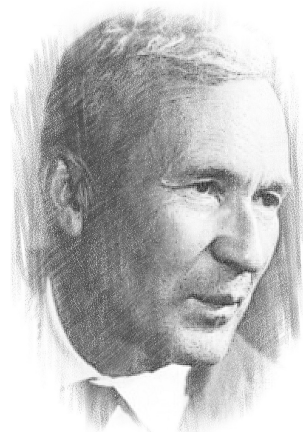
Chapitre 16

Complexité de Kolmogorov et nombres aléatoires

Andrei Nikolaïevitch Kolmogorov (1903 - 1987) est certainement l'un des mathématiciens les plus fameux et les plus prolifiques de *l'école de Moscou*, fondée par Dmitri Egorov et Nikolaï Luzin au début du XX^{e} siècle, et dont nous aurons l'occasion de reparler dans la section 29-1. À l'issue de sa thèse en 1929, effectuée sous la direction de Luzin, Kolmogorov a déjà publié de nombreux articles et acquis une renommée internationale. Il devient en 1931 professeur à l'université de Moscou et y mènera une brillante carrière durant laquelle il participera à la fondation de pans entiers des mathématiques modernes.



Andreï Nikolaïevitch Kolmogorov, 1903–1987

Ses travaux les plus connus concernent sans aucun doute l'axiomatisation, en 1933, de la théorie des probabilités [10], dont nous reparlerons dans le chapitre 17. Trente ans plus tard, Kolmogorov a des contributions importantes en topologie, en théorie des systèmes dynamiques, et a participé à la résolution du treizième problème de Hilbert. Sa carrière n'est pas terminée pour autant. Il est alors sur le point de démarrer un autre champ d'étude mathématique

qui connaîtra une fois de plus un retentissement considérable : la théorie algorithmique de l'aléatoire, complémentaire de la théorie des probabilités, avec cette fois l'utilisation d'un outil nouveau, l'informatique. Il développe notamment sa notion de complexité éponyme, dont nous nous efforcerons au long de ce chapitre de montrer la richesse.

1. Complexité de Kolmogorov

Informellement, la complexité de Kolmogorov d'un objet fini est une mesure de la quantité d'information nécessaire pour calculer cet objet. Si Kolmogorov [121] fut le premier à publier sur le sujet, cette idée remonte en fait aux travaux de Solomonoff [209].

Définition 1.1. On appelle *Machine* une fonction partielle calculable de $2^{<\mathbb{N}}$ vers $2^{<\mathbb{N}}$. La *complexité de Kolmogorov de σ relativement à M*, notée $C_M(\sigma)$, est la longueur de la plus petite chaîne τ telle que $M(\tau) = \sigma$. Formellement, $C_M(\sigma) = \min\{|\tau| \ : \ M(\tau) = \sigma\}$. ♢

La complexité de Kolmogorov d'une chaîne relativement à une machine M peut être vue comme une mesure de sa compression maximale possible via M. C'est donc une notion relative, qui dépend de la machine qui est utilisée, et si la machine en question ne fait rien, la notion n'est pas très intéressante. L'idée est bien sûr d'utiliser des machines qui compressent au maximum l'information.

1.1. Machine universelle

C'est Solomonoff le premier qui comprend la possibilité de définir une machine optimale, que l'on appelle aussi *universelle* : une machine dont le taux de compression sera au moins aussi bon que celui de n'importe quelle autre machine, à constante près.

Définition 1.2. Une machine U est dite *universelle* si, pour toute machine M, il existe une constante $c_M \in \mathbb{N}$ telle que $C_U(\sigma) \leqslant C_M(\sigma) + c_M$, pour toute chaîne σ. ♢

Une machine universelle compresse donc aussi bien que n'importe quelle autre machine M, mais à une certaine constante additive près c_M. Évidemment, si la constante est grande par rapport à la chaîne que l'on veut compresser, la notion perd de sa force, mais le point important est que la constante ne dépend que de la machine M et non de la chaîne que l'on veut compresser. Le poids de cette constante s'amenuise donc à mesure que la taille des chaînes considérées augmente. En introduisant son concept de

machine universelle, Solomonoff a bien évidemment montré qu'une telle machine existait, ce qui a aussi été démontré indépendamment par Kolmogorov.

Théorème 1.3 (Solomonoff [210], Kolmogorov [122])
Il existe une machine universelle.

PREUVE. Soit $(M_e)_{e\in\mathbb{N}}$ une énumération des machines, c'est-à-dire que M_e est la machine de code e. il suffit de définir la fonction calculable

$$U : 2^{<\mathbb{N}} \to 2^{<\mathbb{N}}$$

qui sur la chaîne $0^e 1\sigma$ renvoie la valeur de $M_e(\sigma)$. Étant donné une machine M_e, il est clair que, pour toute chaîne σ, on a

$$C_U(\sigma) \leqslant C_{M_e}(\sigma) + (e+1).$$

∎

Notation

On fixe dès à présent une machine universelle U, et l'on note $C(\sigma)$ la valeur $C_U(\sigma)$. Dès lors, $C(\sigma)$ sera la *complexité de Kolmogorov* de σ.

Remarquons que la machine universelle que nous avons fixée n'a pas d'importance : à constante additive près, toutes les machines universelles compressent les chaînes de manière optimale, et tous les théorèmes qui vont suivre sont indépendants du choix de celle-ci.

1.2. Les chaînes aléatoires

L'idée d'utilisation de la complexité de Kolmogorov comme mesure d'aléatoire est simple : moins une chaîne est compressible, plus elle est aléatoire. On montre facilement que des chaînes incompressibles de toutes les tailles existent.

Proposition 1.4. Pour tout n, il existe une chaîne σ de taille n telle que $C(\sigma) \geqslant n$. ⋆

PREUVE. Il s'agit d'un simple argument de comptage : U est une fonction, et associe donc à une chaîne au plus une autre chaîne. Aussi, pour tout n, le nombre de chaînes de taille strictement inférieure à n est de $\Sigma_{i=0}^{n-1} 2^i = 2^n - 1$. Il existe donc au moins une chaîne de taille n qui n'est calculée via U par aucune chaîne de taille strictement inférieure à n. ∎

Nous nous intéresserons dans la suite aux chaînes incompressibles à constante près : si une chaîne de taille 10 000 est compressible par un programme de taille 9990, mais pas mieux, elle peut être moralement considérée comme

« fortement » aléatoire. L'importance de cette constante s'effacera complètement quand nous poursuivrons dans la section 2 notre étude de l'aléatoire sur les préfixes d'objets infinis.

1.3. Le degré Turing de la complexité de Kolmogorov

Avec l'utilisation d'une machine universelle, la complexité de Kolmogorov d'une chaîne s'apparente à la taille du plus petit programme informatique capable de calculer cette chaîne. Il s'agit en quelque sorte de sa meilleure compression possible, si l'on ne prend pas en considération le temps nécessaire à sa décompression, qui peut s'avérer particulièrement long... Quant au temps nécessaire à sa compression, la situation est encore pire : il ne s'agit même plus d'un processus calculable !

1.3.1. La complexité de Kolmogorov n'est pas calculable

On donne une première démonstration du fait que la complexité de Kolmogorov n'est pas calculable, via une formalisation mathématique du paradoxe de Berry : « soit n le plus petit entier que l'on ne peut pas définir en moins de cinquante mots ». Le paradoxe — qui devrait apparaître clairement au lecteur — vient de ce que le mot « définir » est lui-même mal défini. Il suffit de le remplacer par « calculable ».

Proposition 1.5. La fonction $\sigma \mapsto C(\sigma)$ n'est pas calculable. ⋆

Preuve. Supposons que la fonction $\sigma \mapsto C(\sigma)$ soit calculable. Alors, on peut créer la fonction calculable $f : \mathbb{N} \to 2^{<\mathbb{N}}$ qui sur n renvoie la première chaîne σ — disons lexicographiquement — telle que $C(\sigma) > n$. En utilisant l'écriture en binaire des entiers, on peut définir la machine $M : 2^{<\mathbb{N}} \to 2^{<\mathbb{N}}$ qui sur la chaîne σ_n (qui est l'écriture binaire de n) renvoie donc $f(n)$.

Comme la taille nécessaire pour représenter n en base 2 est de $\log_2(n)$, on a ainsi $C_M(\sigma) \leqslant \log_2(n)$, pour toute chaîne $\sigma = f(n)$. Il y a donc une constante c_M telle que $C(\sigma) < \log_2(n) + c_M$, pour toute chaîne $\sigma = f(n)$. Dans le même temps, chacune de ces chaînes est choisie telle que $C(\sigma) > n$, ce qui donne $n < C(\sigma) < \log_2(n) + c_M$. Pour n suffisamment grand, tel que $n > \log_2(n) + c_M$, on a une contradiction. ∎

La fonction $\sigma \mapsto C(\sigma)$ n'est donc pas calculable. En revanche, elle est *approchable par le dessus*.

Définition 1.6. Une fonction $f : \mathbb{N} \to \mathbb{N}$ est *approchable par le dessus* si elle est la limite d'une suite $(f_n)_{n\in\mathbb{N}}$ de fonctions calculables avec, pour tout n, l'inégalité $f_{n+1} \leqslant f_n$. ♢

On peut définir l'approximation $(f_n)_{n\in\mathbb{N}}$ de la complexité de Kolmogorov en assignant à $f_0(\sigma)$ la taille du premier programme τ trouvé tel que $U(\tau)\downarrow=\sigma$, et en assignant à $f_{n+1}(\sigma)$ la taille du plus petit programme τ tel que $U(\tau)[n+1]\downarrow=\sigma$ si cette taille est plus petite que $f_n(\sigma)$, et $f_n(\sigma)$ sinon.

On montre facilement que les fonctions approchables par le dessus sont calculables avec l'arrêt des programmes informatiques

Exercice 1.7. Montrer que toute fonction approchable par le dessus est $\emptyset'$-calculable. ◇

1.3.2. La complexité de Kolmogorov est Turing-complète

Il est possible de renforcer la proposition 1.5 et de montrer que la connaissance de la complexité de Kolmorogov permet en fait de calculer le problème de l'arrêt. Il s'agit d'un bon exercice, pour lequel nous préparons ci-après le lecteur avec une proposition plus simple, évidente si l'on s'en tient à la construction qui a été faite ci-dessus d'une machine universelle, mais qui demande un peu de travail si l'on considère les machines universelles de manière abstraite.

Proposition 1.8. Soit X une représentation d'une machine universelle

$$U : 2^{<\mathbb{N}} \to 2^{<\mathbb{N}}$$

(par exemple, avec $X(\langle\sigma,\tau\rangle)=1$ ssi $U(\sigma)\downarrow=\tau$ et $X(\langle\sigma,\epsilon\rangle)=1$ ssi $U(\sigma)\uparrow$). Alors, $X\geqslant_T\emptyset'$. ⋆

Preuve. On définit la machine M telle que $M(0^e)\downarrow=0^s$ si s est le plus petit entier tel que $\Phi_e(e)[s]\downarrow$. Si un tel entier n'existe pas, le calcul $M(0^e)$ ne s'arrête pas. Soit une constante d telle que $C_U(\sigma)<C_M(\sigma)+d$, pour toute chaîne σ. Étant donné la connaissance de U, pour savoir si $\Phi_e(e)\downarrow$, il suffit de regarder $U(\sigma)$ pour toute chaîne σ de taille inférieure à $e+d+1$, de récupérer la plus grande valeur s telle que $U(\sigma)\downarrow=0^s$ pour l'une de ces chaînes σ, et de calculer $\Phi_e(e)[s]$. On a alors $\Phi_e(e)\downarrow\leftrightarrow\Phi_e(e)[s]\downarrow$. ∎

L'un des exercices suivants consiste à montrer que la connaissance de la complexité de Kolmogorov permet de calculer la machine universelle U qui lui est associée, et donc le problème de l'arrêt. La considération suivante sera utile, et présente aussi son intérêt propre : la proposition 1.5 ne montre pas seulement que la complexité de Kolmogorov est non calculable, mais aussi que pour toute fonction de compression $I : 2^{<\mathbb{N}} \to 2^{<\mathbb{N}}$ telle que $C_I(\sigma)\leqslant C(\sigma)$, la fonction $\sigma\mapsto C_I(\sigma)$ est non calculable. En revanche, comme nous le montre l'exercice 1.11, une telle fonction ne permet pas nécessairement de calculer le problème de l'arrêt, et *a fortiori* la complexité de Kolmorogov qui lui est associée non plus. Il est nécessaire pour cela d'utiliser la connaissance *exacte* de la complexité de Kolmogorov.