Équirépartition d'une suite de nombres

Thomas Chomette
Farouk Boucekkine

http://dma.ens.fr/culturemath

Observez donc le premier chiffre des 2^n pour $n \leq 30$:

1, 2, 4, 8, 16, 32, 64, 256, 512, 1 024, 2 048, 4 096, 8 192, 16 384, 32768, 65536, 131072, 262144, 524288, 1048576, 2097152, 4194304, 8388608, 16777216, 33554432, 67108864, 134217728, 268435456, 536870912, 1073741824, etc...

Que remarquez-vous ?

1 Un petit paradoxe...

On obtient 9 fois le chiffre "1", 6 fois le chiffre "2", 3 fois le chiffre "3", 3 fois le chiffre "4", 3 fois le chiffre "5", 3 fois le chiffre "6", 0 fois le chiffre "7", 3 fois le chiffre "8", 0 fois le chiffre "9".

L'expérience semble nous montrer que "1" apparaît nettement plus fréquemment que les autres.

Pourtant, il paraît clair qu'il y a « autant » de nombres commençant par "1", que par "2", "3",...Par conséquent, l'intuition voudrait qu'on obtienne, au moins *asymptotiquement*, autant de nombres commençant par "1" que par chacun des autres chiffres.

Pour expliquer ce paradoxe apparent, trois voies naturelles s'offrent à nous :

1. Aller jusqu'à 30 est insuffisant, les effets asymptotiques prévus par l'intuition ne se révèlent clairement que pour n très grand.
2. Cette bizarrerie est une propriété très spécifique de la suite des 2^n, qui se comporte de manière contraire aux probabilités intuitives.
3. L'intuition se trompe et il faut essayer de voir rigoureusement ce qui se passe.

La première hypothèse est aisée à réfuter...si on possède des moyens de calcul suffisants (un ordinateur et un logiciel de calcul, par exemple). Dans ce cas, on constate que le phénomène de prédominance relative des "1" se poursuit, avec un taux tendant vers environs 3,01/10 (contre les 1/10 prévus par l'intuition.) Constatons au passage que c'est déjà à peu près ce qu'on obtient au bout de 30 valeurs.

La seconde hypothèse peut être réfutée en s'intéressant à d'autres suites, par exemple toutes les suites de la forme M^n, pour $M \in \mathbb{R}_*^+$. On constate exactement le même phénomène, et plus surprenant, le taux de "1" comme premier chiffre tend *pour tout M* vers la même valeur approximative de 3,01/10.

Manifestement il y a quelque chose à trouver !

Pour trancher cette question, il va falloir définir proprement cette idée de « probabilité » d'apparition de nombres.

2 Là où l'intuition s'égare

Le problème principal de l'intuition dans ce type de question est le suivant : on voudrait faire des raisonnements probabilistes sur $\mathbb{N}$ *tout entier*, fondés sur le fait que, si l'on tire un nombre au hasard, tous les entiers sont équiprobables. Ce qui est impossible ! !

En effet, tirons au hasard un entier naturel, et considérons des probabilités à valeurs dans $[0,1]$, (1 étant la probabilité de tirer n'importe quel nombre de $\mathbb{N}$.)

Supposons que tous les nombres aient la même probabilité non nulle d'apparaître $\alpha \in]0,1[$, prenons $N > 1/\alpha$. La probabilité d'apparition d'un nombre compris entre 0 et N est alors $(N+1)\alpha$ qui est strictement supérieur à 1 !

Par conséquent, toute probabilité sur $\mathbb{N}$ telle que les entiers soient équiprobables implique que cette probabilité commune soit 0. Et on ne peut donc rien conclure sur le type de problème à partir duquel nous sommes partis...

En fait, la seule définition envisageable de la « proportion » recherchée est *à n fixé* :

$$\frac{Card\{p \leq n, 2^p \text{ a pour premier chiffre "1"}\}}{n}$$

Ce qui nous intéresse alors est de savoir si cette quantité a un comportement asymptotique intéressant quand $n \longrightarrow \infty$.

Pour répondre à cette question, nous allons nous ramener à $[0,1]$, où les probabilités marchent beaucoup mieux, car elles sont données par des intégrales bien définies.

Remarquons d'abord que tout nombre $X \in \mathbb{R}_+^*$ s'écrit sous la forme $X = 10^{log(X)}$, le log étant entendu de base 10, ici.

On a alors $X = 10^{log(X)-E(log(X))}.10^{E(log(X))}$, et le premier chiffre de X est donc celui de $10^{log(X)-E(log(X))}$

De plus, $0 \leq X - E(X) < 1$, et la fonction $t \longmapsto 10^t$ est une fonction croissante. Par conséquent, le premier chiffre de X dépend de la position de $X - E(X)$ dans $[0,1]$, suivant le tableau suivant :

$X - E(log(X))$ appartient à	Le premier chiffre de X est
$[0, log(2)[$	1
$[log(2), log(3)[$	2
$[log(3, log(4)[$	3
$[log(4), log(5)[$	4
$[log(5), log(6)[$	5
$[log(6), log(7)[$	6
$[log(7), log(8)[$	7
$[log(8), log(9)[$	8
$[log(9), 1[$	9

Quand on calcule $log(2)$, on trouve approximativement $0.3010299956639812\ldots$ ceci rappelle furieusement la probabilité asymptotique qu'on a trouvée par nos expérimentations du paragraphe précédent !

Or, les 2^n dont le premier chiffre est "1" sont ceux tels que $2^n - E(2^n) \in [0, log(2)[$.

Imaginons à présent, intuitivement, que les $2^n - E(2^n)$ se « répartissent bien » dans [0,1]. La « proportion » de ceux qui seront dans le segment $[0, log(2)[$ sera alors, *intuitivement* égale à $log(2)$ (rapport entre les longueurs de $[0, log(2)[$ et de $[0, 1]$.)

La suite de ce texte va nous permettre de rendre rigoureuse cette intuition qui, cette fois, n'est pas trompeuse.

3 Equirépartion d'une suite

Considérons une suite $(u_n)_{n\in\mathbb{N}}$ d'éléments du segment $[\,0\,;1\,]$. On cherche à établir un critère qui permet de savoir si la suite se « répartit bien » dans le segment. C'est à dire qu'elle va ou non parcourir « l'ensemble du segment », ce de manière « uniforme ». Ceci n'a *a priori* pas un sens très précis, en effet différents points de vue nous amènent à plusieurs types de réponses.

3.1 Vision ensembliste

La suite $(u_n)_{n\in\mathbb{N}}$ peut-elle prendre toutes les valeurs possibles ?

On sait, depuis Cantor, que ce n'est pas possible : il existe plusieurs types d'infinis distincts, c'est-à-dire plusieurs types d'ensembles infinis non équipotents (tels que l'on ne puisse pas établir de bijection entre eux). En particulier le cardinal du segment $[\,0\,;1\,]$ n'est pas dénombrable. La meilleure façon de le voir est de dire la chose suivante : la suite $(u_n)_{n\in\mathbb{N}}$ construite, on va exhiber un réel de l'intervalle $[\,0\,;1\,]$ qui n'est pas atteint par la suite. Soit donc x le réel dont le développement décimal est défini par : Le $n^{\text{ième}}$ chiffre de x après la virgule est 0 si le $n^{\text{ième}}$ chiffre de u_n après la virgule est non nul, c'est 1 sinon. Alors par définition, x et u_n ont des développement décimaux qui diffèrent au

$n^{\text{ième}}$ chiffre après la virgule, donc des développement décimaux distincts. En particulier $x \neq u_n$, ce pour tout n.

Le point de vue ensembliste n'est donc pas le bon : aucune suite ne parcourt l'ensemble du segment $[\,0\,;1\,]$.

3.2 Vision topologique

L'ensemble des valeurs prises par la suite $(u_n)_{n\in\mathbb{N}}$ peut-il être *dense* dans $[\,0\,;1\,]$? (c'est à dire que tout élément de $[0;1]$ est limite d'une suite prise parmi ces valeurs.)

Ce critère est déjà plus pertinent : il est vérifié par certaines suites, mais pas toutes (une suite constante ne le vérifie pas, par exemple). Un bon exemple de suite dense est construit comme suit : On sait que l'ensemble des rationnels de $[0;1]$ est dense dans $[0;1]$, et que cet ensemble est dénombrable, on peut donc numéroter ses éléments. Considérons donc la suite (u_n) où u_n est le $n-$ième rationnel de $[0;1]$. Cette suite est dense, et notons que 0 n'est atteint qu'une fois.

Mais une telle définition n'est pas dynamique, c'est-à-dire qu'elle ne prend pas en compte l'ordre dans lequel les termes de la suite se suivent.

Par exemple, reprenons notre suite (u_n) précédente, et construisons (v_n) telle que :

$$v_{2n} = 0 \quad et \quad v_{2n+1} = u_n$$

Tous les rationnels de $[0;1]$ sont atteints, la suite est donc dense, mais on se rend bien compte qu'il y aura une accumulation de termes en 0, la répartition ne sera pas du tout uniforme.

Au contraire, si l'on ne prend la valeur 0 qu'en les indices qui sont premiers, on aura une accumulation beaucoup moins rapide, même si au final il y aura une infinité de valeurs nulles (l'opération correspond à une permutation des entiers naturels).

On a donc trois exemples de suites denses qui ne se répartissent pas de la même façon.

3.3 Équirépartition

Finalement, on va demander à la suite de « visiter tous les lieux avec la même fréquence ». C'est une notion bien plus forte que la densité : une suite équirépartie est dense, mais l'inverse n'est pas vrai (cf contre-exemple précédent).

Intuitivement, la notion d'équirépartition se comprend assez facilement en imaginant que notre segment $[\,0\,;1\,]$ est une pellicule photographique, et que la suite laisse sur son passage des petites tâches de lumière. La notion de densité correspond à l'idée qu'au bout d'un temps infini, il n'y a pas de zone noire, aussi petite soit-elle. Alors que la notion d'équirépartition correspond à une pellicule qui tend à être uniformément éclairée.

Définition 3.1 *Une suite $(u_n)_{n\in\mathbb{N}}$ à valeurs dans $[\,0\,;1\,]$ est dite équirépartie si et seulement si pour tout couple de réels $a < b$ de $[\,0\,;1\,]$, la suite d'entiers de terme général $A_n = \mathrm{Card}\,\{p \leq n, u_p \in [\,a\,;b\,]\}$ est asymptotiquement équivalente à $n(b-a)$.*

Une suite $(u_n)_{n\in\mathbb{N}}$ à valeurs dans $\mathbb{R}$ est dite équirépartie modulo 1 si et seulement si la suite de terme général $u_n - E(u_n)$, à valeurs dans $[\,0\,;1\,[$, est équirépartie (E désigne la fonction partie entière).

Remarque 3.2 *On peut remplacer dans les définitions précédentes tous les intervalles fermés par des segments ouverts ou semi-ouverts. Voyez-vous pourquoi ? (*Indication : *regardez ce qui se passe si $a = b$.) Cela nous sera utile lors de la conclusion de ce texte...*

Remarque 3.3 *En fait, on peut définir la notion d'équirépartition sur n'importe quel espace X muni d'une mesure positive finie μ. Pour simplifier, supposons que $\mu(X) = 1$ (c'est-à-dire que μ est une « mesure de probabilités »). Une suite $(u_n)_{n\in\mathbb{N}}$ à valeurs dans X est équirépartie si et seulement si, pour tout sous-ensemble Y de X mesurable, la suite de terme général $y_n = \mathrm{Card}\,\{p \leq n, u_p \in Y\}$ est asymptotiquement équivalente à $n\mu(Y)$.*

4 Le critère de Weyl

La notion d'équirépartition est fortement rattachée à la théorie de l'intégrale de Riemann. L'idée sous-jacente est que la suite $(u_n)_{n\in\mathbb{N}}$ est représentative, qu'elle renvoie la même moyenne sur une fonction que l'intégrale de Riemann. Plus précisément, on a le

Théorème 4.1 *Une suite $(u_n)_{n\in\mathbb{N}}$ à valeurs dans $[\,0\,;1\,]$ est équirépartie si et seulement si pour toute fonction f Riemann-intégrable, on a :*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(u_k) = \int_0^1 f(t)dt$$

Démonstration : Soient $a \leq b$ deux réels de l'intervalle $[\,0\,;1\,]$. soit $\mathbb{1}_{[\,a\,;b\,]}$ la fonction indicatrice de l'intervalle $[\,a\,;b\,]$, c'est-à-dire la fonction :

$$x \mapsto \begin{cases} 1 & \text{si } x \in [\,a\,;b\,] \\ 0 & \text{sinon} \end{cases}$$

La fonction $\mathbb{1}_{[\,a\,;b\,]}$ est en escalier donc Riemann-intégrable, et d'intégrale

$$\int_0^1 \mathbb{1}_{[\,a\,;b\,]}(t)dt = \int_a^b dt = b - a$$

Or $\sum\limits_{k=1}^{n} \mathbb{1}_{[\,a\,;b\,]}(u_k) = \mathrm{Card}\,\{p \leq n, u_p \in [\,a\,;b\,]\}$ par définition de la fonction $\mathbb{1}_{[\,a\,;b\,]}$. Donc $(u_n)_{n\in\mathbb{N}}$ est équirépartie si et seulement si, pour tous réels $a \leq b$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{[\,a\,;b\,]}\left(u_k\right)=\int_0^1\mathbb{1}_{[\,a\,;b\,]}(t)dt$$

Si la suite $(u_n)_{n\in\mathbb{N}}$ est équirépartie, on a donc la propriété annoncée pour toute fonction de type $\mathbb{1}_{[\,a\,;b\,]}$, et l'on veut ceci pour toute fonction f Riemann-intégrable. Toute fonction en escalier pouvant s'écrire comme une combinaisons linéaire de fonctions indicatrices de segments, la propriété est donc encore vraie pour toute fonction en escalier.

Enfin, toute fonction Riemann-intégrable est limite uniforme d'une suite de fonctions en escalier. Autrement dit, si f est une fonction Riemann-intégrable, alors pour tout $\varepsilon > 0$, on peut trouver une fonction en escalier φ_ε telle que pour tout $x \in [\,0\,;1\,]$ on ait $|f(x) - \varphi_\varepsilon(x)| < \varepsilon$.
Soit donc f une telle fonction $\varepsilon > 0$ et φ_ε en escalier vérifiant la propriété d'approximation de f. Alors pour tout n, on a :

$$\left|\frac{1}{n}\sum_{k=1}^{n}f\left(v_k\right)-\frac{1}{n}\sum_{k=1}^{n}\varphi_\varepsilon\left(v_k\right)\right|<\frac{1}{n}\sum_{k=1}^{n}\frac{\varepsilon}{n}=\varepsilon$$

et

$$\left|\int_0^1 f(t)dt-\int_0^1\varphi_\varepsilon(t)dt\right|<\int_0^1\varepsilon dt=\varepsilon$$

Comme on a déjà $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\varphi_\varepsilon\left(v_k\right)=\int_0^1\varphi_\varepsilon(t)dt$

on en déduit qu'à partir d'un rang n_0 on a (pour $n \geq n_0$)

$$\left|\frac{1}{n}\sum_{k=1}^{n}\varphi_\varepsilon\left(v_k\right)-\int_0^1\varphi_\varepsilon(t)dt\right|\leq\varepsilon$$

Par l'inégalité triangulaire, on a donc obtenu un entier n_0 tel que pour $n \geq n_0$

$$\left|\frac{1}{n}\sum_{k=1}^{n}f\left(v_k\right)-\int_0^1 f(t)dt\right|<3\varepsilon$$

Ceci étant vrai pour tout réel ε strictement positif, on en déduit

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f\left(v_k\right)=\int_0^1 f(t)dt$$

Réciproquement, si la suite $(u_n)_{n\in\mathbb{N}}$ vérifie, pour toute fonction f Riemann-intégrable,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f\left(u_k\right)=\int_0^1 f(t)dt$$

alors ceci est vrai en particulier pour les fonction indicatrices d'intervalles (qui sont Riemann-intégrables), et donc la suite est équirépartie.

□

Corollaire 4.2 *Une suite $(v_n)_{n\in\mathbb{N}}$ est équirépartie modulo 1 si et seulement si, pour toute fonction f Riemann-intégrable et $1-$périodique, on a*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f\left(v_k\right) = \int_0^1 f(t)dt$$

Démonstration : En effet, la suite $(v_n)_{n\in\mathbb{N}}$ est équirépartie modulo 1 si et seulement si la suite de terme général $u_n = v_n - E(v_n)$ est équirépartie dans $[\,0\,;1\,]$, si et seulement si pour toute fonction f Riemann-intégrable

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f\left(u_k\right) = \int_0^1 f(t)dt$$

Si f est $1-$périodique, on a alors $f(v_n) = f(u_n)$ pour tout n, et donc on a le résultat annoncé. Inversement, si pour toute fonction f Riemann-intégrable et $1-$périodique, on a

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f\left(v_k\right) = \int_0^1 f(t)dt$$

cela signifie que la suite $(u_n)_{n\in N}$ vérifie la propriété voulue, donc est équirépartie, et donc $(v_n)_{n\in\mathbb{N}}$ est équirépartie modulo 1.

$\square$

Théorème 4.3 (Critère de Weyl) *La suite $(v_n)_{n\in\mathbb{N}}$ est équirépartie modulo 1 si et seulement si, pour tout entier m non nul,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \exp\left(2i\pi m v_k\right) = 0$$

Démonstration : Là encore, l'un des deux sens est évident : si la suite $(v_n)_{n\in\mathbb{N}}$ est équirépartie modulo 1, alors d'après le résultat précédent, pour toute fonction f Riemann-intégrable et $1-$périodique, on a

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f\left(u_k\right) = \int_0^1 f(t)dt$$

C'est donc le cas en particulier pour le fonction $x \mapsto \exp\left(2i\pi m x\right)$ (m entier non nul).

Réciproquement, on utilise le théorème de Stone-Weierstrass périodique (l'ensemble des polynômes trigonométriques est dense dans celui des fonctions périodiques continues, donc dans l'ensemble des fonctions périodiques Riemann-intégrable). Si la suite vérifie, pour tout entier $m > 0$:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \exp\left(2i\pi m v_k\right) = 0$$

alors c'est aussi vrai pour toute fonction $1-$périodique et Riemann-intégrable, d'intégrale nulle. Il suffit pour conclure de voir qu'on peut décomposer toute fonction 1−périodique et Riemann-intégrable en somme d'une fonction constante et d'une fonction d'intégrale nulle sur $[\,0\,;1\,]$.

$\square$

Corollaire 4.4 *Soit α un irrationnel. alors la suite de terme général $n\alpha$ est équirépartie modulo 1.*

Démonstration : Soit α un irrationnel. On va montrer que la suite $(n\alpha)_{n\in\mathbb{N}}$ vérifie le critère de Weyl. Toute la difficulté repose dans la remarque suivante : pour m et k deux entiers, on a $\exp(2i\pi mk\alpha) = (\exp(2i\pi m\alpha))^k$. On obtient

$$\sum_{k=1}^{n} \exp(2i\pi mk\alpha) = \sum_{k=1}^{n} (\exp(2i\pi m\alpha))^k = \exp(2i\pi\ m\alpha)\frac{1-(\exp(2i\pi m\alpha))^n}{1-\exp(2i\pi m\alpha)}$$

(somme d'une suite géométrique, la raison étant différente de 1 puisque α, et donc $m\alpha$, est irrationnel, donc non entier.) Et le rapport $\dfrac{1-(\exp(2i\pi m\alpha))^n}{1-\exp(2i\pi m\alpha)}$ est borné :

$$\left|\frac{1-(\exp(2i\pi m\alpha))^n}{1-\exp(2i\pi m\alpha)}\right| \leq \frac{1+|(\exp(2i\pi m\alpha))^n|}{|1-\exp(2i\pi m\alpha)|} = \frac{2}{|1-\exp(2i\pi m\alpha)|}$$

On en déduit $$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \exp(2i\pi mk\alpha) = 0$$

D'après le critère de Weyl, la suite $(n\alpha)_{n\in\mathbb{N}}$ est donc équirépartie modulo 1.

$\square$

5 Conclusion

Comme on l'a vu dans la vision intuitive en fin du **§2**, la question importante est celle de l'équirépartition modulo 1 de la suite $n.log(2)$.

Utilisons donc le **Corollaire 4.4**, en montrant par l'absurde que $log(2)$ est irrationnel.

Supposons donc que $log(2) = p/q$ avec $p \in \mathbb{Z}$ et $q \in \mathbb{N}^*$. On a alors $2 = 10^{p/q}$, et par suite

$$2^q = 10^p = 5^p.2^p$$

Par conséquent $p = q = 0$, par unicité de la décomposition en nombres premiers. C'est donc impossible, et $log(2)$ est bien irrationnel.

La suite $n.log(2) - E(n.log(2))$ est donc bien équirépartie dans $[0;1]$, et on a donc

$$Card\{p \leq n \ tels\ que\ p.log(2) - E(p.log(2)) \in [0;log(2)[\} \simeq n.log(2)$$

Ce qui équivaut à

$$\lim_{n\to\infty} \frac{Card\{p \leq n, 2^p \text{ a pour premier chiffre "1"}\}}{n} = log(2).$$

Ainsi, la proportion des 2^p de premier chiffre 1 pour p inférieur à un n donné a un sens et tend bien vers la valeur expérimentalement trouvée. Le paradoxe apparent est levé.

Notons par ailleurs que le même raisonnement est valable pour la suite M^n pour presque tous les $M \in \mathbb{R}^+$, et donne alors le même résultat. Quelles sont les exceptions ? Qu'en est-il avec la suite 2^{2^n} ?

Remerciement : Ce texte est issu de la lecture par le premier auteur d'un texte de François Fayard, que nous remercions vivement !