaire $S\alpha S$ de mesure spectrale la mesure de Lebesgue et qui a une entropie maximale; cependant on peut obtenir une majoration de $H(X)$ en majorant $H(W)$ par $h(W_0)$ et $h(W_0)$ est majorée de la manière suivante: pour $1 \leqslant p < \alpha$ on a

$$(E|W_0|^p)^{1/p} = \|W_0\|_p = C(p, \alpha)\,\|W_0\|_\alpha, \tag{92}$$

où $C(p, \alpha)$ est une constante qui ne dépend que de p et α [5]; d'autre part il est bien connu que

$$h(W_0) \leqslant \operatorname{Log}\frac{2\Gamma(1/p)\,\|W_0\|_p}{p^{(p-1)/p}} + \frac{1}{p}, \tag{93}$$

ce qui permet d'obtenir une majoration de $H(X)$.

4.5. Principe du maximum d'entropie. Connaissant les k premières covariations d'un processus X dans $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$, on peut alors se poser deux problèmes:

4.5.1. *Quand l'innovation de X est fixée, trouver la densité spectrale f qui maximise $H(X)$.*

4.5.2. *Trouver le processus X pour lequel $H(X)$ est maximal.*

Qu'on soit dans $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$ le problème 4.5.1 revient à trouver la fonction positive intégrable f qui maximise

$$\int_{-\pi}^{\pi} \operatorname{Log} f(\lambda)\,d\lambda, \tag{94}$$

les k premiers coefficients de Fourier de f étant fixés. C'est la méthode d'estimation de f par maximum d'entropie ([2] et [3]). La solution à ce problème est connue; la densité spectrale cherchée est celle d'un processus autorégressif dans $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$.

La solution au problème 4.5.2 d'après ce qu'on vient de voir, est un processus autorégressif. Il reste à trouver l'innovation W qui maximise $H(W)$ (la contrainte étant $\|W_0\|_\alpha = 1$); comme on l'a vu en 4.4, dans $\mathfrak{C}$ la solution est obtenue en prenant W gaussien. Par contre, dans $\mathfrak{S}_\alpha$, $\alpha < 2$, le problème reste posé.

Enfin si on connaît d'autres moments de X, cela imposera des conditions sur les multispectres de X et la solution du problème 4.5.2 dans $\mathfrak{C}$ ne sera plus un processus gaussien. Pour éclairer cela nous prenons l'exemple simple suivant:

4.5.3. *Trouver le processus X dans $\mathfrak{C}$ dont l'entropie est maximale et les k premières covariances et le moment d'ordre trois sont fixés, le moment d'ordre trois étant non nul.*

La solution est encore un processus autorégressif,

$$X_n = \int_{-\pi}^{\pi} e^{in\lambda}\varphi(\lambda)\,dZ_W, \tag{95}$$

où φ est l'inverse d'un polynôme et W une innovation. Il reste à maximiser $H(W)$. Comme la condition $E(X_0^3) = cste$ peut être satisfaite avec $W = (W_n)_{n\in\mathbf{Z}}$, où les W_n sont indépendants, on choisit une innovation de ce type. Un calcul simple donne

$$\frac{EW_0^3}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\varphi(-\lambda_1-\lambda_2)\,\varphi(\lambda_1)\,\varphi(\lambda_2)\,d\lambda_1\,d\lambda_2 = EX_0^3 \tag{96}$$

et, par conséquent,

$$EW_0^3 = cste. \tag{97}$$

Comme on a aussi $EW_0^2 = 1$ et que $H(W) = h(W_0)$, le maximum de $H(W)$ est obtenu en prenant $W = (W_n)_{n \in \mathbf{Z}}$ avec des W_n indépendants et de densité de probabilité [7]

$$g(w) = c \exp\{a_1 w + a_2 w^2 + a_3 w^3\}, \tag{98}$$

où c, a_1, a_2 et a_3 sont déterminés par

$$\int g(w)\,dw = 1, \quad \int w g(w)\,dw = 0, \quad \int w^2 g(w)\,dw = 1, \quad \int w^3 g(w)\,dw = EW_0^3. \tag{99}$$

4.6. Cas où Y n'a pas de représentation spectrale. Si on ne suppose plus que Y admet une représentation spectrale, on obtient le résultat suivant:

THÉORÈME 4.1. *Soit $Y = (Y_n)_{n \in \mathbf{Z}}$ un processus réel stationnaire en loi vérifiant $E|Y_n| < \infty$. Soit $X = (X_n)_{n \in \mathbf{Z}}$ le processus défini par la série convergente en moyenne*:

$$X_n = \sum_{-\infty}^{+\infty} a_j Y_{n-j}, \quad \text{où} \sum_{-\infty}^{+\infty} |a_j| < \infty. \tag{100}$$

Alors

$$H(X) \geqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Log}|\varphi(\lambda)|\,d\lambda + H(Y), \quad \text{où } \varphi(\lambda) = \sum_{-\infty}^{+\infty} a_j e^{ij\lambda}. \tag{101}$$

De plus, on a l'égalité dans (101), *si $\varphi(\lambda)$ n'a pas de zéros sur le tore $T = [-\pi, \pi]$.*

Preuve. Quand

$$\int_{-\pi}^{\pi} \operatorname{Log}|\varphi(\lambda)|\,d\lambda = -\infty,$$

le théorème est évident. Sinon, $\varphi(\lambda)$ étant une fonction continue, on l'approxime alors uniformément par des polynômes trigonométriques $\varphi_k(\lambda)$ (remarque 3.1 du Lemme 3.2) et (101) s'obtient alors par passage à la limite comme dans le théorème 3.1. Si $\varphi(\lambda) \neq 0$ sur T, comme $\sum |a_j| < \infty$, on a, d'après le théorème de Wiener [12],

$$\frac{1}{\varphi(\lambda)} = \sum_{-\infty}^{+\infty} b_j e^{ij\lambda}, \quad \text{où} \sum_{-\infty}^{+\infty} |b_j| < \infty. \tag{102}$$

Il est alors facile de voir que:

$$Y_n = \sum_{j=-\infty}^{+\infty} b_j X_{n-j} \tag{103}$$

et l'application une nouvelle fois de (101) donne l'égalité voulue.

REFERENCES

[1] R. B. Ash, *Information theory*, Interscience Publishers, 1967.

[2] J. P. Burg, *Maximum entropy spectral analysis*, [in:] *Modern spectrum analysis*, ed. D. G. Childers, Wiley, New York 1978.

[3] — *A new analysis technique for time series data*, ibidem.

[4] S. Cambanis, *Complex symmetric stable variables and processes*, [in:] *Contribution to Statistics*, Essays in Honour of N. L. Johnson, p. 63–79, North-Holland, New York 1982.

[5] S. Cambanis and G. Miller, *Linear problems in p-th order and stable processes*, SIAM J. Appl. Math. 41 (1981), p. 43–69.

[6] S. Cambanis and A. R. Soltani, *Prediction of stable processes: spectral and moving average representations*, Z. Wahr. verw. Gebiete 66 (1984), p. 593–612.

[7] I. Csiszar, *I-divergence geometry of probability distributions and minimization problems*, Ann. Prob. 3.1 (1975), p. 146–158.

[8] Y. Hosoya, *Harmonizable stable processes*, Z. Wahr. verw. Gebiete 60 (1982), p. 517–533.

[9] M. Kanter, *Lower bound for non linear prediction error in moving average processes*, Ann. Prob. 7 (1979), p. 128–138.

[10] M. Pinsker, *Information and information stability of random variables and processes*, Holden Day Series, 1964.

[11] Y. Rozanov, *Stationary random processes*, ibidem, 1967.

[12] W. Rudin, *Real and complex analysis*, McGraw Hill, 1974.

[13] M. Schilder, *Some structure theorems for the symmetric stable laws*, Ann. Math. Stat. 41 (1970), p. 412–421.

[14] C. E. Shannon, *The mathematical theory of communications*, Bell System Technical Journal, 1948.

[15] L. A. Shepp, D. Slepian and A. D. Wyner, *On prediction of moving average processes*, ibidem 59.3 (1980), p. 367–415.

CNRS 743 — Statistique Appliquée
Mathématiques, Bât. 425
Université de Paris Sud
91 405 ORSAY
FRANCE

Received on 2. 2. 1988