CHAPITRE

14

LE CHAOS ALGORITHMIQUE ET LA MÉTHODE D'INCOMPRESSIBILITÉ

par *Paul Vitányi*

De nombreuses théories physiques comme la théorie du chaos traitent fondamentalement de la tension conceptuelle entre déterminisme et hasard. La complexité de Kolmogorov peut exprimer le hasard à l'intérieur du déterminisme et fournit une approche pour formuler le comportement chaotique. Nous utilisons la méthode d'incompressibilité comme outil technique pour quantifier l'imprévisibilité des systèmes chaotiques. Nous présentons la méthode par des exemples : la distribution des nombres premiers et la taille maximale des cliques dans les graphes aléatoires.

1. Introduction

Idéalement, les théories physiques sont des représentations abstraites : ce sont des théories mathématiques axiomatiques d'une réalité physique sous-jacente. Cette réalité, ne pouvant être directement éprouvée, est donc inconnue et même, en principe, inconnaissable. Comme substitut, les scientifiques postulent une description informelle acceptable intuitivement, et formulent ensuite une ou plusieurs théories mathématiques pour décrire les phénomènes.

Le chaos déterministe. De nombreux phénomènes physiques (par exemple météorologiques) satisfont des équations déterministes communément admises. À partir

de données initiales, nous pouvons extrapoler et calculer les états suivants du système. On pensait traditionnellement qu'une précision accrue des données initiales (mesure) et une puissance de calcul accrue conduiraient à une extrapolation de plus en plus précise (prédiction) sur des durées inversement proportionnelles à la précision choisie. Mais il s'est trouvé que pour beaucoup de systèmes (« chaotiques ») l'extrapolation ne peut se faire, au mieux, que sur des durées inversement proportionnelles au logarithme de la précision. En fait, il s'avère qu'il est pratiquement impossible de faire une prédiction à long terme qui soit plus fiable que ce qu'on obtiendrait en jouant à pile ou face avec une pièce équilibrée : ce phénomène est appelé chaos (voir [3] pour une introduction). Il y a deux causes à cela, plus ou moins reliées :

– **Instabilité.** Dans certains systèmes déterministes, une erreur arbitrairement petite sur les conditions initiales peut croître exponentiellement durant l'évolution ultérieure du système, jusqu'à recouvrir le domaine complet des valeurs atteignables par le système. Ce phénomène d'instabilité d'un calcul est bien connu en analyse numérique : les procédures de calcul pour inverser des matrices mal conditionnées (avec un déterminant proche de zéro) vont introduire des erreurs exponentiellement croissantes.
– **Imprévisibilité.** Supposons que nous considérions un système décrit par des équations déterministes finiment représentables (voir ci-après). Même si les segments initiaux de longueur fixée de la représentation binaire infinie des paramètres réels décrivant les états passés du système sont parfaitement connus, et même si la procédure de calcul utilisée n'est entachée d'aucune erreur, il sera encore impossible pour beaucoup de tels systèmes de prédire effectivement (calculer) une quelconque extrapolation significativement longue des états du système avec une fiabilité supérieure à celle d'un tirage à pile ou face parfaitement aléatoire. C'est le cœur des phénomènes chaotiques : l'intrusion du hasard dans le déterminisme.

Remarque 1.1. Dans la suite nous utilisons la notion de « calcul effectif » au sens mathématique bien connu de « calculabilité par une machine de Turing ». De même, nous utilisons de manière interchangeable les notions de « fonction (partielle) récursive » et de « fonction (partielle) calculable ». Dans la théorie de la récursivité, de telles fonctions sont des applications d'une partie de $\mathbb{N}$ dans $\mathbb{N}$ (ou, si on veut, dans $\mathbb{Q}$, par composition avec une énumération explicite de $\mathbb{Q}$). Dans le contexte présent il peut être souhaitable de considérer l'extension aux nombres réels. Une fonction $f : \mathbb{N} \to \mathbb{R}$ est *semi-calculable supérieurement* s'il existe une machine de Turing T calculant une fonction totale $\phi : \mathbb{N} \times \mathbb{N} \to \mathbb{Q}$ telle que $\phi(x,t+1) \leq \phi(x,t)$ et $\lim_{t\to\infty} \phi(x,t) = f(x)$. Cela signifie qu'on peut calculer une approximation par excès de f. Si $-f$ est semi-calculable supérieurement, f est dite semi-calculable inférieurement. Si f est semi-calculable à la fois supérieurement et inférieurement, nous dirons qu'elle est *calculable* (ou récursive, si ses valeurs sont entières ou rationnelles). (Nous pouvons de même considérer des fonctions calculables de variable réelle : $f : \mathbb{R} \to \mathbb{R}$. Cela requiert des définitions soigneuses et il s'avère que la calculabilité implique la continuité. Mais cette sophistication n'est pas nécessaire au présent exposé.) L'extension de la notion de fonction calculable à des variables et valeurs vectorielles est immédiate. Les détails se trouvent dans tout manuel sur la calculabilité ou dans [8].

Remarque 1.2. Il est peut-être utile d'insister sur le fait qu'instabilité et imprévisibilité, bien que proches compagnes, ne sont pas toujours la même chose. Un exemple trivial d'instabilité sans imprévisibilité est un système qui fait un premier choix de manière instable mais se tient ensuite à ce choix. (Un tel système équivaut, par exemple, au jeu de pile ou face avec une « pièce dictatoriale » qui, lancée une première fois, donne un résultat 0 ou 1 avec des probabilités égales, mais donne à chaque lancer suivant le même résultat que celui obtenu la première fois.) Un exemple d'imprévisibilité sans instabilité est une fonction $f_r : \mathbb{N} \to \{0,1\}$ définie par $f_r(n) = r_n$, où $r = r_1 r_2 \ldots$ est une suite binaire infinie aléatoire au sens de Martin-Löf (voir ci-dessous) et donc imprévisible. (Ici $n \in \mathbb{N}$ joue le rôle du temps.)

Probabilité. La théorie classique des probabilités traite du hasard en termes de *variables aléatoires*. Elle ne peut exprimer le concept de données individuelles aléatoires. Cependant notre intuition concernant celles-ci est très forte. Un adversaire prétend avoir une pièce vraiment aléatoire et nous invite à parier sur le résultat. La pièce retombe cent fois de suite sur face. Nous disons qu'elle ne peut pas être aléatoire. L'adversaire, cependant, invoque la théorie des probabilités qui dit que chacune des suites de résultats de cent lancers de pièces a la même probabilité, $1/2^{100}$, et que l'une de ces suites devait bel et bien apparaître. La théorie des probabilités ne nous donne aucune base pour contester un résultat *après* qu'il a été obtenu. Nous pourrions seulement exclure à l'avance le truquage de la pièce en imposant un pari auxiliaire pénalisant un résultat de 100 faces. Mais qu'en est-il de $1010\ldots$? Qu'en est-il d'un segment initial du développement binaire de π ?

Suite régulière :

$$\Pr(00000000000000000000000000) = \frac{1}{2^{26}}$$

Suite régulière :

$$\Pr(01000110110000010100111001) = \frac{1}{2^{26}}$$

Suite aléatoire :

$$\Pr(10010011011000111011010000) = \frac{1}{2^{26}}$$

La première suite est régulière, mais quelle est la différence entre la deuxième suite et la troisième ? La troisième suite a été obtenue en lançant un quarter[1]. La deuxième suite est très régulière : 0, 1, 00, 01, ... La troisième suite passera avec succès des tests montrant son caractère (pseudo-)aléatoire.

[1] Pièce d'un quart de dollar. (N. d. T.)

En fait, la théorie classique des probabilités ne peut pas exprimer la notion de *suite individuelle aléatoire*. Elle peut seulement exprimer l'espérance de propriétés de l'ensemble total des suites obéissant à une certaine distribution.

Cela est analogue à la situation ci-dessus en physique : *« comment un objet individuel peut-il être aléatoire ? »* est un paradoxe de la théorie des probabilités au même titre que *« comment une suite individuelle d'états d'un système déterministe peut-elle être aléatoire ? »* est un paradoxe de la théorie des systèmes physiques déterministes.

Dans la théorie des probabilités, le paradoxe a été résolu de manière satisfaisante en combinant des notions de calculabilité et de théorie de l'information pour exprimer la complexité d'un objet fini. Cette complexité est la longueur du programme binaire le plus court à partir duquel l'objet peut être effectivement reconstruit. On peut l'appeler le *contenu algorithmique d'information* de l'objet. Cette quantité s'avère être un attribut de l'objet seul, et elle est récursivement invariante. C'est la *complexité de Kolmogorov* de l'objet. Il s'avère que cette notion peut être appliquée également au paradoxe physique ci-dessus, comme nous le verrons plus loin.

2. La complexité de Kolmogorov

Pour rendre cet article autonome nous passons rapidement en revue quelques notions et propriétés requises. Le lecteur trouvera les détails et d'autres propriétés dans le manuel [8]. Nous identifions les nombres entiers naturels avec les suites binaires finies en les accouplant deux à deux de la façon suivante :

$$(0,\epsilon),\ (1,0),\ (2,1)\ (3,00),\ (4,01),\ \ldots,$$

où ϵ est la suite vide. La *longueur* $l(x)$ est le nombre de chiffres binaires dans la suite binaire x (par exemple, $l(\epsilon) = 0$) ; cela définit aussi la « longueur » de l'entier naturel correspondant. Si A est un ensemble, alors $|A|$ désigne la cardinalité de A. Soit $\langle\cdot,\cdot\rangle : \mathbb{N}\times\mathbb{N} \to \mathbb{N}$ une fonction bijective « d'appariement » calculable standard. Dans tout cet article, nous supposerons que $\langle x,y\rangle = 1^{l(x)}0xy$.

On définit $\langle x,y,z\rangle$ par $\langle x,\langle y,z\rangle\rangle$.

Il nous faut quelques notions de la théorie des algorithmes (cf. [10]). Soit $\phi_1, \phi_2, \ldots$ une énumération standard des fonctions partielles récursives. La *complexité* (de Kolmogorov) de $x \in \mathbb{N}$, connaissant y, est définie par

$$C(x|y) = \min\{l(\langle n,z\rangle) \ : \ \phi_n(\langle y,z\rangle) = x\}.$$

Cela signifie que $C(x|y)$ est le nombre *minimal* de chiffres binaires d'une description permettant de reconstruire effectivement x en connaissant y. La complexité inconditionnelle est définie par $C(x) = C(x|\epsilon)$. Ces notions ont été introduites originellement dans [6].

Une autre définition est la suivante. Soit

$$C_\psi(x|y) = \min\{l(z) \ : \ \psi(\langle y,z\rangle) = x\} \qquad \textbf{(14.1)}$$

la complexité conditionnelle de x connaissant y relativement à une fonction de décodage ψ. Alors $C(x|y) = C_\psi(x|y)$, où ψ est une fonction partielle récursive universelle qui vérifie $\psi(\langle y,n,z\rangle) = \phi_n(\langle y,z\rangle)$.

Nous aurons besoin des propriétés suivantes. Pour tous $x,y \in \mathbb{N}$, nous avons[2]

$$C(x|y) \leq l(x) + O(1). \tag{14.2}$$

Pour tout $y \in \mathbb{N}$ il existe $x \in \mathbb{N}$ de longueur n tel que $C(x|y) \geq n$. En particulier, nous pouvons poser $y = \epsilon$. De tels x peuvent être qualifiés d'*aléatoires* puisqu'ils n'ont aucune régularité utilisable pour compresser leur description : intuitivement, la description effective la plus courte d'un tel entier x est x lui-même. En général, pour tous n et y, il y a au moins $2^n - 2^{n-c} + 1$ entiers x (distincts) de longueur n tels que

$$C(x|y) \geq n - c. \tag{14.3}$$

Dans certains cas, on souhaite encoder x sous une forme x' *auto-délimitante* (a-d), permettant de décomposer $x'y$ en x et y. Nous utiliserons alors la complexité *préfixale* $K(x)$, introduite dans [7], qui désigne la longueur de la plus courte description *auto-délimitante*. À cette fin, nous considérons des machines de Turing *préfixales*, ainsi dénommées parce qu'elles n'ont que des 0 et des 1 sur leur bande d'entrée et ne peuvent donc pas détecter la fin de l'entrée. Nous définissons alors une entrée comme étant la partie de la bande d'entrée que la machine a lue quand elle s'arrête. Quand $x \neq y$ sont deux telles entrées, il est clair que x ne peut pas être un préfixe de y (c'est-à-dire que y ne peut pas être de la forme xz), donc l'ensemble des entrées forme ce qu'on appelle un *code préfixal* ou un *code sans préfixes*. Nous définissons $K(x|y), K(x)$ de façon similaire aux $C(x|y), C(x)$ ci-dessus, mais relativement à une machine préfixale universelle qui lit d'abord 1^n0 sur la bande d'entrée et simule ensuite la machine préfixale n sur le reste de l'entrée.

On obtient de bonnes majorations de la complexité préfixale de x en itérant la règle simple selon laquelle une description auto-délimitante de la longueur de x suivie de x elle-même est une description a-d de x. Par exemple, $x' = 1^{l(x)}0x$ et $x'' = 1^{l(l(x))}0l(x)x$ sont toutes les deux des descriptions a-d de x, et cela montre que $K(x) \leq 2l(x) + O(1)$ et $K(x) \leq l(x) + 2l(l(x)) + O(1)$.

De même, nous pouvons encoder x sous une forme auto-délimitante de son plus court programme $p(x)$ (de longueur $l(p(x)) = C(x)$) en $2C(x) + 1$ chiffres binaires. En itérant ce schéma, nous pouvons encoder x comme programme auto-délimitant de $C(x) + 2\log C(x) + 1$ chiffres binaires[3], ce qui montre que $K(x) \leq C(x) + 2\log C(x) + 1$, et ainsi de suite.

2.1. La méthode d'incompressibilité

Le secret du succès des arguments de complexité de Kolmogorov comme technique de démonstration réside dans un fait simple : une écrasante majorité de

[2] Partout dans cet article, $O(1)$ (resp. $o(1)$) désignera une quantité bornée (resp. une quantité qui converge vers 0), quel que soit son signe, et $O(f(n))$ signifiera $f(n) \times O(1)$ (resp. $o(f(n))$ signifiera $f(n) \times o(1)$).

[3] Partout dans cet article log désigne le logarithme binaire. (N.d.A.)

chaînes de 0 et de 1 n'ont presque pas de régularités calculables. Nous avons qualifié une telle chaîne d'« aléatoire ». Il n'y a pas de description d'une telle chaîne qui soit plus courte que la description littérale : elle est incompressible.

Les démonstrations traditionnelles font souvent intervenir tous les cas (l'usage est d'utiliser le mot anglais : toutes les *instances*) d'un problème afin de conclure qu'une certaine propriété a lieu pour au moins une instance. La démonstration serait plus simple si cette instance avait pu être utilisée dès le départ. Malheureusement, elle est difficile ou impossible à trouver, et la démonstration doit vraiment faire intervenir toutes les instances. En revanche, dans une démonstration par la méthode d'incompressibilité, on choisit d'abord un objet individuel aléatoire (c'est-à-dire incompressible) dont on sait qu'il existe (même si on ne peut pas le construire). Ensuite on montre que si la propriété présumée n'était pas satisfaite, alors cet objet pourrait être compressé, et ne serait donc pas aléatoire. Donnons quelques exemples simples.

La distribution des nombres premiers. Un nombre premier est un entier naturel qui n'est pas divisible par des entiers naturels autres que lui-même et 1. Au dix-neuvième siècle, Chebychev (on écrit aussi Tchebycheff) montra que le nombre de nombres premiers inférieurs à n croît asymptotiquement comme $n/\log n$[4]. En utilisant la méthode d'incompressibilité, nous ne pouvons pas (encore) prouver cette assertion, mais nous pouvons nous en approcher remarquablement, avec un effort minimal.

Nous démontrons d'abord que pour une infinité de n, le nombre de nombres premiers inférieurs ou égaux à n est au moins $\log n/\log\log n$. La méthode de démonstration est la suivante. Pour chaque n, nous construisons une description permettant de retrouver effectivement n. Cette description fera intervenir les nombres premiers ne dépassant pas n. Pour un certain n, cette description doit être longue, ce qui donnera le résultat désiré.

Soit $p_1, p_2, \ldots, p_m$ la liste de tous les nombres premiers inférieurs ou égaux à n. Alors,

$$n = p_1^{e_1} p_2^{e_2} \cdots p_m^{e_m}$$

peut être reconstruit à partir du vecteur des exposants. Chaque exposant vaut au plus $\log n$ et peut être représenté par $\log\log n$ chiffres binaires. La description de n (connaissant l'ordre de grandeur maximal $\log n$ des exposants) peut être donnée en (approximativement) $m \log\log n$ chiffres binaires. Or, on peut montrer que tout n aléatoire (connaissant $\log n$) ne peut pas être décrit en moins de $\log n$ chiffres binaires, d'où le résultat.

Pouvons-nous faire mieux ? C'est légèrement plus compliqué. Notons $l(x)$ la longueur de la représentation binaire de x. Nous allons montrer que p_m (le m-ième nombre premier) est $\leq m \log^2 m$. (On peut montrer que cela équivaut à dire que le nombre de nombres premiers inférieurs ou égaux à n est plus grand que $n/\log^2 n$.)

[4] Plus précisément, Chebychev a montré (en 1850) que le rapport entre le nombre $\pi(n)$ des nombres premiers $\leq n$, et $n/\log n$, est compris entre deux constantes strictement positives explicites. En fait, ce rapport tend même vers 1 quand n tend vers l'infini : c'est le « théorème des nombres premiers », démontré (en 1896) par Hadamard et la Vallée Poussin. On peut montrer que ce théorème est équivalent à cet énoncé : *si p_n désigne le n-ième nombre premier, alors $\frac{p_n \log e}{n \log n} \to 1$ quand $n \to \infty$.* (N. d. T.)

D'abord, tout nombre entier n est complètement déterminé par la donnée du numéro m de son plus grand facteur premier p_m et du quotient (entier) n/p_m. Nous pouvons donc décrire n par $E(m)n/p_m$, où $E(m)$ est un codage sans préfixe de m. (La description de m doit être auto-délimitante, car sinon on ne saurait pas où s'arrête la description de m et où commence celle de n/p_m.) Pour un n aléatoire, la longueur de cette description, $l(E(m)) + \log n - \log p_m$ doit dépasser $\log n$. Par conséquent, $\log p_m \leqslant l(E(m))$. Or, il est connu (et immédiat d'après la discussion précédente) que nous avons $l(E(m)) \leq \log m + 2\log\log m$. Par suite, $p_m \leqslant m\log^2 m$, comme annoncé.

Graphes aléatoires. En interprétant les chaînes de 0 et de 1 comme des objets combinatoires plus complexes, on est conduit à un nouvel ensemble de propriétés et de problèmes qui n'ont pas d'équivalent direct dans le monde plus « plat » des chaînes. Nous allons maintenant établir des propriétés topologiques, combinatoires et statistiques des graphes ayant une grande complexité de Kolmogorov. Un tel graphe possède simultanément toutes les propriétés hautement probables des graphes engendrés aléatoirement. Ils constituent « presque tous les graphes » et leurs propriétés sont *a fortiori* vérifiées avec une probabilité tendant vers 1 quand le nombre de sommets croît vers l'infini.

Chaque graphe étiqueté $G = (V,E)$ sur n sommets $V = \{1,\ldots,n\}$ (avec au plus une arête, non orientée, entre chaque paire de sommets) peut être représenté (à un automorphisme près) par une chaîne binaire $E(G)$ de longueur $\binom{n}{2}$: nous fixons simplement un ordre, par exemple lexicographique, sur les $\binom{n}{2}$ arêtes possibles d'un graphe à n sommets, et le i-ième chiffre binaire de la chaîne indique la présence (1) ou l'absence (0) de la i-ième arête. Réciproquement, chaque chaîne binaire de longueur $\binom{n}{2}$ encode un graphe à n sommets. Nous pouvons donc identifier chacun de ces graphes avec sa représentation en chaîne binaire.

Nous allons démontrer que G ne contient pas de clique (graphe complet) sur plus de $2\log n + 1 + o(1)$ sommets.

Soit m le nombre de sommets de la plus grande clique $\mathcal{K}$ dans G. Nous essayons de compresser $E(G)$, en l'encodant en $E'(G)$, comme suit :

1. On ajoute en préfixe à $E(G)$ la liste des sommets de $\mathcal{K}$, chaque sommet utilisant $\lceil\log n\rceil$ chiffres binaires[5], donnant un total de $m\lceil\log n\rceil$ chiffres binaires pour le préfixe.

2. On supprime tous les chiffres binaires redondants de la sous-chaîne $E(G)$, représentant les arêtes entre les sommets de $\mathcal{K}$, ce qui économise $m(m-1)/2$ chiffres binaires.
Par suite,

$$l(E'(G)) = l(E(G)) + m\lceil\log n\rceil - \binom{m}{2}. \qquad \textbf{(14.4)}$$

Soit p un programme qui, à partir de n et de $E'(G)$, reconstruit $E(G)$. Alors

$$C(E(G)|n,p) \leq l(E'(G)). \qquad \textbf{(14.5)}$$

[5] $\lceil x\rceil$ désigne le plus petit entier strictement supérieur à x.

Puisqu'il y a $2^{\binom{n}{2}}$ graphes étiquetés sur n sommets et au plus $2^{\binom{n}{2}}-1$ descriptions binaires de longueur inférieure à $\binom{n}{2}$, nous pouvons choisir un graphe étiqueté G sur n sommets qui satisfait

$$C(E(G)|n,p) \geq \binom{n}{2} + o(\log n). \tag{14.6}$$

Les équations (14.6), (14.4), et (14.5) ne peuvent être vraies que si $m \leq 2\log n + 1 + o(1)$.

En fait, le lecteur perspicace va maintenant comprendre que, alors que l'information dans le nouveau préfixe de $E'(G)$ est utilisée par le programme p en insérant des « 1 » dans les trous appropriés de l'ancien suffixe de $E'(G)$ pour reconstruire les arêtes de la clique, l'utilisation d'un autre programme p' nous montrerait que le cardinal maximal d'un ensemble de sommets qui ne sont joints par aucune arête est majoré par la même borne. En fait, tout sous-graphe de G descriptible facilement (en $O(\log n)$ chiffres, connaissant les sommets étiquetés) ne peut avoir plus de $2\log n + 1$ sommets. Cela inclut virtuellement toutes les propriétés potentiellement intéressantes. De plus, l'ensemble des graphes G qui satisfont (14.6) est très grand : un argument facile de dénombrement montre que, parmi les $2^{\binom{n}{2}}$ graphes étiquetés sur n sommets, c'est le cas pour au moins $(1-1/n)2^{\binom{n}{2}}$ graphes. C'est-à-dire qu'en lançant une pièce équilibrée pour déterminer la présence ou l'absence de chacune des $\binom{n}{2}$ arêtes d'un graphe étiqueté à n sommets, nous obtiendrons avec une écrasante probabilité de $1-1/n$ un graphe satisfaisant (14.6). Par suite notre conclusion au sujet de la taille maximale des sous-graphes facilement descriptibles est valable avec une probabilité proche de 1 (c'est-à-dire une probabilité $\geq 1-1/n$) pour les graphes engendrés aléatoirement.

2.2. Les suites aléatoires

On aimerait qualifier d'aléatoire une suite infinie $\omega \in \{0,1\}^\infty$ si $C(\omega_{1:n}) \geq n + O(1)$ pour tout n (où $\omega_{1:n}$ désigne la chaîne composée des n premiers chiffres binaires de ω). Il s'avère que de telles suites n'existent pas. Cette remarque a conduit P. Martin-Löf [9] à créer sa célèbre théorie de « l'aléatoirité ». Le fait qu'une suite ω soit aléatoire au sens de Martin-Löf signifie, en gros, qu'elle va passer avec succès *tous* les tests effectifs d'« aléatoirité » : aussi bien les tests qui sont connus maintenant que ceux qui sont encore inconnus [9].

Plus tard, on s'est rendu compte [1] que l'on peut quand même définir précisément les suites aléatoires de Martin-Löf en utilisant la complexité préfixale de Kolmogorov :

Théorème 2.1. *Une suite binaire infinie ω est aléatoire au sens de Martin-Löf si et seulement si existe un n_0 tel que $K(\omega_{1:n}) \geq n$ pour tout $n > n_0$.*

Des propriétés semblables ont lieu pour les chaînes finies de complexité élevée, quoique dans un sens moins absolu. Pour tout ensemble fini $S \subseteq \{0,1\}^*$ contenant x, nous avons $K(x|S) \leq \log|S| + O(1)$. Il suffit en effet de considérer le codage auto-délimitant de x constitué des $\lceil \log|S| \rceil$ chiffres de son rang dans l'ordre lexicographique dans S. Ce codage est appelé codage *data-to-model*. Le manque de généricité de x par rapport à S est la quantité par laquelle $K(x|S)$ est inférieure

à la longueur du codage data-to-model. Le *défaut d'« aléatoirité »* de x dans S est défini par

$$\delta(x|S) = \log|S| - K(x|S), \tag{14.7}$$

si $x \in S$, et ∞ sinon. Si $\delta(x|S)$ est petit, alors x peut être considéré comme un membre *typique* de S. Il n'y a pas de propriétés spéciales simples qui le distinguent de la majorité des éléments de S. Ce n'est pas seulement de la terminologie : si $\delta(x|S)$ est petit, alors x satisfait *toutes* les propriétés de basse complexité de Kolmogorov, valables sur S avec une probabilité élevée. Considérons par exemple l'ensemble $S = \{0,1\}^n$ des chaînes de longueur n. Alors $\delta(x|S) = n - K(x|n) + O(1)$. Soit $\delta(n)$ une fonction adéquate comme $\log n$ ou $\sqrt{n}$. Alors les propriétés suivantes sont des analogues finis de la notion de suite aléatoire au sens de Martin-Löf, [8] :

(i) Si P est une propriété satisfaite par tous les x tels que $\delta(x|S) \leq \delta(n)$, alors P est vérifiée avec une probabilité d'au moins $1 - 1/2^{\delta(n)}$ par les éléments de S.
(ii) Soit P une propriété quelconque vérifiée par les éléments de S avec une probabilité d'au moins $1 - 1/2^{\delta(n)}$. Alors P a lieu simultanément pour tous les $x \in S$ tels que $\delta(x|S) \leq \delta(n) - K(P|n) + O(1)$.

Faisons un pas de plus. L'« aléatoirité » des suites infinies et des chaînes finies n'a de sens que dans le contexte d'un espace probabilisé dont elles sont des éléments aléatoires. Dans le cas ci-dessus des suites infinies, cet espace est l'ensemble $\{0,1\}^\infty$ muni de la mesure uniforme λ, également appelée mesure « du jeu de pile ou face », puisque $\lambda(\omega_1 \ldots \omega_n) = 1/2^n$ est égale à la probabilité d'obtenir la chaîne de n chiffres $\omega_1 \ldots \omega_n$ en lançant n fois une pièce équilibrée. On peut généraliser l'approche aléatoire à une mesure calculable arbitraire μ. C'est une mesure pour laquelle il y a une machine de Turing T telle que, pour tous n et $\varepsilon > 0$, à partir de chaque entrée $\omega_1 \ldots \omega_n$,ε, la machine s'arrête avec une sortie r telle que $|\mu(\omega_1 \ldots \omega_n) - r| \leq \varepsilon$. (Au lieu de «machine de Turing », nous pouvons aussi dire « programme informatique » dans un langage informatique universel comme LISP ou Java.) Nous pouvons maintenant parler de *suites* μ*-aléatoires*, c'est-à-dire de suites qui satisfont toute propriété (au sens adéquat effectif de Martin-Löf) qui a lieu avec μ-probabilité égale à 1 pour les suites de $\{0,1\}^\infty$. Le théorème suivant est extrait de [8] :

Théorème 2.2. *Soit* μ *une mesure calculable. Une suite binaire infinie* ω *est* μ*-aléatoire au sens de Martin-Löf si et seulement s'il existe un* n_0 *tel que* $K(\omega_{1:n}) \geq -\log\mu(\omega_{1:n})$ *pour tout* $n > n_0$.

Notons que pour $\mu = \lambda$, la distribution uniforme, on a $-\log\lambda(\omega_{1:n}) = n$, ce qui redonne le théorème 2.2. On peut étendre la notion de μ-« aléatoirité » aux chaînes finies d'une manière adéquate.

3. La théorie algorithmique du chaos

Quand les physiciens ont affaire à un système chaotique, ils croient que tout repose sur un système déterministe sous-jacent mais que les trajectoires *individuelles* sont *imprévisibles* à partir des états observables. Malheureusement, dans

le cadre classique cela ne peut pas être exprimé précisément et on doit alors recourir à l'ersatz des ensembles statistiques d'états et de trajectoires, et s'engager dans les subtilités du raisonnement probabiliste, qui est essentiellement en dehors du sujet. Mais en utilisant la complexité de Kolmogorov, on peut exprimer directement la chaoticité du système et les propriétés d'imprédictibilité des trajectoires individuelles, ce qui est précisément l'intuition qu'on veut exprimer. C'est cette idée que nous allons défendre maintenant.

Par simplicité, on va supposer un temps discret, décrivant $\mathbb{N}$. Dans un système déterministe X l'*état* du système au temps t est X_t. L'*orbite* du système est la suite des états consécutifs $X_0, X_1, X_2, \ldots$. Par simplicité, nous supposons que les états sont des éléments de $\{0,1\}$. Les définitions ci-dessous se généralisent facilement. À chaque système, qu'il soit déterministe ou aléatoire, nous associons une mesure μ sur l'espace des orbites $\{0,1\}^\infty$, en définissant $\mu(x)$, pour $x \in \{0,1\}^*$, comme étant la probabilité qu'une orbite commence par x.
Étant donné un segment initial $X_{0:t}$ de l'orbite, nous voulons calculer X_{t+1}. Et même si ce n'est pas possible, nous aimerions en calculer une prédiction meilleure que celle obtenue en lançant une pièce au hasard.

Définition 3.1. *Soit $S = \{0,1\}^\infty$, muni de la mesure μ, l'ensemble des orbites d'un système X. Soit ϕ une fonction partielle récursive et soit $\omega \in S$. Définissons*

$$\zeta_i = \begin{cases} 1 & \textit{si } \phi(\omega_{1:i-1}) = \omega_i \\ 0 & \textit{sinon.} \end{cases}$$

Le système est chaotique *si, pour toute fonction calculable ϕ, on a*

$$\lim_{t\to+\infty} \frac{1}{t} \sum_{i=0}^{t-1} \zeta_i = \frac{1}{2},$$

avec μ-probabilité 1.

Remarque 3.1. Dans le cas d'un système chaotique, aucune fonction calculable ϕ ne prédit mieux les évolutions du système qu'un lancer de pièce. Dans cette définition de la chaoticité, l'exigence essentielle est formulée en termes d'imprévisibilité algorithmique des orbites. Les propriétés d'instabilité du système sont exprimées par la mesure μ (comme dans la définition 3.1) induite par le système. Soit, par exemple, μ la mesure uniforme (habituellement notée λ). Une orbite telle que $\omega = \omega_1 \ldots \omega_n 11 \ldots$ est parfaitement prévisible après les n premiers chiffres. En fait, la prévisibilité par une fonction calculable appropriée est vérifiée par toutes les suites ω qui sont calculables, comme les développements binaires des rationnels mais aussi de nombres transcendants comme $\pi = 3,14\ldots$. Cependant, pour tout $\omega \in S$, et tout $\varepsilon > 0$, les ω' tels que $|\omega - \omega'| \leq \varepsilon$, c'est-à-dire les ω' dans la boule de centre ω et de rayon ε, sont imprévisibles avec une probabilité uniforme égale à 1. Cela vient du fait que l'ensemble des suites aléatoires au sens de Martin-Löf dans la boule de rayon ε a la même mesure uniforme que l'ensemble de toutes les suites de cette boule. Ainsi, d'une infime perturbation aléatoire d'une orbite, résultera presque sûrement une orbite imprévisible.

Néanmoins, il n'est pas vrai que l'imprévisibilité entraîne l'instabilité. Si, par exemple, μ concentre toute sa probabilité en $\mu(\omega) = 1$ pour une suite $\omega = \omega_1\omega_2\ldots$

telle que $K(\omega_{1:n}) \geq n$ pour tout n, c'est-à-dire que ω est aléatoire au sens de Martin-Löf, alors les éléments subséquents ω_i sont complètement imprévisibles à partir de la donnée des éléments antérieurs. Cependant l'orbite est complètement stable, elle est en fait déterministe. Le point clé est bien sûr que l'orbite est une suite fixée, bien que tout à fait non calculable. Nous laissons au lecteur la construction d'exemples semblables, où l'orbite n'est pas complètement fixée, n'est pas instable, mais est cependant complètement imprévisible.

Remarque 3.2. Dans la théorie du chaos, on considère habituellement des systèmes déterministes X dont les états x, appartenant à un certain domaine R, évoluent par pas discrets selon la formule $x_{n+1} = f(x_n)$, où les x_n sont des nombres réels, ou des vecteurs de nombres réels, x_0 est la valeur initiale donnée, et f une fonction transformant l'état présent du système en son état suivant. La considération de nombres réels arbitraires n'a pas de signification physique, puisqu'ils exigent une précision infinie, ce qui n'est pas accessible à la mesure physique. De plus, on ne connaît aucune loi ni constante physique exacte avec une précision supérieure à, disons, dix décimales, et par conséquent la même chose doit être vraie pour l'opérateur f d'évolution du système. Par suite, dans l'analyse du comportement du système, on remplace les vraies valeurs x_n par une approximation finie représentée par une classe d'équivalence contenant x_n. Ces classes d'équivalence représentent les différents états que l'on peut effectivement « distinguer », « observer », ou « mesurer ». Par exemple, si les x_n appartiennent au domaine $[0,1]$, alors nous pouvons choisir de diviser le domaine $[0,1]$ en deux parties égales, $R_0 = [0,\frac{1}{2}[$ et $R_1 = [\frac{1}{2},1]$. Ensuite, nous considérons un système (X_n) défini par $X_n = i$ si $x_n \in R_i$ ($n = 0,1,\ldots$ et $i \in \{0,1\}$). Observons que cela définit à la fois la valeur initiale X_0 et les états suivants du système $X_1, X_2, \ldots$ à partir du système original X de valeur initiale x_0. Le « chaos » est défini pour le système dérivé (X_n) qui représente l'évolution des états « distinguables ». Il devient maintenant clair que pour des états initiaux différents x_0 et x_0', même appartenant à la même classe d'équivalence, disons R_0, de sorte que $X_0 = 0$, les orbites $X_0 = 0, X_1, \ldots$ peuvent être très différentes à partir de X_1. S'il apparaît que l'orbite $X_0 = 0, X_1, \ldots$ est, en un sens approprié, imprévisible, même si x_0 et l'opérateur d'évolution f sont connus, alors nous dirons que le système est « chaotique ».

Notre définition 3.1 est fondée sur ce qui suit. Soit X un système défini par $x_{n+1} = f(x_n)$. Supposons que nous choisissions aléatoirement l'état initial x_0 dans son domaine R selon une mesure ρ. Autrement dit, si $x_0 = x_{0,1}x_{0,2}\ldots$, alors la probabilité de choisir $x_{0,1}\ldots x_{0,r}$ est $\rho(x_{0,1}\ldots x_{0,r})$. (Cela n'exclut pas la possibilité de choisir un x_0 particulier avec la probabilité 1 : il suffit de concentrer toute la probabilité de ρ sur x_0.) Quand on considère le système dérivé des états distinguables, la probabilité de l'état initial $X_0 = i$ est $\rho(R_i)$ ($i \in \{0,1\}$), mais bien que la probabilité de l'état suivant X_1 soit déterminée complètement par ρ et f, elle est sensible aux variations de l'une ou l'autre, et de même à toutes les étapes suivantes $X_2, X_3, \ldots$. Néanmoins, f et ρ déterminent complètement la probabilité de chaque segment initial de chaque orbite des états distinguables. Pour un segment initial $\omega_0\ldots\omega_n$, nous désignons alors cette probabilité par $\mu(\omega_0\ldots\omega_n)$, et cela définit la mesure μ de la définition 3.1. Observons que si f et ρ sont calculables d'une certaine manière, alors il en est de même pour

μ. Par exemple, si $f(\omega_0\omega_1\ldots) = \omega'_0\omega'_1\ldots$ est tel qu'il existe une fonction g calculable, monotone strictement croissante et une fonction calculable h telle que $h(\omega_0\ldots\omega_n) = \omega'_0\ldots\omega'_{g(n)}$, et si, de plus, ρ est calculable, alors

$$\mu(X_0\ldots X_m) = \rho\{\omega_0\ldots\omega_n \ : g(n) = m \text{ et } h(\omega_0\ldots\omega_n) = \omega'_0\ldots\omega'_m \text{ et } \omega'_j \in R_i \text{ si et seulement si } X_j = i \ (j = 0,\ldots,m, i \in \{0,1\})\}.$$

Le système est *uniformément instable* si pour tout ω et tout $\varepsilon > 0$, la boule $\{\omega' \ : |\omega - \omega'| \leq \varepsilon\}$ de rayon ε a un ensemble correspondant d'orbites $\{X'_0X'_1\ldots\}$ qui est, en un sens approprié, «dense » dans l'ensemble de toutes les orbites possibles du système. C'est par exemple le cas si cet ensemble est égal à l'ensemble de toutes les orbites possibles du système.

Le système est *uniformément imprévisible* si pour tout ω et tout $\varepsilon > 0$, la boule $\{\omega' \ : |\omega - \omega'| \leq \varepsilon\}$ de rayon ε produit un ensemble d'orbites dans lequel le sous-ensemble des suites aléatoires au sens de Martin-Löf est de mesure pleine. (Ici nous entendons l'« aléatoirité » au sens de Martin-Löf relativement à la distribution uniforme, et « de mesure pleine » relativement à la mesure μ induite sur l'ensemble des orbites $X_0, X_1, \ldots$ des états distinguables du système.)

Il est clair que certains systèmes peuvent être à la fois uniformément imprévisibles et uniformément instables, mais ils peuvent aussi n'être que l'un des deux. La *chaoticité* de la définition 3.1 peut résulter de l'une quelconque de ces trois possibilités.

3.1. La multiplication par 2 modulo 1

Un exemple bien connu de système chaotique est la *multiplication par 2, modulo* 1 (voir [5]). On considère le système *déterministe* D dont l'état initial est un nombre réel dont la représentation binaire est $x_0 = 0.\omega$ ($\omega \in S$), et défini par

$$x_{n+1} = 2x_n \pmod 1, \tag{14.8}$$

où $(\mathrm{mod}\ 1)$ signifie que l'on supprime la partie entière. Ainsi, toutes les itérées de x_0 sous la transformation (14.8) appartiennent à l'intervalle unité $[0,1]$. Celui-ci correspond à ce qu'en physique on appelle « l'espace des phases ». Nous pouvons partitionner cet espace des phases en deux cellules, la cellule gauche $R_0 = [0, \frac{1}{2}[$ et la cellule droite $R_1 = [\frac{1}{2}, 1]$. Ainsi x_n appartient à la cellule gauche R_0 si et seulement si le n-ième chiffre de ω est 0.

On peut obtenir la multiplication par 2 modulo 1 de la façon suivante. Dans la théorie du chaos, [3], les gens ont étudié pendant des années le système logistique à temps discret L_α

$$y_{n+1} = \alpha y_n(1 - y_n)$$

qui applique l'intervalle unité dans lui-même si $0 \leq \alpha \leq 4$. Quand $\alpha = 4$, en posant $y_n = \sin^2 \pi x_n$, nous obtenons :

$$x_{n+1} = 2x_n \pmod 1.$$

Théorème 3.1. *Il existe un système chaotique (par exemple l'application D de doublement modulo 1 et l'application logistique L_α pour certaines valeurs de α, comme $\alpha = 4$), où la mesure μ de la définition 3.1 est la distribution uniforme (la mesure « pile ou face » λ, où $\lambda(x) = 2^{-l(x)}$, probabilité d'obtenir la chaîne binaire finie x avec une pièce équilibrée).*

Démonstration.

Nous démontrons que D est un système chaotique. Puisque L_4 se ramène à D par spécialisation, cela montre que L_4 est également chaotique. Supposons que ω est aléatoire. Alors, d'après le théorème 2.1,

$$C(\omega_{1:n}) > n - 2\log n + O(1). \tag{14.9}$$

Soit ϕ une fonction partielle récursive quelconque. À partir de ϕ et ω, construisons ζ comme dans la définition 3.1.

Raisonnant par l'absurde, supposons qu'il existe un $\varepsilon > 0$ tel que

$$\left| \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \zeta_i - \frac{1}{2} \right| \geq \varepsilon.$$

Alors, il y a un $\delta > 0$ tel que

$$\lim_{n\to\infty} \frac{C(\zeta_{1:n})}{n} \leq 1 - \delta. \tag{14.10}$$

Nous démontrons cela comme suit. Le nombre de suites binaires de longueur n où les nombres de 0 et de 1 diffèrent par une quantité d'au moins εn est

$$N = 2 \cdot 2^n \sum_{m=(\frac{1}{2}+\varepsilon)n}^{n} b(n,m,\frac{1}{2}) \tag{14.11}$$

où $b(n,m,p)$ est la probabilité de m succès sur n épreuves d'un processus de Bernoulli $(p,1-p)$: la distribution binomiale. Une estimation générale de la probabilité de déviation dans la distribution binomiale, où m est le nombre de résultats favorables de n épreuves dont la probabilité de succès est p, $0 < p < 1$, et $q = 1 - p$, est fournie par la majoration de Chernoff, [2, 4],

$$\Pr(|m - np| \geq \varepsilon n) \leq 2e^{-(\varepsilon n)^2/3n}. \tag{14.12}$$

Par conséquent, nous pouvons décrire chacun des éléments $\zeta_{1:n}$ en question en nous donnant n et εn par des descriptions auto-délimitantes de $2\log n + 4\log\log n$ chiffres, et en désignant la chaîne concernée dans un ensemble restreint d'au plus N éléments, donc en $\log N$ chiffres, où

$$N = 2^n \Pr(|m - np| \geq \varepsilon n) \leq 2^{n+1} e^{-(\varepsilon n)^2/3n}.$$

Par suite,

$$C(\zeta_{1:n}) \leq n - \varepsilon^2 n \log e + 2\log n + 4\log\log n + O(1).$$

Nous pouvons donc choisir

$$\delta = \varepsilon^2 \log e - \frac{2\log n + 4\log\log n + O(1)}{n}.$$

Ensuite, étant donné ζ et ϕ nous pouvons reconstruire ω comme suit :

for $i = 1,2,\ldots$ do :
if $\phi(\omega_{1:i-1}) = a$ and $\zeta_i = 0$ then $\omega_i := \neg a$
else $\omega_i := a$.

Par conséquent,

$$C(\omega_{1:n}) \leq C(\zeta_{1:n}) + K(\phi) + O(1). \tag{14.13}$$

Maintenant, les équations (14.9), (14.10) et (14.13) donnent la contradiction souhaitée. Par le théorème 2.1, on a :

Assertion 1. L'ensemble des ω vérifiant l'équation (14.9) a une mesure uniforme égale à 1.

Dans la définition de la multiplication par 2 modulo 1 nous avons déja noté ceci : partant d'un état initial $x_0 = 0,\omega_1\omega_2\ldots$ la multiplication par 2 donne la trajectoire $X_0, X_1, \ldots$ avec $X_0 = i$ si et seulement si $x_0 \in R_i$ et $X_j = \omega_j$ pour $j = 1,2,\ldots$ et $i \in \{0,1\}$. Par suite,

Assertion 2. Si nous choisissons l'état initial x_0 selon la loi de probabilité uniforme dans $[0,1]$, alors la mesure induite sur l'ensemble des trajectoires des états distinguables $X_0, X_1, \ldots$ de la multiplication par 2 modulo 1, est la mesure uniforme.

Ensemble, les assertions 1 et 2 prouvent le théorème. □

Dans [5] l'argument est le suivant. En supposant que l'état initial est tiré aléatoirement dans $[0,1[$ selon la mesure uniforme λ, nous pouvons utiliser des arguments de complexité pour montrer que l'orbite observable de la multiplication par 2 modulo 1 ne peut pas être mieux prédite qu'un jeu de pile ou face. En effet avec λ-probabilité égale à 1, l'état initial tiré sera une suite infinie aléatoire au sens de Martin-Löf. Par définition, de telles suites ne peuvent pas être plus effectivement prédites qu'un jeu de pile ou face aléatoire, voir [9].

Mais dans ce cas nous n'avons pas besoin d'en faire autant. L'orbite observée est constituée essentiellement des chiffres consécutifs de l'état initial. Choisir l'état initial aléatoirement selon la mesure uniforme est isomorphe à lancer une pièce équilibrée pour l'engendrer. Néanmoins, l'approche adoptée ci-dessus nous permet de traiter la chaoticité même quand la condition initiale est sélectionnée selon une mesure non uniforme. De plus, on peut imaginer de coupler des états initiaux, calculables mais pseudo-aléatoires, avec des algorithmes de prédiction polynomiaux en temps. De tels développements feront partie d'un futur travail.

D'un point de vue pratique, on peut avancer qu'il n'y a pas d'intérêt à étudier les suites infinies : en pratique, l'entrée aura toujours une précision finie. Or une suite infinie qui est aléatoire peut très bien avoir un segment initial fini arbitrairement long qui soit complètement régulier. C'est pourquoi nous examinons la théorie avec entrées de précision finie dans le paragraphe suivant.

3.2. Le chaos avec une entrée de précision finie

Dans le cas des entrées réelles de précision finie, la distinction entre systèmes chaotiques et non chaotiques peut être faite précisément, mais elle est nécessairement graduelle. Cela nous amène à la définition suivante.

Définition 3.2. *Soit S, μ, ϕ, ω et ζ comme dans la définition 3.1. Un système déterministe de* précision d'entrée n *est* (ε, δ)-chaotique *si, pour toute fonction calculable ϕ, on a*

$$\left| \frac{1}{n} \sum_{i=1}^{n} \zeta_i - \frac{1}{2} \right| \leq \varepsilon,$$

avec une μ-probabilité d'au moins $1 - \delta$.

Ainsi un système est chaotique au sens de la définition 3.1, comme la multiplication par 2 modulo 1 ci-dessus, si et seulement s'il est de précision infinie et $(0,0)$-chaotique. Le système est *probablement approximativement imprévisible* : un système *pai*-chaotique.

Théorème 3.2. *Les systèmes D et L_α (pour certaines valeurs de α, comme $\alpha = 4$) ci-dessus sont $(\sqrt{(\delta(n) + O(1)) \ln 2/n}, 1/2^{\delta(n)})$-chaotiques pour toute fonction δ telle que $0 < \delta(n) < n$, où μ, dans la définition 3.2, est la mesure uniforme λ.*

Démonstration.
Montrons que D est (ε, δ)-chaotique. Puisque L_4 se ramène à D, cela implique que L_4 est aussi (ε, δ)-chaotique. Soit x une chaîne binaire de longueur n telle que :

$$C(x) \geq n - \delta(n). \tag{14.14}$$

Soit ϕ une fonction calculable en temps polynomial, et z défini par :

$$z_i = \begin{cases} 1 & \text{si } \phi(x_{1:i-1}) = x_i \\ 0 & \text{sinon.} \end{cases}$$

Alors, x peut être reconstruit à partir de z et de ϕ comme précédemment, et par conséquent :

$$C(x) \leq C(z) + K(\phi) + O(1).$$

Au vu de l'équation (14.14), cela entraîne

$$C(z) \geq n - \delta(n) - K(\phi) + O(1). \tag{14.15}$$

Nous analysons maintenant les nombres de 0 et de 1 dans z (nous noterons $\#\,\mathrm{uns}(z)$ le nombre de chiffres 1 dans z). En utilisant la majoration de Chernoff (équation (14.12)) avec $p = q = \frac{1}{2}$, on voit que le nombre N des z qui ont un excédent des 1 par rapport aux 0 tel que

$$\left| \#\,\mathrm{uns}(z) - \frac{n}{2} \right| \geqslant \varepsilon n,$$

est tel que :

$$N \leq 2^{n+1} e^{-(\varepsilon n)^2/n}.$$

Par conséquent, nous pouvons donner une description effective de z en donnant une description de ϕ, δ et de l'indice de z dans l'ensemble de taille N, en utilisant le nombre de chiffres suivant :

$$n - \varepsilon^2 n \log e + K(\phi) + K(\delta) + O(1). \tag{14.16}$$

Les équations (14.15) et (14.16) nous donnent

$$\varepsilon \leq \sqrt{\frac{\delta(n) + 2K(\phi) + K(\delta) + O(1)}{n \log e}}. \tag{14.17}$$

En faisant l'hypothèse simplificatrice que $K(\phi)$ et $K(\delta)$ sont des $O(1)$, cela fournit

$$\left| \#\,\mathrm{uns}(z) - \frac{n}{2} \right| \leqslant \sqrt{(\delta(n) + O(1)) n \ln 2}. \tag{14.18}$$

Le nombre de chaînes binaires x de longueur n telles que $C(x) < n - \delta(n)$ est au plus $2^{n-\delta(n)} - 1$ (il n'y a pas plus de programmes de longueur inférieure à $n - \delta(n)$). Par conséquent, la probabilité uniforme qu'un nombre réel commence par un segment initial x de longueur n tel que $C(x) \geq n - \delta(n)$ est donnée par :

$$\lambda\{\omega \ : \ C(\omega_{1:n}) \geq n - \delta(n)\} > 1 - \frac{1}{2^{\delta(n)}}. \tag{14.19}$$

Par suite, puisque nous utilisons la même application D que dans le théorème 3.1, la distribution initiale sur les entrées induit une distribution uniforme $\mu = \lambda$ sur l'ensemble correspondant de trajectoires des états distinguables. De plus, chacune de ces trajectoires est représentée par la même suite de chiffres que le développement binaire de l'état initial. Ainsi, d'après les deux assertions presque équivalentes de la démonstration du théorème 3.1, le système D est (ε,δ)-chaotique avec $\varepsilon = \sqrt{(\delta(n) + O(1)) \ln 2/n}$ et $\delta = 1/2^{\delta(n)}$. □

Remerciements

Cet article est issu d'un exposé de l'auteur à l'université de Waterloo, Canada, en 1991. Ce travail est en partie soutenu par l'UE à travers le groupe de travail NeuroColt II, le Réseau d'Excellence PASCAL, et les projets QAIP et RESQ.

L'idée de relier la primalité et la longueur des mots dans le codage préfixal, pour fournir une majoration du n-ième nombre premier par la méthode d'incompressibilité, est due à P. Berman, et la preuve présentée est due à J. Tromp.

BIBLIOGRAPHIE

[1] G. J. Chaitin, *A theory of program size formally identical to information theory*, J. Assoc. Comp. Mach. **22** (1975), p. 329-340.

[2] T. H. Cormen, C. E. Leiserson et R. L. Rivest, *Introduction to algorithms*, MIT Press, Cambridge, 1990.

[3] R. L. Devaney, *An introduction to chaos dynamical systems*, Addison-Wesley, 2nde édition, 1989.

[4] P. Erdös et J. Spencer, *Probabilistic methods in combinatorics*, Academic Press, New-York, 1974.

[5] J. Ford, *How random is a coin toss ?*, Physics Today **36** (1983), April, p. 40-47.

[6] A. N. Kolmogorov, *Three approaches to the definition of the concept "quantity of information"*, Problems in information transmission **1** (1965), p. 1-7.

[7] L. A. Levin, *Laws of information conservation (non-growth) and aspects of the foundation of probability theory*, Problems in information transmission **10** (1974), p. 206-210.

[8] M. Li et P. M. B. Vitányi, *An introduction to Kolmogorov complexity and its applications*, 2nde édition, Springer-Verlag, New York, 1997.

[9] P. Martin-Löf, *On the definition of random sequences*, Information and Control **9** (1966), p. 602-619.

[10] H. J. Rogers, Jr., *Theory of recursive functions and effective computability*, McGraw-Hill, 1967.