Université Louis-Pasteur de Strasbourg
Institut de Mathématique
(janvier 1978)

NOMBRES AU HASARD

de Borel à Martin-Loef

C. Dellacherie

AVERTISSEMENT

Ce qui suit est la rédaction (amplifiée) d'une conférence faite, à l'invitation de D. Foata rt H. Barreau, au Séminaire sur les Fondements des Sciences de l'U.L.P., en février 1976, conférence reprise et complétée au Colloquium Mathematicum Montis Regii des Universités de Montréal, en octobre 1976, à la demande de J.-C. Taylor.

Il s'agissait d'un exposé, à caractère plus ou moins historique, sur le développement de la notion de « suite de nombres au hasard ». Trop technique sans doute pour la part non « scientifique » de l'auditoire du Séminaire de l'U.L.P., il fut par contre « uniformément » bien accueilli par celui, mathématicien, du Séminaire de Montréal. Aussi, si malgré la demande itérée de H. Barreau, je n'avais pu me résoudre à rédiger la première conférence, j'avais rédigé la seconde, à la demande de J.-C. Taylor, à l'intention de la « Gazette des Mathématiques du Québec » (qui ne se réserve pas de copyright). Et c'est cette rédaction, légèrement remaniée en tenant compte de critiques formulées par C. Doléans-Dade et J. Martinet, qui est publiée ici.

L'exposé comporte deux parties distinctes tant du point de vue historique (la première partie, §1 à §4, couvre essentiellement la période 1900~1940, et la seconde, §5 à §9, après une introduction à la théorie des fonctions récursives, couvre la période 1960~…), que du point de vue technique : la première partie devrait être facilement accessible à tout mathématicien (hors quelques passages, écrits en petits caractères, qui, sans être indispensables, peuvent apporter des compléments utiles au lecteur connaissant un peu de théorie des probabilités) ; la seconde partie, sans être ésotérique (du moins, je l'espère), demande sans doute plus d'attention, sinon plus de « métier ».

Enfin, je voudrais, une fois pour toutes, prendre la précaution de dire que je n'ai pas assez consulté les œuvres originales (bel euphémisme !), ni réfléchi suffisamment aux questions soulevées, pour soutenir la validité de mes considérations historiques. J'estimerai en fait avoir atteint mon but si le lecteur a pris un certain plaisir à lire ce qui suit, et la seule validité que j'ose espérer est celle de la plupart (ne soyons pas trop téméraires !) des assertions mathématiques que je serai amené à faire.

INTRODUCTION

Il s'agit donc de voir comment les mathématiciens ont tenté, depuis le début du siècle jusqu'à maintenant (les deux noms dans le titre sont là pour noter deux périodes importantes), de donner un sens précis au mot « hasard » dans l'expression « suite de nombres au hasard ». Pour fixer les idées dans un cadre simple, et néanmoins très riche, nous nous contenterons de regarder les suites composées de 0 et de 1, où 0 et 1 sont « équiprobables » (ce dernier mot étant à peu près aussi vague que « au hasard »).

D'abord, de quelles suites s'agit-il ? Suites finies, ou suites infinies ? Les êtres idéaux étant généralement plus simples à étudier que les êtres concrets, nous nous bornerons ici à étudier les suites infinies (voir cependant la bibliographie). Depuis Hilbert, c'est maintenant un lieu commun de remarquer que le développement des mathématiques modernes se fait souvent en construisant « autour » des êtres concrets, des êtres idéaux plus simples à étudier, et, ensuite (si l'on n'est pas totalement absorbé par la beauté de ces êtres, qui peuvent devenir concrets par la pratique historique), en revenant à l'étude des êtres concrets, pour laquelle on utilise tout l'arsenal d'outils forgés dans (et par) la construction et l'étude des êtres idéaux. Et, quand nous disons « suites infinies », nous entendons[1] bien « suites actuellement infinies », ou nous ne serions pas à la hauteur de notre idéal ! Nous terminerons en fait notre exposé en jetant un coup d'œil (éclairé par le développement antérieur) sur les suites potentiellement infinies.

Ce choix des suites infinies de 0 et de 1 n'a pas été fait au hasard, si vous me permettez ce jeu de mots, puisque, quitte à remplacer 0 par « pile » et 1 par « face », on tombe sur le jeu de pile ou face, qui a été une des grandes sources de la théorie des probabilités. En fait, définir la notion de suite de 0 et de 1 « au hasard », c'est essayer de définir ce qu'est une partie « typique » du jeu de pile ou face, la pièce (non biaisée) étant jetée coup après coup *ad infinitum*.

Faisons à cette occasion une petite digression, pour convaincre le lecteur combien à la fois l'intuition première peut être inconséquente et finalement quand même justifiée. Définissons une suite de 0 et de 1 de la manière suivante : regardons le développement décimal de π

$$\pi = 3,14159265358979323846264338327 90\ldots$$

et codons les décimales 0, 1, …, 4 par « 0 » et les décimales 5, 6, …, 9 par « 1 ». On obtient la suite infinie

0001101101111100010101000100110…

Le lecteur est sans doute enclin, comme moi, à penser que cette suite est « au hasard ». Mais, j'ai aussi envie de dire que c'est le fait même du hasard de ne pouvoir être mécaniquement engendré (encore une bonne raison pour considérer des suites actuellement infinies), si bien qu'une suite au hasard de 0 et de 1 ne doit pouvoir être construite à l'aide d'un algorithme. Or, heureusement, on connaît de bons algorithmes pour calculer π…

La leçon à tirer de cela est qu'il n'existe pas une seule notion de suite de nombres « au hasard », mais toute une hiérarchie, certaines suites étant « plus au hasard » que d'autres. Et c'est cette hiérarchie que nous allons bientôt parcourir en respectant, en bon fonctionnaire,

[1] Il n'est pas nécessaire d'entendre ces subtilités pour la suite.

l'ordre hiérarchique qui, par ailleurs, colle très bien avec l'ordre historique de l'apparition des notions de suites de nombres au hasard.

1. UN PEU DE CALCUL DES PROBABILITÉS

Nous désignons, suivant la coutume, par Ω l'ensemble des suites infinies de 0 et de 1, *i.e.* l'ensemble de toutes les parties du jeu de pile ou face, infinies et logiquement possibles. Pour $\omega \in \Omega$ et $n \in \mathbb{N}$, $X_n(\omega)$ est le n-ième terme de ω ; autrement dit, X_n est la n-ième application coordonnée sur $\Omega = \{0, 1\}^{\mathbb{N}}$. Nous munissons maintenant Ω de la probabilité P du jeu de pile ou face (non biaisé) : en termes savants, que nous n'utiliserons pas, c'est l'unique loi sur la tribu borélienne de Ω telle que les applications X_n soient indépendantes, équidistribuées, et de distribution $P\{X_n = 0\} = P\{X_n = 1\} = ½$. Nous aurons en fait besoin de connaître peu de choses sur l'espace Ω et sa loi de probabilité P, et nous verrons ce peu après les quelques lignes suivantes en petits caractères, réservées aux initiés.

> L'espace $\Omega = \{0, 1\}^{\mathbb{N}} = (\mathbb{Z}/2\mathbb{Z})^{\mathbb{N}}$ est un groupe additif métrisable compact, et P est sa mesure de Haar normalisée. Par ailleurs, on peut identifier Ω à l'espace classique de Cantor (ensemble des $x \in [\![0, 1]\!]$ dont la décomposition en base 3 ne comporte pas de 2) et P en est alors la mesure canonique ; ou encore à $[\![0, 1]\!]$ tout entier, grâce à la décomposition en base 2 (les ambiguïtés bien connues sont négligeables), et P est alors la restriction de la mesure de Lebesgue.

Nous désignons par S l'ensemble des suites finies de 0 et de 1 et nous appelons, pour $s \in S$, <u>îlot d'indice</u> s, la partie I_s de Ω constituée par les suites infinies ω commençant par s : en abrégé, $I_s = \{\omega \mid s \dashv \omega\}$. Les îlots sont manifestement les êtres les plus concrets que l'on puisse définir au sujet des suites infinies. On notera que, si I_s et I_t sont deux îlots, ou bien I_s et I_t sont disjoints, ou bien l'un est contenu dans l'autre (avec égalité ssi $s = t$), suivant que…, le lecteur complètera lui-même. Ceci dit, on dit qu'une partie U de Ω est un <u>ouvert</u> de Ω si elle est égale à la réunion d'une suite (finie ou infinie) d'îlots (les îlots sont donc les ouverts les plus simples) : il est clair que tout ouvert peut s'écrire comme réunion d'une suite d'îlots disjoints, de diverses manières d'ailleurs (nous ne chercherons pas à dégager une « meilleure » écriture).

La probabilité P du jeu de pile ou face est une application $B \mapsto P[B]$), définie pour toutes les parties de Ω (en fait, pas tout à fait toutes, mais ça nous est bien égal ici), à valeurs dans $[\![0, 1]\!]$, et donnant un sens mathématique précis à la phrase « il y a une probabilité P[B] pour que ω appartienne à B » ; en particulier, on pose $P[\varnothing] = 0$ et $P[\Omega] = 1$. Nous n'aurons pratiquement besoin, pour la suite, que de connaître la probabilité d'un îlot, de savoir comment on calcule la probabilité d'un ouvert, et de comprendre ce qu'est un ensemble de probabilité nulle. Commençons par les îlots : si $s = (a_1, \ldots, a_n)$ est une suite finie de 0 et de 1, de longueur $|s| = n$, on pose

$$P[I_s] \;=\; P\{X_1 = a_1, \ldots, X_n = a_n\} \;=\; 2^{-n} \;=\; 2^{-|s|}$$

où l'expression « $P\{\ldots\}$ » peut se lire « probabilité pour que le premier terme de ω soit égal à a_1, et que…, et que le n-ième terme de ω soit égal à a_n ». L'indépendance des X_n s'exprime par le fait que

$$P\{X_1 = a_1, \ldots, X_n = a_n\} \;=\; P\{X_1 = a_1\} \times \ldots \times P\{X_n = a_n\},$$

chacun des $P\{X_i = a_i\}$ étant par ailleurs égal à ½. Maintenant, si U est un ouvert de Ω, égal à la réunion d'une suite (I_{s_n}) d'îlots disjoints, on pose tout simplement

$$P[U] \;=\; \sum_n P[I_{s_n}] \;=\; \sum_n 2^{-|s_n|} ;$$

le lecteur vérifiera sans peine que P[U] ne dépend pas de la décomposition choisie de U en îlots disjoints . Enfin, une partie N de Ω a une probabilité nulle (on dit encore que N est négligeable) si, pour tout $\varepsilon > 0$, il existe un ouvert U_ε contenant N tel que $P[U_\varepsilon] < \varepsilon$: par exemple, toute partie dénombrable[1] de Ω est négligeable. Noter aussi le résultat éminemment important suivant : la réunion d'une suite finie ou infinie d'ensembles négligeables est encore négligeable (petit exercice sur la série géométrique).

Une propriété ($\mathcal{P}$), vérifiée ou non par chaque $\omega \in \Omega$, est dite avoir lieu presque sûrement – en abrégé, p.s. – si l'ensemble de ω qui ne vérifient pas ($\mathcal{P}$) est négligeable. Voici un exemple fondamental d'une telle propriété. Posons, pour tout entier *n*,

$$S_n \;=\; X_1 + \ldots + X_n ;$$

la variable aléatoire S_n est égale au nombre de faces apparues au cours des *n* premiers jets. Nous nous intéressons au comportement de S_n / n quand *n* tend vers l'infini ; on a le résultat suivant, dû à Borel et connu sous le nom de la loi forte des grands nombres :

THÉORÈME. – La limite de S_n / n, quand *n* tend vers l'infini, existe p.s. et vaut p.s. ½. Autrement dit, il existe un ensemble négligeable N tel que, pour tout $\omega \notin N$, on ait $\lim_n (X_1(\omega) + \ldots + X_n(\omega)) / n = ½$.

Cela est tout à fait conforme à notre intuition : « en moyenne », on doit obtenir pile, ou face, une fois sur deux. Il est aussi intuitif que la limite ½ doit être atteinte de manière assez désordonnée. Et, de fait, on a le résultat suivant, qui montre que de grandes fluctuations sont inévitables (il existe un résultat beaucoup plus précis, appelé loi du logarithme itéré, que nous ne donnerons pas ici)

THÉORÈME. – La propriété suivante a lieu p.s. :

$$\forall\, n : \;\; \exists\, p \geq n : \;\; \exists\, q \geq n : \;\; S_p < (n - \sqrt{n})/2 \text{ et } S_q > (n + \sqrt{n})/2$$

Les propriétés énoncées dans les deux théorèmes précédents sont des exemples de propriétés asymptotiques : ($\mathcal{P}$) est asymptotique si le fait que ω vérifie ou non ($\mathcal{P}$) ne dépend que du comportement à l'infini de ω. Il existe un résultat tout à fait général, dû à Kolmogorov et connu sous le nom de loi du tout ou rien, qui assure que, pratiquement, toute propriété asymptotique que l'on peut définir (en particulier, toute propriété asymptotique borélienne) est soit vraie p.s., soit fausse p.s..

Il est intuitif qu'une propriété ($\mathcal{P}$) fausse p.s. (auquel cas la propriété (non $\mathcal{P}$) est vraie p.s.) est très « contraignante », et que donc une suite ω vérifiant une telle propriété est très « particulière ». Aussi serait-il tentant de définir la notion de suite de 0 et de 1 « au hasard » comme suit (nous abrégeons l'expression « si et seulement si » en « ssi », suivant un usage maintenant assez répandu)

---

[1] Un ensemble sera dit dénombrable s'il est vide, fini non vide, ou infini dénombrable (*i.e.* si on peut ranger ses éléments en une suite indexée par $\mathbb{N}$). Par exemple, l'ensemble des suites infinies de 0 et de 1 constantes à partir d'un certain rang est dénombrable.

la suite ω est « aléatoire » ssi elle vérifie toute propriété p.s. vraie

Malheureusement cette définition est inconséquente : si, pour $w \in \Omega$, on définit la propriété $(\mathcal{P}_w)$ en posant que ω vérifie $(\mathcal{P}_w)$ ssi

$$\forall\, m : \ \exists\, n > m : \ \ X_n(\omega) \neq X_n(w)$$

alors $(\mathcal{P}_w)$ est une propriété asymptotique qui a lieu p.s., mais, comme w elle-même ne vérifie pas $(\mathcal{P}_w)$, aucune suite ne peut vérifier à la fois toutes les propriétés $(\mathcal{P}_w)$ quand w parcourt Ω. Nous n'aurions pas rencontré cette difficulté si nous nous étions contentés, dans notre définition d'une suite aléatoire, de considérer une suite $(\mathcal{P}_n)$ de propriétés vraies p.s. car, la réunion d'une suit d'ensembles négligeables étant encore négligeable, la conjonction de toutes les $(\mathcal{P}_n)$ serait encore vraie p.s.. Mais alors, quelle suite de propriétés p.s. vraies choisir ? Nous entrons là dans le vif du sujet, avec une idée un peu plus précise sur la notion de hiérarchie évoquée précédemment : plus la suite des $(\mathcal{P}_n)$ sera compréhensive, plus aléatoire sera une suite ω vérifiant chacune des $(\mathcal{P}_n)$.

Nous allons passer en revue maintenant différentes notions de suites aléatoires apparues historiquement (que le lecteur se rassure, nous ne serons pas exhaustif !). Noter bien, cependant, que le problème ne pouvait, historiquement, se poser dans les termes où nous venons de le poser. L'axiomatique du calcul des probabilités de Kolmogorov, nécessaire pour notre développement précédent, date de 1933, et c'est autour de cette année que s'est cristallisé le débat entre les tenants de la théorie de la mesure comme fondement des probabilités, comme Kolmogorov, et ceux, comme von Mises, pour qui ce devait être la notion de suite aléatoire (à définir correctement) qui devait être le principe premier. On sait que la théorie de Kolmogorov a triomphé : elle était plus « idéale », et, en dernier ressort, comme nous le verrons, c'est elle qui a permis, avec la notion de fonction récursive dégagée par les logiciens – aussi dans les années 30 –, à Martin-Loef de trouver en quelque sorte l'ultime bonne définition de la notion de suite aléatoire, en 1966.

2. LES NOMBRES NORMAUX DE BOREL (1909)

De toutes les propriétés asymptotiques vraies p.s., la plus fondamentale est sans nul doute la loi des grands nombres. Mais, demander à une suite ω, pour être qualifiée d'aléatoire, de vérifier seulement cette loi est, sans nul doute aussi, insuffisant. Personne ne sera enclin à considérer que la suite ω = 010101010101… est aléatoire : la configuration « 01 » revient trop souvent alors que la configuration « 00 » n'apparaît jamais, et de même pour les configurations de longueur > 2. Or, une suite finie s étant donnée, on s'attend à ce que, dans une suite aléatoire ω, la fréquence relative d'apparition de la configuration s, par rapport à celles de même longueur $|s|$, soit « finalement » égale à $2^{-|s|}$, vu qu'il y a $2^{|s|}$ suites finies de longueur $|s|$. Une manière d'écrire cela (il y en a d'autres, un peu différentes, mais finalement équivalentes) est de poser, si $s = (a_1, a_2, \ldots, a_k)$,

$$\begin{aligned} Y_n(\omega) &= 1 \qquad \text{si} \qquad X_n(\omega) = a_1,\, X_{n+1}(\omega) = a_2,\, \ldots,\, X_{n+k}(\omega) = a_k, \\ Y_n(\omega) &= 0 \qquad \text{sinon} \end{aligned}$$

pour tout entier n, et de demander que $\lim_n (Y_1(\omega) + \ldots + Y_n(\omega))/n$ existe et vaille $2^{-k} = 2^{-|s|}$ – et cela, pour toute suite finie $s \in S$. Comme l'ensemble S est dénombrable, on définit ainsi une suite de propriétés asymptotiques, du même type que celle de la loi des grands nombres, qui ont aussi lieu p.s.. Nous dirons que la suite ω est aléatoire au sens de Borel si elle vérifie

ces propriétés, et, d'après ce qui vient d'être dit, presque toute suite dans Ω est aléatoire en ce sens.

En fait, Borel présentait les choses de manière un peu différente. Il prenait le segment $[\![0, 1]\!]$, muni de la mesure de Lebesgue, puis développait $x \in [\![0, 1]\!]$ en base 2, et donnait un critère équivalent à celui donné ci-dessus (revenant à développer x suivant les bases puissances de 2) pour qualifier son nombre x d'entièrement normal ; et, finalement, démontrait que presque tout nombre de $[\![0, 1]\!]$ est entièrement normal. Il considérait aussi, ensuite, toutes les bases possibles, et obtenait alors la notion de nombre absolument normal, dont nous ne parlerons pas.

Les suites aléatoires au sens de Borel sont déjà bien aléatoires, si je puis m'exprimer ainsi : elles résistent bien à de nombreux tests classiques de stochasticité (voir, à ce sujet, le livre de Knuth cité dans la bibliographie). Mais Borel lui-même avait donné une construction explicite d'une telle suite. On a donc des algorithmes pour en fabriquer, et, en fait, Champernowne a donné, en 1935, l'exemple très simple suivant d'algorithme pour fabriquer une suite aléatoire au sens de Borel : ordonner les suites finies de 0 et de 1 suivant leur longueur, puis, pour les suites de même longueur, les ordonner suivant l'ordre lexicographique ; écrire enfin le tout à la queue le leu. Cela donne (les barres sont des lignes imaginaires pour faciliter la lecture… et l'écriture !)

0|1|00|01|10|11|000|001|010|011|100|101|110|111|0000|0…

Pour les suites infinies composées de 0, 1, 2, …, 9, Champernowne a aussi donné l'exemple frappant suivant de suite aléatoire au sens de Borel (nous laissons au lecteur le soin de dégager l'algorithme) :

01234567891011121314151617181920212223242526272829303 13…

Par ailleurs, on ne sait toujours pas, me semble-t-il, si les suites construites à partir de $\pi = 3{,}14…$ sont aléatoires au sens de Borel.

Les suites aléatoires au sens de Borel ont des propriétés paradoxales, que vérifieront aussi les suites, encore plus aléatoires, que nous verrons par la suite. Il est clair que, si ω est une suite de Borel et si s est une suite finie, alors la configuration s apparaît, et même une infinité de fois, dans ω. Or, il est facile d'imagine un codage de l'alphabet français, blanc et ponctuations comprises, du type « alphabet Morse », à l'aide des signes 0 et 1. Et on est alors assuré de trouver quelque part (et même en moult endroits), dans notre suite aléatoire ω, les œuvres complètes de V. Hugo, dans leur ordre chronologique, codées de cette manière. Autrement dit, on ne peut obtenir la stochasticité globale, qui implique en quelque sorte un manque de forme définie, sans avoir localement de grandes régularités. Bien entendu, cette régularité locale est noyée dans un tel « bruit de fond » qu'elle passe en général totalement inaperçue. On peut donner un sens précis, quantitatif, à ce qui vient d'être dit, à l'aide de la notion de complexité que nous verrons apparaître tout à la fin de l'exposé.

3. LES COLLECTIFS DE VON MISES (1919) ET WALD (1937)

Comme nous l'avons dit plus haut, l'idée de von Mises était de fonder le calcul des probabilités sur la notion de suite aléatoire. Voici, en gros, comment il désirait définir la notion de suite aléatoire, appelée par lui « collectif » : une suite infinie ω de 0 et de 1 est un

collectif (*i.e.*, intuitivement, une partie « typique » du jeu de pile ou face) ssi, non seulement ω vérifie la loi des grands nombres, mais également toute sous-suite infinie w de ω déterminée par une règle non anticipante (penser à un joueur qui peut choisir, juste après le n-ième jet, de participer ou non au jeu à l'occasion du $(n+1)$-ième jet, en fonction des n jets précédemment effectués). Plus précisément, définissons une règle de sélection $\mathcal{R}$ sur Ω de la manière suivante : $\mathcal{R}$ est une application $s \mapsto \mathcal{R}(s)$ de l'ensemble S des suites finies (suite vide $\varnothing$ comprise) dans $\{0, 1\}$, et, pour tout $\omega \in \Omega$, la règle de sélection $\mathcal{R}$ détermine une sous-suite $w = \mathcal{R}\omega$, éventuellement finie et même vide, en supprimant dans ω le signe $X_1(\omega)$ si $\mathcal{R}(\varnothing) = 0$ et, de manière générale, en supprimant le signe $X_{n+1}(\omega)$ de ω lorsque $\mathcal{R}(X_1(\omega)), \ldots, X_n(\omega)) = 0$. Dans ces conditions, la tentative de définition de von Mises peut s'écrie

> la suite ω est un collectif ssi, pour toute règle de sélection $\mathcal{R}$, la suite $\mathcal{R}\omega$, si elle est infinie, vérifie la loi des grands nombres.

Or, il est facile de voir que cette définition est inconséquente : la suite ω, pour être un collectif doit comporter une infinité de 0 et de 1, et il existe toujours une règle de sélection $\mathcal{R}$ déterministe (*i.e.* $\mathcal{R}(s)$ ne dépend que de $|s|$) telle que $\mathcal{R}\omega$ soit la suite infinie composée uniquement de 0. Cela ne gênait pas outre mesure von Mises (n'étant pas constructiviste de nature, je n'ai pas pris le temps de bien comprendre ses arguments), et, au fond, il avait raison, comme nous allons le voir maintenant.

C'est à Wald, l'un des fondateurs de la statistique mathématique, que l'on doit la mise en forme, à peu près correcte, de la notion de collectif. Wald considère d'abord le cas où l'on se donne un ensemble dénombrable $\{\mathcal{R}\}$ de règles de sélection, stable par composition (*i.e.* si $\mathcal{R}_1$, $\mathcal{R}_2$ sont des éléments de $\{\mathcal{R}\}$, alors la règle $\mathcal{R}_3$ définie[1] par $\mathcal{R}_3\,\omega = \mathcal{R}_2\,\mathcal{R}_1\,\omega$ appartient encore à $\{\mathcal{R}\}$), et définit la notion de collectif relativement à $\{\mathcal{R}\}$: c'est la définition précédente, dans laquelle on remplace « pour toute règle de sélection » par « pour toute règle de sélection appartenant à $\{\mathcal{R}\}$ ». Il montre alors que, pour $\{\mathcal{R}\}$ donné, presque toute suite ω est un collectif relativement à $\{\mathcal{R}\}$, et donne par ailleurs un « pseudo »-algorithme pour « construire » un tel collectif (nous reviendrons plus loin sur le « pseudo » et les guillemets). Puis vient un argument raisonnable, que nous mettrons en forme plus loin : l'ensemble des règles de sélection que l'on peut effectivement définir (penser à un programme définissant une telle règle) est dénombrable, vu que l'on travaille effectivement, en mathématique (ou en programmation), avec un alphabet dénombrable. Par conséquent, si l'on prend pour $\{\mathcal{R}\}$ l'ensemble des règles de sélection effectivement définissables, il est possible de définir la notion absolue de collectif.

En fait, cette notion absolue de collectif peut être bien mise en forme une fois correctement définie la notion de fonction calculable. Et les logiciens étaient justement en train, dans les années 30, de dégager une telle notion sous le nom de fonction récursive. Aussi n'est-il pas

---

[1] À toute règle $\mathcal{R}$ est associée une application $\widehat{\mathcal{R}}$ de S dans S comme suit : $\widehat{\mathcal{R}}(s)$ est la suite finie obtenue en supprimant le k-ième élément de s, pour $1 \le k \le |s|$, lorsque $\mathcal{R}(X_1(s), \ldots, X_{k-1}(s)) = 0$. Nous laissons au lecteur le soin de définir correctement la règle $\mathcal{R}_3$ à l'aide de $\widehat{\mathcal{R}}_1$ et $\widehat{\mathcal{R}}_2$. Par ailleurs, si $\{\mathcal{S}\}$ est un ensemble de règles de sélection, il existe un plus petit ensemble $\{\mathcal{R}\}$ de règles stable par composition et contenant $\{\mathcal{S}\}$; et $\{\mathcal{R}\}$ est encore dénombrable si $\{\mathcal{S}\}$ est dénombrable.

étonnant que Church, l'un des pionniers de la théorie des fonctions récursives, ait fini de rendre rigoureuse la définition de Wald en 1940. Nous reviendrons là-dessus, plus loin, après avoir esquissé la théorie des fonctions récursives (qui nous sera indispensable pour expliquer la notion de suite aléatoire au sens de Martin-Loef). Nous nous contenterons, pour le moment, de la plausibilité de l'argument de Wald, et de signaler que von Mises accepta la définition de Wald comme étant celle qu'il avait en vue.

Maintenant, est-ce qu'un collectif, au sens absolu, représente bien la notion intuitive d'une suite « vraiment » aléatoire ? D'abord, on peut montrer, assez facilement, que tout collectif est une suite aléatoire au sens de Borel (voir Knuth, p. 141) : on est donc monté dans la hiérarchie. Ensuite, aucune suite w, construite à l'aide d'un algorithme (effectivement spécifié), ne peut être un collectif. En effet, si w est une suite infinie dont les termes successifs sont fournis par un algorithme, on peut associer à w deux règles de sélection $\mathcal{R}_0$ et $\mathcal{R}_1$, déterministes et effectivement définissables, en posant pour $s \in S$ de longueur $(n - 1) \geq 0$,

$$\mathcal{R}_0(s) = 1 \quad \text{ssi} \quad X_n(w) = 0,$$
$$\mathcal{R}_1(s) = 1 \quad \text{ssi} \quad X_n(w) = 1$$

Alors, l'une des suites $\mathcal{R}_0 w$, $\mathcal{R}_1 w$ est infinie, et constante, si bien que w n'est pas un collectif. L'algorithme de Wald, dont nous avons parlé plus haut, est un pseudo-algorithme car il fait intervenir une numérotation de l'ensemble dénombrable des règles effectivement définissables (celles-ci sont rangées en une suite par cette numérotation) et cette numérotation ne peut être effectuée par un algorithme, si bien que l'algorithme de Wald ne peut être effectivement spécifié (il n'existe aucun programme qui pourrait l'effectuer). On aborde là un résultat excessivement important de la théorie des fonctions récursives, que nous retrouverons plus loin.

Un collectif, au sens absolu, est donc déjà quelque chose de bien aléatoire. Cependant, nous allons voir maintenant qu'il existe des collectifs qui sont encore trop réguliers pour être acceptés comme représentants d'une partie « typique » du jeu de pile ou face.

4. LA CRITIQUE DE VILLE (1939)

D'abord Ville montre que, étant donné un ensemble dénombrable $\{\mathcal{R}\}$ de règles de sélection (pas forcément effectivement définissables), il existe toujours un collectif w relativement à $\{\mathcal{R}\}$ tel que l'on ait

$$S_n(w) / n = (X_1(w) + \ldots + X_n(w) / n \geq ½$$

pour tout n, si bien que la limite ½ est atteinte à l'infini sans fluctuation autour d'elle. En conséquence, tout probabiliste rejettera l'idée que la notion de collectif puisse s'identifier avec la notion intuitive de partie typique du jeu de pile ou face.

Ensuite, Ville montre là où le bât blesse, du point de vue théorie du jeu de pile ou face, dans l'idée de von Mises : avec les règles de sélection, von Mises tient compte seulement du fait qu'un joueur peut choisir les moments où il participe au jeu, mais ne dit rien quant à la manière de miser du joueur. Or, il est intuitif qu'au jeu de pile ou face, un joueur partant au début avec n capital > 0 mais fini, et jouant tant que ce capital, qui diminue ou augmente de la valeur de sa mise à chaque coup joué, n'est pas nul, ne peut trouver un « truc » étant apelé, en

termes de joueur professionnel, une martingale, c'est ainsi que Ville a fait entrer et le nom et la notion de martingale en théorie des probabilités.

> Une parenthèse ici pour les professionnels (j'entends, les probabilistes, et non les joueurs !). C'est aussi Ville, dans le même ouvrage, qui a le premier défini et utilisé l'exponentielle du mouvement brownien au sens de la théorie des martingales.

On sait que la notion de martingale est devenue, grâce aux travaux fondamentaux de Doob, l'une des notions les plus importantes de la théorie moderne des probabilités et des processus stochastiques. Nous allons maintenant, en suivant Ville, définir la notion de martingale (en fait, de martingale positive) dans notre contexte.

Supposons que notre joueur possède, au début du jeu, un capital $M_0 \geq 0$, fini. Avant chaque coup n, il peut décider de miser une partie de son capital, qui vaut M_{n-1} à ce moment-là, sur 0 ou sur 1 ; lorsque sa mise est nulle, il ne prend pas part au jeu au n-ième coup : on retrouve les règles de sélection. Lorsque le coup n est joué, son capital devient $M_n = M_{n-1} \pm V_n$ où V_n est sa mise au n-ième coup, suivant qu'il a gagné ou perdu. Si nous prenons maintenant en compte toutes les parties logiquement possibles – *i.e.* tout Ω –, on est amené à poser la définition suivante : une martingale positive $\mathcal{M} = (M_n)_{n\geq 0}$ est une suite de fonctions ≥ 0 sur Ω telle que

1) M_0 est une constante.
2) M_n est, pour tout n, une fonction de $X_1, \ldots, X_n$. Autrement dit, $M_n(\omega)$ ne dépend que des n premiers termes de ω, et nous considérerons que M_n est aussi une fonction sur l'ensemble des suites finies de longueur n.
3) Pour tout $n \geq 0$ et toute suite $s = (a_1, \ldots, a_n)$ de longueur n, on a

$$M_{n+1}(a_1, \ldots, a_n, 0) + M_{n+1}(a_1, \ldots, a_n, 1) \;=\; 2\, M_n(a_1, , a_n)$$

En particulier, si $M_n(\omega) = 0$, alors $M_{n+k}(\omega) = 0$ pour tout $k > 0$.

Avant d'énoncer quelques résultats de Ville sur les martingales, définissons encore la notion, également très importante, de temps d'arrêt, dans notre contexte. Soit $\mathcal{R}$ une règle de sélection et définissons une application $N_{\mathcal{R}}$ de Ω dans $\{0\} \cup \mathbb{N} \cup \{\infty\}$ par

$$N_{\mathcal{R}}(\omega) \;=\; \inf \{n \geq 0 \mid \mathcal{R}(X_1(\omega), \ldots, X_n(\omega)) = 0\}$$

en convenant que $\inf \varnothing = \infty$. La fonction $N_{\mathcal{R}}$, qui est clairement un « temps d'arrêt » pour notre joueur, vérifie la condition suivante : pour $n \geq 0$, le fait que ω vérifie ou non l'égalité $N_{\mathcal{R}}(\omega) = n$ ne dépend que des n premiers termes $X_1(\omega), \ldots, X_n(\omega)$, si bien que $\{\omega \mid N_{\mathcal{R}}(\omega) = n\}$ est ou vide ou la réunion de certains des 2^n îlots I_s tels que $|s| = n$. On appelle temps d'arrêt toute fonction N de Ω dans $\{0\} \cup \mathbb{N} \cup \{\infty\}$ vérifiant cette condition. Par exemple, si $\mathcal{S}$ est une autre règle de sélection (éventuellement égale à $\mathcal{R}$), la fonction

$$N_{\mathcal{R},\mathcal{S}}(\omega) \;=\; \inf \{n > N_{\mathcal{R}}(\omega) \mid \mathcal{S}(X_1(\omega), \ldots, X_n(\omega)) = 0\}$$

est un temps d'arrêt. Par ailleurs, il est facile de voir que tout temps d'arrêt N est de la forme $N_{\mathcal{R}}$ pour une règle de sélection $\mathcal{R}$.

On peut alors énoncer le problème de l'existence d'une martingale, au sens du joueur professionnel, comme suit : étant donné un capital $C > 0$ de départ, existe-t-il une martingale positive $\mathcal{M} = (M_n)_{n\geq 0}$, au sens mathématique, telle que $M_0 = C$ et que

1) l'on puisse trouver une suite croissante (au sens large) de temps d'arrêt finis (N_k) de sorte que $M_{N_k(\omega)}(\omega)$ tende, en croissant, vers l'infini avec k, pour « suffisamment » de ω ?

2) tout au moins, il existe un temps d'arrêt fini N assurant que, « en moyenne », M_N soit $> M_0$? Plus précisément, M_N étant la fonction $\omega \mapsto M_N(\omega)$, est-il possile d'avoir $M_0 < E[M_N]$? l'espérance $E[M_N]$ vaut ici

$$E[M_N] = \sum_n E[M_n \times \mathbf{1}_{\{N=n\}}] = \sum_n 2^{-n} \left(\sum_{s \in S(n)} M_n(s)\right)$$

où $S(n)$ est l'ensemble des s de longueur n telle que I_s soit contenu dans $\{\omega \mid N(\omega) = n\}$ (noter que M_n est constante sur I_s pour tout $s \in S(n)$).

Intuitivement (tout au moins, lorsqu'on connaît déjà le résultat !), on s'attend à ce que la réponse à ces deux questions soit négative. Et c'est ce que Ville démontre en prouvant que, pour toute martingale <u>positive</u> $\mathcal{M} = (M_n)$, on a

1) $\limsup_n M_n = \lim_n \sup M_{n+k} < \infty$ <u>p.s.</u>

2) $E[M_N] = M_0$ pour tout temps d'arrêt borné N, et $E[M_N] \leq M_0$ pour tout temps d'arrêt fini N (en fait, Ville, obnubilé par l'équitabilité du jeu de pile ou face, affirme erronément qu'on a encore égalité lorsque N est fini non borné ; cela arrive – par exemple si $\mathcal{M}$ est bornée, ou encore si les mises V_n sont bornées (uniformément) et si $E[N] < \infty$ – mais ce n'est pas toujours le cas : il faut s'arrêter de jouer à temps…).

Et nos suites aléatoires, dans tout cela ? Nous y revenons, tout de suite après avoir cité encore deux résultats de Ville, qui ont un impact immédiat sur leuur hiérarchie :

a) Il existe une partie négligeable N^0 de Ω contenant, pour tout ensemble dénombrable $\{\mathcal{R}\}$de règles de sélection, un collectif w relativement à $\{\mathcal{R}\}$.

b) Pour toute partie négligeable N de Ω, il existe une martingale positive $\mathcal{M}$ telle que N soit contenu dans l'ensemble $\{\limsup_n M_n = \infty\}$, lui aussi négligeable (autrement dit, les martingales fournissent un procédé exhaustif pour « isoler les suites trop particulières »).

On en déduit, d'une part, l'existence d'une martingale positive $\mathcal{M}^0$ telle que, pour tout ensemble dénombrable $\{\mathcal{R}\}$ de règles, il existe un collectif w relativement à $\{\mathcal{R}\}$ vérifiant $\limsup_n M^0_n = \infty$: ce qui discrédite, une fois de plus, la notion de collectif, du point de vue de la théorie des jeux. Et, d'autre part, une nouvelle définition naïve (parce qu'inconséquente) d'une suite aléatoire, équivalente – si je puis m'exprimer ainsi – à celle vue vers la fin du §1 :

> la suite ω est aléatoire ssi, pour toute martingale positive $\mathcal{M}$, on a $\limsup_n M_n(\omega) < \infty$.

La critique de Ville ne troubla pas outre mesure Wald[1]. Après tout, la notion de collectif voulait refléter certaines propriétés asymptotiques[2] liées à la loi des grands nombres, et il était

---

[1] Les quelques lignes qui suivent s'inspirent, non pas directement de la lecture des articles de Wald, mais de celle d'une conférence de Martin-Loef (*cf.* la bibliographie).

[2] Malgré l'apparence trompeuse (j'en sais quelque chose !), les propriétés « être un collectif » et « $\limsup_n M_n(\omega) < \infty$ » ne sont pas asymptotiques. Il est cependant vrai que tout ensemble

bien possible qu'elle ne reflète pas toutes les propriétés asymptotiques. Par ailleurs, il doit bien n'y avoir qu'une quantité dénombrable d'ensembles négligeables effectivement définissables, et donc rien n'est perdu (hors peut-être la simplicité naturelle de la définition à la von Mises).

Nous verrons, au §7, que Wald n'avait point tort. Cependant, l'opinion de Ville, partagée à l'époque par de nombreux mathématiciens à cause de certains paradoxes fameux (*cf.* le paradoxe de Richard, au paragraphe suivant), était qu'on ne pouvait définir correctement les ensembles négligeables effectivement définissables, et donc que, le mieux que l'on pouvait faire, était de définir la notion de suite aléatoire relativement à une famille dénombrable d'ensembles négligeables, sans qu'il existe de notion absolue.

5. LE PARADOXE DE RICHARD (1905)

Ce paradoxe, qui prouvait, à l'aide d'un argument diagonal à la Cantor, que la notion de « nombre réel définissable en français », était inconséquente, jetait un voile de suspicion sur tout raisonnement reposant sur l'idée d'une « suite des objets d'un certain type, effectivement définissables dans une langue ». Le texte, très court, de Jules Richard, est si beau que je ne puis résister à l'envie de le reproduire *in extenso* avant d'en présenter une version bien actualisée. Il s'agit d'un fac simile des pages 295 et 296 du tome 30 de *Acta Mathematica* (1906) – je n'ai pu trouver la version de 1905, parue dans le tome 16 de la *Revue Générale des Sciences Pures et Appliquées*, mais, vu le titre, il s'agit sans doute du même texte.

> NdT – Le document original inclut effectivement, à cet endroit, un fac simile de la communication de Richard (*Acta Mathematica* **30**, 23 mai 1906) ; il est ici simplement reproduit comme citation.

LETTRE
À Monsieur le Rédacteur en Chef de la Revue Générale des Sciences

par

I. RICHARD
à DIJON

Dans son numéro du 30 mars 1905, la *Revue* signale certaines contradictions qu'on rencontre dans la théorie générale des ensembles.

Il n'est pas nécessaire d'aller jusqu'à la théorie des nombres ordinaux pour trouver de telles contradictions. En voici une qui s'offre dès l'étude du continu, et à laquelle plusieurs autres se ramèneraient probablement :

Je vais définir un certain ensemble de nombres, que j nommerai l'ensemble E, à l'aide des considérations suivantes :

Écrivons tous les arrangements deux à deux des vingt-six lettres de l'alphabet français, en rangeant ces arrangements par ordre alphabétique, puis, à la suite tous les arrangements trois à trois, rangés par ordre alphabétique, puis, à la suite, ceux quatre à quatre, etc. Ces arrangements peuvent contenir la même lettre répétée plusieurs fois, ce sont des arrangements avec répétition.

---

négligeable est contenu dans un ensemble négligeable asymptotique (*i.e.* saturé pour la relation d'équivalence $\omega \sim w$ ssi $\exists\, n : \forall\, k : X_{n+k}(\omega) = X_{n+k}(w)$).

Quel que soit l'entier p, tout arrangement des vingt-six lettres p à p se trouvera dans ce tableau, et comme tout ce qui peut s'écrire avec n nombre fini de mots est un arrangement de lettres, tout ce qui peut s'écrire se trouvera dans le tableau dont nous venons d'indiquer le mode de formation.

La définition d'un nombre se faisant avec des mots, et ceux-ci avec des lettres, certains de ces arrangements seront des définitions de nombres. Biffons de nos arrangements tous ceux qui ne sont pas des définitions de nombres.

Soit u=1 le premier nombre défini par un arrangement, u_2 le second, u_3 le troisième, etc.

On a ainsi, rangés dans un ordre déterminé, *tous les nombres définis à l'aide d'un nombre fini de mots*.

Donc : Tous les nombres qu'on peut définir à l'aide d'un nombre fini de mots forment un ensemble dénombrable.

Voici maintenant où est la contradiction. On peut former un nombre n'appartenant pas à cet ensemble.

« Soit p la n-ième décimal du n-ième nombre de l'ensemble E ; formons un nombre ayant zéro pour partie entière, et pour n-ième décimale $p + 1$, si p n'est égal ni à huit ni à neuf, et l'unité dans le cas contraire. Ce nombre N n'appartient pas à l'ensemble E. S'il était le n-ième nombre de l'ensemble E, son n-ième chiffre serait le n-ième chiffre décimal de ce nombre, ce qui n'est pas. »

Je nomme G le groupe de lettres entre guillemets.

Le nombre N est défini par les mots du groupe G, c'est-à-dire par un nombre fini de mots ; il devrait donc appartenir à l'ensemble E. Or, on a vu qu'il n'y appartient pas.

Telle est la contradiction.

Montrons que cette contradiction n'est qu'apparente. Revenons à nos arrangements. Le groupe de lettres G est un de ces arrangements ; il existera dans mon tableau. Mais, à la place qu'il occupe, il n'a pas de sens. Il y est question de l'ensemble E, et celui-ci n'est pas encore défini. Je devrai donc le biffer. Le groupe G n'a de sens que si l'ensemble E est totalement défini, et celui-ci ne peut l'être que par un nombre infini de mots. *Il n'y a donc pas contradiction*.

On peut encore remarquer ceci : L'ensemble de l'ensemble E et du nombre N forme un autre ensemble. Ce second ensemble est dénombrable. Le nombre N peut être intercalé à un certain rang k dans l'ensemble E, en reculant d'un rang tous les autres nombres de rang supérieur à k. Continuons à appeler E l'ensemble ainsi modifié. Alors le groupe de mots G définira un nombre N′ *différent de* N, puisque le nombre N occupe maintenant le rang k, et que le k-ième chiffre de N′ n'est pas égal au k-ième chiffre du k-ième nombre de l'ensemble E.

Tel qu'il est écrit, ce paradoxe a un air « très sémantique »[1]. C'est d'ailleurs l'avis de Richard lui-même (*cf.* sa résolution du paradoxe), et Peano a pu écrire « Exemplo de Richard non pertine ad Mathematica sed ad Linguistica » (*Rivista di Mat.* **8**, 1906) : réflexion emplie d'un humour bien involontaire quand on sait que Gödel (de son propre aveu) s'est inspiré des paradoxes « richardiens » pour démontrer son fameux théorème sur l'incomplétude de l'arithmétique axiomatisée à la Peano !

Nous allons présenter maintenant une version « moderne » du paradoxe de Richard, qui mettra en valeur son côté syntaxique. Encore un mot avant de s'y mettre. Au lieu de définir des nombres réels, nous allons définir des fonctions de $\mathbb{N}$ dans $\mathbb{N}$, ce qui revient à peu près au même, mais est mieux adapté à la suite de notre exposé. Ainsi, Richard, dans sa lettre, aurait pu écrire partout « fonction de N dans N » au lieu de « nombre » dans la définition de son ensemble $E = (u_n)$, et remplacer son groupe de lettres G par le suivant : « Formons la fonction v qui à l'entier n associe l'entier $u_n(n) + 1$ » : il aurait même obtenu ainsi un énoncé plus simple de son paradoxe (la raison de son choix tient au fait que son expression G formule exactement le procédé diagonal inventé par Cantor pour exhiber des nombres transcendants).

Considérons un langage de programmation pour ordinateur[2] : il a un alphabet fini, et une grammaire bien définie. Tout programme écrit dans ce langage est une suite finie de caractères de l'alphabet, arrangés suivant certaines règles de syntaxe qui font que, étant donnée une suite finie de caractères, on sait vérifier mécaniquement (il existe un programme pour ce faire) si oui ou non cette suite est un programme. Regardons alors les programmes calculant une fonction de $\mathbb{N}$ dans $\mathbb{N}$: il n'y en a qu'une quantité dénombrable, que nous rangeons en une suite (P_n), et nous appelons f_n la fonction calculée par P_n (on peut avoir $f_m = f_n$ pour $m \neq n$). Définissons une fonction g de $\mathbb{N}$ dans $\mathbb{N}$ en posant (comme plus haut : c'est un avatar du procédé de Cantor) :

$$g(n) = f_n(n) + 1$$

On a alors notre paradoxe de Richard : d'une part, la fonction g a bien l'allure d'une fonction programmable (*i.e.* qui peut être calculée à l'aide d'un programme), et d'autre part, si on suppose qu'il existe un entier k tel que $g = f_k$, on a

$$f_k(k) = g(k) = f_k(k) + 1$$

et on tombe sur une contradiction.

Un moment de réflexion convainc qu'ici, l'argument de Richard pour résoudre cette contradiction ne marche plus : car on ne biffe pas, ici, un arrangement de caractères de notre alphabet après en avoir interprété le sens (s'il en a un), mais uniquement pour des raisons syntaxiques (le fait que l'entrée et la sortie d'un programme soient des entiers est inscrit dans les notations même du programme). Alors, pour résoudre ce paradoxe, le lecteur qui a compris quelque chose à nos remarques à propos du « pseudo »-algorithme de Wald nous répondra sans doute que la fonction g n'est pas programmable, car on n'a pas pu ranger nos programmes en une suite (P_n) à l'aide d'un algorithme. Je rétorquerai à cela, pour le plonger

---

[1] Seul ce côté sémantique a été retenu dans le paradoxe de Berry – plus célèbre, mais, à mon avis, bien moins intéressant – qui va chercher « le plus petit entier qui n'est pas définissble en moins de seize mots français » à l'aide d'une expression qui ne comporte que quinze mots.

[2] Que le lecteur qui ne sait pas programmer ne se sente pas trop frustré : je suis dans son cas !

momentanément dans la perplexité, qu'il n'est pas difficile d'imaginer un programme Π pour calculer g :

1) Notre alphabet de programmation a, mettons, $b - 1$ éléments, que nous identifions aux entiers 1, 2, …, $b - 1$; cela permet d'identifier toute suite finie de caractères à un entier que nous appellerons le code de notre suite finie, et de ranger ces suites finies suivant la grandeur de leur code.

2) Ceci fait, on sait vérifier mécaniquement (= il existe un programme Q) si une suite de caractères (= un entier, son code, entrée de Q) est, oui ou non (= sortie 0 ou 1 de Q), un programme écrit pour calculer une fonction de $\mathbb{N}$ dans $\mathbb{N}$ (*i.e.* l'entrée est une variable entière, et la sortie aussi).

3) Il existe alors un programme P qui, pour l'entrée n, dresse la liste des n premiers codes k de suites finies tels que $Q(k) = 1$. Le dernier code écrit, pour l'entrée n, est alors le code de notre programme P_n.

4) Considérons enfin le programme Π défini, en gros, comme suit : si l'entrée vaut n, Π effectue d'abord le programme P sur l'entrée n et obtient ainsi, comme résultat intermédiaire, le code de P_n ; puis, décodant ce code en instructions de programmation, il se met à calculer la sortie de P_n comme résultat intermédiaire ; ceci fait, il ajoute 1 à ce résultat, et le sort. Est-ce que Π ne calcule pas g ?

En fait, la résolution du paradoxe est cachée sous les mots « écrit pour calculer » du point 2), et « ceci fait » du point 4). On sait bien vérifier si, oui ou non, une suite finie S de caractères est un programme ayant « l'intention » de calculer une fonction de $\mathbb{N}$ dans $\mathbb{N}$: le programme Q nous donne la réponse, au bout d'un temps fini. Mais rien ne prouve (et c'est ce qui arrive souvent) que, pour certaines entrées, ce programme S, plein de bonnes intentions, nous fournira un jour sa sortie : il peut y avoir de telles « boucles » à l'intérieur du programme que le calcul – s'il est effectivement mis en route – ne puisse jamais être achevé (nous ne nous posons pas ici de problèmes « pratiques » de limitation des mémoires, ni de la longueur « pratiquement » excessive d'un calcul : notre ordinateur st une machine « idéale »). Par conséquent, il faut considérer qu'en général l'un de nos programmes P_n calcule ce que nous appellerons une semifonction de $\mathbb{N}$ dans $\mathbb{N}$, à savoir une fonction qui n'est définie éventuellement que sur une partie de $\mathbb{N}$: le domaine de définition de cette semifonction est exactement l'ensemble des entrées n pour lesquelles le programme fournit une sortie au bout d'un temps fini de calcul (soit encore les entrées pour lesquelles l'algorithme représenté par le programme converge : on entend ici par convergence l'obtention du résultat au bout d'un n ombre fini d'étapes). Au fond, on retrouve ici, sous le nom de semifonction, ce qu'on appelait (au moins, du temps de ma prime jeunesse) fonction dans l'enseignement secondaire : onnous donnit un algorithme (sous la forme, par exemple, d'une frction rationnelle), et la première chose à faire était d'en trouver le domaine de définition.

Finalement, voici comment est résolu notre paradoxe de Richard :

a) Il n'existe pas d'algorithme pour énumérer les programmes calculant une fonction de $\mathbb{N}$ dans $\mathbb{N}$ parce qu'il n'existe pas d'algorithme permettant de vérifier si, oui ou non, un programme calculant une semifonction calcule en fait une fonction (*i.e.* une semifonction définie sur tout $\mathbb{N}$). Autrement dit, il n'existe pas d'algorithme permettant de vérifier si, oui ou non, un algorithme est convergent. C'est là un

résultat essentiel de la théorie des fonctions récursives, que l'on démontre justement en exhibant une contradiction, à l'aide d'un avatar du procédé diagonal, s'il en était autrement ! En fait, nous avons fourni pratiquement les éléments nécessaires pour écrire une telle démonstration.

b) Par contre, il existe un algorithme pour énumérer les programmes calculant une semifonction de $\mathbb{N}$ dans $\mathbb{N}$. Il n'y a plus ici de paradoxe car si, dans ce cadre, on arrive comme plus haut à l'égalité

$$f_k(k) \;=\; g(k) \;=\; f_k(k) + 1$$

cela signifie justement que la semifonction g n'est pas définie en k ! L'égalité $h_1(x) = h_2(x)$ de deux semifonctions en $x \in \mathbb{N}$ doit être entendue comme suit : ou bien h_1 et h_2 sont toutes deux définies en x, et on a $h_1(x) = h_2(x)$ au sens habituel, ou bien ni h_1 ni h ne sont définies e x.

Notons aussi – nous l'utiliserons plus loin – l'existence d'un programme U, dit universel, permettant de calculer à la foi toutes les semifonctions de $\mathbb{N}$ dans $\mathbb{N}$. C'est en fait unevariante de notre programme Π. En voici une description grossière : l'entrée de U est ici un couple d'entiers (n, k) ; U effectue d'abord le programme P du point 3) appliqué à l'entrée n, ce qui fournit le code du programme P_n (au bout d'un temps fini de calcul) ; puis U suit les instructions du programme P_n appliqué à l'entrée k et sort, si elle existe, la sortie $f_n(k)$ de ce programme.

Nous passons maintenant à l'étude des semifonctions programmables.

6. UN PEU DE THÉORIE DES FONCTIONS RÉCURSIVES

On peut définir la théorie des fonctions récursives comme étant la théorie des fonctions calculables par algorithme. Il existe en fait des manières fort diverses de mathématiser les notions intuitives de « algorithme », « calculabilité », etc. (noter qu'il est nécessaire de mathématiser cette notion si l'on veut pouvoir prouver que certains problèmes n'ont pas de solution algorithmique : tel est le cas du dixième problème de Hilbert, datant de 1901, qui porte sur l'existence d'un algorithme permettant de décider si, oui ou non, une équation diophantienne admet une solution entière : problème résolu négativement entre 1960 et 1970 par M. Davis, H. Putnam, J. Robinson, Ju. Matijasevič et G. Čudnovskiĭ). Nous en avons justement décrit une (grossièrement, mais elle peut être formalisée) en identifiant la notion d'algorithme avec celle de programme écrit dans un langage de programmation. Et c'est un résultat profond des logiciens que toutes les manières dégagées jusqu'ici pour mathématiser cette notion intuitive conduisent à caractériser les mêmes objets : les fpnctions calculables, mettons de $\mathbb{N}$ dans $\mathbb{N}$, sont toujours les mêmes. D'où la thèse de Church (c'est un acte de foi, non un théorème !) : les fonctions intuitivement calculables sont exactement les fonctions récursives.

Nous présentons maintenant une autre méthode pour définir les fonctions récursives, qui a l'avantage sur la première de mieux s'écrire dans le langage mathématique habituel, mais qui a l'inconvénient d'être souvent moins suggestive, en particulier en ce qui concerne la démonstration des deux résultats fondamentaux : existence d'un algorithme universel, et non-existence d'un algorithme pouvant décider si oui ou non un algorithme converge.

D'abord, un peu de terminologie. L'ensemble des entiers $\mathbb{N}$ commencera désormais souvent par 0 ; un k-uple est un élément $(m_1, \ldots, m_k)$ de $\mathbb{N}^k$ et sera noté $\boldsymbol{m}$ s'il n'y a pas d'ambiguïté possible sur k. Nous entendons par le mot fonction une application définie sur $\mathbb{N}^k$ (le nombre k d'arguments dépendant de la fonction), à valeurs dans $\mathbb{N}$. Enfin, nous confondrons tout ensemble A de k-uples (*i.e.* toute partie A de $\mathbb{N}^k$) avec sa fonction indicatrice $\mathbf{1}_A$: on a[1] $\mathbf{1}_A(\boldsymbol{m}) = 1$ si $\boldsymbol{m} \in A$ et $\mathbf{1}_A(\boldsymbol{m}) = 0$ si $\boldsymbol{m} \notin A$; ainsi, un ensemble sera dit récursif si son indicatrice l'est.

L'ensemble des fonctions récursives est défini comme étant le plus petit ensemble de fonctions contenant certaines fonctions, qui seront appelées fonctions de base, et stable pour certaines opérations sur les fonctions, que nous appellerons opérations de base. Alors que les opérations de base choisies sont presque toujours les mêmes, les fonctions de base dépendent souvent (mais peu) de l'humeur de celui qui les donne.

Nous prendrons les fonctions de base suivantes :

a) à un argument : les fonctions constantes (de $\mathbb{N}$ dans $\mathbb{N}$) ;

b) à deux arguments : somme, produit, diagonale de $\mathbb{N}^2$;

c) à k arguments ($k \geq 1$) : fonctions coordonnées sur $\mathbb{N}^k$ (pour $k = 1$, on trouve la fonction « identité ») ;

et les opérations de base suivantes :

a) *composition* : si g est une fonction à n arguments, et $f_1, \ldots, f_n$ des fonctions à $k_1, \ldots, k_n$ arguments, la fonction obtenue par composition de g avec $f_1, \ldots, f_n$ est la fonction h à $k_1 + \ldots + k_n$ arguments, définie comme suit

$$h(\boldsymbol{m}) \;=\; h(\boldsymbol{m}_1, \ldots, \boldsymbol{m}_n) \;=\; g(f_1(\boldsymbol{m}_1), \ldots, f_n(\boldsymbol{m}_n)) \;;$$

b) *schéma de récurrence* : si, pour $k \geq 1$, g est une fonction à $k + 1$ arguments et f une fonction à $k - 1$ arguments (f est une fonction constante pour $k = 1$), la fonction obtenue par application du schéma de récurrence à f et g est la fonction h sur $\mathbb{N}^k = \mathbb{N}^{k-1} \times \mathbb{N}$ définie par récurrence comme suit

$$\begin{aligned} h(\boldsymbol{m}, 0) &= f(\boldsymbol{m}) \\ h(\boldsymbol{m}, n+1) &= g(\boldsymbol{m}, h(\boldsymbol{m}, n), n) \end{aligned}$$

(il faut un peu de pratique pour y reconnaître un objet familier) ;

c) *minimalisation* : si f est une fonction sur $\mathbb{N}^{k+1} = \mathbb{N}^k \times \mathbb{N}$ qui vérifie la condition (*) suivante

$$(*) \quad \forall\, \boldsymbol{m} \in \mathbb{N}^k : \;\; \exists\, n \in \mathbb{N} : \;\; f(\boldsymbol{m}, n) = 0$$

la fonction obtenue par minimalisation à partie de f est la fonction g sur $\mathbb{N}^k$ définie comme suit

$$g(\boldsymbol{m}) \;=\; \text{le plus petit } n \text{ tel que } f(\boldsymbol{m}, n) = 0$$

(cela sert à définir des fonctions « implicites »).

---

[1] On dit aussi fonction caractéristique au lieu d'indicatrice. On prendra garde que, souvent, les logiciens utilisent une autre définition de la fonction caractéristique.

On démontre (mais ce n'est absolument pas évident) qu'on peut faire l'économie, pour définir l'ensemble des fonctions récursives, du schéma de récurrence parmi les opérations de base : il est, en quelque sorte, implicitement contenu dans la minimalisation.

Les fonctions de base sont des fonctions calculables (ou des algorithmes ?) élémentaires, et les opérations de base des articulations élémentaires d'algrithmes. Voyons maintenant ce qui joue ici le rôle d'algorithme. Il est facile de voir qu'une fonction f est récursive ssi il en existe une description récursive, *i.e.* une suite finie $f_1, \ldots, f_n$ de fonctions telles que $f = f_n$ et que, pour tout $i \leq n$, la fonction f_i soit une fonction de base, ou soit obtenue à l'aide d'une opération de base appliquée à certaines des f_j pour $j < i$. Il est alors clair que toute fonction récursive est intuitivement calculable (la réciproque est la thèse de Church), et on en déduit aussi, sans trop de peine, que l'ensemble des fonctions récursives est dénombrable. Noter, au passage, l'analogie entre fonctions de base et axiomes, opérations de base et règles d'inférence, description récursive et démonstration ; et également l'analogie entre description récursive et programme.

Au sujet de la condition (*), le lecteur perspicace aura noté que sa vérification ressemble beaucoup à celle de la convergence d'un algorithme, et en aura déduit l'impossibilité d'énumérer effectivement l'ensemble des fonctions récursives. Plus précisément, on établit aisément (à l'aide de l'argument à la Cantor déjà vu au §5) qu'il n'existe pas, par exemple, de fonction récursive à deux arguments, universelle pour les fonctions récursives à un argument : autrement dit, il n'existe pas de fonction récursive $(m, n) \mapsto f(m, n)$ telle que les fonctions $n \mapsto f_m(n) = f(m, n)$, qui sont récursives, à un argument, pour $m \in \mathbb{N}$, fournissent toutes les fonctions récursives à un argument lorsque m parcourt $\mathbb{N}$.

Par contre, si on appelle semifonction (on dit aussi, plus fréquemment d'ailleurs, fonction partielle) sur $\mathbb{N}^k$ une fonction définie éventuellement sur une partie de $\mathbb{N}^k$ (et à valeurs dans $\mathbb{N}$), et si l'on définit l'ensemble des semifonctions récursives comme étant le plus petit ensemble de semifonctions contenant les fonctions de base (les mêmes que précédemment) et stable pour les opérations de base (les mêmes que précédemment, mais sans restriction (*) quant à l'application de la minimalisation : on a alors, avec les notations introduites plus haut,

$g(\boldsymbol{m}) =$ le plus petit n tel que $f(\boldsymbol{m}, 0), f(\boldsymbol{m}, 1), \ldots, f(\boldsymbol{m}, n)$ soient définis et que $f(\boldsymbol{m}, n) = 0$ si un tel n existe,

$=$ non défini sinon ;

par ailleurs, dans un schéma de récurrence, $h(\boldsymbol{m}, n + 1)$ ne peut être défini que si $f(\boldsymbol{m})$ et $g(\boldsymbol{m}, i, j)$, pour $i \leq h(\boldsymbol{m}, n)$ et $j \leq n$, sont définis). Donc, si l'on définit ainsi l'ensemble des semifonctions récursives (que l'on peut encore définir à l'aide des descriptions récursives), il est possible d'énumérer effectivement cet ensemble ; nous avons vu, à la fin du §5, comment construire, dans un langage de programmation, une énumération effective des semifonctions récursives à un argument. De manière générale, il existe, pour tout entier k, une semifonction récursive $(m, \boldsymbol{n}) \mapsto f_k(m, \boldsymbol{n})$ sur $\mathbb{N}^{k+1} = \mathbb{N} \times \mathbb{N}^k$, universelle pour les fonctions récursives à k arguments : lorsque m parcourt $\mathbb{N}$, la suite de semifonctions $\boldsymbol{n} \mapsto f_{k,m}(\boldsymbol{n}) = f_k(m, \boldsymbol{n})$ pour k

fixé[1], fournit exactement toutes les semifonctions récursives sur $\mathbb{N}^k$ (rappelons que deux semifonctions sont égales ssi elles ont même ensemble de définition, et sont égales sur cet ensemble). En fait, il n'existe pas d'algorithme pour vérifier si, oui ou non, deux semifonctions récursives sont égales (dans notre mathématisation de la notion d'algorithme, cela veut dire que, pour k fixé, l'ensemble $\{(m, m') \mid f_{k,m} = f_{k,m'}\}$ n'est pas récursif) si bien que, dans cetype d'énumération, il est inévitable qu'il y ait des répétitions – nous reviendrons là-dessus à la fin du dernier paragraphe. Ajoutons, pour lever toute ambiguïté, qu'il est vrai (sans que cela soit évident) qu'une semifonction récursive définie partout sur son $\mathbb{N}^k$ est une fonction récursive ; l'inverse est évidemment vrai.

Intéressons-nous maintenant plus particulièrement aux ensembles (un ensemble est, dans ce paragraphe, une partie d'un espace $\mathbb{N}^k$). La notion mathématique d'ensemble récursif correspond à la notion intuitive d'ensemble décidable : on a un algorithme qui nous dit si, oui ou non, un k-uple appartient à cet ensemble, à savoir l'indicatrice de cet ensemble ou, plutôt, cette indicatrice et une description récursive de celle-ci (il arrive souvent, même si c'est parfois gênant, de confondre fonction calculable et mode de calcul d'icelle : c'est ce qu'on fait tous les jours avec les fonctions rationnelles ; lorsqu'on a affaire à une semifonction récursive, les points où elle n'est pas définie correspondent alors aux points où l'algorithme ne converge pas).

Tous les ensembles familiers de l'arithmétique sont récursifs[2] : par exemple, l'ensemble des nombres premiers, ou celui des triplets (p, q, r) où r est le pgcd de p et q. Un peu de réflexion persuade qu'il ne doit pas être facile de nommer effectivement un ensemble non récursif, quoiqu'ils soient légion (il n'y a qu'une quantité dénombrable de parties récursives de $\mathbb{N}$ alors que l'on sait que l'ensemble des parties de $\mathbb{N}$ n'est pas dénombrable) ; en fait, il est facile de donner un « pseudo »-algorithme pour en « construire » un : énumérer les parties récursives de $\mathbb{N}$ (ce qui ne peut être fait par un algorithme), on obtient une suite (B_n), et lui appliquer le procédé diagonal (qui, lui, est algorithmique), on obtient $C = \{n \mid n \notin B_n\}$ qui n'est pas récursif[3] Nous verrons mieux tout à l'heure ; pour le moment, voici un passage, en petits caractères, pour le lecteur qui connaît un peu la topologie générale.

---

[1] Si l'on « code de manière récursive » les suites finies d'entiers par les entiers en prenant, par exemple, pour code de $\boldsymbol{n} = (n_1, \ldots, n_k)$ le nombre n où

$$n = p_1^{n_1+1} \times \ldots p_k^{n_k+1}$$

et où (p_k) est la suite des nombres premiers, dans leur ordre naturel, il est possible de représenter à la fois toutes les semifonctions récursives, k et m variant, par une seule semifonction récursive Ψ à trois arguments : il suffit de poser (encore faut-il le démontrer !)

$$\Psi(k, m, n) = f_{k,m}(\boldsymbol{n}) \quad \text{si } \boldsymbol{n} \in \mathbb{N}^k \text{ et } n = \text{code de } \boldsymbol{n}$$

en laissant $\Psi(k, m, n)$ non défini si n n'est le code d'aucun k-uple.

[2] Bien entendu, tout ensemble fini est récursif.

[3] Si, suivant la démarche de Richard, on prend pour B_n la n-ième partie de $\mathbb{N}$ définie par une expression française, on obtient le paradoxe richardien ayant inspiré Gödel dans son étude de l'arithmétique.

Nous donnons, dans le cadre de notre espace $\{0, 1\}^{\mathbb{N}}$, un analogue topologique de la notion d'ensemble décidable : ce sera celle de partie à la fois ouverte et fermée de Ω (il s'agit en fait de bien plus qu'une simple analogie). Soit A une partie de Ω et regardons quand on peut décider si, oui ou non, une suite ω appartient à A en ne regardant qu'un nombre fini de ses sections commençantes $\omega \mid n = (X_1(\omega), \ldots, X_n(\omega))$. Si A n'est pas à la fois ouvert et fermé, ce n'est manifestement pas possible si ω appartient à la frontière de A (qui n'est pas vide). Par contre, si A est ouvert-fermé, il existe un entier n tel que l'îlot d'indice $\omega \mid n$ soit inclus dans A, ou dans $\complement$A ; alors, si on a une description de A et $\complement$A comme réunions d'une suite d'îlots, il est possible de décider en regardant alternativement les indices des îlots constitutifs de A et $\complement$A et en s'arrêtant quand on tombe sur une section commençante de ω. Dans ce cadre, l'analogue d'une fonction calculable de $\mathbb{N}$ dans $\mathbb{N}$ est une fonction f de Ω dans Ω vérifiant la condition

$$\forall\, n :\ \exists\, m :\ \ \omega \mid m = w \mid m \ \ \Longrightarrow\ \ \omega \mid n = w \mid n$$

Comme Ω est un espace compact, il n'est pas difficile de voir qu'on vient juste d'écrire que f doit être continue. En fait, le <u>bon</u> analogue est obtenu en demandant de plus que la fonction $n \mapsto$ le plus petit m tel que… soit récursive !

La classe des ensembles récursifs est stable pour les réunions finies, les intersections finies, la complémentation, et aussi pour les images réciproques par des applications récursives (une application $\boldsymbol{g}$ de $\mathbb{N}^p$ dans $\mathbb{N}^q$ est dite récursive si la fonction g_i obtenue en composant $\boldsymbol{g}$ avec la i-ième fonction coordonnée sur $\mathbb{N}^q$ est récursive, pour $i \le q$). On montre aussi qu'une application est récursive ssi son graphe est récursif (ici, l'analogie topologique a une faiblesse : il faut bien payer quelque part le fait que $\mathbb{N}$ est discret). Par contre, la projection d'un ensemble récursif n'est pas, en général, récursive (nous illustrerons cela, un peu plus loin, par quelques mots sur le dixième problème de Hilbert).

Un ensemble égal à la projection d'un ensemble récursif est dit <u>semi-récursif</u> (on dit aussi, plus fréquemment d'ailleurs, qu'il est <u>récursivement énumérable</u> : nous verrons bientôt pourquoi). Une partie A de $\mathbb{N}^q$, projection d'une partie B de $\mathbb{N}^p = \mathbb{N}^q \times \mathbb{N}^{p-q}$ où $p > q$, vérifie

$$\boldsymbol{m} \in \mathrm{A} \ \ \Longleftrightarrow\ \ \exists\, \boldsymbol{n} : (\boldsymbol{m}, \boldsymbol{n}) \in \mathrm{B}$$

si bien que, si B est récursif, on a, pour $\boldsymbol{m} \in$ A, un algorithme pour vérifier si $\boldsymbol{m}$ appartient effectivement à A : énumérer de manière effective[1] $\mathbb{N}^k = \mathbb{N}^{p-q}$ par une bijection récursive $\boldsymbol{n}(\cdot)$, et regarder successivement les valeurs de $\mathbf{1}_{\mathrm{B}}(\boldsymbol{m}, \boldsymbol{n}, 0)$, $\mathbf{1}_{\mathrm{B}}(\boldsymbol{m}, \boldsymbol{n}, 1)$, … La notion mathématique d'ensemble <u>semirécursif</u> correspond à la notion intuitive d'ensemble positivement décidable : on a un algorithme qui nous dit oui si un point appartient à l'ensemble, mais peut ne pas dire non, parce qu'il peut ne pas converger, si le point n'appartient pas à l'ensemble. La classe des ensembles semirécursifs est stable pour les réunions finies, les intersections finies, mais <u>non</u> pour la complémentation (un ensemble est récursif ssi il est semirécursif ainsi que son complémentaire) ; elle est aussi stable pour les images réciproques, et directes, par des

[1] On peut obtenir une énumération de $\mathbb{N}$ par une bijection récursive (ainsi que son inverse) de $\mathbb{N}$ sur $\mathbb{N}^k$ de la manière suivante : ordonner les k-uples $(n_1, \ldots, n_k)$ suivant la grandeur de leur norme $n_1 + \ldots + n_k$ et, pour une même valeur de la norme, suivant l'ordre lexicographique. On en déduit que, du point de vue de la récursivité, tous les espaces $\mathbb{N}^k$, $k \ge 1$, sont isomorphes.

applications récursives. On montre aussi qu'une semifonction (et, plus généralement, une « semiapplication ») est récursive ssi son graphe est semirécursif, le graphe G de $\boldsymbol{g}$ étant défini par

$$G = \{(\boldsymbol{m}, \boldsymbol{n}) \mid \boldsymbol{g}(\boldsymbol{m}) \text{ est défini et vaut } \boldsymbol{n}\}$$

– il est récursif si $\boldsymbol{g}$ est définie partout sur son $\mathbb{N}^p$ (et récursive), d'où le fait, non évident, qu'une semifonction récursive partout définie est une fonction récursive.

Revenons, une dernière fois, à notre analogue topologique. Le rôle des ensembles semirécursifs est joué par les ouverts de Ω, et, plus généralement, de Ω^k pour $k \geq 1$. Ici encore, quelques petites faiblesses : une fonction continue n'a jamais son graphe ouvert ! On a cependant les mêmes propriétés de stabilité (ou d'instabilité !), à condition de ne considérer que les applications continues et ouvertes (par exemple, une projection) pour les images directes. Noter que tout ouvert de Ω^{ik} est projection d'un ouvert-fermé de Ω^{k+1} (même de $\Omega^k \times \mathbb{N}$), et qu'il existe un ouvert U de $\Omega \times \Omega^k$ universel pour les ouverts de Ω^k : *i.e.* les coupes de U parallèles à l'espace Ω^k fournissent tous les ouverts de Ω^k. On peut par exemple construire U pour $k = 2$ comme suit : W étant l'ensemble des $w \in \Omega$ comportant une infinité de 0, et (J_m) étant une énumération des îlots de Ω avec $J_0 = \varnothing$, définir un ouvert V de $W \times \Omega$ en posant

$$(w, \omega) \in V \quad \text{ssi} \quad \omega \in \bigcup_n J_{m(w,n)}$$

où $m(w, n)$ est la longueur (éventuellement nulle) du « n-ième paquet de 1 » dans w ; prendre finalement pour U un ouvert de $\Omega \times \Omega$ dont la trace sur $W \times \Omega$ est égale à V. Nous verrons qu'il existe aussi des ensembles semirécursifs universels (mais pas si aisés à construire *ex nihilo* !).

Voici trois caractérisations importantes du fait, pour une partie de $\mathbb{N}^k$, d'être semirécursive :

1) être la projection d'une partie récursive de $\mathbb{N}^{k+1}$;
2) être le domaine de définition d'une semifonction récursive (de plus, la semi-indicatrice de notre partie A, qui vaut 1 sur A et n'est pas définie sur $\complement$A, est alors une telle semifonction) ;
3) être vide, ou bien être l'image directe de $\mathbb{N}$ par une application récursive de $\mathbb{N}$ dans $\mathbb{N}^k$; d'où le nom d'ensemble récursivement énumérable : on a un algorithme pour énumérer effectivement les éléments de notre ensemble (avec répétitions possibles, en fait évitables si l'ensemble est infini).

Ces caractérisations ne sont pas étonnantes si l'on pense à la notion intuitive de « décidabilité positive ». Voyons par exemple comment on peut démontrer 1) $\Leftrightarrow$ 3). Si la partie A de $\mathbb{N}^k$ est récursivement énumérée par une application récursive $\boldsymbol{g}$ de $\mathbb{N}$ dans $\mathbb{N}^k$, alors A est la projection sur $\mathbb{N}^k$ du graphe de $\boldsymbol{g}$, qui est récursif. Réciproquement, si A est la projection sur $\mathbb{N}^k$ d'une partie récursive B de $\mathbb{N}^{k+1}$, on peut énumérer récursivement les éléments de A comme suit : d'abord, on[1] énumère récursivement les éléments de B par une application $\boldsymbol{g}$ en énumérant récursivement les $(k + 1)$-uples et en les « biffant » au fur et à mesure à l'aide de

[1] On suppose A infini, le cas fini étant trivial.

l'algorithme $\mathbf{1}_B$ s'ils n'appartiennent pas à B ; puis on compose $\boldsymbol{g}$ avec la projection π de $\mathbb{N}^{k+1}$ sur $\mathbb{N}^k$, d'où

$$A = \{\pi(\boldsymbol{g}(0)), \pi(\boldsymbol{g}(1)), \ldots\}$$

Ainsi, pour une fois, le langage des mathématiques modernes des lycées et collèges peut être utilisé de manière non stérile : un ensemble récursif est un ensemble que l'on peut définir effectivement par compréhension, tandis qu'un ensemble semirécursif est un ensemble que l'on peut définir effectivement par extension[1].

La caractérisation 2) nous intéresse ici parce qu'elle va nous fournir immédiatement l'existence d'ensembles semirécursifs universels à partir de celle de semifonctions récursives universelles. Ainsi, si, pour tout k, f_k est une semifonction récursive sur $\mathbb{N} \times \mathbb{N}^k$, son domaine de définition U_k est un ensemble semirécursif, universel pour les parties semirécursives de $\mathbb{N}^k$: ses coupes $U_{k,n} = \{\boldsymbol{m} \mid (n, \boldsymbol{m}) \in U_k\}$, lorsque n parcourt $\mathbb{N}$, fournissent une énumération effective des parties semirécursives de $\mathbb{N}^k$. Et l'existence de U_2 assure l'existence d'une partie semirécursive A de $\mathbb{N}$ qui n'est pas récursive : il suffit de poser $A = \{n \mid (n, n) \in U_2\}$, l'argument diagonal assurant que $\complement A$ n'est pas parmi les coupes de U_2, et donc n'est pas semirécursif.

Il n'existe pas d'exemple « simple et naturel » au sens de l'arithmétique élémentaire, de partie semirécursive de $\mathbb{N}$ qui ne soit pas récursive, encore que les exemples abondent en logique (si on formalise la théorie élémentaire de l'arithmétique et si l'on code « de manière récursive » les expressions du langage par des entiers : c'est là la quintessence du théorème d'incomplétude de Gödel[2]). Nous illustrerons cependant les possibilités d'application de la théorie des fonctions récursives à l'étude de l'arithmétique par quelques mots sur le dixième problème de Hilbert.

On appelle équation diophantienne une équation de la forme

$$(\circ) \quad P(\boldsymbol{m}, \boldsymbol{x}) = P(m_1, \ldots, m_p, x_1, \ldots, x_q) = 0$$

où $P(\boldsymbol{m}, \boldsymbol{x})$ est un polynôme à coefficients dans $\mathbb{Z}$, et dépendant des $p + q$ variables $m_1, \ldots, m_p, x_1, \ldots, x_q$: le premier paquet $\boldsymbol{m}$ est un ensemble de paramètres, et le second $\boldsymbol{x}$ l'ensemble des inconnues de l'équation. On appelle ensemble diophantien associé à (°) la partie A de $\mathbb{N}^p$ constituée par les p-uples $\boldsymbol{m}$ tels que l'équation (+) ait au moins une solution $\boldsymbol{x}$ dans $\mathbb{N}^q$:

$$\boldsymbol{m} \in A \quad \text{ssi} \quad \exists\, \boldsymbol{x} : \ P(\boldsymbol{m}, \boldsymbol{x}) = 0$$

Le problème de Hilbert peut alors s'énoncer ainsi : existe-t-il un algorithme qui permette de décider si, oui ou non, une équation diophantienne a au moins une solution ? Une fonction polynôme à coefficients dans $\mathbb{Z}$ pouvant s'écrire comme différence de deux fonctions polynômes à coefficients dans $\mathbb{N}$, un ensemble diophantien A est donc de la forme

---

[1] La distinction entre méthode de décision et méthode de génération remonte au moins à Leibniz, qui distinguait l'*Ars indicandi* et l'*Ars inveniandi*.

[2] Une théorie est complète si tout énoncé clos (*i.e.* sans variables libres) est un théorème, ou la contraposition d'un théorème. Il est facile de montrer qu'une théorie « bien » axiomatisée et complète est décidable (*i.e.* on a un algorithme pour vérifier si oui ou non un énoncé clos est un théorème). C'est le cas pour la théorie des groupes abéliens.

$\{\boldsymbol{m} \mid \exists\, \boldsymbol{x} : f_1(\boldsymbol{m}, \boldsymbol{x}) = f_2(\boldsymbol{m}, \boldsymbol{x})\}$ où f_1 et f_2 sont des fonctions récursives : tout ensemble diophantien est donc semirécursif, et une solution positive au problème de Hilbert impliquerait que tout ensemble diophantien est récursif. Or, c'est l'extrême opposé qui est vrai : tout ensemble semirécursif est diophantien ! Étant donnée l'existence d'un ensemble semirécursif universel dans $\mathbb{N}^2$, on en déduit qu'il existe[1] une équation diophantienne, à deux paramètres, $U(n, m, \boldsymbol{x}) = 0$, telle que les ensembles $A_n = \{n \mid \exists\, \boldsymbol{x} : U(n, n, \boldsymbol{x}) = 0\}$ est alors un ensemble diophantien qui n'est pas récursif, et on n'a pas d'algorithme pour décider pour quelles valeurs de n l'équation diophantienne $U(n, n, \boldsymbol{x}) = 0$ admet une solution[2]. Par ailleurs, associons à tout polynôme $P(m, \boldsymbol{x})$ à un paramètre un autre polynôme $Q(m, \boldsymbol{x})$ par

$$Q(m, \boldsymbol{x}) \;=\; (m + 1)\,(1 - P^2(m, \boldsymbol{x})) - 1\ ;$$

il est facile de voir que $P(m, \boldsymbol{x}) = 0$ ssi $Q(m, \boldsymbol{x}) \geq 0$, auquel cas on a de surcroît $Q(m, \boldsymbol{x}) = m$. On en déduit que toute partie diophantienne de $\mathbb{N}$ (et donc toute partie semirécursive) est exactement l'ensemble des valeurs ≥ 0 prises par un polynôme à coefficients dans $\mathbb{Z}$ (sans paramètre : m est une variable pour Q), lorsque ses variables parcourent $\mathbb{N}$: on peut, par exemple, représenter ainsi l'ensemble des nombres premiers, qui est récursif, et le plus « court » polynôme connu pour ce faire, dû à Jones, n'est pas trop énorme (325 symboles, soit 5 lignes environ).

La notion intuitive de fonction calculable couvre un champ plus grand que celle de fonction récursive telle que nous l'avons définie : on peut parler de « fonction calculable » d'un ensemble E dans un autre F, où E et F sont « algorithmiquement engendrés ». Cela se mathématise à l'aide de la notion de « numérotage », mais il est temps que nous revenions à nos moutons aléatoires (si nous ne voulons pas devoir changer le titre de l'exposé), et nous nous contenterons d'illustrer cela en définissant ci-dessous la notion de fonction récursive de l'ensemble S des suites finies de 0 et de 1 dans $\mathbb{N}$, et, plus loin, dans lui-même.

7. LES SUITES ALÉATOIRES DE CHURCH (1940)

Rappelons qu'une règle de sélection $\mathfrak{R}$ est une application de l'ensemble S des suites finies de 0 et de 1 dans $\{0, 1\}$ (c'est donc l'indicatrice d'une partie de S), et que l'application de la règle $\mathfrak{R}$ à une suite infinie ω fournit une sous-suite (finie ou infinie) $\mathfrak{R}\omega$ de ω obtenue comme suit : $\mathfrak{R}\omega$ est ce qu'il reste de ω lorsqu'on a biffé les $X_n(\omega)$ correspondant aux n tels que $\mathfrak{R}(X_1(\omega), \ldots, X_{n-1}(\omega)) = 0$.

Wald avait proposé de dire que $\omega \in \Omega$ est un collectif, au sens absolu, si pour toute règle effectivement définissable $\mathfrak{R}$ telle que $\mathfrak{R}\omega$ soit infinie, la suite $\mathfrak{R}\omega$ vérifie la loi forte des grands nombres. Et c'est Church qui a fini de rendre rigoureuse cette définition en substituant à la notion intuitive de « règle effectivement définissable » la notion mathématique, alors récente, de « règle récursive ».

Définir la notion de « règle récursive », c'est définir la notion de sous-ensemble récursif de S. Nous définirons, plus généralement, la notion de fonction récursive de S dans $\mathbb{N}$. L'idée est

---

[1] On peut en écrire effectivement une, mais il faut beaucoup de papier !

[2] Cela justifie, *a posteriori*, le caractère « artisanal » de la théorie des équations diophantiennes.

très simple : ayant un « bon » numérotage des éléments de S, *i.e.* une « bonne » application surjective Ψ de $\mathbb{N}$ sur S (l'injectivité n'est pas nécessaire), une fonction f de S dans $\mathbb{N}$ est dite récursive ssi la fonction $f \circ \Psi$ de $\mathbb{N}$ dans $\mathbb{N}$ l'est. Reste à définir « bon » en évitant de retomber sur une notion absolue mais intuitive, ou de n'obtenir qu'une définition de récursivité relative à un numérotage donné Ψ. Eh bien, on montre que l'on obtient la même notion de récursivité pour toutes les applications Ψ de $\mathbb{N}$ sur S vérifiant les conditions (intuitivement justifiées) suivantes :

a) $\Psi(0)$ est la suite vide $\varnothing$ (pour l'esthétique !) ;

b) la fonction $\ell(\cdot)$ de $\mathbb{N}$ dans $\mathbb{N}$ qui, à $n \in \mathbb{N}$ associe la longueur $|\Psi(n)|$ de la suite $\Psi(n)$, est récursive ;

c) la fonctions $c(\cdot, \cdot)$ de $\mathbb{N}^2$ dans $\mathbb{N}$ définie comme suit est récursive : pour tout n, et tout m tel que $l \leq m \leq \ell(n)$, $c(m,$ n*) est la* m-*ième coordonnée* $X_m(\Psi(n))$ de la suite $\Psi(n)$; pour les autres couples (m, n), on pose, par convention, $c(m, n) = 0$.

Un tel numérotage nous est fourni, par exemple, par le classement (déjà vu à l'occasion des nombres normaux de Borel) des suites finies selon leur longueur, et, pour une même longueur, selon l'ordre lexicographique :

$$\Psi(0) = \varnothing,\ \Psi(1) = (0),\ \Psi(2) = (1),\ \Psi(3) = (0, 0),\ \Psi(4) = (0, 1),\ \Psi(5) = (1, 0),\ \text{etc.}$$

Pour fixer les idées, c'est toujours ce numérotage (au demeurant bijectif) de S que nous prendrons par la suite. Noter que cela nous permet de considérer l'entier n comme un nom pour désigner la suite finie $\Psi(n)$.

Ceci fait, la définition de collectif, en un sens absolu, est bien établie. Elle n'en reste pas moins toujours sujette à la critique de Ville : il existe des suites aléatoires au sens de Church qui ne fluctuent pas assez (sans doute en raison de la guerre, il semble que Church n'ait pas eu connaissance des travaux de Ville).

Nous allons passer maintenant à la notion de suite aléatoire au sens de Martin-Loef : elle reposera en définitive sur la possibilité de mathématiser la notion intuitive d'ensemble « effectivement négligeable » à l'aide de la théorie des fonctions récursives.

8. LES SUITES ALÉATOIRES DE MARTIN-LOEF (1966)

La guerre d'une part, le triomphe de l'axiomatique de Kolmogorov d'autre part, ont tous deux contribué à rejeter dans l'ombre, pour un temps, l'étude de la notion de suite aléatoire. Elle resurgit cependant, sans jamais atteindre la même acuité, vers les années 1960 (la coïncidence avec le développement de l'informatique n'est sans doute pas fortuite) avec le problème de la définition d'une suite finie aléatoire, et, plus généralement, avec celui de la complexité d'une suite finie, dans des travaux de Kolmogorov dont nous parlerons peu ici (voir cependant le §9). Soulignons cependant leur importance historique pour notre propos, et, plus généralement, l'importance du rôle de Kolmogorov dans le développement de l'étude des mathématiques des algorithmes en URSS. C'est finalement Martin-Loef qui, prolongeant les travaux de Kolmogorov, a le premier fondé une théorie conséquente des suites infinies « vraiment » aléatoires.

Rappelons la dernière définition intuitive (suggérée par Wald) à laquelle nous étions arrivés à la fin du §4 :

> la suite ω est aléatoire ssi elle n'appartient à aucun ensemble effectivement négligeable.

Maintenant, un sous-ensemble N de Ω est négligeable ssi il existe une suite d'ouverts (V_n) contenant N, que l'on peut supposer décroissante[1] telle que l'on ait $P[V_n] < 2^{-n}$ pour tout n ; et tout ouvert (non vide) est la réunion d'une suite d'îlots. Pour pouvoir dire que N est effectivement négligeable, il faut, intuitivement, que la suite d'ouverts (V_n) puisse être engendrée algorithmiquement (ce qui implique, en particulier, que chacun des ouverts V_n soit lui-même engendré algorithmiquement à partir des îlots, qui sont les ouverts « bien concrets » de départ) ; dans ce cas, on pourra dire que l'intersection des V_n est un ensemble négligeable effectivement définissable.

Or, nous avons un bon numérotage Ψ des éléments de S, qui nous permet de considérer un entier m comme un nom pour la suite finie $\Psi(m)$ et donc, par métonymie, pour l'îlot $I_{\Psi(m)}$. Se donner effectivement un ouvert V, c'est alors se donner une fonction récursive f de N dans N donnant les noms d'îlots constitutifs de V (autrement dit, $V = \bigcup_m I_{f(\Psi(m))}$) et donc se donner une partie semirécursive de $\mathbb{N}$ (à savoir $f(\mathbb{N})$, récursivement énumérée par f). Et, finalement, se donner effectivement une suite (V_n) d'ouverts , c'est se donner une partie semirécursive $\mathcal{O}$ de $\mathbb{N}^2$: la première coordonnée du couple $(n, m) \in \mathcal{O}$ nous dit que c'est l'ouvert V_n que l'on construit (c'est donc le nom de l'ouvert), tandis que la seconde nous dit que l'îlot de nom m est un îlot constitutif de V_n.

D'où la définition d'un <u>test de Martin-Loef</u> : c'est une partie semirécursive $\mathcal{O}$ de $\mathbb{N} \times \mathbb{N}$ telle que la suite d'ouverts (V_n) correspondante vérifie les conditions suivantes :

a) $V_{n+1} \subseteq V_n$

b) $P[V_n] < 2^{-n}$

Le nom « test » vient de la pratique statistique pour tester l'hypothèse de la stochasticité d'une suite : soumettre ω au test de stochasticité $\mathcal{O}$, au seuil $\varepsilon = 2^{-n}$, c'est regarder si oui ou non[2] ω< se trouve dans V_n et accepter que ω est aléatoire si ω est hors de V_n. Est associé au test $\mathcal{O}$ l'ensemble « effectivement négligeable » $N(\mathcal{O})$, égal à l'intersection de tous les V_n. D'où la définition d'une <u>suite aléatoire au sens de Martin-Loef</u> :

> la suite ω est aléatoire ssi, pour tout test $\mathcal{O}$, la suite ω n'appartient pas à $N(\mathcal{O})$.

Bien entendu, il n'y a qu'une quantité dénombrable de tests de Martin-Loef, si bien que presque toute suite ω est aléatoire en ce sens.

On peut se demander si la réunion (dénombrable) des $N(\mathcal{O})$, $\mathcal{O}$ parcourant l'ensemble des tests, est encore un « ensemble effectivement négligeable » qui serait alors maximal. <u>La réponse est oui, et on a même mieux</u>. Étant donnée l'existence d'une partie semirécursive de

---

[1] L'intersection d'un nombre fini d'ouverts est encore un ouvert.

[2] L'ouvert V_n étant « semirécursif », seule la réponse « oui » est, en général, « algorithmique ». On retrouve là, en quelque sorte, une dissymétrie bien connue des statisticiens.

$\mathbb{N}^2$, universelle pour les parties semirécursives de N^2, Martin-Loef a montré qu'il existe un<u>test $\mathcal{U}$ universel au sens suivant</u> : pour tout test $\mathcal{T}$, il existe un nombre entier k tel que, pour tout m, l'ouvert V_{m+k} soit contenu dans l'ouvert U_m, ce qui implique en particulier que $N(\mathcal{T})$ est contenu dans $N(\mathcal{U})$.

On montre facilement que toute suite aléatoire au sens de Martin-Loef est aléatoire au sens de Church : on est monté dans la hiérarchie. En particulier, toute suite infinie de 0 et de 1 construite à l'aide d'un algorithme n'est pas aléatoire au sens de Martin-Loef (*cf.* §3) (dans notre langage mathématique, la suite w est construite par un algorithme ssi il existe une fonction récursive f de N dans $\{0, 1\}$ telle que l'on ait $X_n(w) = f(n)$ pour tout $n > 0$). D'autre part, il est clair, vu la définition, que toute suite aléatoire au sens de Martin-Loef vérifiera les tests concrets de stochasticité imaginés ou imaginables, et ne peut donc être sujette aux critiques de « non fluctuation » de Ville.

Depuis les travaux de Martin-Loef, il est apparu diverses définitions voisines mais distinctes de la notion de suite aléatoire, faisant appel à la théorie des fonctions récursives. Le lecteur intéressé pourra en particulier consulter l'article critique de Schnorr (en anglais), où il verra réapparaître la notion de martingale de Ville et diverses mises en forme de la définition naïve écrite à la fin de notre §4 (une martingale positive est à valeurs dans $\mathbb{R}_+$, et il faut définir ce que l'on peut appeler une martingale « calculable) ; le lecteur passionné pourra se plonger dans la monographie (en allemand) de Schnorr[1] : Schnorr y dégage en particulier une notion de suite un peu moins aléatoire que celle de Martin-Loef, et donne des arguments convaincants comme quoi cette notion représente le « rêve » de von Mises.

Par ailleurs, nous avons interprété la notion intuitive de « définition semirécursive » dans un sens très constructiviste : la notion d'ensemble semirécursif n'épuise pas, loin de là, la notion intuitive d'ensemble « nommable », même s'il n'existe, intuitivement, qu'une quantité dénombrable de tels ensembles[2]. En fait, les logiciens ont dégagé toute une hiérarchie de l'effectivité (la récursivité, qui correspond à la « plus grande effectivité », se trouve au bas de cette échelle) : cela permet de définir des notions de suites aléatoires encore plus fortes que celle de Martin-Loef. De plus, les logiciens utilisent aussi une notion de « réel aléatoire »[3], dégagée par Solovay des travaux de Cohen sur le forcing, et qui est en quelque sorte le *nec plus ultra* en la matière ; mais, le manque de place et, surtout, le manque de science, m'empêchent d'en dire plus ici.

Revenant dans le domaine de la théorie des fonctions récursives et des suites aléatoires au sens de Martin-Loef, nous terminons maintenant cet exposé en disant, comme promis, quelques mots sur la notion de complexité d'une fonction semiréursive et de complexité monotone d'une suite.

---

[1] Les travaux cités, datant de 1971, sont trop vieux pour tenir compte des problèmes d'algorithmes policiers pour tester systématiquement l'anarchie des suites. Dans ce cas, la notion de martingale, au sens de l'équitation, peut s'avérer très utile.

[2] Chassez le paradoxe de Richard, il revient au galop !

[3] Par le truchement du développement en base 2, cela donne une notion de suite aléatoire de 0 et de 1.

9. UN PEU DE THÉORIE DE LA COMPLEXITÉ[1]

C'est Kolmogorov qui, le premier, a défini une notion de complexité d'une suite finie : relativement à un algorithme à entrée et sortie dans S, la complexité de Kolmogorov de $t \in S$ est la longueur de la plus courte suite $s \in S$ telle que t soit la sortie correspondant à l'entrée s (la complexité est infinie si t ne peut sortir de l'algorithme). La notion de semifonction récursive permet de mathématiser cette définition, et l'existence de semifonctions récursives universelles permet de donner, en un certain sens, une valeur absolue à cette notion de complexité ; nous verrons cela, à la fin du §9, dans un cadre élargi, en suivant les idées de Manin.

Cette notion a été reprise par Martin-Loef dans son étude de la notion de suite finie aléatoire (dont nous n'avons pas parlé) : Martin-Loef montre, en un sens précis, que les suites finies ayant une grande complexité sont les mêmes que celles qui résistent bien à ses tests de stochasticité.

La complexité de Kolmogorov, qui a des rapports avec la théorie de l'information de Shannon, a été beaucoup étudiée, et nous invitons le lecteur intéressé à lire l'article de synthèse de Levin et Zvonkin. Mais, comme elle ne permet pas de caractériser exactement ls suites infinies aléatoires au sens de Martin-Loef, nous commençons par présenter une variante, due à Levin, qui n'a pas cet inconvénient.

Voici ce que nous voulons d'abord faire : donner un sens précis à la notion d'algorithme monotone de S dans $S \cup \Omega$. Intuitivement, un tel algorithme A, pour une entrée, $s \in S$, sort, s'il converge, une suite $A(s)$ finie ou infinie[2]. de sorte que si l'on a $s_1 \dashv s_2$ (on entend par là que s_1 st une suite commençante de s_2) alors on ait $A(s_1) \dashv A(s_2)$ ou $A(s_1) = A(s_2)$. Nous allons exprimer cela en regardant les couples (s, t) dans $S \times S$ tels que $t \dashv A(s)$ ou $t = A(s)$ et en exigeant, pour dire que l'on a un algorithme, que l'ensemble de ces couples soit récursivement énumérable.

Nous identifions cette fois $\mathbb{N}$ et S grâce à notre bonne numérotation Ψ (qui est bijective) : on sait alors ce qu'est une fonction récursive de S dans S, une partie semirécursive (ou récursivement énumérable) de $S \times S$, etc. Nous dirons qu'une partie A de $S \times S$ est un algorithme monotone de S dans $S \cup \Omega$ si les conditions suivantes sont vérifiées :

a) A est une partie semirécursive de $S \times S$;

b) $((s, t) \in A \text{ et } (s, t') \in A) \implies (t \dashv t' \text{ ou } t' \dashv t \text{ ou } t = t')$;

c) $(s' \dashv s \text{ et } (s', t) \in A) \implies (s, t) \in A$.

---

[1] (Note pour les informaticiens). La complexité dont il va être question mesure en quelque sorte le temps mis pour trouver un programme calculant une fonction et non le temps mis par l'exécution d'un programme calculant la fonction. Et il ne s'agit pas d'une mesure récursive de complexité à la Blum, mais plutôt de borne inférieure de telles mesures : notre mesure de complexité à la Kolmogorov sera bien loin d'âtre calculable !

[2] Il peut sembler paradoxal qu'un algorithme puisse converger et fournir une suite infinie. C'est pourtant ce que l'on fait lorsqu'on écrit « soit $\omega = 010101\ldots$ ». Le paradoxe disparaît dans la mathématisation.

À toute suite finie s, l'algorithme A associe alors, si elle existe, la suite finie ou infinie A(s) définie comme suit : on regarde tous les t tels que $(s, t) \in$ A et on compose A(IsI) en « collant » ces t les uns sur les autres (ce qui est possible d'après b)).

Maintenant, étant donné un algorithme monotone A, on définit pour toute suite finie t la complexité monotone $\mathbf{km}_A(t)$ relativement à A comme étant la longueur de la plus courte suite s telle que $t \dashv A(s)$ ou $t = A(s)$, (la complexité de t est infinie s'il n'existe pas de telle suite s). Noter que, si A est tout bêtement la diagonale de S × S, alors on a $\mathbf{km}(t) = |t|$, la longueur de t.

À première vue, il semble impossible de définir une notion de complexité (minimale) absolue ; en fait, c'est possible, à condition de se contenter d'une « minimalité asymptotique ». Plus précisément, de l'existence d'ensembles semirécursifs universels on déduit [1] l'existence d'un algorithme monotone U universel au sens suivant : on a $\mathbf{km}_U(t) \leq |t| \dot{+} 2$ et, pour tout algorithme monotone A, il existe une constante k_A telle que l'on ait $\mathbf{km}_U(t) \leq \mathbf{km}_A(t) + k_A$ pour tout $t \in$ S (la meilleure constante k_A mesure en quelque sorte la complexité monotone de l'algorithme A par rapport à l'algorithme U). Ayant choisi une fois pour toutes notre algorithme universel U (il n'y a pas unicité), nous écrirons $\mathbf{km}(t)$ au lieu de $\mathbf{km}_U(t)$.

Soit maintenant ω une suite infinie ; pour tout entier n, nous désignerons par $\omega \mid n$ la suite finie $(X_1(\omega), \ldots, X_n(\omega))$ des n premiers termes de ω. Étant donnée la définition de la complexité monotone, il est clair que $\mathbf{km}(\omega \mid n)$ est une fonction croissante[2] de n. Et Levin démontre les résultats suivants, bien conformes à l'intuition :

a) la suite ω est engendrée par un algorithme (*i.e.* $n \mapsto X_n(\omega)$ est une fonction récursive ssi la fonction $n \mapsto \mathbf{km}(\omega \mid n)$ est bornée ;

b) la suite ω est aléatoire au sens de Martin-Loef ssi il existe une constante C_ω telle que l'on ait $|\mathbf{km}(\omega \mid n) - n| < C_\omega$ pour tout n.

Par ailleurs, on vérifie aisément que, pour m et n fixés, l'ensemble

$$\{t \in S \mid |t| = n \text{ et } \mathbf{km}(t) < n - m\}$$

a au plus 2^{n-m} éléments, et donc que $P\{\omega \mid \mathbf{km}(\omega \mid n) < n - m\}$ est $< 2^{-m}$: ainsi, la plupart des suites finies de longueur n ont une grande complexité, ce qui est une manière de dire qu'elles sont aléatoires.

Nous transitons maintenant vers la définition de la complexité de Kolmogorov selon Manin. Nous avons tout à l'heure identifié S à ℕ, ce qui nous a permis de parler de fonction récursive de S dans S. Mais, à rebours, l'identification de N à S permet de définir la complexité d'un entier n comme étant celle de la suite $\Psi(n)$; on notera que $\ell(n) = |\Psi(n)|$ est égal à

---

[1] Il s'agit là d'un résultat analogue à celui de l'existence d'un test universel au sens de Martin-Loef. Nous verrons une démonstration pour le cas de la complexité de Kolmogorov.

[2] Dans le cas de la complexité de Kolmogorov, il n'y a pas croissance. Il y a, au contraire, pour la plupart des suites ω, des « chutes de complexité imprévisibles » qui traduisent l'occurrence inopinée de grandes régularités dans ω (du moins, c'est mon interprétation). Voir à ce sujet l'article de Levin et Zvonkin.

$\lfloor \log_2 (n+1) \rfloor$, *i.e.* la partie entière du logarithme en base 2 de $(n+1)$[1]. Un algorithme A de S dans S devient alors une énumération de fonctions constantes A(0), A(1), …

De manière générale, si $(f_n)_{n\geq 0}$ est une énumération récursive de semifonctions récursives à p arguments (*i.e.* la semifonction $(n, \boldsymbol{m}) \mapsto f_n(\boldsymbol{m})$ est une semifonction récursive f à $p+1$ arguments), nous définissons la complexité de Kolmogorov $\mathbb{k}_f(g)$ de la fonction semirécursive g à p arguments relativement à $f = (f_n)$ coome suit : nous posons

$$c_f(g) \;=\; \inf \{n \mid g = f_n\}$$

en convenant que $c_f(g)$ est infini si $\{\ldots\}$ est vide, et nous prenons

$$\mathbb{k}_f(g) \;=\; \lfloor \log_2 (c_f(g) + 1) \rfloor$$

On est assuré que $\mathbb{k}_f(g)$ est finie pour toute semifonction récursive g à p arguments ssi f est une semifonction universelle. Et il existe de meilleures semifonctions universelles, fournissant une complexité asymptotique minimale[2] :

THÉORÈME. – Pour tout entier p, il existe une semifonction récursive Φ à $(p+1)$ arguments vérifiant la condition suivante : pour toute énumération récursive $f = (f_n)$ de semifonctions récursives à p arguments, il existe une constante $k_{\Phi, f}$ telle que l'on ait

$$\mathbb{k}_\Phi(g) \;\leq\; \mathbb{k}_f(g) + k_{\Phi, f}$$

pour toute semifonction récursive g à p arguments.

Démonstration : Soit $b(\cdot, \cdot)$ une bijection récursive de $\mathbb{N} \times \mathrm{N}$ sur $\mathbb{N}$, de croissance linéaire en son second argument : il existe une fonction Θ de $\mathbb{N}$ dans $\mathbb{N}$ telle que l'on ait $b(i, j) \leq \Theta(i) \times j$ pour tout couple (i, j). On peut par exemple prendre[3]

$$\begin{aligned} b(i,j) &= 2^i\,(2j+1) - 1 \\ \Theta(i) &= 2^{i+1} + 1. \end{aligned}$$

Soit d'autre part U une semifonction récursive sur$\mathbb{N}^{p+2} = \mathbb{N} \times \mathbb{N} \times \mathbb{N}^p$, universelle pour les semifonctions récursives sur $\mathbb{N}^{p+1} = \mathbb{N} \times \mathbb{N}^p$. Nous définissons alors notre semifonction Φ sur $\mathbb{N}^{p+1}$ par

$$\Phi(n, \boldsymbol{m}) \;=\; \mathrm{U}(b^{-1}(n), \boldsymbol{m})$$

Maintenant, si $f = (f_n)$ est une énumération récursive de semifonctions à p arguments et si, pour g semifonction récursive à p arguments,

$$c_f(g) \;=\; \inf \{n \mid g = f_n\}$$

---

[1] On sait que le logarithme en base 2 sert, en théorie de l'information, à calculer, les « bits ». Les complexités introduites dans ce §9 mesurent des longueurs de programme et donc des quantités d'information pour construire certains objets.

[2] Le premier théorème de ce type a été démontré indépendamment par Kolmogorov et Solomonoff (lequel doit être américain).

[3] La bijection utilisée par Kolmogorov est meilleure en ce qui concerne la croissance de Θ – elle est en i^2.

est fini (auquel cas $c_f(g)$ est un code de g pour f), on a pour tout $\boldsymbol{m}$,

$$\begin{aligned} g(\boldsymbol{m}) &= f(c_f(g), \boldsymbol{m}) \\ &= \mathrm{U}(c_{\mathrm{U}}(f), c_f(g), \boldsymbol{m}) \\ &= \Phi(b(c_{\mathrm{U}}(f), c_f(g)), \boldsymbol{m}) \end{aligned}$$

D'où l'on a

$$c_\Phi(g) \leq b(c_{\mathrm{U}}(f), c_f(g)) \leq c_f(g) \times \Theta(c_{\mathrm{U}}(f))$$

et, finalement, l'inégalité voulue en passant aux logarithmes.

La complexité de Kolmogorov $\mathbb{k}_\Phi(\cdot)$ est bien loin d'être une fonction calculable. Plus précisément, on peut démontrer le résultat suivant. Soit (f_n) une énumération récursive de semifonctions récursives à p arguments et soit $\mathbb{k}(\cdot)$ la fonction de $\mathbb{N}$ dans $\mathbb{N}$ définie par

$$\mathbb{k}(n) = \mathbb{k}_\Phi(f_n) ;$$

alors, si A est une partie semirécursive de $\mathbb{N}$ telle que $\mathbb{k}(\cdot)$ ne soit pas bornée sur A (soit encore, telle qu'il existe une infinité de semifonctions distinctes f_n avec $n \in \mathrm{A}$), il n'existe aucune semifonction récursive g définie au moins sur A et vérifiant la condition suivante : il existe une constante C telle que l'on ait $|\mathbb{k}(n) - g(n)| < \mathrm{C}$ sur A. Cela est en particulier vrai si $\mathrm{A} = \mathbb{N}$ et si $f = (f_n)$ est tout bêtement l'énumération naturelle des fonctions constantes (*i.e.* $f_n = n$) ; comme on sait que l'on a $\mathbb{k}_\Phi(n) \leq \mathbb{k}_f(n) + k_{\Phi, f} = n + k_{\Phi, f}$, on en déduit l'existence de « chutes importantes et imprévisibles » de $\mathbb{k}_\Phi(n)$ lorsque n parcourt les entiers.

Revenons, une dernière fois, à nos suites infinies ω de 0 et de 1. Nous avons déjà dit que la complexité de Kolmogorov ne permet pas de caractériser exactement les suites aléatoires au sens de Martin-Loef. On a cependant, entre autres, les deux résultats suivants (dus d'ailleurs à Martin-Loef) : l'entier n étant identifié à la suite finie $\Psi(n)$, et $\omega \mid n$ désignant la suite finie des n premiers termes de ω,

a) pour toute suite ω, il existe une constante C_ω telle qu'on ait

$$\mathbb{k}_\Phi(\omega \mid n) \leq n - \log_2(n+1) + \mathrm{C}_\omega$$

pour une infinité d'entiers n ;

b) pour toute suite ω aléatoire au sens de Martin-Loef, il existe une constante C'_ω telle que l'on ait

$$\mathbb{k}_\Phi(\omega \mid n) \geq n - 4\log_2(n+1) - \mathrm{C}'_\omega$$

pour tout entier n.

BIBLIOGRAPHIE

D'abord, quelques jalons historiques, de Borel à Martin-Loef :

1909 E. BOREL : Les probabilités dénombrables et leurs applications arithmétiques (*Rend. Circ. Mat. Palermo* **27**, 247–271).

1919 R. VON MISES : Grundlagen der Wahrscheinlichkeitsrechnung (*Math. Z.* **5**, 52–99).

1933 D.G. CHAMPERNOWNE : The construction of decimals normals in the scale of ten (*J. London Math. Soc.* **8**, 254–260).

1937 A. WALD : Die Widerspruchsfreiheit des Kollektivbegriffes (*Ergebnisse eines mathematischen Kolloquiums* **8**, 38–72; et aussi *in Actualités Sci. Indust.* **735**, 79–99, 1938).

1939 J. VILLE : Étude critique de la notion de collectif (Gauthier-Villars, Paris).

1940 A. CHURCH : On the concept of a random sequence (*Bull. Amer. Math. Soc.* **46**, 130–135).

1963 A.N. KOLMOGOROV : On tables of random numbers (*Sankhyā* **25**, 369–376).

1964 R.J. SOLOMONOFF : A formal theory of inductive inference (*Information and Control* **7**, 1–22).

1965 A.N. KOLMOGOROV : Tri podhoda k opredeleniju ponjatija « količestvo informaciĭ » (*Probl. peredači inform.* **1**, 3–11).

1966 P. MARTIN-LOEF : The definition of random sequences (*Information and Control* **9**, 602–619).

Ensuite, les articles et ouvrages que j'ai vraiment consultés (il est heureux qu'il y ait quelques répétitions!), par ordre alphabétique des auteurs, en ce qui concerne les « nombres au hasard » :

D.E. KNUTH : The Art of Computer Programming, Vol. 2 : Seminumerical Algorithms (Addison-Wesley, Reading, 1969).

> [C'est un ouvrage magnifiquement écrit, auquel le lecteur pourra se reporter, en particulier, pour avoir une idée de ce qu'on entend par « table de nombres au hasard ».]

L.A. LEVIN : On the notion of a random sequence (*Soviet. Math. Dokl.* **14**, 1413–1416, 1973).

L.A. LEVIN : Various measures of complexity for finite objects (*Soviet. Math. Dokl.* **17**, 522–525, 1976).

YU.I. MANIN : A Course in Mathematical Logic (Springer, New-York - Heidelberg - Berlin, 1977).

P. MARTIN-LOEF : The definition of random sequences (*Information and Control* **9**, 602–619, 1966).

P. MARTIN-LOEF : The literature on von Mises' kollektivs revisited (mars 1966, 36 pages dont 6 de bibliographie).

> [Il s'agit de la rédaction d'une conférence à caractère historique, allant essentiellement de Borel à Church (mais la bibliographie, très substantielle, va plus loin), nettement plus documentée que la nôtre. J'ignore si elle a été jamais publiée. Je remercie ici mon collègue A. Fuchs qui m'a prêté un « preprint » en sa possession.]

M. MIGNOTTE : Deuxième thèse, non publiée.

> [Je remercie aussi mon collègue M. Mignotte qui m'a communiqué ses notes, ainsi que son « dossier bibliographique ».]

C.P. SCHNORR : A unified approach to the definition of random sequences (*Mathem. Syst. Theory* **5**, 246–258, 1971).

C.P. SCHNORR : Zufälligkeit und Wahrscheinlichkeit (*Lecture Notes in Mathematics* n° **218**, Springer, Heidelberg, 1971).

J. VILLE : Étude critique de la notion de collectif (Gauthier-Villars, Paris, 1939).

A.K. ZVONKIN, L.A. LEVIN : The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms (*Russian Math. Surveys* **25** n° 6, 83–124, 1970).