ANNALES DE L'I. H. P., SECTION B

JACQUES-ÉDOUARD DIES

Information et complexité

Annales de l'I. H. P., section B, tome 12, n° 4 (1976), p. 365-390

<http://www.numdam.org/item?id=AIHPB_1976__12_4_365_0>

Ann. Inst. Henri Poincaré, Section B :
Vol. XII, n° 4, 1976, p. 365-390. *Calcul des Probabilités et Statistique.*

Information et complexité

par

Jacques-Édouard DIES
4, rue des Violettes Beauzelle, 31700 Blagnac

RÉSUMÉ. — Étant donnée une mesure de probabilité sous-calculable sur l'ensemble des objets binaires finis, on peut associer, à chacun de ces objets, une information hartleyenne entière ; d'autre part, étant donné un algorithme d'un certain type, on peut associer, par la méthode de Kolmogorov, une complexité à chaque objet fini. Il est naturel d'étudier les liens existant entre ces deux mesures de l'information.

Nous introduisons une famille spéciale d'algorithmes (décodeurs) étroitement liée à la théorie classique du codage sans bruit ; nous donnons une évaluation numérique de la complexité par rapport à ces algorithmes en démontrant que chaque décodeur peut être considéré comme le « totalisé » d'une machine de Chaitin.

Les deux principaux résultats que nous obtenons sont les suivants. D'abord, toute information hartleyenne relative à une mesure de probabilité sous-calculable est exprimable en termes d'une mesure de complexité relative à un décodeur et réciproquement. Ensuite, l'étroitesse de l'écart entre ces deux mesures de l'information permet de déterminer le caractère aléatoire d'une suite infinie. Ce dernier résultat met en évidence l'importance des concepts d'information dans la définition algorithmique de la notion de hasard.

ABSTRACT. — Given a subcalculable probability measure over the set of finite binary objects, an integer hartleyan information can be associated with each of these objects ; on the other hand, given an algorithm of a certain type, a complexity can be associated with each finite object using Kolmo-

gorov's method. It is natural to study the links between these two definitions of information.

We introduce a special family of algorithms (decoders) tightly connected with the classical theory of noiseless coding; we give a precise numerical estimate of the complexity relative to those algorithms, prooving that each decoder can be considered as the « totalized » of a Chaitin's machine.

The two main results we get are the following: first, any hartleyan information relative to a subcalculable probability measure can be expressed in terms of a complexity measure relative to a decoder and *vice versa.*

Then, the uniform proximity of these two measures of information allows us to determine the random character of an infinite sequence. This last result underlines the importance of the concepts of information in order to define algorithmically the notion of randomness.

---

1. INTRODUCTION ET NOTATIONS

Connaissant un objet fini, par exemple une suite finie de 0 et de 1, on cherche à définir l'information contenue dans cet objet en associant à tout objet x, un nombre *entier*, son « information ». Deux voies se présentent à nous.

D'abord, la méthode « classique ». Prenons une mesure de probabilité μ sur l'ensemble des suites binaires; en identifiant l'objet x à l'événement élémentaire des suites binaires commençant par x, noté $x\mathrm{X}^{\infty}$, nous pouvons définir l'*information hartleyenne* de x comme étant $-\log \mu(x\mathrm{X}^{\infty})$, et, puisque nous voulons définir un nombre entier, il est naturel de prendre $[-\log \mu(x\mathrm{X}^{\infty})]$. Si, de plus, nous voulons que cette information soit calculable, dans un certain sens, il faut nous limiter à étudier les familles de mesures de probabilité « calculables ». Toutefois, si l'on prend la famille M_0 de toutes les mesures de probabilités calculables, au sens où l'on définit les réels calculables, cette famille ne possède pas deux propriétés importantes; il nous faut alors « encadrer » M_0 par deux familles M_- et M_+ possédant ces propriétés et qui soient « proches » de M_0; ceci fait l'objet de la partie 2.

Ensuite, la méthode algorithmique de Kolmogorov. Étant donné un algorithme A d'un certain type, on définit la *complexité* d'un objet fini x par rapport à cet algorithme A par

$$\mathrm{K}_{\mathrm{A}}(x) = \begin{cases} \min \{ \mid p \mid : \mathrm{A}(p) = x \} & \text{si un tel } p \text{ existe.} \\ \infty & \text{sinon.} \end{cases}$$

Nous introduirons, pour notre part, une famille D d'algorithmes, appelés *décodeurs* car ils sont étroitement liés à la théorie classique du codage sans bruit.

On a donc défini la complexité d'un objet fini x et cette complexité (relative à un algorithme universel) vérifie, à un terme négligeable près, toutes les propriétés usuelles de l'information (nous ne les démontrons pas ici afin de limiter l'exposé et parce que les preuves sont semblables à celles de Kolmogorov).

A ce niveau, puisque nous disposons de deux mesures de l'information, il est naturel d'étudier leurs liens éventuels ; en fait, et ceci est montré dans la partie 3, il existe une correspondance entre M_+, ensemble des mesures de probabilité « calculables » et D, l'ensemble des décodeurs mais, en outre, on arrive à montrer que l'information hartleyenne est exprimable dans les termes d'une mesure de complexité et réciproquement. Ce type de résultat, atteint par d'autres techniques par Willis (1970) mais uniquement au niveau M_-, est donc élargi aux niveaux théoriquement intéressants M_0 et M_+.

De plus, et c'est ce que montre la partie 4, l'étroitesse de l'écart entre ces deux mesures de l'information permet de déterminer le caractère aléatoire d'une suite infinie. Ce dernier résultat met en évidence l'importance des concepts d'information pour approcher la notion de hasard.

NOTATIONS. — N désigne l'ensemble des entiers positifs.

$X = \{ 0, 1 \}$.
X^* désigne l'ensemble des suites binaires finies.
X^∞ désigne l'ensemble des suites binaires infinies.
$X^n = \{ x \in X^* : |x| = n \}$.

$|x|$ désigne la longueur de $x \in X^*$.
$x \sqsubset y$: x est un préfixe de $y \in X^* \cup X^\infty$.
Si $\omega \in X^\infty$, ω_n désigne les n premières lettres de ω.
Λ désigne le mot vide.
$x = 0^\alpha 1^\beta 0^\gamma \ldots$ Par exemple $x = 000110100 = 0^3 1^2 0 1 0^2$.
xy désigne la concaténation de $x \in X^*$ et de $y \in X^* \cup X^\infty$; par exemple, si $x = x_1 x_2 \ldots x_n$ et $y = y_1 y_2 \ldots y_p$, $xy = x_1 x_2 \ldots x_n y_1 y_2 \ldots y_p$.
Si $A \subset X^*$ et $B \subset X^* \cup X^\infty$, on pose

$$AB = \{ xy : x \in A, y \in B \}.$$

∂E désigne le cardinal de l'ensemble fini E.
$\bar{F}(E_1, E_2)$ désigne l'ensemble des fonctions récursives générales de E_1 dans E_2.

$F(E_1, E_2)$ désigne l'ensemble des fonctions récursives partielles de E_1 dans E_2.

Si $f \in F$, Δf désigne le domaine de définition de f.

Si f et g sont deux fonctions à valeurs entières, on note

$$f \preccurlyeq g \quad \text{si} \quad \exists C \quad \forall x : f(x) \leqslant g(x) + C .$$
$$f \asymp g \quad \text{si} \quad f \preccurlyeq g \quad \text{et} \quad g \preccurlyeq f .$$

Soit $\omega \in X^\infty$; on note

$$f(\omega) \preccurlyeq g(\omega) \quad \text{si} \quad \exists C \quad \forall n : f(\omega_n) \leqslant g(\omega_n) + C .$$

$[x]$ désigne la partie entière du réel x.

λ désigne la mesure de Lebesgue sur X ; autrement dit

$$\forall x \in X^n \qquad \lambda(xX^\infty) = 2^{-n} .$$

2. MESURES DE PROBABILITÉ SUR X^∞

D'après le théorème d'extension de Kolmogorov, si μ est une mesure de probabilité sur X^∞, muni de la tribu naturelle de ses sous-ensembles, elle est caractérisée par sa valeur sur les événements élémentaires. Autrement dit, il suffit de connaître

$$\mu(xX^\infty) \qquad (x \in X^*)$$

et d'avoir

$$\sum_{x \in X^n} \mu(xX^\infty) = 1 \qquad \text{(pour tout } n \in N) .$$

Définissons quelques notions attachées à ces mesures de probabilité :

a) Comme il a été dit dans l'introduction, si μ est une mesure de probabilité sur X^∞, à tout objet fini $x \in X^*$ on peut associer un nombre entier $J_\mu(x)$, appelé *information hartleyenne* (entière) de x et défini par :

$$J_\mu(x) \stackrel{\Delta}{=} [-\log \mu(xX^\infty)] .$$

b) Soit μ une mesure de probabilité sur X^∞.

On appelle *test* un ensemble récursivement énumérable

$$Y \subset N \times X^*$$

tel que

$$\mu(Y_i X^\infty) \leqslant 2^{-i} .$$

où

$$Y_i = \{ x : (i, x) \in Y \} .$$

A tout test Y on peut associer l'ensemble μ-négligeable

$$o(Y) = \bigcap_{i \in N} Y_i X^\infty .$$

Une suite $\omega \in X^\infty$ est dite *μ-aléatoire* (au sens de Martin-Löf (1966)) si, quel que soit le test Y,

$$\omega \notin o(Y) .$$

Nous désignerons par X_μ^∞ l'ensemble des suites μ-aléatoires.

Comme on le sait, une suite est aléatoire au sens de Martin-Löf si elle vérifie toutes les lois *effectives* de probabilité (loi des grands nombres, théorème limite central, ...), mais on n'envisage pas formellement d'autres lois ; autrement dit, on peut s'attendre à ce que deux mesures de probabilité suffisamment « proches » soient « indiscernables » du point de vue des suites aléatoires par rapport à ces mesures ; nous poserons donc

c) Définition. — Deux mesures de probabilité μ et μ' sont équivalentes ($\mu \sim \mu'$) si $J_\mu \asymp J_{\mu'}$.

PROPOSITION. — $\mu \sim \mu' \Leftrightarrow \exists C \quad \forall x : 2^{-C} \leqslant \mu'(xX^\infty)/\mu(xX^\infty) \leqslant 2^C$.

Exemple. — Soit λ la mesure de Lebesgue et définissons μ de la façon suivante :

$$\mu(xX^\infty) = \begin{cases} 2^{-n} & x \in X^n - \{ 0^n, 1^n \} . \\ 2^{-n-1} & x = 0^n . \\ 2^{-n} + 2^{-n-1} & x = 1^n . \end{cases}$$

Donc,

$$\forall n \in N \qquad \mu(0^n X^\infty)/\lambda(0^n X^\infty) = 1/2$$
$$\mu(1^n X^\infty)/\lambda(1^n X^\infty) = 3/2 .$$

Donc

$$\forall x \in X^* \qquad 2^{-1} \leqslant \mu(xX^\infty)/\lambda(xX^\infty) \leqslant 2,.$$

Donc

$$\mu \sim \lambda .$$

THÉORÈME 2.1. — $\mu \sim \nu \Rightarrow X_\mu^\infty = X_\nu^\infty$.

Démonstration. —

$$\mu \sim \nu \Rightarrow \exists C \quad \forall x \in X^* \qquad \nu(xX^\infty) \leqslant 2^C \mu(xX^\infty) .$$

Soit $\omega \notin X_\mu^\infty$; alors, par définition, il existe une famille récursivement énumérable $\{ Y_i \}$, telle que

$$\mu(Y_i X^\infty) \leqslant 2^{-i} \qquad \text{et} \qquad \omega \in o(Y) .$$

Alors

$$\nu(Y_i X^\infty) \leqslant 2^C \mu(Y_i X^\infty) \leqslant 2^{C-i} .$$

Soit

$$Y'_i = Y_{i+C},$$

Alors $Y' = \{ Y'_i \}$ est récursivement énumérable avec

$$\nu(Y'_i X^\infty) \leqslant 2^{-i} \qquad \text{et} \qquad \omega \in o(Y')$$

Donc

$$\omega \notin X_\nu^\infty .$$

En échangeant les rôles de μ et ν, on obtient le résultat. □

d) Nous allons introduire, avec Zvonkin-Levin (1970), une dernière notion, celle de *mesure universelle* dans une famille de mesures de probabilité.

Soit donc M une famille de mesures de probabilité sur X^∞. Nous dirons que $\mu_\infty \in M$ est une mesure universelle dans M si :

$$\forall \mu \in M \quad \exists C : \mu(xX^\infty) \leqslant C . \mu_\infty(xX^\infty) .$$

Autrement dit, μ est absolument continue par rapport à μ_∞ et la dérivée de Radon-Nikodym est majorée par C.

En statistiques, le problème suivant se pose : étant donnée une suite $\omega \in X^\infty$, déterminer par rapport à quelle mesure de M, ω est aléatoire ; l'assertion la plus faible que l'on puisse faire est de supposer que ω est μ_∞-aléatoire ; autrement dit, μ_∞ peut être considérée comme une mesure *a priori*.

Formellement, en utilisant la démonstration du théorème 2.1, on a

$$\omega \notin X_{\mu_\infty}^\infty \Rightarrow \omega \notin X_\mu^\infty .$$

Donc, si ω est μ-aléatoire ($\mu \in M$), ω est nécessairement μ_∞-aléatoire.

Nous avons attribué, au point *a*), une information $J_\mu(x)$ à l'objet $x \in X^*$. Nous désirons, en outre, que cette correspondance se fasse de manière effective. Ceci nous oblige à considérer des familles de mesures de probabilité « calculables ».

Considérons donc M_0, la famille des mesures *calculables* ainsi définies ; pour tout $x \in X^*$, $\mu(xX^\infty)$ sera un réel calculable (au sens usuel, e. g. Schoenfield (1967)).

Autrement dit nous avons

Définition. —

$$\mu \in M_0 \Leftrightarrow \begin{cases} \exists \varphi \in F(N \times X^*, X) & \text{tel que} \\ \mu(xX^\infty) = \sum_{i \in N} \varphi(i, x) 2^{-i} . \end{cases}$$

Cette définition est la plus naturelle ; il faudrait que M_0 possède deux propriétés essentielles :

— d'abord, il faudrait que $\forall \mu \in M_0$, $J_\mu \in \bar{F}(X^*, N)$,

— ensuite, que M_0 possède une mesure universelle.

Or, la famille « naturelle » M_0 ne possède aucune de ces propriétés ; nous sommes donc conduits à encadrer M_0 par deux familles possédant chacune une des propriétés indiquées et qui seront proches, au sens de *c*).

Définition. —

$$\mu \in M_+ \Rightarrow \begin{cases} \exists \varphi \in \bar{F}(N \times X^*, N) & \text{tel que} \\ \forall i \in N, \quad \forall x \in X^* & \varphi(i, x) \in [0, 2^i] \quad \text{et} \\ \mu(xX^\infty) = \sum_{i \in N} \varphi(i, x) 2^{-i}. \end{cases}$$

Nous avons

$$M_0 \subset M_+$$

et nous avons le résultat suivant, démontré dans la 3e partie :

THÉORÈME 3.6. — M_+ possède une mesure universelle μ_∞.

Remarque. — On peut démontrer que M_+ coïncide avec l'ensemble des mesures « sous-calculables » définies, de façon différente, par Zvonkin-Levin (1970).

D'autre part nous posons

Définition (Willis (1970))

$$\mu \in M_- \Leftrightarrow \begin{cases} \exists \varphi \in \bar{F}(N \times X^*, N) & \\ \exists \psi \in \bar{F}(X^*, N) & \text{tels que} \\ \mu(xX^\infty) = \sum_{i=1}^{\psi(|x|)} \varphi(i, x) 2^{-i}. \end{cases}$$

Il est clair que $M_- \subset M_0$

et, en outre, nous avons la proposition suivante :

PROPOSITION 2.2. — $\forall \mu \in M_- \quad J_\mu \in \bar{F}(X^*, N)$.

Démonstration. — Soit $\mu \in M_-$; alors

$$J_\mu(x) = [-\log \mu(xX^\infty)].$$

Soient φ et ψ les fonctions récursives attachées à μ. Alors

$$J_\mu(x) = \min \{ |p| : |p| \leqslant \psi(|x|) \quad \text{et} \quad \varphi(|p|, x) = 1 \}.$$

Donc

$$J_\mu \in \bar{F}(X^*, N). \qquad \square$$

Nous avons donc trouvé deux familles encadrant M_0 :

$$M_- \subset M_0 \subset M_+ .$$

Il nous reste à prouver que ces trois familles sont proches ; d'une part, nous montrerons dans les théorèmes 3.2 et 3.3 que toute mesure de M_+ est l'image de la mesure de Lebesgue par un algorithme. D'autre part, nous avons le

THÉORÈME 2.3. —

$$\forall \mu_0 \in M_0 \quad \exists \mu_- \in M_- : \mu_0 \sim \mu_- .$$

Démonstration. — Soient

$$\mu_0 \in M_0 \qquad \text{et} \qquad m \in N .$$

Puisque

$$\sum_{x \in X^m} \mu_0(xX^\infty) = 1 ,$$

$$\exists \xi \in X^m \qquad \text{t. q.} \qquad \mu_0(\xi X^\prime) \geqslant 2^{-m} .$$

Soit $x \in X^m - \xi$, et

$$\mu_0(xX^\infty) = 2^{-i_0} + \ldots + \alpha 2^{-i_0-m} + \ldots$$

et soit

$$\mu_-(xX^\infty) \overset{\Delta}{=} 2^{-i_0} + \ldots + 2^{-i_0-m} .$$

Alors

$$\mu_0(xX^\infty) = \mu_-(xX^\infty) + r(xX^\infty) .$$

Or

$$r(xX^\infty) \leqslant \sum_{k \in N} 2^{-i_0-m-k} = 2^{-i_0-m}$$

Mais $2^{-i_0} \leqslant \mu_0(xX^\prime)$; donc

$$0 \leqslant r(xX^\infty) \leqslant 2^{-m} \mu_0(xX^\prime) .$$

Par conséquent nous avons :

(1) $\qquad \forall x \in X^m - \xi, \qquad (1 - 2^{-m}) \mu_0(xX^\infty) \leqslant \mu_-(xX^\infty) \leqslant \mu_0(xX^\infty) .$

Posons maintenant :

$$\mu_-(\xi X^\infty) \overset{\Delta}{=} 1 - \sum_{x \in X^m - \xi} \mu_-(xX^\infty) .$$

En utilisant (1) :

$$\mu_0(\xi X^\prime) \leqslant \mu_-(\xi X^\infty) \leqslant 1 - (1 - 2^{-m}) \sum_{x \in X^m - \xi} \mu_0(xX^\infty) .$$

Or, $1 - 2^{-m} \geqslant 1 - \dfrac{1}{2^m - 1}$, d'où

$$\mu_-(\xi X^\infty) \leqslant (1 - 1/2^m - 1)\mu_0(\xi X^\infty) + 1/(2^m - 1).$$

Par hypothèse,

$$\mu_0(\xi X^\infty) \geqslant 2^{-m},$$

Or

$$2^{-m} = \frac{1 - (1 - 1/2^m - 1)}{2 - (1 - 1/2^m - 1)}$$

Donc

$$2 - \left(1 - \frac{1}{2^m - 1}\right)\mu_0(\xi X^\infty) \geqslant 1/2^m - 1.$$

Soit

$$\left(1 - \frac{1}{2^m - 1}\right)\mu_0(\xi X^\infty) + \frac{1}{2^m - 1} \leqslant 2\mu_0(\xi X^\infty).$$

Finalement nous avons :

$$\mu_0(\xi X^\infty) \leqslant \mu_-(\xi X^\infty) \leqslant 2\mu_0(\xi X^\infty). \tag{2}$$

En rassemblant (1) et (2), il vient :

$$\forall m \in N \quad \forall x \in X^m \qquad 2^{-1}\mu_0(xX^\infty) \leqslant \mu_-(xX^\infty) \leqslant 2\mu_0(xX^\infty).$$

Par conséquent

$$\mu_0 \sim \mu_-.$$

Mais par construction $\mu_- \in M_-$. □

3. INFORMATION ET COMPLEXITÉ

3.1. Approche algorithmique du codage sans bruit

Soit $\omega \in X^\infty$; si nous nous représentons ω comme un message débité par une source d'information d'alphabet X, nous considérons que ω reflète l'ensemble des propriétés statistiques de la source. Si la source est indépendante, il suffit d'examiner les lettres une à une, puis de déterminer les fréquences de 0 et de 1, si la source est markovienne d'ordre 1, il faut en outre examiner les couples de lettres, ... Dans le cas général nous sommes amenés à décomposer $\omega \in X^\infty$ en fragments d'ordre n, pour tout n entier.

Examinons donc un large fragment ω_N de ω. Pour transmettre ce fragment dans un canal sans bruit il faut coder les sous-fragments d'ordre n, $n \leqslant N$. On adopte en général des *codes instantanés* ainsi définis.

Définition. — $S \subset X^*$ est un code instantané si

$$\forall p_1, p_2 \in S \qquad p_1 \neq p_2 \Rightarrow p_1 X^* \cap p_2 X^* = \emptyset .$$

Nous savons (Ash, 1965) que la condition nécessaire et suffisante d'existence de codes instantanés est donnée par l'inégalité de Kraft : S code instantané $\Leftrightarrow \sum_{p \in S} 2^{-|p|} \leqslant 1$.

Introduisons la définition suivante,

Définition. — Nous dirons qu'un code instantané S est *total* si

$$\sum_{p \in S} 2^{-|p|} = 1 .$$

Notons qu'un code instantané est total si, en particulier,

$$\lambda\text{-p. s.} \qquad \sum_{p \in S} p X^\infty = X^\infty .$$

Ce codage étant choisi, il faut ensuite se donner un décodeur f_n permettant de recouvrer les fragments codés de longueur n. Autrement dit, le codage sans bruit pourra se formuler à l'aide d'une famille de décodeurs comme suit.

Définition. — D désigne l'ensemble des *décodeurs*, et

$$f \in D \Leftrightarrow \begin{cases} f \in F(X^* \times N, X^*) \\ \forall n \in N \quad \Delta f(., n) \quad \text{est un code instantané total.} \\ \forall n \in N \quad \forall x \in \Delta f(., n) \quad f(x, n) \in X^n . \end{cases}$$

La théorie classique du codage sans bruit exige que les codes soient « efficaces », i. e. de longueur minimale ; ceci nous conduit à adopter la définition de la complexité de Kolmogorov (1965) ; cf. Introduction.

Définition. — Soit $f \in D$ et $x \in X^*$; on appelle *complexité* de x relativement à f le nombre

$$K_f(x) = \begin{cases} \min \{ |p| : f(p, |x|) = x \} & \text{si un tel } p \text{ existe} \\ \infty & \text{sinon.} \end{cases}$$

Dans le cadre de la théorie classique, Huffman (1952) a introduit la notion de code optimal ; la notion algorithmique équivalente est celle d'algorithme optimal dans une famille d'algorithmes.

Définition. — Soit F une famille d'algorithmes ; un algorithme f_∞ est dit *optimal* si :

$$\forall f \in \mathrm{F} \quad \exists \mathrm{C} \qquad \forall p \in \Delta f \quad \exists p' \in \Delta f_\infty$$

tel que

$$f_\infty(p') = f(p) \qquad \text{avec} \qquad |\, p' \,| \leqslant |\, p \,| + \mathrm{C}\,.$$

Nous avons le résultat suivant

THÉORÈME 3.1. — D possède un décodeur optimal f_∞.

Démonstration. — Soit $\{ f_\alpha \}$ une énumération effective de D.

Si $f = f_\alpha$, on pose $\alpha = \nu(f)$.

Définissons $f_\infty(x, n)$ de la manière suivante : si

$$x = 0^\alpha 1 y\,,$$

on pose

$$f_\infty(x, n) = f_\infty(0^\alpha 1 y, n) \stackrel{\Delta}{=} f_\alpha(y, n)\,.$$

Pour montrer que $f_\infty \in \mathrm{D}$, il faut surtout montrer que

$$\Delta f_\infty(., n) \qquad \text{est un code instantané total}\,.$$

Soient donc $x_1, x_2 \in \Delta f_\infty(., n)$ avec $x_1 \neq x_2$. Par définition de f_∞ :

$$\begin{array}{llll} \exists\alpha & \text{t. q.} \quad x_1 = 0^\alpha 1 y_1 & \text{avec} & y_1 \in \Delta f_\alpha \\ \exists\beta & \text{t. q.} \quad x_2 = 0^\beta 1 y_2 & \text{avec} & y_2 \in \Delta f_\beta \end{array}$$

Prenons

$$z \in x_1 \mathrm{X}^* \cap x_2 \mathrm{X}^*$$

alors

$$0^\alpha 1 y_1 \sqsubset z \qquad \text{et} \qquad 0^\beta 1 y_2 \sqsubset z$$

donc

$$\alpha = \beta\,.$$

Par conséquent, y_1 et y_2 appartiennent à Δf_α et, de plus, $y_1 \neq y_2$ puisque $x_1 \neq x_2$.

Mais $f_\alpha \in \mathrm{D}$ donc

$$y_1 \mathrm{X}^* \cap y_2 \mathrm{X}^* = \emptyset$$

d'où

$$x_1 \mathrm{X}^* \cap x_2 \mathrm{X}^* = \emptyset\,.$$

Il faut encore montrer que $\mathrm{S} = \Delta f_\infty(., n)$ est total.

Par définition de f_∞,

$$\forall p_i \in \mathrm{S} \qquad p_i = 0^\alpha 1 q_i \qquad \text{avec} \qquad q_i \in \Delta f_\alpha(., n)\,.$$

Donc

$$\sum_{p_i \in \mathrm{S}} p_i \mathrm{X}^\infty = \sum_{\substack{\alpha \\ q_i \in \Delta f_\alpha}} 0^\alpha 1 q_i \mathrm{X}^\infty = \sum_{\alpha} 0^\alpha 1 \sum_{q_i \in \Delta f_\alpha} q_i \mathrm{X}^\infty$$

$$= \sum_{\alpha} 0^\alpha 1 \mathrm{X}^\infty$$

$$= 0\mathrm{X}^\infty + 1\mathrm{X}^\infty = \mathrm{X}^\infty \qquad \lambda\text{-p. s.}$$

Soit maintenant

$$f \in \mathrm{D} \qquad \text{et} \qquad \alpha = v(f);$$

si $p \in \Delta f(., n)$, alors par définition

$$p' = 0^{\alpha}1p \in \Delta f_{\infty}(., n)$$

et de plus

$$| p' | = | p | + 1 + \alpha .$$

f_{∞} est donc un décodeur optimal. □

Notation. — Nous poserons

$$\mathrm{K}(x) \stackrel{\Delta}{=} \mathrm{K}_{f_{\infty}}(x).$$

Nous allons dans la proposition et le corollaire qui suivent indiquer, pour la famille D, des résultats familiers en théorie de la complexité (cf. par exemple Zvonkin-Levin (1970)) qui nous seront utiles par la suite.

PROPOSITION. —

$$\forall f \in \mathrm{D} \qquad \mathrm{K} \leqslant \mathrm{K}_f .$$

Démonstration. — Soit donc $f \in \mathrm{D}$ et $\alpha = v(f)$.
Prenons $p_0(x)$ ainsi défini :

$$p_0(x) \begin{cases} 1) & f(p_0(x), | x |) = x . \\ 2) & | p_0(x) | = \mathrm{K}_f(x) . \end{cases}$$

Alors, par définition de f_{∞} :

$$f_{\infty}(0^{\alpha}1p_0(x), | x |) = x$$

d'où

$$\mathrm{K}(x) \leqslant | 0^{\alpha}1p_0(x) | = \mathrm{K}_f(x) + \alpha + 1 . \qquad \square$$

COROLLAIRE. — 1) $\forall x \in \mathrm{X}^* \quad \mathrm{K}(x) \leqslant | x |$.
2) La proportion des $x \in \mathrm{X}^n$ tels que $\mathrm{K}(x) \leqslant n - m$, est inférieure à 2^{-m+1}.

Démonstration. — 1) Soit $i \in \mathrm{D}$, le décodeur identique ainsi défini :

$$\forall n \in \mathrm{N} \quad \forall x \in \mathrm{X}^n \qquad i(x, n) = x .$$

Nous avons

$$\mathrm{K}_i(x) = \min \{ | p | : i(p, | x |) = x \} = | x | .$$

D'après la proposition, $\mathrm{K}(x) \leqslant | x |$. □
2) Si $x \in \mathrm{X}^n$ et $\mathrm{K}(x) \leqslant n - m$, on peut trouver $p \in \mathrm{X}^*$ tel que

$$f_{\infty}(p, n) = x \qquad \text{et} \qquad | p | = \mathrm{K}(x) \leqslant n - m .$$

Donc

$$\partial \{ p : | p | \leqslant n - m \} = 2^{n-m+1} - 1 \geqslant \partial \{ x \in X^n : K(x) \leqslant n - m \} .$$

Or

$$\partial \{ x \in X^n \} = 2^n ,$$

Donc la proportion cherchée est inférieure à

$$(2^{n-m+1} - 1)/2^n = 2^{-m+1} . \qquad \square$$

Le corollaire signifie que la majorité des $x \in X^n$ a une complexité voisine de n ; mais il faut noter que ceci a été établi lorsqu'on ne sait *rien* des $x \in X^n$.

Considérons par exemple les mots x de la forme 0^n1 ($n \in N$) ; ils sont de longueur $n + 1$, mais ils sont entièrement caractérisés par n, donc la complexité des mots 0^n1 est de l'ordre de $| n |$, i. e. de l'ordre de $\log_2 n$; montrons par exemple le point 1) relatif à ceci (le point 2) se démontrerait de même).

Soit θ la bijection récursive de X* dans X* définie par

$$\theta(n) = 0^n1 .$$

Considérons alors $j \in D$ défini par

$$j(x, | \theta(x) |) = \theta(x) .$$

Alors

$$K_j(0^n1) = \min \{ | p | : j(p, n + 1) = 0^n1 \} = | n | \asymp \log_2 n .$$

D'où, d'après la proposition :

$$K(0^n1) \preccurlyeq \log_2 n . \qquad \square$$

3.2. **Totalisation des machines de Chaitin**

Nous allons examiner les liens existant, du point de vue de la complexité, entre les décodeurs D et la famille D_0 des machines introduites par Chaitin (1975).

Ces dernières sont ainsi définies :

Définition. —

$$f \in D_0 \Leftrightarrow \begin{cases} f \in F(X^*, X^*) \\ \Delta f \qquad \text{est un code instantané} . \end{cases}$$

Notons dès à présent la généralisation par Chaitin, et grâce à la famille D_0, de la condition nécessaire et suffisante d'existence d'un code instantané (condition de Kraft) ; cette généralisation sera pour nous un lemme technique essentiel que nous appelons

LEMME DE CHAITIN-KRAFT. — Soit $U \subset X^* \times N$ un ensemble récursivement énumérable ; si U est *consistant*, autrement dit si

$$\sum_{(x,n)\in U} 2^{-n} \leqslant 1 .$$

Alors

$$\exists f \in D_0 \quad \text{t. q.} \quad \partial \{ p \in X^n : f(p) = x \} = \partial \{ (x_i, n_i) \in U : (x_i, n_i) = (x, n) \} .$$

On peut également définir D_0^2, ensemble des calculatrices de Chaitin à deux paramètres ; D_0^2 possède, comme D, une calculatrice optimale, notée φ_∞.

Nous poserons alors

$$K_0(x \mid y) = \min \{ \mid p \mid : \varphi_\infty(p, y) = x \} ,$$
$$K_0(x) = K_0(x \mid \Lambda) .$$

Nous arrivons à relier D et D_0^2 de façon étroite grâce à un procédé technique que nous appelons *totalisation.*

THÉORÈME DE TOTALISATION. — Pour tout $f \in D_0^2$ on peut effectivement construire $\bar{f} \in D$ tel que $K_{\bar{f}}(x) = K_f(x \mid \mid x \mid)$.

Démonstration. — Soit

$$f(x, m) = f_m(x) \in D_0^2 .$$

Définissons comme suit $\bar{f}(p, m)$: on se donne (p, m) et on pose $l = \mid p \mid$. On définit

et
$$X_l = \{ \xi \in X^* : \mid \xi \mid \leqslant l \}$$
$$Y_m = \{ p \in X_l : f_m(p) \in X^m \} .$$

1) Si $f_m(Y_m) \neq X^m$ et $p \in Y_m$, on pose

$$\bar{f}(p, m) \overset{\Delta}{=} f_m(p) .$$

2) Si $f_m(Y_m) = X^m$, on peut alors déterminer effectivement

$$M = \max_{x \in X^m} K_{f_m}(x) .$$

a) Si $M = l$ et $p \in Y_m$, on pose

$$\bar{f}(p, m) \overset{\Delta}{=} f_m(p) .$$

b) Soit $\bar{Y}_m = \{ \xi \in X^M : \exists p \in Y_m, p \sqsubset \xi \}$.

Si $M = l$ et $p \in X^M - \bar{Y}_m$, on pose

$$\bar{f}(p, m) \stackrel{\Delta}{=} 0^m .$$

3) En dehors de ces cas, $\bar{f}$ n'est pas définie. Par construction,

$\bar{f}(p, m)$ est récursive partielle et $\operatorname{Im} \bar{f}(., m) \subset X^m$.

Considérons M précédemment défini ; alors, par construction

$$p \in \Delta\bar{f}(., m) \Leftrightarrow \begin{cases} |p| \leqslant M & \text{et} \quad p \in \Delta f(., m) \\ & \text{ou} \\ |p| = M & \text{et} \quad p \in X^M - \bar{Y}_m . \end{cases}$$

Alors il est clair que $\Delta\bar{f}(., m)$ est un code instantané. De plus,

$$\sum_{p \in \Delta f(., m)} pX^\infty = X^M X^\infty = X^\infty$$

Donc

$$\bar{f} \in D .$$

$\bar{f} \in D$ est appelée la *totalisée* de $f \in D_0^2$.

En outre, la construction implique

$$(*) \qquad K_{\bar{f}}(x) = K_f(x \,|\, |x|) . \qquad \square$$

COROLLAIRE. —

$$K_0(x \,|\, |x|) \asymp K(x) .$$

Démonstration. — Soit φ_∞ une machine optimale de D_0^2 et $\overline{\varphi_\infty}$ sa totalisée. D'après (*),

$$K_{\overline{\varphi_\infty}}(x) = K_0(x \,|\, |x|)$$

Or

$$K(x) \leqslant K_{\overline{\varphi_\infty}}(x)$$

Donc

$$K(x) \leqslant K_0(x \,|\, |x|) .$$

D'autre part, soit f_∞ un décodeur optimal de $D \subset D_0^2$; alors

$$\begin{aligned} K(x) &= \min \{ |p| : f_\infty(p, |x|) = x \} \\ &= K_{f_\infty \in D_0^2}(x \,|\, |x|) \geqslant K_0(x \,|\, |x|) . \end{aligned}$$

Remarque. — Lorsque nous considérons des fragments *de longueur spécifiée* d'une suite $\omega \in X^\infty$, on peut, dans une certaine mesure, nous intéresser à $K_0(x)$ plutôt qu'à $K(x)$; c'est ce que nous ferons dans la 4e partie.

3.3. **Mesures de probabilité associées aux décodeurs**

Il existe un lien complet, déterminé par les deux théorèmes suivants, entre la famille M_+ des mesures de probabilité « sous-calculables » et la famille D des décodeurs.

THÉORÈME 3.2. — A partir de tout décodeur $f \in D$ on peut construire une mesure de probabilité, notée μ_f et appartenant à M_+.

Démonstration. — Soit $f \in D$ et $x \in X^n$.

Posons

$$\mu_f(xX^\infty) \stackrel{\Delta}{=} \sum_{p:f(p,n)=x} 2^{-|p|}.$$

Si on pose

$$f^{-1}x \stackrel{\Delta}{=} \{ p : f(p, n) = x \}$$

et λ désignant la mesure de Lebesgue, nous avons

$$\mu_f(xX^\infty) = \lambda \circ f^{-1}x ;$$

De plus, $\Delta f(., n)$ étant total,

$$\sum_{x \in X^n} \mu_f(xX^\infty) = 1 .$$

Donc μ_f est une mesure de probabilité.

En outre, si

$$\varphi(i, x) \stackrel{\Delta}{=} \partial \{ p \in X^i : f(p, n) = x \}$$

on a

$$\begin{cases} \varphi \in \bar{F}(N \times X^*, N) \\ \varphi(i, x) \in [0, 2^i] . \end{cases}$$

Or

$$\mu_f(xX^\infty) = \sum_{i \in N} \varphi(i, x) 2^{-i}$$

Donc

$$\mu_f \in M_+ . \qquad \square$$

Nous avons un théorème réciproque, associant un décodeur à chaque mesure de probabilité de M_+.

THÉORÈME 3.3. —

$$\forall \mu \in M_+ \quad \exists f \in D \qquad \text{t. q.} \qquad \mu = \mu_f .$$

Démonstration. — Soit $\mu \in M_+$; alors, par définition

$$\exists \varphi \in \bar{F}(N \times X^*, N)$$

tel que

$$\begin{cases} \varphi(i, x) \in [0, 2^i] \\ \mu(xX^\infty) = \sum_{i \in N} \varphi(i, x) 2^{-i} . \end{cases}$$

Fixons $m \in \mathrm{N}$ et définissons

$$\mathrm{U}_m = \{ (x_i, n_i) \in \mathrm{X}^m \times \mathrm{N} \}$$

tel que

$$(1) \qquad \begin{cases} (x, n) \in \mathrm{U}_m & \text{si} \quad \varphi(n, x) \neq 0 \\ \partial \{ (x_i, n_i) \in \mathrm{U}_m : (x_i, n_i) = (x, n) \} = \varphi(n, x) . \end{cases}$$

Alors il est clair que U_m est récursivement énumérable et, de plus, U_m est consistant. En effet

$$(2) \qquad \sum_{(x,n)\in \mathrm{U}_m} 2^{-n} = \sum_{x\in \mathrm{X}^m} \sum_{n} \varphi(n, x) 2^{-n} = \sum_{x\in \mathrm{X}^m} \mu(x\mathrm{X}^\infty) = 1 .$$

En appliquant le lemme de Chaitin-Kraft, il vient

$$\exists f_m \in \mathrm{D}_0$$

tel que

$$\partial \{ (x_i, n_i) \in \mathrm{U}_m : (x_i, n_i) = (x, n) \} = \partial \{ p \in \mathrm{X}^n : f_m(p) = x \} .$$

Donc, d'après (1),

$$\partial \{ p \in \mathrm{X}^n : f_m(p) = x \} = \varphi(n, x) .$$

Par conséquent, puisque d'après (2), $f_m^{-1}(\mathrm{X}^m)$ est total,

$$\begin{aligned} \forall x \in \mathrm{X}^m \qquad \mu(x\mathrm{X}^\infty) &= \sum_{n\in \mathrm{N}} \varphi(n, x) 2^{-n} \\ &= \sum_{n\in \mathrm{N}} 2^{-n} . \partial \{ p \in \mathrm{X}^n : f_m(p) = x \} \\ &= \sum_{p : f_m(p) = x} 2^{-|p|} = \mu_{f_m}(x\mathrm{X}^\infty) . \end{aligned}$$

Soit alors f définie par :

$$f(x, m) \overset{\Delta}{=} f_m(x) .$$

Autrement dit, soit $(x, m) \in \mathrm{X}^* \times \mathrm{N}$; on calcule $f_m(x)$; si $f_m(x)$ existe et si $| f_m(x) | = m$, alors f est définie en (x, m) par $f(x, m) = f_m(x)$.

Il est clair, d'après notre construction, que $f \in \mathrm{D}$. Soit maintenant $x \in \mathrm{X}^*$; posons $m = | x |$. Alors

$$\begin{aligned} \mu_f(x\mathrm{X}^\infty) &= \sum_{p : f(p,m) = x} 2^{-|p|} \\ &= \sum_{p : f_m(p) = x} 2^{-|p|} \\ &= \mu_{f_m}(x\mathrm{X}^\infty) = \mu(x\mathrm{X}^\infty) . \end{aligned} \qquad \square$$

3.4. **Information hartleyenne et complexité**

Nous venons d'examiner les liens existant entre les décodeurs D et les mesures de probabilité de M_+. Il nous reste à en montrer un rapport plus étroit : à tout décodeur f on peut associer une mesure de complexité K_f et à toute mesure de probabilité de M_+ on peut associer l'information hartleyenne J_μ ; il nous faut montrer que les J_μ et les K_f sont liés ; explicitement, nous allons montrer qu'à un terme constant près, toute information hartleyenne peut être considérée comme une mesure de complexité et réciproquement, dans des limites presque certaines.

THÉORÈME 3.4. —

$$\forall \mu \in M_+ \quad \exists \bar{f} \in D \qquad \text{t. q.} \qquad K_{\bar{f}}(x) \asymp J_\mu(x).$$

Démonstration. — Soit $\mu \in M_+$ et $m \in N$.

Considérons

$$U_m \stackrel{\Delta}{=} \{ (x, n) \in X^m \times N : n \geqslant J_\mu(x) + 2 \}.$$

Comme

$$[n > -\log \mu(xX^\infty)] \Leftrightarrow [\mu(xX^\infty) > 2^{-n}]$$

$$\Leftrightarrow \left[\exists t : \sum_{i=1}^{t} \varphi(i, x) 2^{-i} > 2^{-n} \right]$$

$$\Leftrightarrow [\exists t : (t, x, n) \in R, \ R \text{ récursif}],$$

U_m est récursivement énumérable ; en outre U_m est consistant :

$$\sum_{(x,n)\in U_m} 2^{-n} = \sum_{x \in X^m} 2.2^{-[-\log \mu(xX^\infty)]-2} \leqslant \sum_{x \in X^m} \mu(xX^\infty) = 1.$$

En utilisant le lemme de Chaitin-Kraft, il vient

$$\exists f_m \in D_0$$

tel que

$$\partial \{ p \in X^n : f_m(p) = x \} = \begin{cases} 1 & \text{si} \quad J_\mu(x) + 2 \leqslant n. \\ 0 & \text{si} \quad J_\mu(x) + 2 > n. \end{cases}$$

Soit $f \in D_0^2$ définie par :

$$f(x, m) \stackrel{\Delta}{=} f_m(x).$$

Alors, les résultats précédents nous permettent d'affirmer que

$$K_f(x \mid \mid x \mid) \asymp J_\mu(x). \tag{1}$$

Soit alors $\bar{f} \in D$ la totalisée de f.

D'après (*),

$$K_f(x \mid |x|) \asymp K_{\bar{f}}(x).$$

Donc, d'après (1) et (2),

$$K_{\bar{f}}(x) \asymp J_\mu(x). \qquad \square$$

THÉORÈME 3.5. —

$$\forall \hat{f} \in D \quad \exists \mu \in M_+ \qquad \text{t. q.} \qquad \forall x \in X^* - \{0^n\} \qquad K_{\hat{f}}(x) \asymp J_\mu(x).$$

Démonstration. — Soit $f \in D$; d'après le théorème 3.1, $\mu_{\hat{f}} \in M_+$ et

$$\mu_{\hat{f}}(xX^\infty) \geqslant 2^{-K_{\hat{f}}(x)}. \tag{1}$$

Fixons $m \in N$ et considérons

$$U_m \stackrel{\Delta}{=} \{(x, n) \in X^m \times N : n \geqslant K_{\hat{f}}(x) + 1\}.$$

Alors U_m est récursivement énumérable et consistant car

$$\sum_{(x,n)\in U_m} 2^{-n} = \sum_{x\in X^m} 2.2^{-K_{\hat{f}}(x)-1} \leqslant \sum_{x\in X^m} \mu_{\hat{f}}(xX^\infty) = 1 \qquad \text{(d'après (1))}.$$

On peut donc appliquer le lemme de Chaitin-Kraft, d'où

$$\exists f_m \in D_0$$

tel que

$$\partial\{p \in X^n : f_m(p) = x\} = \partial\{(x_i, n_i) \in U_m : (x_i, n_i) = (x, n)\} \in X.$$

Donc si $f \in D_0^2$ est définie par

$$f(x, m) \stackrel{\Delta}{=} f_m(x)$$

On a

$$K_f(x \mid |x|) = K_{\hat{f}}(x). \tag{2}$$

Soit $\bar{f} \in D$ la totalisée de f.

D'après les propriétés de f et d'après la construction de $\bar{f}$, on voit que

$$\forall x \in X^{|x|} - \{0^{|x|}\} \qquad \mu_{\bar{f}}(xX^\infty) = 2^{-K_f(x\mid |x|)} + \ldots + 2^{-M}$$

où

$$M \stackrel{\Delta}{=} \max_{x\in X^m} K_{f_m}(x) \qquad (m = |x|).$$

Donc

$$2^{-K_f(x\mid |x|)} \leqslant \mu_{\bar{f}}(xX^\infty) \leqslant 2^{-K_f(x\mid |x|)-1}$$

d'où

$$J_{\mu_{\bar{f}}}(x) \asymp K_f(x \mid |x|). \tag{3}$$

En combinant (2) et (3), le théorème est démontré. $\square$

3.5. **Mesure de probabilité universelle dans** M_+

Nous allons maintenant montrer l'existence d'une mesure universelle dans M_+ et évaluer son comportement numérique.

THÉORÈME 3.6 (amélioration du théorème de Levin (1970))

$$\exists \mu_\infty \in M_+ \qquad \text{t. q.} \qquad K(x) \asymp J_{\mu_\infty}(x).$$

Démonstration. — Soit $f_\infty \in D$ un décodeur optimal; alors, d'après le théorème 3.2,

$$\mu_\infty = \mu_{f_\infty} \in M_+ .$$

donc, d'après le théorème 3.4,

$$\exists f \in D \qquad \text{t. q.} \qquad K_f(x) \asymp J_{\mu_\infty}(x)$$

or

$$K(x) \leqslant K_f(x)$$

d'où

$$K(x) \leqslant J_{\mu_\infty}(x).$$

D'autre part, puisque

$$\mu_\infty(xX^\infty) = \sum_{p: f_\infty(p, |x|) = x} 2^{-|p|}$$

et que $K(x) = \min \{ |p| : f_\infty(p, |x|) = x \}$, on a

$$\mu_\infty(xX^\infty) \geqslant 2^{-K(x)}. \tag{1}$$

D'où $J_{\mu_\infty}(x) \leqslant K(x)$. □

THÉORÈME 3.7. — La mesure $\mu_\infty \in M_+$ précédemment définie est une mesure universelle dans M_+.

Démonstration. — Soit $\mu \in M_+$; alors, d'après le théorème 3.4,

$$\exists f \in D \qquad \text{t. q.} \qquad K_f(x) \asymp J_\mu(x).$$

Or

$$K(x) \leqslant K_f(x)$$

donc

$$\exists k \quad \forall x \qquad K(x) \leqslant -\log \mu(xX^\infty) + k,$$

donc

$$\exists k \quad \forall x \qquad \mu(xX^\infty) \leqslant 2^k . 2^{-K(x)} \triangleq C . 2^{-K(x)},$$

or on sait, d'après le (1) du théorème 3.6, que

$$2^{-K(x)} \leqslant \mu_\infty(xX^\infty).$$

Donc

$$\exists C \quad \forall x \qquad \mu(xX^\infty) \leqslant C . \mu_\infty(xX^\infty). \qquad \square$$

Conformément à la 1re partie, μ_∞ est donc, d'un point de vue statistique,

une mesure de probabilité *a priori* dans M_+ ; nous allons donner une estimation numérique de μ_∞.

THÉORÈME 3.8. — μ_∞ est approximativement d'ordre $1/n$.

Démonstration. — 1) D'abord, il existe $C > 0$ tel que la probabilité d'un 1 après $n0$ soit supérieure à $1/Cn$.

En effet, d'après le théorème 3.6,

$$\exists k_1 \quad \text{t. q.} \quad -k_1 \leqslant K(0^n1) + \log \mu_\infty(0^n1X^\infty) \leqslant +k_1 .$$

Donc

$$\mu_\infty(0^n1X^\infty) \geqslant 2^{-k_1}.2^{-K(0^n1)}.$$

Or, d'après la remarque de la page 377,

$$\exists k_2 \quad \text{t. q.} \quad K(0^n1) \leqslant \log n + K_2 .$$

donc

$$\mu_\infty(0^n1X^\infty) \geqslant 1/Cn \qquad \text{avec} \qquad C = 2^{k_1+k_2}. \qquad \square$$

2) D'autre part, d'après le théorème 3.6,

$$(1) \qquad \exists k \quad \forall n \qquad \log \mu_\infty(0^n1X^\infty) \leqslant -K(0^n1) + k .$$

et nous poserons

$$L \triangleq 2^{k+1} .$$

Nous allons montrer que pour tout $C > 0$, la proportion des n tels que la probabilité d'apparition d'un 1 après $n0$ soit supérieure à C/n est inférieure à L/C.

En effet, si

$$\mu_\infty(0^n1X^\infty) \geqslant C/n ,$$

alors

$$\log \mu_\infty(0^n1X^\infty) \geqslant -\log n + \log C .$$

Donc, d'après (1),

$$K(0^n1) \leqslant \log n - (\log C - k).$$

Le résultat est obtenu en utilisant la remarque de la page 377. $\square$

4. CARACTÉRISATION DE LA NATURE ALÉATOIRE D'UNE SÉQUENCE PAR L'ÉCART ENTRE LES DEUX MESURES D'INFORMATION

Soient données une mesure μ calculable ($\mu \in M_0$) et une séquence $\omega \in X^\infty$; on a vu que, d'un point de vue théorique, l'information hartleyenne et la complexité sont des notions très proches ; toutefois on ne connaît encore

aucun lien lorsqu'on applique ces deux mesures de l'information aux fragments *d'une même* séquence $\omega \in X^\infty$. Considérons donc $\omega_n \sqsubset \omega$; on peut calculer $J_\mu(\omega_n)$ d'une part, et $K(\omega_n)$ d'autre part.

Conformément à la remarque de la page 379, puisque nous nous intéressons à des fragments de longueur donnée, nous pouvons, dans une certaine mesure, remplacer $K(\omega_n)$ par $K_0(\omega_n)$.

Le résultat intéressant que l'on obtient est que la *proximité* des deux mesures d'information K_0 et J_μ sur ω, caractérise le fait que ω est μ-aléatoire.

De façon plus précise, on a le

THÉORÈME 4.1. — Soit $\mu \in M_0$:

$$\omega \in X_\mu^\infty \Leftrightarrow J_\mu(\omega) \preccurlyeq K_0(\omega) .$$

On peut « affaiblir » ce théorème en ne considérant que des mesures de M_- ; on obtient alors le

THÉORÈME 4.2. — Soit $\mu \in M_-$:

$$\omega \in X_\mu^\infty \Leftrightarrow J_\mu(\omega) \preccurlyeq K_0(\omega) .$$

Or ces deux théorèmes sont équivalents ; en effet :

1) Le théorème 4.1 implique en particulier le théorème 4.2.

2) D'autre part, supposons le théorème 4.2 vrai et soit $\mu \in M_0$. D'après le théorème 2.3, $\exists \nu \in M_-$: $\mu \sim \nu$. Si $\omega \in X_\mu^\infty$, alors, d'après le théorème 2.1 $\omega \in X_\nu^\infty$. Donc, d'après le théorème 4.2 $J_\nu(\omega) \preccurlyeq K_0(\omega)$. Or $J_\mu \asymp J_\nu$; donc $J_\mu(\omega) \preccurlyeq K_0(\omega)$.

Réciproquement, supposons encore le théorème 4.2 vrai et soit $\mu \in M_0$ et $\omega \in X^\infty$ tels que $J_\mu(\omega) \preccurlyeq K_0(\omega)$. D'après le théorème 2.3 $\exists \nu \in M_-$ tel que $J_\mu \asymp J_\nu$. Donc $J_\nu(\omega) \preccurlyeq K_0(\omega)$; donc, d'après le théorème 4.2, $\omega \in X_\nu^\infty = X_\mu^\infty$. □

Il nous reste donc à démontrer le théorème 4.2 ; nous utiliserons certaines techniques que Schnorr (1973) a utilisées pour démontrer un cas très particulier de notre théorème.

Démonstration :

Condition nécessaire. — Soit $\mu \in M_-$; il nous faut montrer que

$$\omega \in X_\mu^\infty \Rightarrow J_\mu(\omega) \preccurlyeq K_0(\omega) .$$

Posons

$$Y_i \overset{\Delta}{=} \{ x : K_0(x) \leqslant J_\mu(x) - i \} .$$

Supposons que $\mu(Y_i X^\infty) > 2^{-i}$. Si

$$Y_i X^\infty = \bigcup_k x_k X^\infty ,$$

alors

$$\sum_k \mu(x_k X^\infty) \geqslant \mu\left\{\bigcup_k x_k X^\infty\right\} = \mu(Y_i X^\infty) > 2^{-i}.$$

Donc

$$\exists n \in N : \sum_{k=1}^{n} \mu(x_k X^\infty) > 2^{-i}.$$

φ_∞ étant une calculatrice optimale de D_0, il existe un code *instantané* $p_1, p_2, \ldots, p_n$ tel que :

$$\begin{cases} \varphi_\infty(p_k) = x_k \\ \quad | p_k | = K_0(x_k) \leqslant J_\mu(x_k) - i. \end{cases}$$

Soit λ la mesure de Lebesgue ; alors

$$\begin{aligned} \lambda\left\{\bigcup_{k=1}^{n} p_k X^\infty\right\} &= \sum_{k=1}^{n} \lambda(p_k X^\infty) \\ &= \sum_{k=1}^{n} 2^{-|p_k|} \\ &\geqslant \sum_{k=1}^{n} 2^{i - J_\mu(x_k)} \\ &\geqslant 2^i \sum_{k=1}^{n} \mu(x_k X^\infty) > 1. \end{aligned}$$

Il y a contradiction, donc $\mu(Y_i X^\infty) \leqslant 2^{-i}$; de plus, $\mu \in M_-$ donc J_μ est récursive (proposition 2.2) ; donc Y_i est récursivement énumérable.

Par conséquent $Y = \{ Y_i X \}_{i \in N}$ est un μ-test.

Si on n'a pas $J_\mu(\omega) \leqslant K_0(\omega)$, alors

$$\forall i \quad \exists n_0 : J_\mu(\omega_{n_0}) - K_0(\omega_{n_0}) \geqslant i,$$

donc

$$\forall i \quad \exists n_0 : \omega_{n_0} \in Y_i$$

d'où

$$\forall i \qquad \omega \in Y_i X^\infty.$$

Par conséquent,

$$\omega \in \bigcap_i Y_i X^\infty = o(Y) \qquad \text{donc} \qquad \omega \notin X_\mu^\infty. \qquad \square$$

Condition suffisante

Soit $\mu \in M_-$ et supposons que $\omega \notin X_\mu^\infty$; alors il existe un μ-test Y tel que $\omega \in o(Y)$.

Or il est toujours possible de choisir un μ-test $Y \subset N \times X^*$ tel que $\forall i \in N$, Y_i soit un code instantané. Alors

$$\sum_{x \in Y_i} \mu(xX^\infty) = \mu(Y_i X^\infty) \leqslant 2^{-i}. \tag{1}$$

Définissons alors

$$U_i \overset{\Delta}{=} \{ (x, n) \in Y_i \times N : n \geqslant J_\mu(x) + 2 - i \}.$$

U_i est récursivement énumérable et consistant car

$$\begin{aligned} \sum_{(x,n) \in U_i} 2^{-n} &= \sum_{x \in Y_i} \sum_{k \geqslant 0} 2^{-J_\mu(x) - 2 + i + k} \\ &= \sum_{x \in Y_i} 2^{i+1} . 2^{-[-\log \mu(xX^\infty)] - 2} \\ &\leqslant \sum_{x \in Y_i} 2^i \mu(xX^\infty) \leqslant 1 \qquad \text{(d'après (1))}. \end{aligned}$$

En utilisant le lemme de Chaitin-Kraft, il vient

$$\exists f_i \in D^0$$

tel que

$$\partial \{ p \in X^n : f_i(p) = x \} = \begin{cases} 1 & \text{si} \quad n \geqslant J(x) + 2 - i \\ 0 & \text{si} \quad n < J(x) + 2 - i. \end{cases}$$

Alors,

$$\forall x \in Y_i \qquad K_{f_i}(x) = J_\mu(x) + 2 - i. \tag{2}$$

Soit maintenant l'application bijective

$$\xi \in X^* \to \bar{\xi} \in X^*$$

définie par $\xi = x_1 \ldots x_n$, $\bar{\xi} = x_1 x_1 \ldots x_n x_n 01$.

En identifiant canoniquement N et X, on a une application de $N \to X^* : i \to \bar{i}$, et, évidemment

$$| \bar{i} | \leqslant 2 \log i + 4. \tag{3}$$

Définissons alors l'algorithme f par

$$f(\bar{i}x) \overset{\Delta}{=} f_i(x) \qquad \text{pour} \qquad x \in \Delta f_i.$$

Soient $y_1, y_2 \in \Delta f$ et $y_1 \neq y_2$.

Alors, par définition de f :

$$\exists i_1 \quad \exists x_1 \in \Delta f_{i_1} : y_1 = \bar{i}_1 x_1$$
$$\exists i_2 \quad \exists x_2 \in \Delta f_{i_2} : y_2 = \bar{i}_2 x_2$$

Soit $z \in y_1 X^* \cap y_2 X^*$; alors

$$\bar{i}_1 x_1 \sqsubset z \qquad \text{et} \qquad \bar{i}_2 x_2 \sqsubset z .$$

La forme même des $\bar{i}$ implique $\bar{i}_1 = \bar{i}_2$ donc $i_1 = i_2 = i$.

Donc $x_1, x_2 \in \Delta f_i$ et $x_1 \neq x_2$ puisque $y_1 \neq y_2$.

Or $f_i \in D_0$ donc $x_1 X^* \cap x_2 X^* = \emptyset$, d'où $y_1 X^* \cap y_2 X^* = \emptyset$; par conséquent $f \in D_0$. Soit donc $i \in N$ et $x \in Y_i$.

Calculons

$$K_f(x) = \min \{ | p | : f(p) = x \}$$

comme

$$\exists q \in \Delta f_i \qquad \text{t. q.} \qquad p = \bar{i} q$$

on a

$$K_f(x) = | \bar{i} | + K_{f_i}(x) .$$

En utilisant (2) et (3), il vient

$$K_f(x) \leqslant J_\mu(x) - i + 2 \log i + 6$$

Soit à présent $i \in N$ et $\omega \in Y_i X^\infty$; alors

$$\exists n : J_\mu(\omega_n) - K_f(\omega_n) \geqslant i - 2 \log i - 6$$

donc, si $\omega \in o(Y)$, on a

$$\forall i \quad \exists n : J_\mu(\omega_n) - K_f(\omega_n) \geqslant i - 2 \log i - 6$$

or

$$\exists C \quad \forall x \quad K_0(x) \leqslant K_f(x) + C .$$

Donc

$$\forall i \quad \exists n : J_\mu(\omega_n) - K_0(\omega_n) \geqslant i - 2 \log i - 6 - C = \theta(i) .$$

$\theta(i)$ étant strictement croissante on a finalement :

$$\forall i \quad \exists n_0 \qquad J_\mu(\omega_{n_0}) - K_0(\omega_{n_0}) \geqslant i . \qquad \square$$

Ce théorème est une voie nous permettant d'aborder une question importante de la théorie de l'Information, la propriété d'équipartition asymptotique (P. E. A.).

Si on a une source régie par une probabilité $\mu \in M_+$, la P. E. A. est définie par :

$$\mu\text{-p. s.} \qquad \lim_{n \to \infty} - \frac{\log \mu(\omega_n X^\infty)}{n} = H .$$

Le théorème précédent nous permet de définir la P. E. A. par :

$$\mu\text{-p. s.} \qquad \lim_{n \to \infty} K_0(\omega_n)/n = H .$$

et ces deux définitions sont équivalentes sur M_- ; en effet, soit $\mu \in M_-$; alors, si l'on reprend la démonstration du théorème 3.4,

$$\forall i \quad \forall x \in X^i \quad \exists f_i \in D_0 : K_{f_i}(x) = J_\mu(x) + 2 .$$

Définissons ensuite $f \in D_0$, comme page 388, par

$$f(\bar{i}x) = f_i(x) \qquad \text{pour} \qquad x \in \Delta f_i$$

Alors, comme page 389 ;

$$K_f(x) \leqslant J_\mu(x) + 2 \,||\, x \,||\, .$$

Or

$$K_0(x) \leqslant K_f(x) .$$

Donc

$$J_\mu(x) \leqslant K_0(x) \Leftrightarrow J_\mu(x) \leqslant K_0(x) \leqslant J_\mu(x) + 2 \,||\, x \,||\, .$$

L'équivalence annoncée est donc démontrée du fait que

$$||\, \omega_n \,|| \asymp |\, n \,| \asymp \log n = o(n) .$$

BIBLIOGRAPHIE

[1] ASH, Information Theory. *Interscience Publishers*, 1965.

[2] CHAITIN, « A theory of program size formally equivalent to information theory », *J. A. C. M.*, t. **22**, n° 3, 1975, p. 329-340.

[3] HUFFMAN, « A method for the construction of minimum redundancy codes », *Proc. IRE*, t. **40**, n° 10, 1952, p. 1098-1101.

[4] KOLMOGOROV, « Three approaches to the quantitative definition of information ». *Inform. Transmission*, t. **1**, 1965, p. 3-11.

[5] MARTIN-LOF, « The definition of random sequences ». *Information and Control*, t. **9**, 1966, p. 602-619.

[6] SCHNORR, « Process complexity and effective random tests ». *JCSS*, t. **7**, 1973, p. 376-388.

[7] SCHOENFIELD, Mathematical Logic. Addison-Wesley, 1967.

[8] WILLIS, « Computational complexity and probability constructions ». *J. A. C. M.*, t. **17**, n° 2, 1970, p. 241-259.

[9] ZVONKIN I LEVIN, « Complexité d'un objet fini... ». *Uspehi Matematiceskih Nauk.*, t. **156**, 1970.

(Manuscrit reçu le 14 octobre 1976)