Les Nombres premiers, entre l'ordre et le chaos

Gérald Tenenbaum
Professeur à l'Institut Elie Cartan à Nancy

Michel Mendès France
Professeur à l'Université Bordeaux 1

DUNOD

Une précédente édition de cet ouvrage a été publiée
en 1997 aux Presses Universitaires de France
dans la collection « Que sais-je ? », rééditée en 2000.

ISBN : 978-2-1005-5936-7

Ce livre est dédié à la mémoire de Paul Erdős, qui nous a quittés en septembre 1996, au moment même où nous achevions la rédaction de la première édition. Il fut pour nous un oncle (ainsi que l'appelaient ses amis) et un maître. Un géant des mathématiques a disparu, mais son empreinte marquera bien des siècles à venir.

Table des matières

Avant-propos

Issu de notre ouvrage *Les Nombres premiers*, paru dans la collection *Que sais-je ?* et à présent épuisé, ce petit livre représente un pari sans doute ambitieux : fournir au grand public scientifique une description concise de la théorie analytique moderne des nombres premiers — à l'exclusion toutefois des apports de la théorie des formes modulaires.

Répondant à des questions posées depuis l'Antiquité — y a-t-il beaucoup de nombres premiers ?, comment se répartissent-ils ?, etc. —, ce domaine connaît depuis un siècle un essor sans précédent, dû notamment aux interactions avec la théorie des probabilités. Les tables de nombres premiers mettent en évidence un aspect chaotique, dont le désordre apparent s'accorde finalement avec des modèles aléatoires classiques, issus, par exemple, de phénomènes physiques. Là est précisément l'objet de cet opuscule : décrire, puis tenter de comprendre, comment une suite aussi hautement déterminée que celle des nombres premiers peut renfermer une telle part de hasard.

Insistons un peu sur ce point. Le hasard total, le chaos, c'est la complexité infinie. Par ailleurs, la complexité d'un nombre entier croît manifestement avec sa taille, et répondre à des questions de base devient souvent difficile : le nombre $2^{43112609} - 1$ est-il premier ?[(1)] Combien de fois son développement décimal contient-il le chiffre 7 ? etc. Au voisinage de l'infini, la suite des nombres entiers, et partant celle des nombres premiers, mime le hasard. Les directions modernes de la théorie analytique des nombres tentent de rendre compte des modalités de cette tendance.

1. Oui, d'après Edson Smith (2008).

Physiciens et philosophes discutent encore de l'hypothétique « variable cachée », malgré les travaux convaincants d'Alain Aspect, qui semblent en proscrire l'existence. L'école de Copenhague, avec Niels Bohr, défend la thèse que le monde subatomique est régi par le hasard. Einstein, quant à lui, rêve d'une explication sub-subatomique totalement déterministe. La suite des nombres premiers ne pourrait-elle servir de modèle aux idées d'Einstein, lui qui prétendait que Dieu ne joue pas aux dés au moment même où Mark Kac, éminent arithméticien, professait l'opinion que les nombres premiers s'adonnent secrètement au jeu de pile ou face ?

La dialectique ordre/désordre occupe les théoriciens des nombres depuis que Legendre et Gauss ont conjecturé une répartition harmonieuse des nombres premiers, à savoir que le n-ième nombre premier p_n est proche de $n \ln n$.[1] Une telle régularité dans l'aléatoire ne doit pas surprendre : quoi de plus imprévisible que le jet d'une pièce de monnaie alors que la probabilité qui régit l'événement, constamment égale à $\frac{1}{2}$, prévoit une tendance à l'équilibre entre les piles et les faces ? Tous ceux qui ont observé la suite des nombres premiers ont remarqué à la fois l'irrégularité et la régularité de leur distribution. À l'instar du jeu de pile ou face, le comportement en moyenne est régulier alors que les fluctuations locales, ici le passage de p_n à p_{n+1}, demeure très complexe.

Nous avons choisi de décrire ces phénomènes de répartition en nous appuyant sur le cheminement historique et l'évolution graduelle de la philosophie — c'est-à-dire la représentation archétypale — des nombres premiers. Les Chapitres 1, 2 et 4 sont principalement consacrés aux résultats de régularité, alors que le Chapitre 3 traite surtout des aspects aléatoires de la répartition. Au Chapitre 5, nous décrivons les conjectures principales qui sous-tendent la théorie et nous comprenons, *in fine*, que cette dichotomie n'est qu'apparente : hasard et nécessité se conjuguent harmonieusement pour produire de la structure, chacune des

1. Cette assertion, qui porte aujourd'hui le nom de *théorème des nombres premiers*, a été démontrée par Jacques Hadamard et Charles de La Vallée-Poussin en 1896.

deux perspectives éclairant et expliquant l'autre ; compte-tenu des contraintes liées à la structure d'ordre des entiers, la répartition des nombres premiers s'avère aussi harmonieuse que possible.

La présente édition diffère notablement des précédentes. Nous avons, en particulier, significativement enrichi les introductions des différents chapitres de manière à fournir au non-spécialiste une description aussi fidèle que possible du contenu mathématique et des idées-forces sous-jacentes. Ces développements, largement métaphoriques, peuvent, en toute rigueur, être omis par un lecteur rompu au langage mathématique et aux notions principales utilisées dans le texte : logarithmes, congruences, nombres complexes, convergences, interversion de sommations, etc. En revanche, ils peuvent constituer l'essentiel de l'apport pour celui qui ne possède pas (ou pas encore) le bagage scientifique nécessaire au décryptage des démonstrations. Notre espoir est qu'ils recèlent également un intérêt pour le scientifique chevronné, soit en fournissant une piste de vulgarisation — les scientifiques parlent aussi aux profanes —, soit en mettant en évidence une « philosophie mathématique » implicite.

Nous avons également tenu compte des récentes avancées de la théorie en incluant la recension de deux résultats remarquables.

Le premier, dû à Goldston, Pintz et Yıldırım, affirme que, pour tout $\varepsilon > 0$, la différence $p_{n+1} - p_n$ entre deux nombres premiers consécutifs est inférieure à $\varepsilon \ln p_n$ pour une infinité d'indices n.[1] On est loin de la conjecture des nombres premiers jumeaux selon laquelle cette différence est infiniment souvent égale à 2, mais cela représente un saut qualitatif considérable.

Le second résultat est dû à Green et Tao. Il confirme une importante conjecture d'Erdős selon laquelle la suite des nombres premiers contient des progressions arithmétiques arbitrairement longues. Ainsi $\{3, 5, 7\}$ est une progression de raison 2 et de longueur 3, alors que

1. Plus précisément, ces auteurs établissent que, pour une constante convenable c, on a

$$p_{n+1} - p_n \leqslant c\sqrt{\ln p_n}(\ln \ln p_n)^2$$

pour une infinité d'indices n.

{199, 409, 619, 829, 1039, 1249, 1459, 1669, 1879, 2089} est une progression de raison 210 et de longueur 10. On sait dorénavant qu'il en existe de longueur supérieure à toute borne donnée par avance.

Nous ne donnerons pas les preuves de ces résultats. Elles sont longues et profondes et les détailler nous conduirait loin hors du cadre de cet ouvrage. Nous tenterons toutefois de satisfaire la curiosité du lecteur en fournissant des indications aussi précises que possible sur les idées essentielles. Signalons que Terence Tao a été récompensé en 2006 par la Médaille Fields, la plus haute distinction internationale pour les mathématiques.

Nous avons cru intéressant de parsemer cette édition d'illustrations de mathématiciens cités ici. Si le texte nous fait plonger dans une abstraction souvent exigeante, les portraits apportent une dimension sensible à l'exposé : les mathématiciens sont des êtres humains et voir leurs visages permet d'approcher leurs personnalités. Derrière leurs regards, on pourra peut-être deviner ce minuscule décalage qui souvent détermine une vie entière.

Tout choix est par nature restrictif : notre parti pris narratif a aussi ses dangers et ses désavantages. Rompant délibérément (certains diront scandaleusement) avec une tradition séculaire dans ce type d'ouvrage, nous ne fournissons pas de table de nombres premiers[1] et nous ne donnons pas notre démonstration favorite de la loi de réciprocité quadratique. Plus grave, les divers aspects de la théorie du crible sont seulement esquissés, bien que cette approche ait fourni de remarquables avancées, et nous n'abordons pas les diverses et profondes généralisations des nombres premiers en algèbre commutative : idéaux premiers des corps de nombres, polynômes irréductibles sur un anneau ou un corps fini, etc. On consultera à cet effet les ouvrages classiques disponibles en français.[2] Nous occultons aussi, quasi totalement, l'aspect « diviseurs » des nombres

1. Une carence à laquelle suppléeront aisément les calculatrices de bureau.

2. Voir, par exemple : Samuel, *Théorie algébrique des nombres,* Hermann, 1967 ; Borevitch & Chafarevitch, *Théorie des nombres*, Gauthier–Villars, 1967 ; Serre, *Corps locaux*, Hermann, 1968.

premiers qui fournit pourtant aux méthodes probabilistes de la théorie des nombres un champ d'investigation privilégié.[1] Enfin, nous ne faisons qu'aborder très succinctement (au § 1.5) les aspects cryptographiques et algorithmiques de la théorie, dont les saisissantes applications ont franchi, depuis plusieurs décennies, les frontières de la médiatisation grand public.[2]

La science en général, et, en son sein, les mathématiques, constitue une part sans cesse grandissante de la culture générale. Par ailleurs, les ouvrages « plaisants et délectables »[3] consacrés aux aspects spectaculaires des nombres premiers ne manquent pas, et certains sont d'ailleurs tout à fait remarquables.[4] Nous avons donc choisi d'assumer notre différence, selon une expression aujourd'hui consacrée, en visant un peu plus haut qu'il n'est d'usage dans un ouvrage de vulgarisation. Nous sommes conscients que certains développements pourront sembler ardus — ils le sont. Nous avons parfois préféré un court calcul (le dessin des mathématiciens) à de longues explications, et le style est volontairement dense, voire, de place en place, allusif. Cela nous a paru nécessaire à la mise en évidence des arguments essentiels. Ainsi, nous espérons qu'un lecteur assidu et tenace verra sa curiosité satisfaite par des preuves localement complètes — c'est en particulier avec ce souci que nous avons rédigé le Chapitre 4, essentiellement autonome. Mais nous encourageons aussi le lecteur plus pressé, ou moins désireux d'entrer dans les détails, à lire cet opuscule « en diagonale », tant il est vrai que seules les définitions comptent, pourvu qu'on les comprenne, et avec elles l'émergence d'une logique/musique

1. Ce point de vue est développé dans quelques ouvrages à présent classiques, en particulier : Elliott, *Probabilistic number theory* (2 vol.), Springer Verlag, 1979-1980 ; Hall & Tenenbaum, *Divisors*, Cambridge University Press, 1988 ; Montgomery& Vaughan, *Multiplicative number theory I*, Cambridge University Press, 2007 ; Tenenbaum, *Introduction à la théorie analytique et probabiliste des nombres*, Belin 2008.

2. Pour en savoir plus, voir par exemple les ouvrages de Robin, de Menezes, van Oorschot & Vanstone, et de Koblitz, cités dans la bibliographie.

3. Selon l'expression de Bachet (1612).

4. L'exhaustif *Nombres premiers : mystères et records* (PUF, 1994) de Ribenboim, et le lumineux *Merveilleux nombres premiers* (Belin, 2000) de Delahaye en font partie.

intrinsèque où elles s'appellent mutuellement. Le reste n'est que bavardage.

À la lecture analytique scolaire consistant à ne lire la ligne $n+1$ que lorsque l'on a compris et assimilé la ligne n, nous voulons opposer (ce qui est rendu possible précisément par la relative tension de l'exposé), une lecture synthétique, un *glissando*, où le fil directeur est sans cesse apparent. La synthèse faite, rien n'empêche (et, on l'aura compris, c'est en fait nécessaire pour progresser encore) de reprendre la lecture plume en main et d'affronter la rigueur, voire la rugosité, des démonstrations. Le jeu en vaut la chandelle.

Ce livre n'est donc pas facile, mais nous espérons qu'empreint de mystère, il engendrera une discrète poésie. De la complexité naît le rêve. Ni Stéphane Mallarmé ni Umberto Eco ne nous contrediront.

La maison Dunod, que nous avons plaisir à remercier ici, s'est engagée à nos côtés pour maintenir, sous une forme attrayante, la disponibilité de notre texte initial, qui a connu un certain succès depuis 1997, avec notamment la reconnaissance de l'Académie des Sciences,[1] une traduction anglaise,[2] et une traduction chinoise.[3]

Pour cette édition nous tenons évidemment à exprimer notre gratitude tous ceux qui nous ont aidé dès 1997 : Jean-Paul Allouche, Jean-Philippe Anker, Michel Balazard, Daniel Barlet, Régis de la Bretèche, Éric Charpentier, Hédi Daboussi, Cécile Dartyge, Jean-Marc Deshouillers, Jean-Claude Fort, Andrew Granville, Jerzy Kaczorowski, Bernard Landreau, Pierre Marchand, Gérard Mathieu, Jean-Louis Nicolas, Emmanuel Pedon, Patrick Sargos, Jacques Sicherman, André Sef, Jie Wu, et Paul Zimmermann.

1. Prix Paul Doistau-Émile Blutet de l'information scientifique 1999.

2. *The prime numbers and their distribution*, Student mathematical library 6, American Mathematical Society, 2000

3. *Les nombres premiers,* Mathematics series for graduate students 9, Tsinghua University Press, 2007.

À cette liste, il convient à présent d'ajouter Vitaly Bergelson, Daniel Goldston, Guillaume Hanrot, Charles Mozzochi, János Pintz, Jia-Yan Yao et Cem Yıldırım pour leurs précieux conseils. Régis de la Bretèche, Cécile Dartyge, et Jie Wu se sont à nouveau amicalement et efficacement mobilisés pour améliorer la présente édition.

Nancy et Bordeaux, mai 2010,
G. T. & M. M.F.

Notations et conventions

Nous indiquons ici les principales notations et conventions utilisées dans l'ensemble de l'ouvrage. Celles qui n'apparaissent que dans un chapitre ou un paragraphe sont définies localement.

La lettre $\mathbb{N}$ désigne l'ensemble des entiers naturels $\{1, 2, \ldots\}$ et $\mathcal{P}$ celui des nombres premiers. Les ensembles des entiers relatifs, des nombres réels et des nombres complexes sont désignés respectivement par $\mathbb{Z}$, $\mathbb{R}$, et $\mathbb{C}$. La lettre p, avec ou sans indice désigne toujours un élément de $\mathcal{P}$. On écrit $a \mid b$ (resp. $a \nmid b$) pour signifier que a divise (resp. ne divise pas) b et $p^\nu \| a$ indique que $p^\nu \mid a$ et $p^{\nu+1} \nmid a$.

Le pgcd de deux entiers $a,\ b$ est noté (a, b). Lorsque $(a, b) = 1$, on dit que a et b sont premiers entre eux. Le nombre d'éléments d'un ensemble fini A est désigné, selon les circonstances, par $|A|$ ou $\sum_{a\in A} 1$. On désigne par $P^+(a)$ (resp. $P^-(a)$) le plus grand (resp. le plus petit) facteur premier d'un entier $a \in \mathbb{N}$, avec la convention $P^+(1) = 1$, $P^-(1) = \infty$.

Le logarithme népérien est noté ln.[1] Les itérés ln ln, ln ln ln, etc., sont notés $\ln_2$, $\ln_3$, etc. La constante d'Euler γ est définie comme la limite

$$\gamma = \lim_{N\to\infty} \Big(\sum_{n\leqslant N} 1/n - \ln N \Big).$$

On a $\gamma \approx 0,577215664$. Il est à noter, cependant, que nous suivrons l'usage traditionnel en désignant également par γ, au chapitre II, la partie imaginaire d'un zéro générique non trivial de la fonction zêta de Riemann. Aucune confusion ne sera à craindre.

1. $\ln a$ est donc, pour $a \geqslant 1$, l'aire du domaine plan limité par les axes $x = 1$, $x = a$, $y = 0$ et la courbe $y = 1/x$. On a, lorsque a est « grand », $\ln a \sim \sum_{n\leqslant a} 1/n$.

La partie entière et la partie fractionnaire d'un nombre réel x sont notées respectivement $\lfloor x \rfloor$ et $\langle x \rangle$. Ainsi

$$\lfloor 5/3 \rfloor = 1, \quad \langle -3, 15 \rangle = 0, 85.$$

Le signe d'affectation $:=$ indique que le membre de gauche d'une égalité est défini par celui de droite.

La fonction *logarithme intégral* est définie par

$$\mathrm{li}(x) := \int_2^x \frac{\mathrm{d}t}{\ln t} \qquad (x \geqslant 2).$$

Lorsque la lettre s désigne un nombre complexe, nous définissons implicitement ses parties réelle et imaginaire par $s = \sigma + i\tau$.

Étant données des fonctions f, g, de variable réelle ou complexe, nous employons indifféremment la notation de Landau $f = O(g)$ ou celle de Vinogradov $f \ll g$ pour signifier qu'il existe une constante positive C telle que $|f| \leqslant Cg$ dans le domaine de définition commun à f et g. Une éventuelle dépendance de C en fonction d'un paramètre α pourra être indiquée sous la forme $f = O_\alpha(g)$, ou $f \ll_\alpha g$. La notation de Landau $f = o(g)$ est utilisée dans son sens habituel de $\lim f/g = 0$.[1]

Nous désignons par *fonction indicatrice* d'un ensemble A la fonction qui vaut 1 sur A et 0 sur son complémentaire. Enfin, $\mathcal{C}^k[a, b]$ désigne l'espace des fonctions k fois continûment dérivables sur l'intervalle $[a, b]$.

Par ailleurs, nous utiliserons souvent la manipulation suivante pour estimer une moyenne pondérée par des coefficients complexes a_n $(n \in \mathbb{N})$ d'une somme aux valeurs entières d'une fonction $f \in \mathcal{C}^1[1, x]$:

$$\begin{aligned}\sum_{1 \leqslant n \leqslant x} a_n f(n) &= \sum_{1 \leqslant n \leqslant x} a_n \Big\{ f(x) - \int_n^x f'(t)\,\mathrm{d}t \Big\} \\ &= f(x) \sum_{1 \leqslant n \leqslant x} a_n - \int_1^x f'(t) \Big\{ \sum_{1 \leqslant n \leqslant t} a_n \Big\}\,\mathrm{d}t.\end{aligned}$$

1. Ainsi $O(1)$ désigne une quantité bornée alors que $o(1)$ dénote une quantité qui tend vers 0.

Chapitre 1

La genèse : d'Euclide à Tchébychev

1. Introduction

Compter, c'est d'abord compter sur soi. Au sens figuré, bien sûr, mais aussi au sens propre : compter sur ses doigts, sur ses pieds, sur ses épaules, sur ses genoux, etc. L'étymologie des noms de nombres fait en effet apparaître qu'ils sont des vestiges de langues très anciennes dans lesquelles ils désignaient les différentes parties du corps. Archétypes de notre représentation du monde, les nombres font, au sens le plus fort, partie de nous. À tel point que l'on peut légitimement se demander si l'objet d'étude de l'arithmétique n'est pas l'esprit humain lui-même. De là, naît une étrange fascination : comment ces nombres, que nous portons si profond, engendrent-ils des énigmes aussi redoutables ? Parmi ces mystères, celui des nombres premiers est sans doute l'un des plus anciens et des plus résistants. Notre objectif dans ce petit livre est d'initier le lecteur à quelques-unes des méthodes inventées par l'homme pour appréhender cette récalcitrante intimité. Puisse-t-il mesurer notre ignorance à l'aune de ces arcanes imbriqués et en concevoir un insatiable appétit de connaissance.

Il y a, fondamentalement, deux manières de conjuguer les entiers. On peut les ajouter et les multiplier. Mais alors que, par l'addition, on peut retrouver chaque entier fixé à l'avance à l'aide d'entiers plus petits, on s'aperçoit rapidement que, pour la multiplication, il est nécessaire d'introduire, de-ci de-là, des éléments nouveaux, irréductibles à ceux qui les précèdent. Ces éléments sont appelés des nombres premiers et, depuis la nuit des temps, l'humanité cherche à préciser le « de-ci de-là »...

L'ensemble des nombres premiers débute donc ainsi :

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, . . ., 571, . . .

Il y a près de vingt-trois siècles, Euclide démontrait l'infinitude de l'ensemble des nombres premiers. D'une grande beauté et d'une grande simplicité, sa preuve, avec les notations modernes, tient en quatre caractères :

$$n! + 1.$$

En effet, ce nombre n'est divisible par aucun entier d tel que $2 \leqslant d \leqslant n$; il ne possède donc que des facteurs premiers excédant n. Cela établit l'existence d'au moins un nombre premier plus grand que toute limite fixée à l'avance.

Pour tout nombre réel $x > 0$, on note $\pi(x)$ le nombre des nombres premiers p qui ne dépassent pas x, soit

$$\pi(x) := \sum_{p \leqslant x} 1.^{(1)}$$

Le résultat d'Euclide signifie que $\pi(x)$ tend vers l'infini avec x. Le problème se pose donc d'étudier la vitesse de croissance de cette fonction.

Il faut attendre Euler, au dix-huitième siècle, pour qu'une véritable percée soit faite dans ce domaine. Il découvre la formule

1. Cette écriture signifie simplement que l'on compte 1 pour chaque nombre premier n'excédant pas x. La notation peut sembler étrange à première vue, mais elle a fait ses preuves, notamment parce qu'elle se prête à diverses manipulations analogues à celles des intégrales.

fondamentale

$$\zeta(\sigma) := 1 + \frac{1}{2^\sigma} + \frac{1}{3^\sigma} + \cdots$$
$$= \frac{1}{1-1/2^\sigma}\,\frac{1}{1-1/3^\sigma}\,\frac{1}{1-1/5^\sigma}\cdots \qquad (\sigma > 1),$$

qui relie une somme étendue à tous les entiers à un produit infini portant sur tous les nombres premiers. La formule n'est pas valable pour $\sigma = 1$, mais Euler n'hésitait pas à lui donner le sens suivant[1]

$$\prod_p \big(1 - 1/p\big) := \big(1-1/2\big)\big(1-1/3\big)\big(1-1/5\big)\cdots = 0.$$

Sachant que $\ln(1-1/p)$ est de l'ordre de $-1/p$, Euler en concluait que la somme des inverses des nombres premiers diverge :

$$\sum_p \frac{1}{p} := \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \cdots = \infty.$$

Ce calcul sera justifié et précisé au § 7.

Gauss (1792), puis Legendre (1798, 1808) conjecturent que l'on a, lorsque x tend vers l'infini,

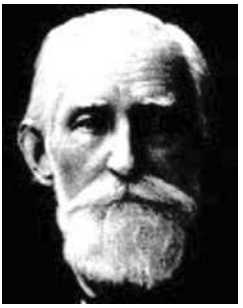
$$\pi(x) \sim x/\ln x.^{[2]}$$

Quelques décennies plus tard, en 1852, Tchébychev établit une forme faible de cette conjecture : il existe des constantes strictement positives a et b telles que l'on ait

$$ax/\ln x < \pi(x) < bx/\ln x \qquad (x \geqslant 2).$$

Carl Friedrich Gauss
(1777–1855)

1. Qu'en langage moderne on pourrait qualifier de prolongement par continuité.
2. Il est utile de garder à l'esprit que cela équivaut à la validité de la formule asymptotique $p_n \sim n \ln n$, où p_n désigne le n-ième nombre premier.



Pafnouti Tchébychev
(1821–1894)

Numériquement, le résultat constitue un argument très convaincant en faveur de la conjecture de Gauss-Legendre : pour x assez grand, on peut choisir $a = 0{,}921$ et $b = 1{,}106$.

Un petit détour est nécessaire pour décrire le progrès suivant, dû à Riemann.

Le lecteur a sans doute, sinon une connaissance précise, du moins une certaine idée de ces nombres particuliers, qualifiés d'« imaginaires » par Descartes, et inventés par les mathématiciens italiens du XV^{e} siècle pour résoudre les équations de degré 3 et 4. Toute l'histoire consiste à admettre l'existence de racines carrées de nombres négatifs et à constater que le calcul formel fonctionne. Ainsi, selon Cardan, si l'on note $i = \sqrt{-1}$, les solutions de l'équation $x(10 - x) = 40$ sont $5 \pm i\sqrt{15}$: par exemple

$$\begin{aligned}(5 + i\sqrt{15})(5 - i\sqrt{15}) &= 25 - (i\sqrt{15})^5 \\ &= 25 - i^2 15 = 25 + 15 = 40.\end{aligned}$$

Le trouble vient de ce qu'on sait depuis toujours que le carré d'un nombre réel est nécessairement positif, et donc qu'il n'existe aucun nombre réel dont le carré soit négatif. Descartes, Newton et Bombelli considéraient que l'apparition de ces nombres comme solutions d'équations signifiait que le problème était en réalité impossible.

René Descartes
(1596–1650)

Mais, en terre de mathématiques, l'interdiction peut être contournée sans contrevenir aux règles de la logique. À la suite de Wallis et Gauss, les mathématiciens ont trouvé un espace nouveau dans lequel les nombres réels et imaginaires pouvaient cohabiter paisiblement : il suffisait de penser à introduire une dimension supplémentaire

pour l'univers des nombres — une révolution. À la droite numérique, où les nombres réels sont rangés à la queue-leu-leu, on a substitué un pré carré où la longueur était mesurée à l'aune du nombre réel 1, et la largeur à celle du nombre imaginaire i. Dans ce nouveau monde, un nombre est une combinaison du type

$$a + ib$$

où a et b sont réels. Sous cette forme, aucune simplification n'est possible, mais le calcul suit les règles ordinaires, compte tenu de la loi $i^2 = -1$: par exemple

$$(a + ib)^2 = a^2 - b^2 + 2iab.$$

L'univers des combinaisons $a+ib$ de nombres réels et imaginaires est désigné comme le champ des *nombres complexes*, habituellement noté par la lettre $\mathbb{C}$. Dans $\mathbb{C}$, on peut additionner, soustraire, diviser, multiplier selon les lois usuelles. On dit que $\mathbb{C}$ est un *corps*.

Le corps $\mathbb{C}$ des nombres complexes offre une surprise de taille : non seulement *toutes* les équations polynomiales à coefficients réels y ont des solutions, mais il en va de même des équations à coefficients complexes. Ainsi, l'équation $x^3 + 2i = 0$ a pour solutions

$$i\sqrt[3]{2}, \quad -\tfrac{1}{2}\sqrt[3]{2}\sqrt{3} - i\tfrac{1}{2}\sqrt[3]{2}, \quad \tfrac{1}{2}\sqrt[3]{2}\sqrt{3} - i\tfrac{1}{2}\sqrt[3]{2}.$$

Riemann (1859) exploite l'idée géniale de prolonger la fonction zêta

$$\zeta(s) = \sum_{n \geqslant 1} n^{-s}$$

en une fonction de la variable complexe. Cela permet de donner un sens à $\zeta(s)$ même hors du domaine de convergence de la série et, comme le démontre Riemann, l'étude du prolongement se prête remarquablement bien à la déduction de renseignements asymptotiques concernant $\pi(x)$.

Enfin, en 1896, Hadamard et La Vallée-Poussin, s'appuyant l'un et l'autre sur les travaux de Riemann, démontrent indépendamment le théorème des nombres premiers

$$\pi(x) \sim x/\ln x \qquad (x \to \infty).$$

Bernhard Riemann
(1826–1866)

Jacques Hadamard
(1865–1963)

Charles de La Vallée-Poussin
(1866–1962)

Ce livre est un peu l'histoire de cette aventure. Nous verrons qu'elle n'est pas terminée.

2. Une brève histoire de ce qui va suivre

2.1. Les lettres et les mots

Comme annoncé dans l'avant-propos, ce paragraphe, ainsi que tous ceux qui portent le même titre aux chapitres suivants, invite à une promenade initiatrice ouvrant sur la suite du chapitre.

La raison essentielle de l'intérêt des mathématiciens pour les nombres premiers est qu'ils constituent un alphabet (infini) permettant d'attribuer un nom unique à chaque nombre entier : ainsi le « nom » de 12 est 2·2·3, et celui de 2010 est 2·3·5·67. Dans ce dictionnaire à jamais inachevé, chaque mot possède un sens exclusif et toutes les combinaisons de lettres constituent des mots.

C'est d'ailleurs précisément pour cette raison que le nombre 1 (qui satisfait à la règle de n'être divisible que par lui-même et par 1) n'est pas considéré comme un nombre premier par les mathématiciens : s'il l'était, il y aurait une infinité de mots, ne différant que par le nombre des 1 y apparaissant, pour représenter chaque nombre entier.

Euclide
(∼330 – ∼275)

Aussi naturelle qu'elle paraisse être, la preuve de cette correspondance univoque et exhaustive entre les nombres entiers et les produits de nombres premiers n'est pas une évidence : comment peut-on être certain que la seule manière d'obtenir 48613 en multipliant entre eux deux entiers plus grands que 1 consiste à combiner 173 et 281 ? C'est au mathématicien d'Alexandrie (sans doute) d'origine grec Euclide que l'on doit la première démonstration rigoureuse.

Gauss a généralisé l'énoncé. Sa démonstration, reposant sur un résultat de Bachet datant de 1624, est encore celle qui est la plus communément utilisée pour démontrer le résultat d'Euclide.

Une version imagée de cette preuve est la suivante : supposons que vous pariez à pile ou face avec un ami selon la règle suivante : « tu me donnes x euros à chaque fois que je gagne, et je te donne y euros lorsque c'est toi qui gagnes » Le théorème est alors le suivant : si y est premier et si x n'est pas multiple de y, il existe toujours une suite de gains et pertes conduisant à un solde de 1 euro. Exemple : si vous recevez 20 euros à chaque gain et donnez 17 euros à chaque perte, il vous restera exactement 1 euro après 6 gains et 7 pertes.

2.2. Des horloges particulières

On peut déduire de ce qui précède le résultat suivant, dont nous donnons ici encore une formulation analogique. Supposons que le cadran d'une pendule soit gradué de telle sorte qu'une heure (c'est-à-dire un tour complet) corresponde à un nombre premier p

de minutes. Alors pour tout nombre a de minutes ne correspondant pas à un nombre entier d'heures, le nombre a^{p-1}, une fois reporté sur la pendule, aboutira à exactement une minute après l'heure pile. Exemple : $p = 13$, $2^{p-1} = 2^{12} = 315 \times 13 + 1$, soit 315 tours complets et une minute.

Ce phénomène est en fait plus général : pour tout nombre b de minutes ne correspondant pas à un nombre exact d'heures, on pourra trouver un exposant c tel que la puissance a^c coïncide avec b sur le cadran de l'horloge. Il se trouve que le calcul de c connaissant a et b est extrêmement difficile, alors que celui de a^c est très peu coûteux en temps de calcul. Ce mécanisme constitue donc un verrou facile à fermer et difficile à ouvrir : c'est le principe (très simplifié) des méthodes de cryptographie modernes.

Un autre champ d'étude, également dû à Gauss, consiste à se demander, parmi des graduations de l'horloge, lesquelles sont des carrés parfaits — autrement peuvent être obtenus en faisant parcourir a^2 graduations à l'aiguille — et lesquelles n'en sont pas. Surprise : alors qu'il existe 10 carrés parfaits jusqu'à 100, on constate que sur l'horloge avec 101 graduations, il y a 50 carrés ! Ce phénomène se révèle général : sur les horloges à nombre premier, essentiellement une graduation sur deux est un carré parfait.

Une grande découverte de Gauss, connue sous le nom de *loi de réciprocité quadratique*, consiste à donner une formule simple permettant de décider si un nombre premier q est un carré sur l'horloge à p minutes dès que l'on sait si p est ou non carré sur l'horloge à q minutes.

Les nombres premiers, d'apparence si irrégulière, renferment d'incroyables symétries.

2.3. Compter les nombres premiers

L'un des mystères les plus captivants concernant les nombres premiers consiste à évaluer leur nombre $\pi(x)$ dans l'intervalle $[2, x]$. Comme indiqué au paragraphe précédent, il est assez facile

de montrer que $\pi(x)$ devient arbitrairement grand lorsque x croît indéfiniment. Mais à quelle vitesse ?

Le génial Euler a donné une version quantitative du résultat d'Euclide sur l'infinitude de l'ensemble des nombres premiers : la somme des inverses des nombres premiers dans l'intervalle $[2^u, 2^v]$ est proche de celle des inverses des nombres entiers dans l'intervalle $[u, v]$. Exemple :

Leonhard Euler
(1707–1783)

$$\sum_{100<n\leqslant 200} \frac{1}{n} \approx 0{,}70065, \quad 0{,}70064 < \sum_{2^{100}<p\leqslant 2^{200}} \frac{1}{p} < 0{,}70066.$$

Cette formule constitue un soutien heuristique majeur pour le *théorème des nombres premiers*, conjecturé par Legendre et Gauss, et qui permet d'approcher la taille de $\pi(x)$ ou, ce qui revient au même, celle du n-ième nombre premier en fonction de n seulement. Nous y reviendrons.

Adrien-Marie Legendre
(1752–1833)

Une autre approche quantitative de l'ensemble des nombres premiers consiste à produire des tables. La technique la plus ancienne est due à Ératosthène, qui propose de commencer par cocher, dans l'ensemble des entiers de 1 à N, tous les multiples de 2 : le premier nombre non coché, soit 3, est premier ; on coche ensuite tous les multiples de 3 non déjà cochés ; le premier nombre non coché est 5, qui est premier, etc.

Ératosthène de Cyrène
(∼276 – ∼194)

Cette méthode, pratique pour produire des tables de taille modeste, trouve rapidement ses limites lorsque l'on cherche à la formaliser en vue d'une approche théorique. Il fallut attendre Tchébychev et l'année 1852 pour voir apparaître des techniques de comptage des nombres premiers fournissant des évaluations asymptotiques numériques effectives. Tchébychev utilise astucieusement des outils bien connus des analystes, comme la formule de Stirling, datant de 1730, et démontre le postulat de Bertrand (1845), stipulant essentiellement qu'entre un entier et son double se place toujours au moins un nombre premier.

C'est ensuite Franz Mertens qui, au trois-quarts du XIX^{e} siècle, reprend le flambeau de la course à la conjecture de Gauss-Legendre. L'une de ses formules, très surprenante, est passée à la postérité : à l'instar du résultat d'Euler mentionné plus haut, il montre qu'une constante bien connue des mathématiciens, portant également le nom d'Euler et apparaissant naturellement dans l'étude de la somme des inverses des nombres entiers, scelle une formule basique concernant les nombres premiers. Le lien structurel entre les nombres premiers et les nombres entiers possède décidément de multiples facettes.

2.4. Le crible

La méthode d'Ératosthène est trop belle pour être à ce point inefficace. C'est sans doute ce que se disait *in petto* le mathématicien norvégien Viggo Brun dans les années 1920 lorsqu'il a mis au point sa méthode, consistant à remplacer la formule exacte du crible par une formule approchée plus maniable. Ce faisant, il a découvert beaucoup plus : trouver les nombres premiers par une méthode de crible, autrement dit chercher la non-divisibilité en étudiant la divisibilité, met en œuvre des principes très généraux qui s'appliquent largement au-delà de la théorie des nombres premiers. Son plus grand succès a concerné les nombres premiers dits jumeaux, comme 41 et 43 ou 107 et 109, parce que leur écart est le plus petit possible. Alors que la série des inverses des

nombres premiers diverge, Brun a montré que la série des inverses des nombres premiers jumeaux est convergente, autrement dit que la somme

$$\frac{1}{3}+\frac{1}{5}+\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\frac{1}{13}+\frac{1}{17}+\frac{1}{19}+\cdots+\frac{1}{107}+\frac{1}{109}+\cdots$$

reste en dessous d'une valeur fixe même lorsque l'on étend la sommation à tous les couples de nombres premiers jumeaux.

La théorie du crible est aujourd'hui une branche à part entière de l'arithmétique. Elle a contribué efficacement à ses succès récents, comme le théorème de Goldston, Pintz et Yıldırım sur les petits écarts entre nombres premiers consécutifs — cf. § 3.6.

3. Décomposition canonique

Un nombre premier est un nombre entier $p > 1$ qui ne possède aucun diviseur d satisfaisant à $1 < d < p$. On notera que 1 n'est pas premier et que 2 est le seul nombre premier pair.

Ainsi que nous l'avons signalé dans le paragraphe précédent, la situation centrale des nombres premiers en arithmétique est justifiée par le fait que *chaque entier $n > 1$ se décompose de manière unique, à l'ordre des facteurs près, en un produit de nombres premiers*

$$n = \prod_{j=1}^{k} p_j^{\nu_j}.$$

L'existence de la décomposition est presque évidente : il suffit de raisonner par récurrence. Si la propriété est vraie pour tout entier $< n$ et si n n'est pas premier, alors n possède des diviseurs $d \in]1, n[$. Pour chacun d'entre eux, on peut factoriser n sous la forme $n = dm$ avec $m = n/d \in]1, n[$. L'hypothèse de récurrence appliquée à d et m fournit alors le résultat souhaité pour n.

La question de l'unicité est beaucoup moins facile. Elle nécessite le *premier théorème d'Euclide : si un nombre premier p divise un produit ab, alors il divise a ou b.* Admettons cela pour un instant.

Alors, si n possède deux décompositions en produits de facteurs premiers, disons $n = \prod_{j=1}^{k} p_j^{\nu_j}$ et $n = \prod_{i=1}^{\ell} q_i^{\sigma_i}$, chaque p_j est un q_i et réciproquement. En particulier $k = \ell$ et, quitte à réordonner les facteurs, nous en déduisons que $p_j = q_j$ pour $1 \leqslant j \leqslant k$. Supposons par exemple $\nu_1 > \sigma_1$. Alors $m = n/p^{\sigma_1}$ posséderait deux décompositions correspondant à des ensembles distincts de nombres premiers, ce qui contredirait la première partie de la démonstration.

La preuve moderne du premier théorème d'Euclide passe par la structure des idéaux de l'ensemble $\mathbb{Z}$ des entiers relatifs. On dit qu'un sous-ensemble non vide I de $\mathbb{Z}$ en est un idéal s'il est stable par soustraction et par multiplication par un entier quelconque, en d'autres termes :

$$(x \in I, y \in I) \Rightarrow x - y \in I, \quad (x \in \mathbb{Z}, y \in I) \Rightarrow xy \in I.$$

Il est facile de constater que les ensembles de multiples

$$\alpha\mathbb{Z} = \{\alpha m : m \in \mathbb{Z}\}$$

sont des idéaux de $\mathbb{Z}$. Plus surprenante est la réciproque : *tout idéal I de $\mathbb{Z}$ est un ensemble de multiples $I = \alpha\mathbb{Z}$*. En effet, un idéal I non réduit à $\{0\}$ contient des éléments positifs : l'une ou l'autre des propriétés de la définition impliquent que I est stable par l'application $x \mapsto -x$. Soit α le plus petit élément positif de I. Alors $\alpha\mathbb{Z} \subset I$. Réciproquement, pour tout $\beta \in I$, on peut effectuer la division euclidienne $\beta = \alpha q + r$ avec $0 \leqslant r < \alpha$. Or $\alpha q \in I$, donc $r = \beta - \alpha q \in I$. D'après la propriété de minimalité de α, il faut que $r = 0$ et $\beta = \alpha q \in \alpha\mathbb{Z}$.

Nous sommes maintenant en mesure d'établir le premier théorème d'Euclide. Soient $a,\ b,\ p$ tels que $p \mid ab$ et $p \nmid a$. L'ensemble $a\mathbb{Z} + p\mathbb{Z}$ de toutes les combinaisons linéaires $ax + py$ avec $x, y \in \mathbb{Z}$ est un idéal $\alpha\mathbb{Z}$ de $\mathbb{Z}$. Cet idéal contient a et p, donc $\alpha \mid a$ et $\alpha \mid p$. Cette dernière condition implique $\alpha = 1$ ou $\alpha = p$ et la seconde éventualité est exclue puisque $p \nmid a$. On a donc $\alpha = 1$,

d'où $a\mathbb{Z} + p\mathbb{Z} = \mathbb{Z}$. En particulier, il existe $x, y \in \mathbb{Z}$ tels que

$$ax + py = 1.^{(1)}$$

En multipliant les deux membres de cette égalité par b, il suit $b = abx + pby = p\{(ab/p)x + by\}$, d'où $p \mid b$, ce qu'il fallait démontrer.

4. Congruences

La notion de congruence est due à Gauss. Soit m un nombre entier $\geqslant 1$. On dit que deux nombres entiers x et y sont congrus modulo m si la différence $x - y$ est divisible par m. On note dans ce cas

$$x \equiv y \,(\mathrm{mod}\, m).$$

L'ensemble des entiers congrus à un entier donné a s'appelle la classe de a, notée $\overline{a}$. On dit qu'un élément de $\overline{a}$ est un *représentant* de $\overline{a}$. Chaque classe possède un unique représentant dans l'ensemble $\{0, 1, \ldots, m-1\}$, et le représentant d'un entier quelconque n est alors égal à son reste dans la division par m. L'ensemble des classes d'équivalence est noté $\mathbb{Z}/m\mathbb{Z}$,[2] et l'on peut vérifier que, pour tous entiers a, b, les classes $\overline{a+b}$ et $\overline{ab}$ ne dépendent que des classes respectives de a et b. On peut donc munir $\mathbb{Z}/m\mathbb{Z}$ d'une addition et d'une multiplication définies par

$$\overline{a} + \overline{b} = \overline{a+b}, \quad \overline{a}\,\overline{b} = \overline{ab}.$$

Ces opérations[3] confèrent à $\mathbb{Z}/m\mathbb{Z}$ une structure d'anneau commutatif unitaire : autrement dit, les règles de calcul sont

1. Cette assertion porte le nom de *théorème de Bachet* (1624). Elle est souvent attribuée à tort à Bézout.

2. $\mathbb{Z}/m\mathbb{Z}$ est donc en bijection naturelle avec tout intervalle d'entiers de longueur m, par exemple $\{0, 1, \ldots, m-1\}$.

3. Il est d'usage d'omettre les barres de surlignement pour désigner les classes de $\mathbb{Z}/m\mathbb{Z}$. Il est également fréquent que, même lorsqu'ils sont en fait relatifs aux classes, les calculs soient présentés en termes des représentants dans $\mathbb{Z}$. Le signe d'égalité est alors remplacé par celui de congruence et on ajoute aux relations formulées la mention $(\mathrm{mod}\, m)$.

les mêmes que dans $\mathbb{Z}$. Cependant, ces règles ne s'étendent pas à la division : $\mathbb{Z}/m\mathbb{Z}$ n'est en général pas *intègre*, c'est-à-dire que le produit de deux éléments non nuls peut être nul. Par exemple, $\overline{2} \times \overline{3} = \overline{0}$ dans $\mathbb{Z}/6\mathbb{Z}$. On désigne par $(\mathbb{Z}/m\mathbb{Z})^*$ l'ensemble des éléments inversibles de $\mathbb{Z}/m\mathbb{Z}$, c'est-à-dire l'ensemble des classes $\overline{a}$ pour lesquelles l'équation $\overline{a}\,\overline{x} = \overline{1}$ possède une solution. La *fonction indicatrice d'Euler*, traditionnellement notée $\varphi(m)$, est définie par

$$\varphi(m) = |(\mathbb{Z}/m\mathbb{Z})^*|.^{(1)}$$

Nous reviendrons plus loin sur les propriétés de cette fonction, mais nous pouvons observer immédiatement que *pour chaque entier* $m \geqslant 1$, $\varphi(m)$ *est égal au nombre des entiers* a, $1 \leqslant a \leqslant m$, *qui sont premiers à* m.

Soit, en effet, $N(m)$ ce nombre. On a $\varphi(m) \leqslant N(m)$ puisque $ax \equiv 1 \pmod{m}$ implique $(a, m) = 1$. Réciproquement, si $(a, m) = 1$, alors, en appliquant le théorème de Bachet à tous les facteurs premiers (comptés avec multiplicité) de m et en faisant le produit des relations $ax + py = 1$ correspondantes, on obtient $aX + mY = 1$ pour des entiers convenables X, Y. Cela implique bien $\overline{a} \in (\mathbb{Z}/m\mathbb{Z})^*$.

Une conséquence importante de cette application du théorème de Bachet est que, *pour tout nombre premier* p, *l'ensemble* $\mathbb{Z}/p\mathbb{Z}$ *est un corps,* autrement dit : pour tout entier a non multiple de p, la classe de a est inversible dans $\mathbb{Z}/p\mathbb{Z}$. Cela résulte immédiatement de la démonstration précédente puisque $a \not\equiv 0 \pmod{p}$ équivaut à $(a, p) = 1$. En particulier, $\mathbb{Z}/p\mathbb{Z}$ est intègre[(2)] et $\varphi(p) = p - 1$ pour tout nombre premier p. Nous verrons au § 8 que l'on a plus généralement

$$\varphi(n) = n \prod_{p|n} (1 - 1/p)$$

pour tout entier $n \geqslant 1$.

1. Rappelons que nous désignons par $|A|$ le nombre d'éléments d'un ensemble fini A.
2. Voir p. 14.

De ce qui précède, on déduit encore le résultat fondamental qu'*une équation polynomiale à coefficients entiers de degré d*

$$P(x) \equiv 0 \,(\mathrm{mod}\, p)$$

possède au plus d racines distinctes dans $\mathbb{Z}/p\mathbb{Z}$.

Supposons, en effet, que l'équation possède k racines distinctes $x_1, \ldots, x_k$. On voit en écrivant $P(x) = P(x) - P(x_1)$ et en utilisant de manière systématique l'identité

$$x^j - x_1^j = (x - x_1) \sum_{0 \leqslant i < j} x_1^i x^{j-1-i} \qquad (1 \leqslant j \leqslant d),$$

que l'on peut écrire $P(x) \equiv (x - x_1)^{a_1} Q_1(x) \,(\mathrm{mod}\, p)$, où $a_1 \geqslant 1$, et $Q_1(x)$ est un polynôme de degré $d - a_1$, dont x_1 n'est pas racine. Ainsi x_2 doit être racine de $Q_1(x)$. En itérant le procédé, on obtient

$$P(x) \equiv \prod_{1 \leqslant j \leqslant k} (x - x_j)^{a_j} Q_k(x) \,(\mathrm{mod}\, p),$$

où Q_k est un polynôme sans racine dans $\mathbb{Z}/p\mathbb{Z}$ et dont le coefficient du terme de plus haut degré est le même que celui de $P(x)$. Cela implique bien que $k \leqslant \sum_{1 \leqslant j \leqslant k} a_j \leqslant d$.

Soient p un nombre premier, et a un entier non multiple de p. Considérons l'application $x \mapsto ax$ dans $(\mathbb{Z}/p\mathbb{Z})^*$. C'est une injection, donc, puisque l'ensemble est fini, une bijection. Il suit

$$\prod_{x \in (\mathbb{Z}/p\mathbb{Z})^*} x = \prod_{x \in (\mathbb{Z}/p\mathbb{Z})^*} ax = a^{p-1} \prod_{x \in (\mathbb{Z}/p\mathbb{Z})^*} x,$$

d'où

$$a^{p-1} \equiv 1 \,(\mathrm{mod}\, p) \qquad (a \in (\mathbb{Z}/p\mathbb{Z})^*).$$

Cette importante identité est connue sous le nom de *petit théorème de Fermat.*[1] On peut d'ailleurs la généraliser facilement au cas de l'anneau $(\mathbb{Z}/m\mathbb{Z})^*$ pour m quelconque : si $a \in (\mathbb{Z}/m\mathbb{Z})^*$, l'application $x \mapsto ax$ permute les éléments de $(\mathbb{Z}/m\mathbb{Z})^*$. On obtient avec les mêmes calculs (en utilisant cette fois que le produit d'éléments inversibles de $(\mathbb{Z}/m\mathbb{Z})^*$ est encore inversible)

$$a^{\varphi(m)} \equiv 1 \,(\mathrm{mod}\, m) \qquad (a \in (\mathbb{Z}/m\mathbb{Z})^*).$$

C'est la célèbre *formule d'Euler*.

Soit n un nombre entier impair. Nous verrons au paragraphe suivant que le calcul numérique de $2^{n-1} (\mathrm{mod}\, n)$ est rapide même pour de grandes valeurs de n. Si cette quantité n'est pas égale à 1, alors n n'est pas premier. Si elle vaut 1, cela n'implique pas que n soit premier, mais les exceptions, appelées *nombres pseudo-premiers en base 2,* sont rares. Cette remarque est à la base des tests modernes de primalité.

Le petit théorème de Fermat n'a pas à proprement parler de réciproque.[2] On dit qu'un nombre entier $n > 1$ est un *nombre de Carmichael* si n est composé (i.e. non premier) et vérifie $a^{n-1} \equiv 1 \,(\mathrm{mod}\, n)$ pour tout a premier avec n. Le plus petit nombre de Carmichael est 561. Alford, Granville et Pomerance ont démontré en 1992 qu'il existe une infinité de nombres de Carmichael.

Une autre conséquence du fait que $\mathbb{Z}/p\mathbb{Z}$ est un corps est le *théorème de Wilson* (1770, dans un article de Waring) :

$$p \text{ premier } \Leftrightarrow\ (p-1)! + 1 \equiv 0 \,(\mathrm{mod}\, p).$$

1. Par opposition au fameux *grand théorème de Fermat*, selon lequel l'équation $x^n + y^n = z^n$ n'a pas de solution en entiers non nuls x, y, z dès que $n \geqslant 3$. Cette sibylline assertion datant approximativement de 1638 a été récemment démontrée par Wiles et Taylor (1994)... mais c'est une tout autre histoire.

2. Cependant Lucas (1891) a montré que si $(a, n) = 1$, $a^{n-1} \equiv 1 \,(\mathrm{mod}\, n)$ et $a^{(n-1)/q} \not\equiv 1 \,(\mathrm{mod}\, n)$ pour tout facteur premier q de $n - 1$, alors n est premier. Ce résultat est utilisé pour produire de grands nombres premiers.

La condition est suffisante : si $q|p$ et $1 \leqslant q < p$, de sorte que $q|(p-1)!$, elle implique $q|1$, d'où $q = 1$. Elle est aussi nécessaire : chaque terme x du produit

$$(p-1)! \equiv \prod_{x \in (\mathbb{Z}/p\mathbb{Z})^*} x \,(\mathrm{mod}\, p)$$

possède un inverse x' modulo p et l'on a $x = x'$ si, et seulement si, $x \equiv 1 \,(\mathrm{mod}\, p)$ ou $x \equiv p-1 \,(\mathrm{mod}\, p)$. En regroupant les termes de façon à associer chaque élément x à son inverse x' lorsque $x \neq x'$, on obtient bien que le produit est congru à $1 \times (p-1) \equiv -1 \,(\mathrm{mod}\, p)$.

5. Intermezzo cryptographique : systèmes à clefs publiques

L'intérêt des pouvoirs civils et militaires pour les nombres premiers s'est considérablement accru, depuis une trentaine d'années, en raison de la mise au point de nouvelles méthodes de cryptographie où les mathématiques (et en particulier celles des nombres premiers) jouent un rôle déterminant.

Avec le formidable développement de l'informatique et la nécessité d'assurer la confidentialité de transmissions de toutes natures, les enjeux de ces applications sont économiquement considérables. L'implacable loi de la rentabilité imposant de promouvoir la recherche dans ce domaine, les résultats, y compris théoriques, ont rapidement suivi la mise en place des moyens, et les tests de primalité, les algorithmes de factorisation, les méthodes d'engendrement de nombres premiers et autres élaborations conceptuelles *ad hoc*, ont fleuri sur les campus à l'ombre des machines rutilantes.

Rien de moral dans tout cela, bien sûr, mais peut-être une (double) moralité à déduire : les crédits font, effectivement, progresser la science et, si les avancées ont été aussi performantes lorsqu'une rentabilité immédiate était en vue, que ne peut-on espérer de dotations, même plus faibles, pour une recherche fondamentale à moyen ou long terme ?

Nous nous contenterons ici de décrire succinctement le « système RSA »[1] qui fournit une technique simple et pratiquement inviolable de codage.

Considérons un réseau d'individus $\{I_1, I_2, \ldots, I_n\}$ qui veulent communiquer entre eux de manière *secrète* (messages codés illisibles par d'autres que le ou les destinataires) et/ou *authentifiée* (messages éventuellement lisibles par tous mais infalsifiables). Le système RSA consiste à attribuer à chaque membre I_j du réseau un couple de nombres premiers (p_j, q_j) — c'est-à-dire un entier $n_j = p_j q_j$ produit de deux nombres premiers — et un entier c_j premier à $\varphi(n_j) = (p_j - 1)(q_j - 1)$. Le succès de la confidentialité du système repose sur le fait qu'en l'état actuel des moyens de calcul il est assez facile de produire de « grands » nombres premiers (donc de pourvoir les membres du réseau en p_j et q_j) mais qu'il est quasiment impossible de factoriser un entier de la taille du carré de ces nombres premiers (donc de restituer p_j et q_j à partir de n_j). À l'heure où ce livre est écrit, on peut considérer que des nombres premiers de l'ordre de 10^{100} sont effectivement « grands » en ce sens particulier.

Le réseau se dote d'un annuaire (public) où chaque nom I_j est associé à n_j et c_j. Cependant, seul I_j, qui connaît la factorisation de n_j est à même de trouver l'inverse d_j de c_j modulo $\varphi(n_j)$.[2]

Un *message* est un entier M n'excédant pas la taille (essentiellement commune) des n_j, ou, ce qui revient au même, une succession de tels entiers : il suffit par exemple de remplacer chaque lettre par

1. Par référence à Rivest, Shamir et Adleman, dont l'article de 1978 marque un tournant capital dans le domaine.

2. Le calcul de l'inverse est numériquement très rapide, et ne nécessite qu'un nombre d'opérations de l'ordre du nombre de chiffres. Considérons, en toute généralité, deux entiers a et b tels que $1 \leqslant a < b$ et $(a, b) = 1$. Si q_1 est l'entier le plus proche de b/a alors $r_1 := b - aq_1$ est tel que $|r_1| \leqslant a/2$. De même, il existe un entier q_2 tel que $r_2 := a - r_1 q_2 = a(1 + q_1 q_2) - bq_2$ vérifie $|r_2| \leqslant a/4$. On itère le procédé en écrivant $r_{j+1} = r_{j-1} - r_j q_{j+1}$, $|r_{j+1}| \leqslant a/2^{j+1}$. Après au plus $(\ln 2a)/\ln 2$ étapes on parvient à $r_{k+1} = 0$, donc $r_k | r_{k-1}$. On vérifie facilement que cela implique que $r_k | (a, b)$, donc $r_k = \pm 1$. Or r_k est, par construction, une combinaison linéaire de a et b. On a donc trouvé des entiers u et v tels que $au + bv = \pm 1$. L'inverse de a modulo b vaut u ou $b - u$.

son rang dans l'alphabet et de concaténer les entiers à deux chiffres ainsi obtenus.

Lorsque, disons, I_1 veut envoyer un message codé à I_2, il lui adresse $M_2 = M^{c_2} \pmod{n_2}$, c'est-à-dire le plus petit entier positif $\equiv M^{c_2} \pmod{n_2}$. Lorsque I_2 reçoit M_2 il lui suffit de calculer $M_2^{d_2} \equiv M \pmod{n_2}$, d'après la formule d'Euler.[1] Il peut donc ainsi déchiffrer le message de I_1. Personne d'autre n'est en mesure de le faire car d_2 est (en pratique) incalculable à partir des données de l'annuaire.

Lorsque I_1 veut envoyer un message authentifié M à I_2, il lui adresse $M_1 = M^{d_1} \pmod{n_1}$. Ainsi I_2 (et en fait n'importe quel autre membre du réseau) peut calculer $M_1^{c_1} \equiv M \pmod{n_1}$ (toujours grâce à la formule d'Euler), mais le message est bien authentifié car le fait que cette procédure de décodage fournisse un résultat cohérent atteste d'un codage que seul I_1 était en mesure d'effectuer.

Lorsque I_1 veut envoyer un message à la fois crypté et authentifié à I_2, et si, par exemple, $n_1 < n_2$, il lui adresse $M_{12} = M_1^{c_2} \pmod{n_2}$. Ainsi I_2 peut déchiffrer le message en calculant successivement $M_{12}^{d_2} \equiv M_1 \pmod{n_2}$ et $M_1^{c_1} \equiv M \pmod{n_1}$ — l'authentification étant assurée par le simple fait que cette seconde procédure fournisse un résultat lisible.

Il reste à observer que le calcul de grandes puissances M^c modulo un entier n à N chiffres (avec donc, actuellement, $N \approx 200$) ne nécessite pas d'outils informatiques très sophistiqués. En effet, le calcul de M^2 modulo n correspond à la multiplication de deux nombres à N chiffres, et fournit évidemment un résultat à N chiffres. Cela implique que M^c correspond à k multiplications de deux nombres à N chiffres si $c = 2^k$, et donc, en utilisant l'écriture en base deux $c = \sum_{0 \leqslant j \leqslant k} \varepsilon_j 2^j$ (avec $\varepsilon_j = 0$ ou 1), à au plus $k \ll N$ multiplications dans le cas général. Avec les ordinateurs actuels, et pour des valeurs de N n'excédant pas quelques centaines, le codage est quasiment instantané.

1. On a $c_2 d_2 \equiv 1 \pmod{\varphi(n_2)}$, donc $M_2^{d_2} \equiv M \pmod{n_2}$ si $(M, n_2) = 1$. On peut montrer que cette relation persiste même si $(M, n_2) > 1$.

6. Résidus quadratiques

Soit p un nombre premier impair. Considérons l'application

$$q : (\mathbb{Z}/p\mathbb{Z})^* \to (\mathbb{Z}/p\mathbb{Z})^*$$

définie par $q(x) = x^2$. Alors q est un homomorphisme multiplicatif, c'est-à-dire que l'on a identiquement

$$q(xy) = q(x)q(y).$$

Comme l'équation $q(x) = 1$ possède exactement deux solutions, à savoir 1 et $p-1$ (on pourrait aussi dire ± 1 car $p-1 = -1$ dans $(\mathbb{Z}/p\mathbb{Z})^*$), on voit que si l'équation $q(x) = a$ possède une solution, disons x_1, elle en possède exactement deux, x_1 et $p - x_1$. Un tel nombre a est appelé *résidu quadratique modulo p*. Il découle de ce qui précède qu'il y a exactement $\frac{1}{2}(p-1)$ résidus quadratiques modulo p. En effet, si Q_p dénote l'ensemble des résidus quadratiques, alors les $|Q_p|$ sous-ensembles à deux éléments $q^{-1}\{a\} = \{x \in (\mathbb{Z}/p\mathbb{Z})^* : q(x) = a\}$ $(a \in Q_p)$ constituent une partition de $(\mathbb{Z}/p\mathbb{Z})^*$, d'où $p - 1 = |(\mathbb{Z}/p\mathbb{Z})^*| = 2|Q_p|$.

Considérons alors l'homomorphisme multiplicatif

$$f : (\mathbb{Z}/p\mathbb{Z})^* \to \{-1, 1\}$$

défini par $f(x) = x^{(p-1)/2}$. Le petit théorème de Fermat implique que $f(a) = 1$ pour tout a de Q_p et comme l'équation polynomiale $f(x) = 1$ possède au plus $\frac{1}{2}(p-1)$ solutions dans $(\mathbb{Z}/p\mathbb{Z})^*$, on voit que Q_p est exactement l'ensemble de ces racines. Autrement dit, $f(a)$ vaut 1 si a est résidu quadratique modulo p, et vaut -1 sinon. On exprime habituellement cette propriété en termes du *symbole de Legendre* :

$$\left(\frac{a}{p}\right) := \begin{cases} 1 & \text{si } a \in Q_p, \\ -1 & \text{si } a \notin Q_p. \end{cases}$$

Pour tout $a \in (\mathbb{Z}/p\mathbb{Z})^*$, on a

$$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}.$$

Lorsqu'on applique ce résultat au cas $a = -1$, on obtient que -1 est résidu quadratique modulo p si, et seulement si, $p \equiv 1 \pmod 4$. Sous cette condition, il existe donc deux entiers m et x tels que $pm = x^2 + 1$. Girard a énoncé en 1632 et Fermat a démontré en 1654 qu'un tel nombre premier p est somme de deux carrés. La condition est donc nécessaire et suffisante puisqu'une somme de deux carrés est congrue à 0, 1 ou 2 modulo 4.

Montrons le théorème de Girard–Fermat. Étant donné un nombre premier p tel que $p \equiv 1 \pmod 4$, posons $N = \lfloor \sqrt{p} \rfloor$ et donnons-nous une solution x de l'équation $x^2 \equiv -1 \pmod p$. Parmi les $(N+1)^2 > p$ nombres $u + vx$ avec $0 \leqslant u, v \leqslant N$, deux au moins ont la même classe résiduelle modulo p. Par différence, nous en déduisons que la congruence $a \equiv bx \pmod p$ possède une solution non triviale en entiers a, b tels que $\max(|a|, |b|) < \sqrt{p}$. Cela implique $a^2 \equiv x^2 b^2 \equiv -b^2 \pmod p$, d'où $p \mid a^2 + b^2$. Comme $0 < a^2 + b^2 < 2p$, on a bien $p = a^2 + b^2$.

Euler a prouvé que

$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8} \quad (p > 2)$$

et Gauss a rigoureusement établi une propriété remarquable énoncée par Legendre : si p et q sont des nombres premiers impairs, ils ont même caractère quadratique l'un modulo l'autre sauf s'ils sont tous deux $\equiv 3 \pmod 4$, auquel cas c'est l'inverse qui a lieu. C'est la fameuse *loi de réciprocité quadratique*. En formule :

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4} \qquad (p > 2, q > 2).$$

7. Retour sur l'infinitude de l'ensemble des nombres premiers

La preuve du théorème d'Euclide donnée au § 1 peut être rendue effective, c'est-à-dire qu'elle peut fournir une majoration explicite du n-ième nombre premier. Il suffit pour cela de remplacer, dans l'argument, $n!$ par $\prod_{1 \leqslant j \leqslant n} p_j$. On obtient ainsi

$$p_n \leqslant 1 + \prod_{1 \leqslant j < n} p_j,$$

d'où par une récurrence élémentaire

$$p_n \leqslant \exp(2^{n-1}) \qquad (n \geqslant 1).$$

Un autre raisonnement très simple fournit une estimation bien meilleure. Considérons les n plus petits nombres premiers

$$p_1 < p_2 < \cdots < p_n.$$

Chaque entier $m \leqslant p_n$ s'écrit de manière unique sous la forme $m = \prod_{j=1}^{n} p_j^{\nu_j}$. Si t est l'unique nombre entier tel que $2^t \leqslant p_n < 2^{t+1}$, il y a, pour chaque indice j, au plus $t+1$ choix possibles de ν_j. Le nombre total de valeurs possibles pour m est donc $\leqslant (t+1)^n$. Or, il y a (trivialement !) exactement $p_n \geqslant 2^t$ entiers m n'excédant pas p_n. Donc $2^t \leqslant (t+1)^n$. Cette inégalité implique $t \leqslant n^2 - 1$ dès que $n \geqslant 5$. On en déduit

$$p_n \leqslant 2^{n^2} \qquad (n \geqslant 1).$$

La preuve d'Euler est plus précise encore, et fournit de fait un résultat en un certain sens optimal. Elle s'appuie également sur l'unicité de la décomposition d'un nombre entier en produit de nombres premiers. En remarquant que chaque entier n est aussi décomposable de manière unique sous la forme $n = qm^2$, où q est

sans facteur carré excédant 1,[1] on peut écrire (en désignant par la lettre q un entier générique sans facteur carré)

$$\sum_{n\leqslant x}\frac{1}{n}=\sum_{q\leqslant x}\frac{1}{q}\sum_{m\leqslant\sqrt{x/q}}\frac{1}{m^2}\leqslant\sum_{q\leqslant x}\frac{1}{q}\sum_{m\geqslant 1}\frac{1}{m^2}.$$

Maintenant, on observe que

$$\sum_{m\geqslant 1}\frac{1}{m^2}\leqslant 1+\sum_{m=2}^{\infty}\Big(\frac{1}{m-1}-\frac{1}{m}\Big)=2,$$

puisque la dernière somme est « télescopique », et que

$$\sum_{q\leqslant x}\frac{1}{q}\leqslant\prod_{p\leqslant x}\Big(1+\frac{1}{p}\Big)\leqslant\exp\Big\{\sum_{p\leqslant x}1/p\Big\},$$

où la première inégalité s'obtient en développant le produit en p, et la seconde en utilisant la majoration classique $1+u\leqslant \mathrm{e}^u$ avec $u=1/p$. En insérant encore la minoration

$$\sum_{n\leqslant x}\frac{1}{n}\geqslant\sum_{n\leqslant x}\int_n^{n+1}\frac{\mathrm{d}t}{t}\geqslant\ln x,$$

et en prenant les logarithmes, on déduit de ce qui précède

$$\sum_{p\leqslant x}\frac{1}{p}\geqslant\ln_2 x-\ln 2\qquad(x\geqslant 2).^{[2]}$$

Cela établit le théorème d'Euclide sous une forme forte : *la série des inverses des nombres premiers diverge.* Comme nous le verrons par la suite, la minoration trouvée fournit le bon ordre de grandeur (et même un équivalent asymptotique) pour la somme $\sum_{p\leqslant x}1/p$.

1. On a en fait $q=\prod_{p^{2\nu+1}\|n}p$.
2. Rappelons que $\ln_2$ signifie $\ln\ln$.

En effet, on peut montrer que le théorème des nombres premiers équivaut à la relation asymptotique

$$p_n \sim n \ln n \qquad (n \to \infty).$$

Un calcul standard permet alors d'en déduire que l'on a

$$\sum_{p \leqslant x} \frac{1}{p} \sim \ln_2 x$$

lorsque $x \to \infty$.

8. Le crible d'Ératosthène

Connu depuis l'Antiquité, le crible d'Ératosthène est une méthode pour compter le nombre $N_m(x)$ des entiers n n'excédant pas x et qui sont relativement premiers à un nombre fixé m.

La version naïve consiste à écrire la liste des entiers $n \leqslant x$ et à barrer successivement les multiples des différents facteurs premiers de m. Les nombres restants sont ceux qui sont comptés dans $N_m(x)$.

Pour obtenir une formule générale, on peut raisonner comme suit. Soit $p_1, \ldots, p_k$ la suite (finie) des facteurs premiers distincts de m. On calcule $N_m(x)$ par récurrence sur k en observant que le cas $k = 0$ est trivial : on a $N_1(x) = \lfloor x \rfloor$ pour tout $x > 0$. Posons $m_0 = 1$ et $m_j := p_1 \cdots p_j$ $(1 \leqslant j \leqslant k)$. Pour $0 \leqslant j < k$, les seuls entiers n comptés dans $N_{m_j}(x)$ et non comptés dans $N_{m_{j+1}}(x)$ sont de la forme $p_{j+1} n \leqslant x$ avec $(n, m_j) = 1$. En d'autres termes, on a

$$N_{m_j}(x) - N_{m_{j+1}}(x) = N_{m_j}(x/p_{j+1}).$$

Ainsi

$$\begin{aligned} N_{m_1}(x) &= N_1(x) - N_1(x/p_1) = \lfloor x \rfloor - \lfloor x/p_1 \rfloor, \\ N_{m_2}(x) &= N_{m_1}(x) - N_{m_1}(x/p_2) \\ &= \lfloor x \rfloor - \lfloor x/p_1 \rfloor - \lfloor x/p_2 \rfloor + \lfloor x/p_1 p_2 \rfloor, \end{aligned}$$

et ainsi de suite. On obtient finalement

$$N_m(x) = \sum_{d \mid m^*} (-1)^{\omega(d)} \lfloor x/d \rfloor$$

où m^* désigne le *noyau* sans facteur carré de m (soit $m^* = \prod_{j=1}^{k} p_j$) et où $\omega(d)$ désigne le nombre des facteurs premiers distincts de d (noter qu'en particulier $\omega(1) = 0$). On donne un aspect plus canonique encore à cette formule en introduisant la *fonction de Möbius*

$$\mu(d) = \begin{cases} (-1)^{\omega(d)} & \text{si } d \text{ est sans facteur carré } > 1, \\ 0 & \text{dans le cas contraire.} \end{cases}$$

On obtient alors la *formule de Legendre* (1808) du crible d'Ératosthène :

$$N_m(x) = \sum_{d|m} \mu(d) \lfloor x/d \rfloor .$$

Lorsque l'on choisit $m = \prod_{p \leqslant \sqrt{x}} p$, seuls sont comptés dans $N_m(x)$ le nombre 1 et les nombres premiers p de l'intervalle $]\sqrt{x}, x]$. Donc $N_m(x) = \pi(x) - \pi(\sqrt{x}) + 1$ et l'on obtient

$$\pi(x) = -1 + \pi(\sqrt{x}) + \sum_{P^+(d) \leqslant \sqrt{x}} \mu(d) \lfloor x/d \rfloor .^{(1)}$$

On observera que la somme en d est finie puisqu'il y a au plus $2^{\pi(\sqrt{x})}$ entiers d sans facteur carré tels que $P^+(d) \leqslant \sqrt{x}$. Ce résultat montre que la fonction de Möbius et les nombres premiers sont intimement liés. Nous aurons l'occasion de revenir, aux Chapitres 2 et 4, sur cette interdépendance.

Une autre application de la formule de Legendre est le calcul de la fonction d'Euler

$$\varphi(n) = N_n(n),$$

rencontrée au § 4. Nous obtenons dans ce cas

$$\varphi(n) = \sum_{d|n} \mu(d) n/d = n \prod_{p|n} (1 - 1/p),$$

où la seconde égalité est obtenue en développant le dernier membre.

1. On rappelle que $P^+(d)$ désigne le plus grand facteur premier de d avec la convention $P^+(1) = 1$.

9. Les théorèmes de Tchébychev

Les premiers progrès significatifs concernant l'évaluation de $\pi(x)$ sont dus à Tchébychev (1852). Il n'utilise pas le crible d'Ératosthène, mais une forme faible de la formule de Stirling,[1] à savoir

$$\ln n = \sum_{1\leqslant m\leqslant n} \ln m = n\ln n - n + O(\ln n) \qquad (n \geqslant 2).$$

Cela découle directement de l'estimation suivante, valable pour $m \geqslant 1$,

$$\begin{aligned} 0 &\leqslant \int_m^{m+1} (\ln t)\,\mathrm{d}t - \ln m \\ &= \int_m^{m+1} \ln\left(\frac{t}{m}\right)\mathrm{d}t \leqslant \int_m^{m+1} \left(\frac{t}{m} - 1\right)\mathrm{d}t = \frac{1}{2m}. \end{aligned}$$

En sommant pour $m = 1, 2, \ldots, n-1$ et en utilisant le fait qu'une primitive de $\ln t$ est $t\ln t - t$, on obtient bien la formule annoncée.

L'idée de Tchébychev consiste essentiellement à exploiter la décomposition du nombre $n!$ en produit de facteurs premiers. Pour chaque m on peut écrire

$$\ln m = \sum_{p^\nu \mid m} \ln p,$$

la somme étant étendue à tous les couples (p, ν) tels que $p^\nu \mid m$ avec p premier et $\nu \geqslant 1$. En remplaçant dans l'expression de $\ln n!$ et en intervertissant les sommations, il suit

$$\begin{aligned} \ln n! &= \sum_{1\leqslant m\leqslant n} \ln m = \sum_{p^\nu \leqslant n} \ln p \sum_{\substack{1\leqslant m\leqslant n \\ p^\nu \mid m}} 1 \\ &= \sum_{p^\nu \leqslant n} \ln p \,\lfloor n/p^\nu \rfloor . \end{aligned}$$

1. Démontrée en 1730 : on a $n! \sim n^n \mathrm{e}^{-n}\sqrt{2\pi n}$ $(n \to \infty)$.

Cela suggère d'introduire la fonction

$$\Lambda(d) := \begin{cases} \ln p & \text{si } \exists\, \nu \geqslant 1 :\ d = p^{\nu}, \\ 0 & \text{dans le cas contraire.} \end{cases}$$

Cette fonction a été étudiée par le mathématicien allemand von Mangoldt à la fin du dix-neuvième siècle et porte aujourd'hui son nom. Avec cette notation, nous déduisons de ce qui précède la formule asymptotique

$$\sum_{d \leqslant n} \Lambda(d)\, \lfloor n/d \rfloor = n \ln n - n + O(\ln n) \quad (n \geqslant 2).$$

Désignons par $B(n)$ le membre de gauche et posons $B(x) = B(\lfloor x \rfloor)$ pour tout $x > 0$. Nous allons utiliser notre estimation de $B(x)$ pour établir un encadrement de la fonction sommatoire de Λ,

$$\psi(x) = \sum_{d \leqslant x} \Lambda(d).$$

Nous verrons par la suite que cela implique très facilement un encadrement de même qualité pour $\pi(x)$.

Nous exploitons l'idée originale de Tchébychev, mais sous une forme rudimentaire qui fournit un résultat numériquement moins précis quoique de même nature. Posons

$$\chi(u) = \lfloor u \rfloor - 2\, \lfloor u/2 \rfloor \qquad (u > 0).$$

Alors χ est une fonction 2-périodique qui vérifie

$$\chi(u) = \begin{cases} 0 & \text{si } 0 \leqslant u < 1, \\ 1 & \text{si } 1 \leqslant u < 2. \end{cases}$$

Nous allons maintenant calculer de deux manières la quantité

$$B_2(x) = B(x) - 2B(x/2).$$

D'une part on a

$$\begin{aligned} B_2(x) &= x\ln x - x - 2\{\tfrac{1}{2}x\ln(\tfrac{1}{2}x) - \tfrac{1}{2}x\} + O(\ln x) \\ &= x\ln 2 + O(\ln x), \end{aligned}$$

et d'autre part

$$B_2(x) = \sum_{d\leqslant x} \Lambda(d)\chi(x/d).$$

De cette dernière expression, on déduit, compte tenu de la propriété de χ indiquée plus haut,

$$\psi(x) - \psi(x/2) \leqslant B_2(x) \leqslant \psi(x).$$

La majoration fournit donc immédiatement une minoration de $\psi(x)$, soit

$$\psi(x) \geqslant x\ln 2 + O(\ln x) \qquad (x \geqslant 2).$$

La minoration de $B_2(x)$ est utilisée de manière inductive : on a

$$\begin{aligned} \psi(x) &\leqslant B_2(x) + \psi(x/2) \leqslant B_2(x) + B_2(x/2) + \psi(x/4) \\ &\leqslant \cdots \leqslant \sum_{0\leqslant j\leqslant k} B_2(x/2^j) + \psi(x/2^{k+1}). \end{aligned}$$

Ici k est un entier arbitraire. Choisissons

$$k = K(x) := \lfloor (\ln x)/\ln 2 \rfloor,$$

de sorte que $\psi(x/2^{k+1}) = 0$. Il suit

$$\begin{aligned} \psi(x) &\leqslant \sum_{0\leqslant j\leqslant K(x)} \Big\{ \frac{x\ln 2}{2^j} + O(\ln x) \Big\} \\ &\leqslant 2x\ln 2 + O((\ln x)^2). \end{aligned}$$

À fins de référence ultérieure, nous rassemblons les estimations obtenues dans un énoncé formel.

Théorème 9.1 (Tchébychev) *On a pour* $x \geqslant 2$

$$x \ln 2 + O(\ln x) \leqslant \psi(x) \leqslant x \ln 4 + O((\ln x)^2).$$

Comme annoncé, nous pouvons déduire aisément de ce résultat une information comparable relative à $\pi(x)$. On a $\psi(x) = \sum_{p^\nu \leqslant x} \ln p$ où la somme porte sur tous les couples (p, ν) avec p premier et $\nu \geqslant 1$. Pour chaque p fixé, il y a exactement $\lfloor \ln x / \ln p \rfloor$ valeurs admissibles de ν, donc

$$\psi(x) = \sum_{p \leqslant x} \left\lfloor \frac{\ln x}{\ln p} \right\rfloor \ln p.$$

Grâce à l'encadrement $\lfloor u \rfloor \leqslant u < \lfloor u \rfloor + 1 \leqslant 2 \lfloor u \rfloor$ $(u \geqslant 1)$, il suit d'abord

$$\psi(x) \leqslant \pi(x) \ln x \leqslant 2\psi(x),$$

puis, en utilisant la majoration du Théorème 9.1,

$$\begin{aligned}
\pi(x) \ln x - \psi(x) &= \sum_{p \leqslant x} \left(\ln x - \left\lfloor \frac{\ln x}{\ln p} \right\rfloor \ln p \right) \\
&\leqslant \sum_{p \leqslant \sqrt{x}} \ln p + \sum_{\sqrt{x} < p \leqslant x} \ln(x/p) \\
&\leqslant \psi(\sqrt{x}) + \sum_{\sqrt{x} < p \leqslant x} \int_p^x \mathrm{d}t/t \\
&= O(\sqrt{x}) + \int_{\sqrt{x}}^x \frac{\pi(t)}{t} \, \mathrm{d}t \ll \frac{x}{\ln x}.
\end{aligned}$$

On peut donc finalement énoncer que

$$\pi(x) = \psi(x)/\ln x + O\big(x/(\ln x)^2\big) \quad (x \geqslant 2),$$

de sorte que le Théorème 9.1 implique le corollaire suivant.

Corollaire 9.2 *On a lorsque* x *tend vers l'infini*

$$\{\ln 2 + o(1)\} \frac{x}{\ln x} \leqslant \pi(x) \leqslant \{\ln 4 + o(1)\} \frac{x}{\ln x}.$$

Les résultats de Tchébychev étaient à vrai dire plus précis. Au lieu de la fonction $\chi(u)$ employée dans la preuve ci-dessus, il utilisait la fonction

$$\chi_1(u) = \lfloor u \rfloor - \lfloor u/2 \rfloor - \lfloor u/3 \rfloor - \lfloor u/5 \rfloor + \lfloor u/30 \rfloor$$

qui est 30-périodique et satisfait à $\chi_1(u) = 1$ pour $1 \leqslant u < 6$. Cela lui a permis de prouver l'encadrement asymptotique

$$\{c_1 + o(1)\}\frac{x}{\ln x} \leqslant \pi(x) \leqslant \{c_2 + o(1)\}\frac{x}{\ln x} \quad (x \to \infty)$$

avec

$$c_1 = \ln\left(2^{1/2}3^{1/3}5^{1/5}30^{-1/30}\right) \approx 0{,}92129,$$
$$c_2 = \tfrac{6}{5}c_1 \approx 1{,}10555.$$

Comme $c_2 < 2c_1$, cela implique en particulier que $\pi(2n-3) > \pi(n)$ pour n assez grand. En explicitant une version effective de ses estimations, Tchébychev a pu prouver que cette inégalité stricte persiste pour tout $n > 3$, confirmant ainsi une conjecture fameuse connue sous le nom de *postulat de Bertrand* (1845) *: pour tout entier $n > 3$, il y a au moins un nombre premier p tel que*

$$n < p < 2n - 2.$$

Il est à noter que le choix plus simple

$$\chi_2(x) := \lfloor x \rfloor - \lfloor x/2 \rfloor - 2\lfloor x/3 \rfloor + \lfloor x/6 \rfloor$$

conduit également à une preuve du postulat de Bertrand, mais avec des constantes c_1 et c_2 légèrement moins bonnes.

Une autre conséquence, indiquée par lui-même, du travail de Tchébychev est l'encadrement

$$\liminf_{x\to\infty} \frac{\pi(x)}{x/\ln x} \leqslant 1 \leqslant \limsup_{x\to\infty} \frac{\pi(x)}{x/\ln x},$$

autrement dit : *si le rapport $\pi(x)\ln x/x$ tend vers une limite, cette limite est* 1.

Cela résulte de la formule asymptotique

$$B(x) = \sum_{d \leqslant x} \Lambda(d) \lfloor x/d \rfloor = x \ln x - x + O(\ln x)$$

qui est à la base de la preuve de Tchébychev. En remplaçant $\lfloor x/d \rfloor$ par $x/d + O(1)$ et en utilisant l'estimation $\psi(x) \ll x$, on obtient

$$\sum_{d \leqslant x} \Lambda(d)/d = \ln x + O(1) \qquad (x \geqslant 1).$$

En observant maintenant que

$$\sum_{d \leqslant x} \frac{\Lambda(d)}{d} = \sum_{d \leqslant x} \Lambda(d) \int_d^\infty \frac{\mathrm{d}t}{t^2} = \frac{\psi(x)}{x} + \int_1^x \frac{\psi(t)}{t^2}\,\mathrm{d}t,$$

on déduit de ce qui précède que

$$\int_1^x \frac{\psi(t)}{t^2}\,\mathrm{d}t = \ln x + O(1) \qquad (x \geqslant 1).$$

Soit alors $\alpha = \limsup_{x\to\infty} \pi(x) \ln x/x$. On a aussi

$$\alpha = \limsup_{x\to\infty} \psi(x)/x$$

puisque, comme nous l'avons vu plus haut,

$$\psi(x) = \pi(x) \ln x + O(x/\ln x).$$

Pour chaque $\varepsilon > 0$, il existe donc un $x_0 = x_0(\varepsilon)$ tel que $\psi(t) \leqslant (\alpha + \varepsilon)t$ pour tout $t \geqslant x_0(\varepsilon)$. D'où

$$\begin{aligned}\int_1^x \frac{\psi(t)}{t^2}\,\mathrm{d}t &\leqslant \int_1^{x_0} \frac{\psi(t)}{t^2}\,\mathrm{d}t + (\alpha+\varepsilon)\int_{x_0}^x \frac{\mathrm{d}t}{t} \\ &\leqslant (\alpha+\varepsilon)\ln x + O_\varepsilon(1).\end{aligned}$$

Cela implique $1 \leqslant \alpha + \varepsilon$ et donc $\alpha \geqslant 1$ puisque ε est arbitraire. On raisonne semblablement pour

$$\beta = \liminf_{x \to \infty} \pi(x) \ln x/x.$$

La preuve du second résultat de Tchébychev est ainsi complétée.

Lorsque Tchébychev publia ses travaux en 1852, il pouvait sembler que la conjecture de Gauss–Legendre

$$\pi(x) \sim x/\ln x \qquad (x \to \infty)$$

était à portée de main. Il fallut pourtant attendre 44 ans pour prouver, grâce aux idées entièrement différentes de Riemann, le théorème des nombres premiers. La preuve élémentaire,[1] héritière de l'approche de Tchébychev, ne devait, quant à elle, pas voir le jour avant près d'un siècle.

10. Les théorèmes de Mertens

Dans la continuation immédiate de Tchébychev, Mertens a établi, avec des outils analytiques de même degré de sophistication, des *formules asymptotiques*[2] pour des sommes portant sur des nombres premiers. Les deux théorèmes mentionnés dans ce paragraphe datent de 1874.

Le résultat connu sous le nom de *premier théorème de Mertens* est l'estimation

$$\sum_{p \leqslant x} \frac{\ln p}{p} = \ln x + O(1) \qquad (x \geqslant 1).$$

Nous avons montré au paragraphe précédent la validité de cette formule lorsque le membre de gauche est remplacé par la somme

1. On dit qu'une preuve est élémentaire si les outils déployés sont construits avec les mêmes éléments que la question posée, ici l'analyse réelle. Voir le § 4.1.

2. C'est-à-dire des évaluations avec terme principal et terme d'erreur, le second étant asymptotiquement négligeable devant le premier.

$\sum_{d\leqslant x}\Lambda(d)/d$. Or, la différence entre cette somme et celle qui nous intéresse maintenant vaut

$$\sum_{p^\nu\leqslant x,\,\nu\geqslant 2}\frac{\ln p}{p^\nu}\leqslant\sum_p\frac{\ln p}{p(p-1)}\ll 1.$$

Le résultat est donc acquis.

Le second théorème de Mertens est connu sous le nom de *formule de Mertens* : on a

$$\prod_{p\leqslant x}(1-1/p)^{-1}=\mathrm{e}^\gamma\ln x+O(1)\quad(x\geqslant 1),$$

où γ désigne la constante d'Euler.

La preuve de ce résultat nécessite plusieurs étapes. On commence par évaluer

$$\begin{aligned}\sum_{p\leqslant x}\frac{1}{p}&=\sum_{p\leqslant x}\frac{\ln p}{p}\int_p^\infty\frac{\mathrm{d}u}{u(\ln u)^2}\\&=\int_2^x\sum_{p\leqslant u}\frac{\ln p}{p}\frac{\mathrm{d}u}{u(\ln u)^2}+\frac{1}{\ln x}\sum_{p\leqslant x}\frac{\ln p}{p}.\end{aligned}$$

Posons $R(u)=\sum_{p\leqslant u}(\ln p)/p-\ln u\,(u\geqslant 2)$. Le premier théorème de Mertens énonce que $R(u)$ est une fonction bornée. Il découle donc du calcul précédent que

$$\begin{aligned}\sum_{p\leqslant x}\frac{1}{p}&=\int_2^x\frac{\mathrm{d}u}{u\ln u}+\int_2^x R(u)\frac{\mathrm{d}u}{u(\ln u)^2}+1+\frac{R(x)}{\ln x}\\&=\ln_2 x+a+O\left(\frac{1}{\ln x}\right)\end{aligned}$$

avec

$$a=\int_2^\infty R(u)\frac{\mathrm{d}u}{u(\ln u)^2}+1-\ln_2 2\approx 0{,}2614972128.$$

Nous allons voir que cette formule est équivalente à celle de Mertens, à la valeur près de la constante a. Partant, elle est parfois également désignée sous le nom de *second théorème de Mertens*.[1]

On a $e^{-1/p}(1-1/p)^{-1} = 1 + O(1/p^2)$. Il découle donc de ce qui précède que

$$\prod_{p\leqslant x}(1-1/p)^{-1} = \prod_{p\leqslant x} e^{1/p}e^{-1/p}(1-1/p)^{-1} = b\ln x + O(1)$$

avec $b = e^a \prod_p e^{-1/p}(1-1/p)^{-1}$.

Il reste à calculer la constante b. C'est la partie la plus difficile de la démonstration, mais aussi la plus stimulante, puisque, sans conteste, c'est l'occurrence de la constante d'Euler qui suscite ici curiosité et intérêt. Le calcul repose sur la fameuse *formule d'Euler* évoquée dans l'introduction

$$\sum_{n\geqslant 1} 1/n^\sigma = \prod_p (1-p^{-\sigma})^{-1} \qquad (\sigma > 1).$$

Il est clair que les deux membres sont convergents. Lorsque l'on développe le produit fini

$$\prod_{p\leqslant x}(1-p^{-\sigma})^{-1} = \prod_{p\leqslant x}\sum_{\nu\geqslant 0} p^{-\nu\sigma}$$

on obtient, grâce à l'unicité de la décomposition en produit de facteurs premiers, la somme $\sum_{P^+(n)\leqslant x} n^{-\sigma}$, étendue à tous les entiers dont le plus grand facteur premier n'excède pas x. Cette somme vérifie évidemment

$$\sum_{n\leqslant x} n^{-\sigma} \leqslant \sum_{P^+(n)\leqslant x} n^{-\sigma} \leqslant \sum_{n\geqslant 1} n^{-\sigma}.$$

On obtient donc la formule d'Euler en faisant tendre x vers l'infini.

1. Notamment dans le Chapitre 4, où nous faisons spécifiquement référence à la *formule de Mertens* lorsque la valeur de la constante e^γ est nécessaire, et au *second théorème de Mertens* lorsque seule l'estimation de $\sum_{p\leqslant x} 1/p$ est utile.

Maintenant, on a classiquement

$$\sum_{n\geqslant 1}\frac{1}{n^\sigma}=\sum_{n\geqslant 1}\int_n^{n+1}\frac{\mathrm{d}t}{t^\sigma}+O\left(\frac{1}{n^\sigma}-\frac{1}{(n+1)^\sigma}\right)$$
$$=\int_1^\infty\frac{\mathrm{d}t}{t^\sigma}+O(1)=\frac{1}{\sigma-1}+O(1).$$

On déduit donc de la formule d'Euler et de notre évaluation du produit $\prod_{p\leqslant x}(1-1/p)^{-1}$ que

$$b=\lim_{h\to 0+}\prod_{p\leqslant\exp(1/h)}\frac{1-p^{-1-h}}{1-p^{-1}}\prod_{p>\exp(1/h)}(1-p^{-1-h}).$$

Un calcul de routine, reposant sur l'estimation $\ln(1+u)=u+O(u^2)$ ($|u|\leqslant\frac{1}{2}$) et l'inégalité $1-p^{-h}\leqslant h\ln p$, fournit alors

$$\ln b=\lim_{h\to 0+}\{A(h)-B(h)\}$$

avec

$$A(h)=\sum_{p\leqslant\exp(1/h)}\frac{1-p^{-h}}{p}=\sum_{p\leqslant\exp(1/h)}\frac{h}{p}\int_1^p\frac{\mathrm{d}t}{t^{1+h}}$$
$$=h\int_1^{\exp(1/h)}\sum_{t<p\leqslant\exp(1/h)}\frac{1}{p}\frac{\mathrm{d}t}{t^{1+h}},$$

et

$$B(h)=\sum_{p>\exp(1/h)}p^{-1-h}=h\sum_{p>\exp(1/h)}\frac{1}{p}\int_p^\infty\frac{\mathrm{d}t}{t^{1+h}}$$
$$=\int_{\exp(1/h)}^\infty\sum_{\exp(1/h)<p\leqslant t}\frac{1}{p}\frac{\mathrm{d}t}{t^{1+h}}.$$

À ce stade, nous utilisons l'évaluation précédemment établie des sommes $\sum_{p\leqslant x} 1/p$ sous la forme

$$\sum_{u<p\leqslant v} \frac{1}{p} = \sum_{\ln u<n\leqslant \ln v} \frac{1}{n} + O\Big(\frac{1}{\ln 2u}\Big) \quad (1 \leqslant u \leqslant v).$$

En remplaçant dans les expressions obtenues pour $A(h)$ et $B(h)$ et en remontant les calculs, nous obtenons lorsque $h \to 0+$

$$A(h) = \sum_{n\leqslant 1/h} \frac{1-\mathrm{e}^{-hn}}{n} + O\big(h\ln(1/h)\big),$$
$$B(h) = \sum_{n>1/h} \frac{\mathrm{e}^{-hn}}{n} + O(h).$$

Ainsi

$$\begin{aligned} A(h) - B(h) &= \sum_{n\leqslant 1/h} \frac{1}{n} - \sum_{n\geqslant 1} \frac{\mathrm{e}^{-hn}}{n} + o(1) \\ &= \ln(1/h) + \gamma + \ln(1-\mathrm{e}^{-h}) + o(1) \\ &= \gamma + o(1). \end{aligned}$$

Cela implique bien que $b = \mathrm{e}^{\gamma}$ et termine la preuve de la formule de Mertens.

11. Le crible de Brun et le problème des nombres premiers jumeaux

Bien que largement postérieur à Tchébychev, le crible combinatoire de Brun, développé entre 1917 et 1924, appartient à la même sphère d'influence. Il fournit dans un cadre très général des estimations de même nature que celles de Tchébychev.

La motivation de Brun est de rendre utilisable la formule de Legendre pour le crible d'Ératosthène, qui, sous sa forme basique, comporte trop de termes pour autoriser des calculs flexibles. Considérons par exemple le cas des nombres premiers

$$\pi(x) - \pi(\sqrt{x}) + 1 = \sum_{P^+(d) \leqslant \sqrt{x}} \mu(d) \lfloor x/d \rfloor .$$

Si nous estimons $\lfloor x/d \rfloor$ par $x/d + O(1)$, nous obtenons

$$\pi(x) - \pi(\sqrt{x}) + 1 = x \prod_{p \leqslant \sqrt{x}} \Big(1 - \frac{1}{p}\Big) + O\big(2^{\pi(\sqrt{x})}\big).$$

D'une part, le terme d'erreur est exponentiellement plus grand que le terme principal ; d'autre part la formule de Mertens nous montre que ce terme principal est équivalent à $2\mathrm{e}^{-\gamma} x/\ln x$, et n'est donc pas asymptotiquement égal à la valeur conjecturée.

Brun parviendra essentiellement à surmonter ces deux difficultés. Sa méthode fournit, dans de nombreux cas, une véritable formule asymptotique ; dans les autres, elle produit des inégalités optimales à une constante multiplicative près.

L'idée de base est simple. Il s'agit de remplacer dans la formule du crible d'Ératosthène

$$N_m(x) = \sum_{n \leqslant x,\, (n,m)=1} 1 = \sum_{d \mid m} \mu(d) \lfloor x/d \rfloor$$

la fonction de Möbius par une fonction s'annulant plus souvent et fournissant cependant une majoration ou une minoration pour $N_m(x)$.

Lorsqu'on interprète, dans le membre de droite de la formule de Legendre, $\lfloor x/d \rfloor$ comme le nombre des entiers $t \leqslant x/d$ et qu'on ordonne la somme double selon les valeurs de $n = td$, on obtient

$$N_m(x) = \sum_{d \mid m} \sum_{t \leqslant x/d} \mu(d) = \sum_{n \leqslant x} \sum_{d \mid (n,m)} \mu(d).$$

Regroupons maintenant, dans la somme intérieure, les valeurs de d ayant le même nombre de facteurs premiers. Posant $K = \omega((n, m))$, il suit

$$\sum_{d|(n,m)} \mu(d) = \sum_{k=0}^{K} (-1)^k \binom{K}{k} = (1-1)^K$$
$$= \begin{cases} 1 & \text{si } K = 0, \text{ i.e. } (n, m) = 1, \\ 0 & \text{si } K > 0, \text{ i.e. } (n, m) > 1. \end{cases}$$

On voit ainsi que la formule du crible se ramène à celle du binôme. Cela fournit la clef de la méthode de Brun : au lieu d'utiliser le développement complet de $(1-1)^K$, il a recours au calcul de la somme partielle

$$\sum_{k \leqslant \ell} (-1)^k \binom{K}{k} = (-1)^\ell \binom{K-1}{\ell} \qquad (\ell \geqslant 0,\ K \geqslant 1).$$

On obtient ainsi

$$\sum_{d|m} \mu_1(d) \lfloor x/d \rfloor \leqslant N_m(x) \leqslant \sum_{d|m} \mu_2(d) \lfloor x/d \rfloor$$

pour toutes fonctions $\mu_1,\ \mu_2$ définies par

$$\mu_1(d) = \begin{cases} \mu(d) & \text{si } \omega(d) \leqslant 2r+1, \\ 0 & \text{si } \omega(d) > 2r+1, \end{cases}$$

$$\mu_2(d) = \begin{cases} \mu(d) & \text{si } \omega(d) \leqslant 2s, \\ 0 & \text{si } \omega(d) > 2s, \end{cases}$$

où r et s sont des paramètres entiers arbitraires.

La méthode, qui pose un difficile et profond problème d'optimisation, a été développée et raffinée, notamment par Rosser, Halberstam, Richert, puis par Iwaniec, qui lui a donné sa forme quasi définitive.

Il nous entraînerait largement hors du cadre de cet ouvrage de développer plus avant la théorie du crible combinatoire. Contentons-nous de donner une application simple.

Théorème 11.1 *Soient $M \in \mathbb{N}$, $N \in \mathbb{N}^*$ et $\mathcal{A}$ un ensemble d'entiers inclus dans l'intervalle $]M, M+N]$. On suppose que, pour chaque nombre premier p, $\mathcal{A}$ est exclu de $w(p)$ classes résiduelles modulo p et que $w(p) \ll 1$. Alors on a*

$$|\mathcal{A}| \ll N \prod_{p \leqslant N} \Big(1 - \frac{w(p)}{p}\Big).$$

Dans certaines circonstances, le crible peut également fournir des bornes inférieures. Une situation standard de ce type est celle de l'évaluation du nombre des entiers n'excédant pas x et sans facteur premier $\leqslant y$. On obtient que

$$\Phi(x,y) := \sum_{n \leqslant x,\, P^-(n) > y} 1 \sim x \prod_{p \leqslant y} \Big(1 - \frac{1}{p}\Big)$$

uniformément pour $x \to \infty$ et $(\ln y)/\ln x \to 0$.

Brun a développé dès 1919 une spectaculaire application de sa méthode au problème fameux des *nombres premiers jumeaux*. On dit que deux nombres premiers impairs (p, q) sont jumeaux si $q = p + 2$, autrement dit si p et q sont aussi proches que le permet leur état de nombres premiers. Ainsi

$$(3,5),\ (5,7),\ (11,13),\ (17,19),\ (29,31),\ (41,43),\ (59,61)$$

sont des couples premiers gémellaires.

Le Théorème 11.1 a pour conséquence immédiate la majoration

$$\pi_2(n) \ll n/(\ln n)^2,$$

pour le nombre $\pi_2(n)$ des couples de nombres premiers jumeaux n'excédant pas n. Brun avait initialement obtenu une estimation très légèrement moins précise mais cependant suffisante pour impliquer que la série des inverses des nombres premiers jumeaux converge — une manière saisissante de mettre en évidence la rareté des nombres premiers jumeaux dans l'ensemble de tous les nombres premiers. On a en effet, notant $\mathcal{J}$ l'ensemble des nombres premiers jumeaux,

$$\sum_{p\in\mathcal{J}} \frac{1}{p} = \sum_{n\geqslant 2} \frac{\pi_2(n) - \pi_2(n-1)}{n} = \sum_{n\geqslant 2} \frac{\pi_2(n)}{n(n+1)}.$$

Le raisonnement probabiliste heuristique suggère que l'ensemble $\mathcal{J}$ est infini et que la majoration de Brun est pratiquement optimale. Essentiellement, on peut supposer que la probabilité pour qu'un nombre entier de taille n choisi au hasard soit premier est $1/\ln n$. Les nombres n et $n+2$ ne sont pas indépendants, mais on peut simuler le hasard en prétendant que les différentes conditions de congruence modulo les nombres premiers $p \leqslant n$ se comportent de manière statistique. Pour que $(n, n+2)$ soit un couple de nombres premiers jumeaux, il faut et il suffit que n soit exclu d'une classe modulo 2 et, pour chaque nombre premier p avec $2 < p \leqslant \sqrt{n}$, de deux classes modulo p. S'il y avait parfaite indépendance, on obtiendrait une probabilité

$$\tfrac{1}{2} \prod_{2<p\leqslant\sqrt{n}} \Big(1 - \frac{2}{p}\Big).$$

Cependant, nous savons que les grands nombres premiers (de l'ordre d'une puissance de n) ne sont pas indépendants puisque l'on a, d'après le théorème des nombres premiers et la formule de Mertens,

$$\pi(n)/n \sim \tfrac{1}{2}\mathrm{e}^{\gamma} \prod_{p\leqslant\sqrt{n}} (1 - 1/p) \sim 1/\ln n.$$

Cela incite à introduire dans le calcul précédent un facteur correctif $(\frac{1}{2}e^{\gamma})^2$ — une fois $\frac{1}{2}e^{\gamma}$ pour chaque nombre premier — et à conjecturer, avec Hardy et Littlewood, que le nombre $\pi_2(n)$ de couples de nombres premiers jumeaux n'excédant pas n vérifie

$$\frac{\pi_2(n)}{n} \sim (\tfrac{1}{2}e^{\gamma})^2 \tfrac{1}{2} \prod_{2<p\leqslant\sqrt{n}} \left(1-\frac{2}{p}\right) \sim \frac{C}{(\ln n)^2},$$

avec $C := 2\prod_{p>2}\left(1-1/(p-1)^2\right) \approx 1{,}320323$.

La constante C est connue sous le nom de *constante des nombres premiers jumeaux*. Par des méthodes de crible, Jie Wu a montré en 2004 que l'on a pour n assez grand

$$\pi_2(n) \leqslant 3{,}4\, C\frac{n}{(\ln n)^2}.$$

Jie Wu

Chapitre 2
La fonction zêta de Riemann

1. Introduction

Riemann n'a écrit qu'un seul article sur la théorie des nombres, publié en 1859. Ce mémoire a bouleversé définitivement le paysage de la discipline. L'approche spécifique de la répartition des nombres premiers qui y est développée, à la fois simple et révolutionnaire, consiste à faire appel à la théorie de Cauchy des fonctions holomorphes, alors relativement récente.[1]

La théorie de Cauchy est traditionnellement enseignée dans les seconds cycles universitaires, et il n'est pas question d'exposer ici ne serait-ce que les rudiments de ce vaste domaine de l'analyse moderne. Nous nous contentons d'indiquer que l'objet central est celui de *fonction analytique d'une variable complexe*, autrement dit une fonction $f(s)$ ($s \in \mathbb{C}$) dont les variations locales « ressemblent » à celles d'un polynôme.

Ainsi, pour tout entier naturel $m \geqslant 0$, une fonction analytique satisfait à une approximation du type

$$f(s_0 + w) \approx a_0 + a_1 w + a_2 w^2 + \cdots + a_m w^m$$

1. La formule intégrale de Cauchy, clef de voûte de la théorie, date de 1825.

lorsque s_0 est un point où f est bien définie et w est un nombre complexe assez petit, i.e. assez proche de l'origine. En termes consacrés, on dit que f développable en série entière

$$f(s) = \sum_{n \geqslant 0} a_n (s - s_0)^n$$

au voisinage de tout point s_0 où elle est définie.

Le saut qualitatif entre l'analyse réelle et l'analyse complexe provient d'une propriété fondamentale découverte par Cauchy : en chaque point, la valeur d'une fonction analytique peut être calculée en effectuant la « moyenne » de ses valeurs en des points voisins. Bien entendu, il faut définir une notion adaptée de voisinage et expliquer comment on calcule la moyenne. C'est là que la représentation géométrique des nombres complexes trouve son utilité la plus éclatante : sans entrer dans une description rigoureuse, on peut dire que les points à prendre en compte dans le calcul de la moyenne peuvent être choisis de manière à constituer dans leur ensemble *toute* courbe entourant le point sélectionné. Tournez autour de 0 selon un cercle, un rectangle, un triangle ou un polygone à 17 côtés, vous obtiendrez toujours $f(0)$, pour peu que vous ayez défini la moyenne selon les règles de l'art.

Formellement, on dit que cette moyenne est obtenue grâce à une *intégrale curviligne*, le long d'une courbe fermée du plan complexe.

Ce procédé très puissant est aussi très souple car il autorise, sous certaines conditions, des déformations du chemin d'intégration qui facilitent l'approximation de l'intégrale.

Riemann a compris très rapidement que la fonction zêta d'Euler

$$\zeta(\sigma) := 1 + \frac{1}{2^\sigma} + \frac{1}{3^\sigma} + \frac{1}{4^\sigma} + \cdots,$$

déjà présentée au Chapitre 1 (voir p. 3) et qui actualise le lien structurel entre les entiers et les nombres premiers, pouvait être prolongée en une fonction analytique de la variable complexe $s \neq 1$, et partant bénéficier de la liberté de manœuvre octroyée par la théorie de Cauchy.

Dans la suite de ce chapitre, nous nous attachons à décrire plus avant ce contexte et l'apport révolutionnaire de Riemann à la théorie analytique des nombres.

2. Une brève histoire de ce qui va suivre

Le calcul numérique suggère facilement les deux formules suivantes

$$1 + \tfrac{1}{4} + \tfrac{1}{9} + \tfrac{1}{16} + \tfrac{1}{25} + \cdots = \pi^2/6$$
$$(1 - \tfrac{1}{4})(1 - \tfrac{1}{9})(1 - \tfrac{1}{25})(1 - \tfrac{1}{49}) \cdots = 6/\pi^2.$$

Dans la somme, apparaissent les inverses des carrés des entiers ; dans le produit, ce sont les inverses des carrés des nombres premiers qui entrent en jeu.

Ces formules ne sont évidemment pas dues au hasard. On peut les justifier sans trop d'effort à l'aide du théorème fondamental de l'arithmétique, qui énonce qu'un nombre entier s'écrit de manière unique comme le produit de nombres premiers : pour un carré, il faudra simplement élever les facteurs premiers au carré. On a par exemple

$$\frac{1}{1 - \frac{1}{4}} = 1 + \tfrac{1}{2^2} + \tfrac{1}{2^4} + \tfrac{1}{2^6} + \cdots$$

de sorte que

$$\begin{aligned}
&\frac{1}{1 - \frac{1}{4}} \frac{1}{1 - \frac{1}{9}} \frac{1}{1 - \frac{1}{25}} \\
&\quad = \left(1 + \tfrac{1}{2^2} + \tfrac{1}{2^4} + \cdots\right)\left(1 + \tfrac{1}{3^2} + \tfrac{1}{3^4} + \cdots\right)\left(1 + \tfrac{1}{5^2} + \tfrac{1}{5^4} + \cdots\right) \\
&\quad = 1 + \tfrac{1}{2^2} + \tfrac{1}{3^2} + \tfrac{1}{5^2} + \tfrac{1}{2^2 \cdot 3^2} + \cdots
\end{aligned}$$

Lorsque le produit de gauche est étendu à tous les carrés de nombres premiers, tous les carrés des entiers apparaissent dans la somme de droite.

Dans cette formule, le fait que les nombres soient élevés au carré est contingent : on pourrait remplacer, formellement, l'exposant 2 par n'importe quel nombre réel, mais, pour des raisons de convergence, les deux membres n'ont de sens que si cet exposant est > 1.

L'idée de Riemann consiste à étendre le champ de la variable au plan complexe. Dans un premier temps, les mêmes raisons de convergence imposent de se limiter au demi-plan $\sigma = \Re e\, s > 1$.

Augustin-Louis Cauchy
(1789–1857)

Cependant, Cauchy a montré que les fonctions analytiques de la variable complexe possèdent une qualité extraordinaire : lorsque deux fonctions coïncident sur un ensemble suffisamment dense, elles coïncident partout. Une telle propriété est totalement en défaut pour les fonctions de variable réelle ; même indéfiniment dérivable (c'est-à-dire infiniment « lisse »), rien n'empêche une telle fonction d'être nulle sur l'intervalle $[0, 1]$ de vivre indépendamment sa vie sur la demi-droite $]1, \infty[$. Mais lorsqu'une fonction analytique est nulle sur le segment $[0, 1]$, elle est identiquement nulle.

Cette découverte étonnante possède un corollaire non moins remarquable. Supposons que nous disposions de deux formules dépendant d'un nombre complexe *s* et ayant respectivement un sens lorsque *s* décrit deux parties distinctes du plan complexe. Si, en outre, ces deux parties ont une intersection assez « dense », et si les valeurs produites par les deux formules coïncident sur cette intersection, nous pouvons les considérer comme définissant, chacune dans son domaine, une fonction unique.

Les mathématiciens désignent ce phénomène comme le *prolongement analytique*.

Cela ouvre des possibilités jusque-là insoupçonnées : on peut, par exemple, établir une propriété concernant les valeurs d'une fonction à l'infini en ne travaillant que sur le disque unité — allumer une bougie à Rome sans quitter Paris : vous en aviez rêvé, Cauchy l'a rendu possible !

Riemann a apporté à la théorie des nombres ce merveilleux cadeau : prolonger la fonction zêta d'Euler au plan complexe tout entier, privé seulement du point $s = 1$.

Il a ensuite montré que les zéros de cette fonction, autrement dit les points du plan complexe où elle s'annule, jouent un rôle essentiel dans la répartition des nombres premiers. Mieux : il y a une correspondance précise entre la situation des zéros dans le plan et le comportement asymptotique de la fonction $\pi(x)$ qui dénombre les nombres premiers n'excédant pas x.

Ainsi, comme nous le verrons plus loin, prouver que la fonction $\zeta(s)$ ne s'annule pas sur la droite $\sigma = \Re e\, s = 1$ équivaut à établir la conjecture de Gauss–Legendre

$$\pi(x) \sim x/\ln x \qquad (x \to \infty).$$

L'hypothèse de Riemann est probablement le plus célèbre des problèmes ouverts en mathématiques : elle affirme que, si $\zeta(s) = 0$ et $0 < \sigma < 1$, alors $\sigma = \frac{1}{2}$, c'est-à-dire que, dans la bande verticale (appelée aussi *bande critique*)

$$B := \{s : 0 < \sigma < 1\},$$

tous les zéros sont alignés sur la droite $\sigma = \frac{1}{2}$.

La conséquence essentielle de cette hypothèse est qu'elle permet d'estimer $\pi(x)$ avec une erreur « essentiellement » de l'ordre de $\sqrt{x}$. Nous donnerons plus loin une formulation plus précise de cette assertion et une expression du terme principal de l'approximation, qui nécessite des notions mathématiques plus élaborées. Cependant, l'essentiel est là : le mystère de la répartition de nos si concrets et si familiers nombres premiers se cache tout entier dans la manière dont sont disposés les zéros d'une certaine fonction analytique dans le plan complexe.

Avant d'aller plus loin et d'expliciter, avec les outils mathématiques nécessaires, la situation décrite plus haut, donnons un exemple simple de fonction analytique dont les zéros sont répartis sur une droite.

On apprend en classe de Terminale la formule dite de Moivre :

$$e^{i\tau} = \cos\tau + i\sin\tau \qquad (\tau \in \mathbb{R}).$$

On en déduit, pour $s := \sigma + i\tau$, que

$$f(s) := e^s - 1 = e^\sigma \cos\tau - 1 + ie^\sigma \sin\tau.$$

L'équation $f(s) = 0$ équivaut donc à $e^\sigma \cos\tau = 1$ et $e^\sigma \sin\tau = 0$, d'où $e^{2\sigma}\{(\cos\tau)^2 + (\sin\tau)^2\} = e^{2\sigma} = 1$, donc $\sigma = 0$, et $\tau = 2k\pi$ $(k \in \mathbb{Z})$.

Ainsi, les zéros de $f(s)$ sont régulièrement disposés sur l'axe imaginaire, de 2π en 2π.

3. Produit eulérien

Comme souligné plus haut, l'idée fondamentale de Riemann consiste à étendre la formule d'Euler (cf. § 1.10, p. 34) à la variable complexe. Il pose ainsi

$$\zeta(s) := \sum_{n\geqslant 1} n^{-s} = \prod_p \left(1 - p^{-s}\right)^{-1}$$

pour tout nombre complexe s de partie réelle $\sigma > 1$.[1] La série du membre de gauche est absolument convergente, et, par conséquent, sa somme $\zeta(s)$ est analytique dans le demi-plan $\sigma > 1$. La convergence du produit infini du membre de droite est également absolue et ne pose aucun problème. L'égalité des deux membres, que l'on peut voir comme la formule de base de la théorie analytique des nombres, peut être établie de deux manières différentes : soit on étend la preuve du Chapitre 1 au cas d'une variable complexe, soit on utilise la formule d'Euler d'argument réel en faisant appel au *principe du prolongement analytique* décrit au § 2.

1. Ici et dans toute la suite de ce chapitre nous posons systématiquement $s = \sigma + i\tau$.

Nous avons vu au chapitre précédent qu'en faisant tendre s vers 1 dans la formule d'Euler réelle, on obtient la divergence de la série $\sum_p 1/p$, et même l'évaluation quantitative plus précise[1]

$$\sum_{p \leqslant x} 1/p = \ln_2 x + a + o(1),$$

où a est une constante adéquate (cf. p. 33).

La formule d'Euler *complexe* possède la conséquence spécifique remarquable que $\zeta(s) \neq 0$ pour tout nombre complexe s du demi-plan $\sigma > 1$: cela découle immédiatement de la convergence du produit infini, dont aucun facteur ne s'annule dans cette région.

L'importance de la localisation des zéros de $\zeta(s)$ apparaîtra plus loin. Pour l'heure, contentons-nous de remarquer que la formule d'Euler implique immédiatement

$$\log \zeta(s) = \sum_p \log\left(\frac{1}{1-p^{-s}}\right) = \sum_p \frac{1}{p^s} + h(s) \ (\sigma > 1)$$

où $h(s) = \sum_p \sum_{\nu \geqslant 2} 1/\nu p^{\nu s}$ est une série convergente dans le demi-plan $\sigma > \frac{1}{2}$.[2] Or, nous verrons que la fonction $\zeta(s)$ est prolongeable en une fonction analytique dans un domaine plus grand que le demi-plan de convergence $\sigma > 1$. Au vu de la formule précédente, les zéros de $\zeta(s)$ (i.e. de son prolongement) dans le demi-plan $\sigma > \frac{1}{2}$ sont également des valeurs remarquables (le terme consacré est celui de *singularités*) du prolongement analytique de la série $\sum_p p^{-s}$. Il s'avère que, grâce à la théorie des fonctions

1. Équivalente à la formule de Mertens (cf. § 1.10), elle-même établie grâce à la formule d'Euler réelle.

2. Ici et dans la suite nous utilisons librement l'existence d'un prolongement complexe de la fonction logarithme satisfaisant à la propriété suivante : étant donnés un nombre réel $\alpha > 0$ et une fonction analytique $f(s)$, sans zéro pour $\Re e\, s > \alpha$ et réelle positive pour $s \in]\alpha, +\infty[$, alors $\log f(s)$ est défini dans le demi-plan $\Re e\, s > \alpha$ et y vérifie $\Re e \log f(s) = \ln |f(s)|$. Le lecteur néophyte pourra se contenter d'admettre que, moyennant certaines précautions qui seront toujours prises dans la suite, ce prolongement satisfait aux règles de calcul usuelles.

holomorphes, la localisation de ces singularités est essentiellement équivalente à l'élucidation du comportement asymptotique de la fonction de comptage $\pi(x)$. Nous reviendrons amplement sur ce point capital.

Avant de poursuivre notre étude des liens entre la fonction zêta de Riemann et la répartition des nombres premiers, signalons quelques autres conséquences remarquables de la formule d'Euler.

On a clairement pour $\sigma > 1$

$$\zeta(s)^{-1} = \prod_p \left(1 - p^{-s}\right).$$

Lorsque l'on développe le produit infini (la preuve donnée au Chapitre 1 de la formule d'Euler s'étend *mutatis mutandis* à ce cadre), on obtient

$$\zeta(s)^{-1} = 1 - \sum_p p^{-s} + \sum_{p<q} (pq)^{-s} - \sum_{p<q<r} (pqr)^{-s} + \cdots$$

où $p, q, r, \ldots$ désignent des nombres premiers. Ainsi seuls des entiers sans facteur carré apparaissent dans cette sommation, et, pour un tel entier n, le coefficient de n^{-s} vaut ± 1 selon que le nombre de facteurs premiers de n est pair ou impair. En d'autres termes, on a

$$\zeta(s)^{-1} = \sum_{n \geqslant 1} \mu(n) n^{-s} \qquad (\sigma > 1)$$

où $\mu(n)$ est la fonction de Möbius, déjà rencontrée au Chapitre 1 — cf. p. 25.

Semblablement, on peut écrire

$$\zeta(s)^2 = \sum_{\ell \geqslant 1, m \geqslant 1} \ell^{-s} m^{-s} = \sum_{n \geqslant 1} \tau(n) n^{-s} \qquad (\sigma > 1)$$

où $\tau(n)$ est le nombre de décompositions de n sous la forme ℓm, c'est-à-dire le nombre de diviseurs de n. Cette relation est un cas particulier de *convolution des fonctions arithmétiques*, que nous décrivons plus en détail au Chapitre 4.

Par dérivation logarithmique, la formule d'Euler fournit

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_p \sum_{\nu \geqslant 1} \frac{\ln p}{p^{\nu s}} = \sum_{n \geqslant 1} \frac{\Lambda(n)}{n^s}$$

où $\Lambda(n)$ est la fonction de von Mangoldt déjà introduite au Chapitre 1 (§ 1.9, p. 27).

On voit ainsi que les fonctions arithmétiques de la théorie élémentaire des nombres premiers apparaissent naturellement comme des coefficients de séries dérivées de $\zeta(s)$ par des transformations analytiques simples. Cela illustre à nouveau le rôle central de la fonction zêta dans la théorie analytique des nombres.

4. Prolongement analytique

Ainsi que nous l'avons signalé plus haut, l'intérêt majeur de considérer $\zeta(s)$ comme fonction d'une variable complexe est que l'on peut en étendre la définition au delà du domaine de convergence de la série. Les renseignements relatifs au prolongement peuvent ensuite être utilisés dans le cadre de la théorie générale des fonctions analytiques, en particulier pour l'évaluation d'intégrales curvilignes.

Nous allons tout d'abord étendre la définition de $\zeta(s)$ au demi-plan $\sigma > 0$. Une manipulation très simple est suffisante. On a pour $\sigma > 1$

$$\begin{aligned}\zeta(s) = \sum_{n \geqslant 1} n^{-s} &= \sum_{n \geqslant 1} s \int_n^\infty \frac{\mathrm{d}t}{t^{s+1}} = s \int_1^\infty \Big(\sum_{n \leqslant t} 1 \Big) \frac{\mathrm{d}t}{t^{s+1}} \\ &= s \int_1^\infty \frac{\lfloor t \rfloor}{t^{s+1}}\, \mathrm{d}t = \frac{s}{s-1} - s \int_1^\infty \frac{\langle t \rangle}{t^{s+1}}\, \mathrm{d}t,\end{aligned}$$

où $\lfloor t \rfloor$ désigne la partie entière du nombre réel t et $\langle t \rangle$ sa partie fractionnaire.

Comme $\langle t \rangle \in [0, 1[$, la dernière intégrale converge pour $\sigma > 0$. Le membre de droite définit donc un prolongement de $\zeta(s)$ au demi-plan $\sigma > 0$ privé du point $s = 1$. Ce prolongement est

analytique (la théorie nous apprend qu'il suffit pour cela que ce soit une fonction dérivable de s), donc, d'après le principe mentionné dans l'introduction, il est uniquement déterminé. Autrement dit, toute autre méthode d'extension de la définition de $\zeta(s)$ (et il y en a de nombreuses !) aurait conduit à la même fonction.

Une des propriétés les plus frappantes et les plus importantes découvertes par Riemann est l'existence d'une *équation fonctionnelle* pour la fonction zêta, de la forme

$$\zeta(s) = \chi(s)\zeta(1-s) \qquad (0 < \sigma \leqslant 1)$$

où χ est définie au moyen de certaines fonctions classiques de l'analyse *et possède un sens pour toute valeur complexe de s non égale à un entier positif impair*. Nous expliciterons plus loin $\chi(s)$. L'essentiel est d'abord de constater que, puisque la transformation $s \mapsto 1 - s$ possède le point $s = \frac{1}{2}$ pour centre de symétrie, l'équation fonctionnelle permet (grâce au principe du prolongement analytique) de définir $\zeta(s)$ dans le demi-plan $\sigma \leqslant \frac{1}{2}$[1] dès lors que l'on en connaît la valeur pour $\sigma \geqslant \frac{1}{2}$. Or, nous venons précisément de définir $\zeta(s)$ pour $\sigma > 0$: *l'équation fonctionnelle fournit donc le prolongement de* $\zeta(s)$ *au plan complexe tout entier*.

Il existe beaucoup de démonstrations différentes de l'équation fonctionnelle de zêta. L'une des plus limpides, que nous présentons ici, repose sur la *formule de Poisson* qui énonce que, si l'on définit la transformée de Fourier d'une fonction f intégrable sur $\mathbb{R}$ par

$$\widehat{f}(x) = \int_{-\infty}^{+\infty} f(y)\mathrm{e}^{-2\pi i x y}\,\mathrm{d}y,$$

on a

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \widehat{f}(n)$$

1. Sauf, *a priori*, $s = 0$, mais la valeur $\zeta(0)$ s'obtient aisément par passage à la limite.

pour toute fonction f dérivable sur $\mathbb{R}$ et vérifiant certaines conditions de décroissance.[1]

Il est très facile de vérifier que les conditions d'application de la formule sommatoire de Poisson sont satisfaites par $f(x) = \mathrm{e}^{-\pi u x^2}$ pour chaque valeur du paramètre positif u. On a alors

$$\widehat{f}(x) = \mathrm{e}^{-\pi x^2/u} u^{-1/2}.$$

On obtient ainsi que la fonction $\vartheta :]0, \infty[\to]0, \infty[$ définie par la formule

$$\vartheta(u) := \sum_{n \in \mathbb{Z}} \mathrm{e}^{-\pi n^2 u}$$

satisfait à l'équation fonctionnelle

$$\vartheta(1/u) = \sqrt{u}\,\vartheta(u) \qquad (u > 0).^{(2)}$$

Nous verrons plus loin comment cette identité implique à son tour l'équation fonctionnelle de la fonction zêta.

La fonction $\chi(s)$ intervenant dans l'équation fonctionnelle de $\zeta(s)$ est définie (à l'aide de la fonction Γ d'Euler sur laquelle nous reviendrons dans un instant) par la formule

$$\chi(s) = 2^s \pi^{s-1} \sin(\tfrac{1}{2}\pi s)\Gamma(1-s),$$

de sorte que l'on a finalement

$$\zeta(s) = 2^s \pi^{s-1} \sin(\tfrac{1}{2}\pi s)\Gamma(1-s)\zeta(1-s) \qquad (s \neq 0, 1).$$

Ici, la fonction sinus représente en fait le prolongement analytique de la fonction trigonométrique usuelle et est définie comme somme de la série

$$\sin z := \sum_{n \geqslant 0} \frac{(-1)^n z^{2n+1}}{(2n+1)!}$$

qui converge pour tout $z \in \mathbb{C}$.

1. On peut montrer, par exemple, que la formule de Poisson est valable si $\sum_{n \in \mathbb{Z}} f(n)$ converge et si $\sum_{n \in \mathbb{Z}} f'(n+x)$ converge uniformément pour $x \in [0, 1]$.
2. La fonction $\vartheta(u)$ est connue sous le nom de *fonction thêta de Jacobi.*

La fonction $\Gamma(s)$ a été découverte par Euler, qui l'a initialement définie par la formule

$$\Gamma(s) = \frac{1}{s} \prod_{n \geqslant 1} \frac{(1 + 1/n)^s}{1 + s/n} \quad (s \in \mathbb{C},\ s \neq 0, -1, -2, \ldots).$$

C'est une fonction omniprésente en analyse mathématique, où elle est la seule à pouvoir contester la position dominante du couple infernal constitué de l'exponentielle et du logarithme. Elle possède de nombreuses définitions équivalentes. La plus fréquemment utilisée dans l'enseignement moderne est

$$\Gamma(s) = \int_0^\infty x^{s-1} \mathrm{e}^{-x}\, \mathrm{d}x,$$

qui a l'inconvénient de n'avoir *a priori* de sens que pour $\sigma > 0$ mais permet facilement, grâce à une banale intégration par parties, d'établir l'équation fonctionnelle

$$s\Gamma(s) = \Gamma(s + 1),$$

qui, à son tour, fournit le prolongement analytique.

En particulier, $\Gamma(n + 1) = n!$, ce qu'on exprime parfois en disant que *la fonction* Γ *interpole la factorielle.* Parmi les autres propriétés classiques de la fonction Gamma, mentionnons encore la jolie *formule des compléments*

$$\Gamma(s)\Gamma(1 - s) = \pi / \sin(\pi s)$$

et la non moins esthétique *formule de duplication* de Legendre

$$\Gamma(s)\Gamma(s + \tfrac{1}{2}) = \sqrt{\pi}\, 2^{1-2s} \Gamma(2s).$$

Après ces préliminaires, il est facile d'établir l'équation fonctionnelle de $\zeta(s)$ selon une méthode indiquée par Riemann lui-même. On commence par observer qu'un simple changement de variable

(soit $x = \pi n^2 y$) dans l'intégrale de définition de $\Gamma(\frac{1}{2}s)$ fournit, pour $\sigma > 0$,

$$\Gamma(\tfrac{1}{2}s)\pi^{-s/2}n^{-s} = \int_0^\infty y^{s/2-1}\mathrm{e}^{-\pi n^2 y}\,\mathrm{d}y \quad (n = 1, 2, \ldots).$$

Pour $\sigma > 1$, on peut sommer cette identité sur toutes les valeurs entières positives de n. On obtient

$$\zeta(s)\Gamma\left(\tfrac{1}{2}s\right)\pi^{-s/2} = \int_0^\infty \vartheta_1(y)y^{s/2-1}\,\mathrm{d}y,$$

où l'on a posé $\vartheta_1(y) := \frac{1}{2}(\vartheta(y) - 1)$.

Maintenant, nous scindons l'intégrale au point $y = 1$ et effectuons le changement de variable $z = 1/y$ pour évaluer la contribution de l'intervalle $0 \leqslant y \leqslant 1$. L'équation fonctionnelle de ϑ fournit

$$\int_0^1 \vartheta_1(y)y^{s/2-1}\,\mathrm{d}y = \frac{1}{s(s-1)} + \int_1^\infty \vartheta_1(z)z^{-(s+1)/2}\,\mathrm{d}z.$$

En reportant dans la formule initiale, il suit

$$\zeta(s)\Gamma\left(\tfrac{1}{2}s\right)\pi^{-s/2} = \frac{1}{s(s-1)} + \int_1^\infty \vartheta_1(x)\{x^{-(s+1)/2} + x^{s/2-1}\}\,\mathrm{d}x.$$

On a $\vartheta_1(x) = O(\mathrm{e}^{-\pi x})$, une majoration plus que suffisante pour assurer la convergence uniforme de l'intégrale précédente sur tout domaine borné de $\mathbb{C}$. Cela définit une fonction analytique de s, manifestement invariante par la transformation $s \mapsto 1 - s$. On en déduit *ipso facto* le prolongement à $\mathbb{C} \setminus \{0, 1\}$ de la fonction $\zeta(s)\Gamma\left(\frac{1}{2}s\right)\pi^{-s/2}$ en une fonction invariante par cette même transformation : c'est l'équation fonctionnelle cherchée, que l'on ramène sans peine à la forme annoncée grâce aux formules des compléments et de duplication.

Ce résultat, d'une grande beauté, est aussi très riche de conséquences. Nous n'en indiquons que quelques-unes.

Faisons d'abord, dans la formule

$$\zeta(s) = \chi(s)\zeta(1-s),$$

tendre s vers $2n+1$, où n est un entier positif. Le membre de gauche tend vers la valeur finie $\zeta(2n+1)$, alors que le module du facteur $\chi(s)$ tend vers l'infini — comme l'atteste, par exemple, la définition d'Euler de $\Gamma(s)$. Cela n'est possible que si $\zeta(1-s)$ tend vers 0. Nous avons donc établi que *la fonction $\zeta(s)$ s'annule aux points $-2, -4, -6, \ldots$, qu'on appelle les zéros triviaux de $\zeta(s)$.*

L'équation fonctionnelle fournit aussi que *le point $s = 1$ est la seule singularité de zêta dans tout le plan complexe.* Cela résulte immédiatement de la validité de cette assertion pour le demi-plan $\sigma \geqslant \frac{1}{2}$, en remarquant que les seules singularités de $\chi(s)$ sont les entiers positifs impairs. En effectuant un développement limité de l'équation fonctionnelle au voisinage de l'origine, on obtient facilement que $\zeta(0) = -\frac{1}{2}$ et, un peu moins facilement,[1] que

$$\zeta'(0) = -\tfrac{1}{2}\log 2\pi.$$

Une troisième conséquence de l'équation fonctionnelle concerne le calcul de $\zeta(2n)$ pour $n \geqslant 1$. Il faut pour cela faire appel à un résultat auxiliaire, issu d'une autre méthode de prolongement analytique de $\zeta(s)$, qui précise que *les valeurs $\zeta(-2n-1)$ $(n \geqslant 0)$ sont des nombres rationnels.* Plus précisément, notant B_n le n-ième nombre de Bernoulli,[2] on a

$$\zeta(-2n-1) = -\frac{B_{2n+2}}{2n+2} \qquad (n \geqslant 0).$$

1. Ce calcul fait en particulier appel à la formule de Stirling citée en note p. 26.
2. Les nombres de Bernoulli sont définis par le développement de Taylor

$$\frac{x}{e^x - 1} = \sum_{n \geqslant 0} \frac{B_n x^n}{n!}.$$

On a $B_1 = -\frac{1}{2}$, $B_2 = \frac{1}{6}$, $B_4 = -\frac{1}{30}$, $B_6 = \frac{1}{42}$, ..., et $B_{2n+1} = 0$ pour tout $n \geqslant 1$.

En reportant dans l'équation fonctionnelle, on obtient

$$\zeta(2n) = (-1)^{n-1} 2^{2n-1} \frac{B_{2n}}{(2n)!} \pi^{2n} \qquad (n \geqslant 1).$$

Ainsi

$$\zeta(2) = \sum_{n \geqslant 1} \frac{1}{n^2} = \tfrac{1}{6}\pi^2, \quad \zeta(4) = \sum_{n \geqslant 1} \frac{1}{n^4} = \tfrac{1}{90}\pi^4, \quad \text{etc.}$$

On sait depuis Lindemann (1882) que le nombre π est *transcendant*, c'est-à-dire qu'il est non seulement irrationnel mais n'annule aucun polynôme à coefficients entiers non identiquement nul.[1] En particulier, *tous les nombres $\zeta(2n)$ sont donc transcendants.*

La nature arithmétique des nombres $\zeta(2n+1)$ pose un problème ouvert. À la suite du résultat de Roger Apéry, qui a montré en 1978 que $\zeta(3)$ est irrationnel,[2] Tanguy Rivoal a établi en 2000 que $\zeta(2n+1)$ est irrationnel pour une infinité de valeurs de n. Cependant ces valeurs de la fonction zêta continuent de braver la sagacité des spécialistes... et la curiosité des amateurs !

5. La droite $\sigma = 1$ et le théorème des nombres premiers

Une fonction analytique $f(s)$ a nécessairement un comportement local simple, puisqu'elle est, au voisinage de chaque point, très bien approchable par un polynôme. En particulier, si $f(s_0) = 0$,

1. Lindemann a ainsi montré que la quadrature du cercle (c'est-à-dire la construction, avec pour seuls instruments la règle et le compas, d'un carré de surface égale à celle du cercle de rayon unité, soit π) est impossible. Il a de ce fait apporté une réponse négative à un problème posé par les Grecs et plus de deux fois millénaire.

2. Apéry donne en particulier la jolie formule

$$\zeta(3) = \tfrac{5}{2} \sum_{n \geqslant 1} \frac{(-1)^{n-1}}{n^3 \binom{2n}{n}}.$$

alors on a, pour une constante c et un entier $k \geqslant 1$ convenables,

$$f(s) \sim c(s - s_0)^k \qquad (s \to s_0).$$

Nous avons vu au § 4 que $\zeta(s)$ est prolongeable en une fonction analytique dans $\mathbb{C} \setminus \{1\}$ et que le comportement de cette fonction est également simple au voisinage de la singularité $s = 1$, puisque l'on a

$$\zeta(s) \sim 1/(s-1) \qquad (s \to 1).$$

En considérant la relation issue du produit eulérien

$$\log \zeta(s) = \sum_p p^{-s} + h(s) \qquad (\sigma > 1)$$

où $h(s)$ est une fonction analytique dans le demi-plan $\sigma > \frac{1}{2}$, on déduit donc qu'au voisinage de $s = 1$ la fonction $\log \zeta(s)$ « ressemble » à $\log\{1/(s-1)\}$, et partant que sa dérivée se comporte comme $-1/(s-1)$. Autrement dit, on a

$$Z(s) := -\frac{\zeta'(s)}{\zeta(s)} = \sum_{n \geqslant 1} \frac{\Lambda(n)}{n^s} \sim \frac{1}{s-1}$$

lorsque s tend vers 1 en restant dans le demi-plan $\sigma > 1$.

Hadamard et La Vallée-Poussin ont établi indépendamment en 1896 que *la fonction $\zeta(s)$ ne s'annule pas sur la droite $\sigma = 1$*, ce que l'on peut interpréter comme le fait que *l'approximation de $Z(s)$ par $1/(s-1)$ reste valable, en un sens faible, tout le long de la droite $\sigma = 1$*. Ainsi que nous allons le voir immédiatement, cela implique le théorème des nombres premiers sous la forme

$$\pi(x) \sim x/\ln x \qquad (x \to \infty).$$

Commençons par une justification heuristique, à laquelle le lecteur peu familiarisé avec les techniques avancées de l'analyse pourra se limiter. Rappelons la définition de la fonction de Tchébychev

$$\psi(x) = \sum_{n \leqslant x} \Lambda(n).$$

Nous avons établi au Chapitre 1 (§ 1.9, p. 29) la formule asymptotique

$$\pi(x) = \frac{\psi(x)}{\ln x}\Big\{1 + O\Big(\frac{1}{\ln x}\Big)\Big\} \qquad (x \to \infty).$$

Donc le théorème des nombres premiers équivaut à

$$\psi(x) \sim x \qquad (x \to \infty).$$

Cela signifie que la fonction Λ se comporte en moyenne comme la fonction $\mathbf{1}$. Or, les séries en s[1] associées à ces fonctions sont respectivement $Z(s)$ et $\zeta(s)$. Observons maintenant que la fonction $\zeta(s)$ est régulière en tout point $s \neq 1$ alors que

$$Z(s) = -\,\mathrm{d}\log\zeta(s)/\,\mathrm{d}s = -\zeta'(s)/\zeta(s)$$

possède une singularité non seulement en $s = 1$ mais aussi en tout zéro de $\zeta(s)$. Ainsi, l'absence de zéro de $\zeta(s)$ sur la droite $\sigma = 1$ signifie que les fonctions $Z(s)$ et $\zeta(s)$ sont de même nature dans le demi-plan *fermé* $\sigma \geqslant 1$. Ce renseignement étaie donc bien qualitativement l'hypothèse que Λ et $\mathbf{1}$, partant $\psi(x) = \sum_{n\leqslant x}\Lambda(n)$ et $\lfloor x\rfloor = \sum_{n\leqslant x} 1$, se comportent de manière similaire.

Techniquement, le lien entre une série de série de Dirichlet

$$F(s) := \sum_{n\geqslant 1} a_n n^{-s}$$

et la fonction sommatoire de ses coefficients

$$A(x) = \sum_{1\leqslant n\leqslant x} a_n$$

est fourni par une classe de formules appelées *formules de Perron.* Dans le cas qui nous occupe, et conformément à l'argument

1. Communément désignées sous le nom de *séries de série de Dirichlet.*

heuristique précédemment développé, il est naturel d'introduire la série

$$F(s) = Z(s) - \zeta(s) = \sum_{n \geqslant 1} \frac{\Lambda(n) - 1}{n^s}.$$

Une formule de Perron adaptée au problème peut alors s'écrire, avec $s = \sigma + i\tau$,

$$\int_0^x (\psi(y) - \lfloor y \rfloor)\, \mathrm{d}y = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{F(s) x^{1+s}}{s(1+s)}\, \mathrm{d}\tau \qquad (\sigma > 1).$$

Sous l'hypothèse que $\zeta(s)$ ne s'annule pas pour $\sigma = 1$, le nombre $F(s)$ tend, lorsque $\sigma \to 1+$, vers une limite finie $F(1 + i\tau)$ pour toute valeur de $\tau \in \mathbb{R}$: on a $F(1 + i\tau) = -\zeta'(1 + i\tau)/\zeta(1 + i\tau) - \zeta(1 + i\tau)$ si $\tau \neq 0$, et la propriété persiste pour $\tau = 0$ puisque $\zeta(s)$ et $Z(s)$ sont toutes deux, dans un voisinage de $s = 1$, de la forme $1/(s-1) + h(s)$, où $h(s)$ est analytique.

Nous admettons alors que l'on peut passer à la limite sous le signe d'intégration dans la formule de Perron : cela est aisément justifié en pratique sous l'hypothèse que $|\zeta(1 + i\tau)|$ ne devient pas trop petit lorsque $|\tau| \to \infty$ et toutes les preuves du fait que cette quantité ne s'annule pas fournissent effectivement ce renseignement — avec d'ailleurs une grande marge de sécurité. Nous obtenons ainsi

$$\int_0^x (\psi(y) - \lfloor y \rfloor)\, \mathrm{d}y = \frac{x^2}{2\pi} \int_{-\infty}^{+\infty} \frac{F(1 + i\tau) x^{i\tau}}{(1 + i\tau)(2 + i\tau)}\, \mathrm{d}\tau.$$

Le membre de droite peut être réécrit sous la forme

$$x^2 \widehat{f}\left(-\frac{\ln x}{2\pi}\right) \quad \text{avec} \quad f(\tau) := \frac{F(1 + i\tau)}{2\pi(1 + i\tau)(2 + i\tau)}.$$

Or un résultat classique d'analyse de Fourier, connu sous le nom de *lemme de Riemann–Lebesgue*, stipule que la transformée de Fourier d'une fonction intégrable tend nécessairement vers 0 à l'infini.[1]

Nous avons donc obtenu

$$\int_0^x (\psi(y) - \lfloor y \rfloor)\,\mathrm{d}y = o(x^2) \qquad (x \to \infty),$$

d'où

$$\int_0^x \psi(y)\,\mathrm{d}y = \tfrac{1}{2}x^2 + o(x^2) \qquad (x \to \infty).$$

Une manipulation très simple, reposant sur la croissance de la fonction ψ, permet alors de conclure. Pour tout $\varepsilon > 0$, on déduit en effet de ce qui précède que

$$\begin{aligned}\tfrac{1}{2}x^2 - \tfrac{1}{2}x^2(1-\varepsilon)^2 + o(x^2) &= \int_{x(1-\varepsilon)}^{x} \psi(y)\,\mathrm{d}y \leqslant \varepsilon x \psi(x) \\ &\leqslant \int_x^{x(1+\varepsilon)} \psi(y)\,\mathrm{d}y = \tfrac{1}{2}x^2(1+\varepsilon)^2 - \tfrac{1}{2}x^2 + o(x^2),\end{aligned}$$

d'où, en divisant par εx,

$$(1 - \tfrac{1}{2}\varepsilon)x + o(x) \leqslant \psi(x) \leqslant (1 + \tfrac{1}{2}\varepsilon)x + o(x).$$

Comme ε peut être choisi arbitrairement petit, on obtient bien la conclusion requise, sous la forme $\psi(x) \sim x$.

Nous avons donc montré comment l'absence de zéro de $\zeta(s)$ sur la droite $\sigma = 1$ implique le théorème des nombres premiers. Il est à noter que la réciproque peut également être établie facilement. Si

$$\zeta(1 + i\tau_0) = 0$$

1. Le lecteur pourra s'en convaincre aisément dans le cas des fonctions en escalier. Le cas général s'en déduit en observant (ou en admettant !) que la transformée de Fourier d'une fonction intégrable peut être rendue uniformément proche de la transformée d'une fonction en escalier.

l'analyticité de $\zeta(s)$ fournit immédiatement, pour un nombre entier $m \geqslant 1$ et une constante $c \neq 0$ convenables, le développement

$$\zeta(s) = c(s - 1 - i\tau_0)^m + \cdots$$

au voisinage de $s_0 = 1 + i\tau_0$, d'où

$$\lim_{\sigma \to 1+} (\sigma - 1) Z(\sigma + i\tau_0) = -m.$$

Par ailleurs, une simple intégration par parties permet d'écrire

$$Z(s) = \frac{s}{s-1} + s \int_1^\infty \big(\psi(y) - y\big) \frac{\mathrm{d}y}{y^{s+1}} \qquad (\sigma > 1),$$

d'où l'on déduit, sous l'hypothèse $\psi(y) \sim y$ $(y \to \infty)$ et en posant $s = \sigma + i\tau_0$,

$$(\sigma - 1)|Z(s)| \leqslant (\sigma - 1)|s| \Big\{ \frac{1}{|\tau_0|} + \int_1^\infty o(y^{-\sigma})\, \mathrm{d}y \Big\} = o(1)$$

lorsque $\sigma \to 1+$. Cela implique $m = 0$ et contredit ainsi l'existence d'un zéro sur la droite $\sigma = 1$.

Il nous reste à montrer que $\zeta(s)$ ne s'annule pas sur la droite $\sigma = 1$. Les preuves initiales de Hadamard et La Vallée-Poussin sont d'apparences assez différentes mais reposent toutes deux sur un argument de même nature, à savoir que la relation $\zeta(1 + i\tau_0) = 0$ impliquerait que $1 + 2i\tau_0$ soit une singularité de $\zeta(s)$. En effet, ainsi que nous l'avons vu au § 3, on a pour $\sigma > 1$

$$\log \zeta(s) = \sum_p p^{-s} + O(1)$$

donc

$$\ln |\zeta(s)| = \sum_p \frac{\cos(\tau \ln p)}{p^\sigma} + O(1).$$

Si cette quantité tend vers $-\infty$ lorsque $\tau = \tau_0$ et $\sigma \to 1+$, elle doit le faire au moins comme $\ln(\sigma - 1)$ puisque ζ est analytique.

Or, le comportement de la fonction zêta sur l'axe réel nous apprend que

$$\sum_p \frac{1}{p^\sigma} \sim \ln\Big(\frac{1}{\sigma-1}\Big).$$

Cela signifie que les $\cos(\tau_0 \ln p)$ sont majoritairement proches de -1, c'est-à-dire que les nombres $\tau_0 \ln p$ sont majoritairement proches de π modulo 2π. Mais alors les nombres $2\tau_0 \ln p$ devraient être majoritairement proches de 0 modulo 2π, de sorte que

$$\lim_{\sigma\to 1+} |\zeta(\sigma + 2i\tau_0)| \to \infty,$$

ce qui contredit la continuité de $\zeta(s)$ en $s = 1 + 2i\tau_0$.

Mertens (1898) a formalisé l'argument précédent sous une forme particulièrement limpide en introduisant l'identité trigonométrique

$$3 + 4\cos\vartheta + \cos 2\vartheta = 2(1+\cos\vartheta)^2 \geqslant 0 \qquad (\vartheta \in \mathbb{R}).$$

Lorsqu'on applique, comme le fait La Vallée-Poussin dans un mémoire subséquent, cette inégalité pour $\vartheta = \tau_0 \ln p^\nu$ et que l'on somme sur toutes les valeurs de p et $\nu \geqslant 1$ après avoir multiplié par $\nu^{-1} p^{-\sigma\nu}$, on obtient, compte tenu de ce qui précède,

$$3\ln|\zeta(\sigma)| + 4\ln|\zeta(\sigma + i\tau_0)| + \ln|\zeta(\sigma + i2\tau_0)| \geqslant 0$$

d'où

$$|\zeta(\sigma)|^3 |\zeta(\sigma + i\tau_0)|^4 |\zeta(\sigma + i2\tau_0)| \geqslant 1.$$

Si $1 + i\tau_0$ était un zéro de $\zeta(s)$, on devrait avoir par analyticité $|\zeta(\sigma + i\tau_0)| \sim c(\sigma - 1)$ lorsque $\sigma \to 1+$, avec une constante convenable $c \geqslant 0$.[1] Compte tenu de la régularité de $\zeta(s)$ en

1. Si $c = 0$, nous interprétons la relation précédente comme

$$|\zeta(\sigma + i\tau_0)| = o(\sigma - 1),$$

ce qui n'altère pas le raisonnement qui suit.

$s = 1 + 2i\tau_0$, le membre de gauche de l'inégalité précédente devrait donc être $O(\sigma - 1)$, et en particulier tendre vers 0, ce qui contredit le fait qu'il doive constamment demeurer $\geqslant 1$.

6. L'hypothèse de Riemann

Nous avons vu au paragraphe précédent que l'absence de zéro de $\zeta(s)$ sur la droite $\sigma = 1$ suffit « essentiellement » à établir le théorème des nombres premiers.[1] Les preuves originales de Hadamard et La Vallée-Poussin peuvent facilement être rendues quantitatives. Par des calculs standard et sans introduire d'idée nouvelle, elles fournissent que tout zéro $\varrho = \beta + i\gamma$ de $\zeta(s)$ satisfait

$$\beta \leqslant 1 - c/(\ln|\gamma|)^9$$

pour une constante convenable $c > 0$.[2] Dès 1898, La Vallée-Poussin, en utilisant d'ailleurs un théorème de factorisation des fonctions analytiques dû à Hadamard, a amélioré ce résultat en

$$\beta \leqslant 1 - c/\ln|\gamma|.$$

Cela fournit une majoration effective pour le terme d'erreur du théorème des nombres premiers. Une technique classique d'intégration complexe permet en effet de déduire facilement de l'estimation précédente que l'on a, pour une constante convenable $a > 0$,

$$\psi(x) = x + O\Big(x\mathrm{e}^{-a\sqrt{\ln x}}\Big),$$
$$\pi(x) = \mathrm{li}(x) + O\Big(x\mathrm{e}^{-a\sqrt{\ln x}}\Big).$$

1. La restriction vient du fait que nous avons également dû supposer que $|\zeta(1+i\tau)|$ ne prend pas de trop petites valeurs lorsque $|\tau| \to \infty$. Ikehara a montré en 1931 qu'aucune minoration n'est en réalité nécessaire et a fourni un cadre général où l'on peut déduire *stricto sensu* le théorème des nombres premiers de l'absence de zéro d'abscisse 1 sans introduire d'information supplémentaire.

2. On peut vérifier numériquement que, si $\zeta(\beta + i\gamma) = 0$, alors $|\gamma| > 14$.

La fonction $\mathrm{li}(x)$, ou *logarithme intégral*, est définie dans les *Notations et conventions.* On a bien sûr

$$\mathrm{li}(x) \sim x/\ln x \qquad (x \to \infty),$$

donc $\pi(x) \sim \mathrm{li}(x)$, mais l'approximation est en réalité de bien meilleure qualité puisque le terme d'erreur est $O\big(x/(\ln x)^k\big)$ pour tout $k > 0$. Cela confirme en particulier une conjecture de Gauss qui, dans la première moitié du dix-neuvième siècle[1] et sur la foi des tables de nombres premiers alors disponibles, avait précisé l'intuition de Legendre en émettant l'hypothèse que le logarithme intégral constituait une excellente approximation de la loi de raréfaction des nombres premiers.

Ainsi, l'absence de zéro de $\zeta(s)$ sur la droite $\sigma = 1$ fournit déjà une très bonne évaluation asymptotique de $\pi(x)$. Dans son mémoire de 1859, Riemann émet l'opinion que beaucoup plus est vrai, à savoir que *tous les zéros non triviaux de $\zeta(s)$ sont situés sur l'axe de symétrie $\sigma = \frac{1}{2}$.*[2]

Cette conjecture, connue sous le nom d'*hypothèse de Riemann*, n'a toujours pas été résolue aujourd'hui, malgré les efforts conjugués de centaines de mathématiciens. Elle est riche d'implications dans toutes les branches de la théorie des nombres et possède des généralisations dans de nombreux domaines des mathématiques.

Pour bien comprendre l'incidence de l'hypothèse de Riemann sur la loi de répartition des nombres premiers, il faut introduire un nouvel outil, également décrit par Riemann, et connu sous le nom de *formules explicites de la théorie des nombres*.

Il s'agit ici de pousser jusqu'à sa limite naturelle la description de la fonction $\psi(x)$ — une formule analogue, mais plus compliquée, est valable pour $\pi(x)$ — en termes des zéros de $\zeta(s)$. Qu'une telle

1. Dans une lettre datant de 1849, en réponse à une question de l'astronome et mathématicien suisse Encke.

2. Gram a montré dès 1903 que les 15 premiers zéros (i.e. les 15 zéros de plus petites ordonnées positives) sont bien d'abscisse $\frac{1}{2}$. On sait aujourd'hui qu'il en va de même pour plus d'un milliard de zéros.

entreprise soit seulement possible peut surprendre. On peut la justifier *a priori* en observant que la théorie de Hadamard permet de définir essentiellement une fonction analytique par ses zéros et ses singularités. Dans le cas de $\zeta(s)$ on a

$$\zeta(s) = \frac{e^{As}}{2(s-1)\Gamma(\frac{1}{2}s+1)} \prod_{\varrho} (1 - s/\varrho) e^{s/\varrho}$$

avec $A = \ln(2\pi) - 1 - \frac{1}{2}\gamma$ [1] et où le produit porte sur tous les zéros non triviaux de $\zeta(s)$. Les éventuels zéros multiples[2] sont répétés avec leur ordre de multiplicité, et l'on peut montrer que le produit infini converge absolument. Comme la fonction zêta, à son tour, définit complètement $\psi(x)$, il est raisonnable de rechercher un lien direct entre $\psi(x)$ et la suite des zéros de $\zeta(s)$.

La formule explicite est

$$\psi(x) = x - \lim_{T\to\infty} \sum_{|\Im m\, \varrho| \leqslant T} \frac{x^{\varrho}}{\varrho} - \frac{\zeta'(0)}{\zeta(0)} - \tfrac{1}{2}\ln(1 - x^{-2}),$$

où la somme porte sur les zéros non triviaux ϱ de $\zeta(s)$, avec la même convention que précédemment concernant les racines multiples. La formule n'est valable sous cette forme que si x n'est pas une puissance d'un nombre premier, mais on peut montrer que l'on a sans restriction sur x

$$\psi(x) = x - \sum_{|\Im m\, \varrho| \leqslant T} \frac{x^{\varrho}}{\varrho} - \frac{\zeta'(0)}{\zeta(0)} - \tfrac{1}{2}\ln(1 - x^{-2}) + R_T(x),$$

avec, notant $\langle\langle x \rangle\rangle := \min_{p,\nu \geqslant 1} |x - p^{\nu}|$,

$$R_T(x) \ll x\frac{(\ln xT)^2}{T} + \frac{x \ln x}{x + T\,\langle\langle x \rangle\rangle}.$$

1. Ici γ désigne la constante d'Euler.
2. On conjecture généralement que tous les zéros sont simples.

Au vu des formules explicites, il n'est pas très surprenant que la taille du terme d'erreur du théorème des nombres premiers soit liée au nombre

$$\Theta := \sup_{\zeta(\varrho)=0} \Re e\, \varrho,$$

qui satisfait $\frac{1}{2} \leqslant \Theta \leqslant 1$ à cause de la symétrie des zéros autour de l'axe $\sigma = \frac{1}{2}$, résultant de l'équation fonctionnelle. On a en fait

$$\psi(x) = x + O(x^{\Theta}(\ln x)^2),$$
$$\pi(x) = \mathrm{li}(x) + O(x^{\Theta} \ln x).$$

Réciproquement, on peut montrer que *le nombre* Θ *est la borne inférieure de l'ensemble des nombres réels* ϑ *tels que*

$$\pi(x) - \mathrm{li}(x) = O(x^{\vartheta}).$$

Ainsi l'hypothèse de Riemann équivaut à la majoration asymptotique

$$(\forall\, \varepsilon > 0) \qquad \pi(x) = \mathrm{li}(x) + O_{\varepsilon}(x^{1/2+\varepsilon}),$$

ou encore à l'égalité

$$\Theta = \tfrac{1}{2}.$$

Hardy a établi en 1914 que la fonction $\zeta(s)$ possède une infinité de zéros sur la droite critique $\sigma = \frac{1}{2}$. Toute majoration du type $\Theta \leqslant 1 - \delta$ avec $\delta > 0$ serait une nouvelle fantastique dans le monde de l'arithmétique.

Une énergie considérable a été déployée pour élucider les mystères de la fonction zêta de Riemann. Donnons une idée des progrès les plus significatifs accomplis depuis les travaux de Hadamard et La Vallée-Poussin en mentionnant, parmi beaucoup, deux classes de résultats.

D'une part, à la suite du théorème de Hardy, on a cherché à affiner les minorations quantitatives du nombre des zéros situés sur la droite critique. De tels résultats doivent être mesurés à l'aune de la formule suivante annoncée par Riemann et rigoureusement

établie par von Mangoldt : notant traditionnellement $\varrho = \beta + i\gamma$ un zéro non trivial générique de $\zeta(s)$, on a pour $T \to \infty$,

$$N(T) := |\{\varrho : \zeta(\varrho) = 0,\ 0 < \gamma \leqslant T\}| \sim \frac{T}{2\pi} \ln \frac{T}{2\pi}\cdot$$

Cela étant, Selberg a montré en 1942 que l'on a, pour une constante convenable $c > 0$,

$$N_0(T) := |\{\varrho : \zeta(\varrho) = 0, \beta = \tfrac{1}{2}, 0 < \gamma \leqslant T\}| \geqslant cN(T)$$

pour tout $T > 0$, établissant ainsi *qu'une proportion positive des zéros non triviaux de* $\zeta(s)$ *sont situés sur la droite critique* $\sigma = \frac{1}{2}$. Ce fut un grand choc, assorti d'un certain tapage médiatique, lorsque Levinson annonça en 1974 que $c > \frac{1}{3}$. Le dernier record en date est dû à Conrey (1989) : $c > \frac{2}{5}$.

Dans l'autre direction, la région sans zéro de La Vallée-Poussin a été étendue par Vinogradov et Korobov qui ont montré indépendamment en 1958 que l'on a, pour tout zéro non trivial de $\zeta(s)$,

$$\beta \leqslant 1 - c_0/\{(\ln|\gamma|)^{2/3}(\ln_2|\gamma|)^{1/3}\}$$

pour une constante absolue convenable $c_0 > 0$.[1] La conséquence sur le terme d'erreur est l'estimation

$$\pi(x) = \mathrm{li}(x) + O\big(x\mathrm{e}^{-c_1(\ln x)^{3/5}/(\ln_2 x)^{1/5}}\big),$$

où c_1 est une constante absolue strictement positive.

Toutes les améliorations subséquentes de la méthode de Vinogradov–Korobov, spectaculaires pour certaines applications, ont une signification marginale dans ce problème, où l'on peut les interpréter comme portant essentiellement sur la valeur de la constante positive c_1.

1. Voir note 2, p. 64.

7. Conséquences arithmétiques des renseignements sur les zéros

À côté de l'application phare à la taille du terme d'erreur dans le théorème des nombres premiers, les renseignements sur les zéros sont susceptibles de nombreuses répercussions arithmétiques, dont certaines ne concernent d'ailleurs pas directement les nombres premiers.

Nous nous bornons à mentionner ici deux champs d'applications.

La première concerne les théorèmes dits *d'oscillation*, qui ont pour objet de fournir une limitation théorique à la qualité du terme reste d'une formule asymptotique. Phragmén (1891) et Landau (1905) ont montré que l'existence de singularités de séries de série de Dirichlet permet d'obtenir effectivement des théorèmes de ce type. Il nous entraînerait trop loin de décrire leur méthode, mais nous pouvons facilement en expliciter certaines conséquences dans le cas de la fonction zêta.[1] Nous emploierons la notation

$$f(x) = \Omega_{\pm}(g(x))$$

pour signifier qu'il existe une constante $c > 0$ et une suite réelle $\{x_n\}_{n=1}^{\infty}$ telle que $x_n \to \infty$ et

$$f(x_{2n}) > c\big|g(x_{2n})\big|, \quad f(x_{2n+1}) < -c\big|g(x_{2n+1})\big|.$$

L'existence d'au moins un zéro dans la bande critique $0 < \sigma < 1$ implique grâce au théorème de Phragmén–Landau que

$$\pi(x) = \mathrm{li}(x) + \Omega_{\pm}\Big(\frac{\sqrt{x}}{\ln x}\Big).$$

Littlewood a montré que l'on peut améliorer ce résultat en

$$\pi(x) = \mathrm{li}(x) + \Omega_{\pm}\Big(\frac{\sqrt{x}\ln_3 x}{\ln x}\Big).$$

1. Ce sont donc bien les zéros de $\zeta(s)$ qui apparaissent ici, en tant que singularités de $\log\zeta(s)$, qui « ressemble » fort à $\sum_p p^{-s}$.

Beaucoup d'autres théorèmes d'oscillation sont liés à la localisation des singularités de fonctions analytiques. L'un des plus célèbres est celui de Hardy relatif à la fonction $\tau(n)$, égale au nombre des diviseurs de n. Dirichlet a montré (cf. § 4.3, p. 122) que l'on a

$$\sum_{n\leqslant x} \tau(n) = x(\ln x + 2\gamma - 1) + O(\sqrt{x})$$

et c'est un des grands problèmes ouverts de la théorie analytique des nombres que de déterminer avec précision la taille du terme d'erreur. En considérant le prolongement analytique de la série

$$\sum_{n} \tau(n)\mathrm{e}^{-s\sqrt{n}}/\sqrt{n},$$

Hardy a montré en 1915, par une méthode beaucoup plus sophistiquée qu'une simple application du théorème de Phragmén–Landau mais reposant sur un principe semblable, que l'on a

$$\sum_{n\leqslant x} \tau(n) = x(\ln x + 2\gamma - 1) + \Omega_{\pm}(x^{1/4}).$$

On conjecture que le terme d'erreur du problème de Dirichlet est effectivement $O(x^{1/4+\varepsilon})$ pour chaque $\varepsilon > 0$.

Une seconde classe d'applications arithmétiques de la connaissance des zéros de la fonction $\zeta(s)$ est celle de la localisation d'entiers soumis à certaines contraintes multiplicatives dans de *petits intervalles*.

Considérons le cas exemplaire des nombres premiers. Le théorème des nombres premiers implique immédiatement la forme forte du postulat de Bertrand

$$\pi(x + y) > \pi(x),$$

pour $y = \varepsilon x$, $x > x_0(\varepsilon)$, où ε est un nombre positif arbitrairement petit. Le terme d'erreur de La Vallée-Poussin fournit la validité du résultat pour

$$y = x\mathrm{e}^{-c\sqrt{\ln x}}$$

avec une constante positive convenable c, et l'hypothèse de Riemann autorise le choix $y = C\sqrt{x}(\ln x)^2$ avec C assez grande. Cependant, il est naturel de conjecturer que l'on peut prendre $y = x^{\varepsilon}$ pour tout $\varepsilon > 0$. Un tel résultat ne découle pas directement de l'hypothèse de Riemann et doit donc être attaqué par des méthodes spécifiques. Les formules explicites décrites au paragraphe précédent sont une approche possible et déjà fructueuse.

Nous avons en effet, d'après les résultats indiqués plus haut,

$$\psi(x+y) - \psi(x) = y + \sum_{|\gamma| \leqslant T} \frac{x^{\varrho} - (x+y)^{\varrho}}{\varrho} + O\Big(\frac{x \ln^2 x}{T}\Big)$$

si $1 \leqslant T \leqslant x$ et, par exemple, x est un demi-entier. À ce stade la stratégie est évidente : choisir T assez grand et majorer la somme sur les zéros pour montrer que $\psi(x+y) - \psi(x)$ est proche de y, ou, à tout le moins $\geqslant \frac{1}{2}y$ pour x, y assez grands.

La réalisation de ce programme nécessite deux ingrédients : il faut savoir, d'une part, que les parties réelles β des zéros ne peuvent pas s'approcher trop de la valeur 1, et, d'autre part, il faut pouvoir majorer le nombre des zéros ϱ pour lesquels $1 - \beta$ a une taille « intermédiaire ». La majoration $\beta \leqslant 1 - c/\ln|\gamma|$ de La Vallée-Poussin est à peine trop faible pour remplir la première condition : toute majoration du type

$$\beta \leqslant 1 - c/(\ln|\gamma|)^{1-a}$$

avec $a > 0$ est suffisante. On peut donc en particulier faire appel aux théorèmes de Vinogradov et Korobov. Le second renseignement fait partie des *théorèmes de densité*, dont l'objet est de majorer le nombre

$$N(\sigma, T) = |\{\varrho : \zeta(\varrho) = 0,\ \sigma \leqslant \beta \leqslant 1, 0 < \gamma \leqslant T\}|$$

de zéros dont l'écart à la droite critique est contrôlé. Ingham a montré en 1940 que

$$N(\sigma, T) \ll T^{3(1-\sigma)/(2-\sigma)}(\ln T)^5$$

pour $\frac{1}{2} \leqslant \sigma \leqslant 1$. En améliorant cette estimation pour la bande $\frac{3}{4} \leqslant \sigma \leqslant 1$, Huxley a montré (1972) que *l'intervalle* $]x, x + x^c]$ *contient* $\{1 + o(1)\}x^c / \ln x$ *nombres premiers pour tout* $c > \frac{7}{12}$ *et* $x > x_0(c)$. Il est à noter que ce résultat est bien plus fort que ce que l'on aurait pu déduire de notre connaissance actuelle de la répartition globale des nombres premiers. Des progrès ont été accomplis plus récemment dans cette direction à la suite d'un article d'Iwaniec & Jutila (1979) qui obtinrent, pour tout $c > \frac{5}{9}$ et x assez grand, l'existence d'au moins un nombre premier dans l'intervalle $]x, x + x^c]$ en introduisant dans la technique, à côté des renseignements analytiques concernant la fonction zêta de Riemann, des développements modernes de la théorie du crible linéaire. Le meilleur résultat en date est dû à Baker, Harman & Pintz (2001), qui prouvent que

$$\pi(x + y) - \pi(x) \geqslant 9y/(100 \ln x) \qquad (x > x_0)$$

pour $y = x^{0,525}$.

Chapitre 3
Répartition stochastique des nombres premiers

1. Introduction

Un des aspects remarquables de la répartition des nombres premiers est cette tendance à la régularité globale et à l'irrégularité locale. Les nombres premiers se comportent comme les gaz parfaits chers aux physiciens. Appréhendée d'un point de vue externe, la distribution est — pour ainsi dire — déterministe, mais dès que l'on cherche à décrire la situation en un point donné, on constate des fluctuations statistiques comme dans un jeu de hasard où l'on sait qu'en moyenne les faces équilibreront les piles mais où, à aucun moment, on ne peut prédire le coup suivant. Les nombres premiers occupent tout l'espace (entendez : le hasard) disponible, c'est-à-dire compatible avec la contrainte drastique qui pèse sur eux : engendrer la suite ultra-régulière des nombres entiers.

Cette idée est sous-jacente dans la plupart des conjectures concernant les nombres premiers : tout ce qui n'est pas trivialement interdit est en fait réalisé.

Dans ce chapitre, nous allons énoncer les principaux résultats établis à l'appui de cette « philosophie », et en présenter les méthodes.

2. Une brève histoire de ce qui va suivre

Divisez un nombre impair par 4 : le reste sera évidement 1 ou 3. Comme tous les nombres premiers sauf 2 sont impairs, on obtient ainsi deux classes de nombres premiers, ceux qui ont un reste 1 et ceux qui ont un reste 3 dans la division par 4.

Conformément à notre modèle statistique des nombres premiers, il est logique de supposer que chacune de ces deux classes comporte à peu près autant d'éléments, ce que confirme l'expérience. Il y a, par exemple, 609 nombres premiers n'excédant pas 10000 dont le reste vaut 1, alors que, dans le même intervalle, 620 donnent le reste 3.

Les nombres positifs ou nuls dont le reste dans la division par q vaut r sont exactement les entiers de la forme $r + qm$ $(m = 0, 1, 2, \ldots)$. Ils forment ce que l'on appelle une *progression arithmétique* de premier terme r et de raison q : on passe d'un terme au suivant en ajoutant la raison. Dans la suite, nous convenons de noter $r + q\mathbb{N}$ une telle progression arithmétique.

Cette question de bonne répartition des nombres premiers dans les progressions arithmétiques a intrigué les mathématiciens pendant de longs siècles ; malgré de spectaculaires avancées, elle continue de le faire.

Il est facile, par exemple, de montrer l'infinitude de l'ensemble des nombres premiers dans chacune des progressions arithmétiques $1 + 4\mathbb{N}$ et $3 + 4\mathbb{N}$.[1] Mais même un résultat en apparence aussi peu ambitieux devient délicat lorsque l'on généralise le problème en remplaçant la division par 4 par une autre. Si l'on choisit la raison 11, tous les restes de 1 à 10 sont envisageables, alors que si l'on divise par 10, seuls 1, 3, 7 et 9 persistent. Comment prouver alors qu'il existe une infinité de nombres premiers dont le développement décimal se termine par le chiffre 7 ou dont le reste dans la division par 11 est 4 ?

1. Voir le § 3.

Peter Gustav Lejeune-Dirichlet
(1805–1859)

C'est le mathématicien allemand Dirichlet, grand admirateur de Gauss et fin connaisseur de la langue française, qui a résolu la question vers 1840. Il a pour cela inventé un outil splendide de simplicité et d'efficacité : les *caractères*, qui portent aujourd'hui son nom. Nous détaillons le concept au paragraphe suivant. Pour l'heure, contentons-nous de signaler que les caractères sont des applications définies sur l'ensemble des entiers naturels et à valeurs complexes, qui respectent la structure multiplicative, autrement dit qui vérifient

$$f(mn) = f(m)f(n) \qquad (m \geqslant 1,\ n \geqslant 1),$$

et permettent cependant de repérer les nombres entiers d'une progression arithmétique fixée.

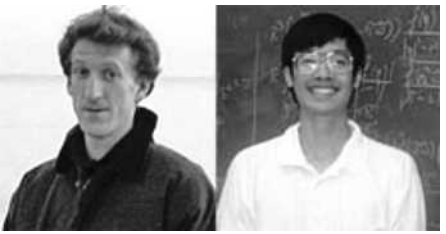
Dirichlet a non seulement établi l'existence d'une infinité de nombres premiers dans toute progression arithmétique admissible, mais il a montré qu'elles contiennent toutes approximativement le même nombre d'éléments. Ce théorème profond vient à l'appui de notre conception de la nature aléatoire des nombres premiers : en l'absence d'objection structurelle, l'équirépartition prévaut.

La question de la bonne répartition des nombres premiers dans les progressions arithmétiques peut être généralisée d'une façon différente. Supposons donné un nombre irrationnel α. Alors tous les nombres αp ($p \in \mathcal{P}$) sont également irrationnels. Si les propriétés statistiques des nombres premiers sont indépendantes de α (et nous n'avons a priori aucune raison de penser le contraire), il est donc logique de conjecturer que les parties fractionnaires[1] $\langle \alpha p \rangle$ sont statistiquement bien réparties dans l'intervalle $[0, 1[$: après

1. La fonction partie fractionnaire est définie p. xx.

tout, un nombre irrationnel n'est jamais qu'une limite de rationnels et, lorsque $\alpha = a/q$, $p = r + mq$, n'a-t-on pas $\langle \alpha p \rangle = \langle ar/q \rangle$?

Nous verrons au § 7 que les nombres premiers se comportent effectivement de manière essentiellement aléatoire au regard de ce critère comme des précédents.



Ben Green et Terence Tao

Une autre version de l'adage selon lequel les nombres premiers empruntent toutes les formes qui ne sont pas trivialement interdites a été rendue effective par un résultat récent dû à Ben Green et Terence Tao (2004) : pour tout entier k, on peut trouver une progression arithmétique de k termes entièrement composée de nombres premiers. La démonstration, dont nous donnons un aperçu sommaire[1] au § 4, est un impressionnant mélange de méthodes issues de la théorie des nombres, la théorie ergodique, l'analyse harmonique, la géométrie discrète et la combinatoire.

Un aspect fascinant du caractère aléatoire de la suite des nombres premiers concerne les grandes et les petites différences entres nombres premiers consécutifs.

Il y a $\pi(N) \sim N/\ln N$ nombres premiers n'excédant pas N, donc l'écart moyen est $\ln N$; mais y a-t-il beaucoup d'écarts $< \frac{1}{2} \ln N$ ou $> 2 \ln N$?

Le mathématicien suédois Harald Cramér a imaginé vers 1936 un modèle statistique reposant sur une hypothèse d'indépendance du caractère de primalité des nombres entiers dans leur ensemble.

1. Inspiré des présentations de Host (2005) et Kra (2005).

Bien sûr, n et $n+1$ ne sont (presque) jamais simultanément premiers, et il en va de même de n et $3n+1$. Cependant, le modèle est pertinent pour beaucoup de questions et suggère par exemple que les plus grands écarts entre nombres premiers de taille N sont de l'ordre de $(\ln N)^2$. Nous énoncerons au § 5 une version plus précise de cette conjecture.

Les meilleurs résultats obtenus à ce jour par les arithméticiens sont encore très éloignés de la conjecture de Cramér.

En ce qui concerne les petits écarts, le modèle statistique prévoit également une répartition harmonieuse : par exemple, le nombre des écarts compris entre $a \ln N$ et $b \ln N$ serait proche de $(\mathrm{e}^{-b} - \mathrm{e}^{-b})\pi(N)$.

Cette dernière conjecture implique en particulier que, pour tout $\varepsilon > 0$, une proportion positive des nombres premiers $p_n \leqslant N$ vérifie $p_{n+1} - p_n \leqslant \varepsilon \ln N$.

Harald Cramér
(1893–1985)

Un tel résultat a longtemps été considéré comme hors de portée. Cependant, Goldston, Pintz et Yıldırım l'ont démontré rigoureusement en 2005. Nous donnerons quelques idées sur leur approche au § 6.

3. Progressions arithmétiques

Soient a et q des nombres entiers premiers entre eux. Il n'y a, *a priori,* aucun obstacle à l'existence d'une infinité de nombres premiers $p \equiv a \pmod q$. Il est même raisonnable, en l'absence d'information contraire, de supputer que les nombres premiers se répartissent équitablement entre les $\varphi(q)$ classes possibles.[1]

1. Voir le Chapitre 1, p. 14, pour la définition de la fonction d'Euler $q \mapsto \varphi(q)$.

Posant

$$\pi(x; a, q) := |\{p \leqslant x : p \equiv a \,(\mathrm{mod}\, q)\}|,$$

cette hypothèse d'équirépartition nous conduit donc naturellement, compte tenu du théorème des nombres premiers, à la conjecture

$$\pi(x; a, q) \sim \frac{x}{\varphi(q) \ln x} \qquad (x \to \infty).$$

Ce problème, qui possède une histoire propre passionnante, parallèle à celle de l'étude de la loi de répartition des nombres premiers, a été finalement résolu en 1896 par La Vallée-Poussin, en combinant sa méthode d'attaque du théorème des nombres premiers (reposant fortement, comme on l'a vu au Chapitre 2, sur les idées de Riemann) avec des outils spécifiques forgés par Dirichlet dans la première moitié du dix-neuvième siècle.

Le premier pas dans cette direction consiste à d'établir rigoureusement l'existence d'une infinité de nombres premiers p dans chaque progression admissible $a \,(\mathrm{mod}\, q)$, avec $(a, q) = 1$.

Considérons par exemple le cas $a = 3$, $q = 4$. On étend sans trop de difficulté la preuve d'Euclide. Supposons, en effet, qu'il n'y ait qu'un nombre fini de nombres premiers, disons $p_1 = 3, \ldots, p_r$, de la forme $4m + 3$. Comme un produit d'entiers de la forme $4m + 1$ est encore de ce type, le nombre $n = 4p_1 \cdots p_r - 1$ possède au moins un facteur premier $p \equiv 3 \,(\mathrm{mod}\, 4)$. On a manifestement $p \neq p_j$ $(1 \leqslant j \leqslant r)$, ce qui fournit la contradiction souhaitée.

Le cas de la progression $p \equiv 1 \,(\mathrm{mod}\, 4)$ est un peu plus délicat mais relève encore d'une technique similaire. S'il n'y a qu'un nombre fini de nombres premiers, disons $p_1 = 5, \ldots, p_r$, de la forme $4m + 1$, alors tout facteur premier p du nombre $n = 4(p_1 \cdots p_r)^2 + 1$ est tel que -1 est résidu quadratique modulo p. D'après ce que nous avons vu au § 1.6 (p. 21), cela

implique $p \equiv 1 \pmod 4$, ce qui est suffisant puisque p ne peut pas faire partie de la liste des p_j.[1]

On pourrait imaginer que ce type d'approche, qui présuppose l'existence d'un polynôme à coefficients entiers dont les valeurs aux arguments entiers sont divisibles par une infinité de nombres premiers de la progression requise, est généralisable à toute progression arithmétique admissible. Il n'en est rien : Ram Murty a établi en 1988 qu'une preuve « euclidienne » de l'infinité de nombres premiers de la forme $a+mq$ est possible si, et seulement si, on a $a^2 \equiv 1 \pmod q$. Il en va ainsi, par exemple, de la progression $4 \pmod{15}$, mais pas de la progression $2 \pmod 7$.

C'est Dirichlet qui, en 1837, a fourni la solution générale de ce problème.[2] Son point de départ était la preuve d'Euler de l'infinitude de l'ensemble des nombres premiers, dont nous avons vu au § 1.7 qu'elle fournissait également la divergence de la série $\sum_p 1/p$.

Considérons d'abord, pour simplifier l'exposé, le cas d'une progression $a \pmod q$ avec q premier et $1 \leqslant a < q$. Dirichlet introduit des fonctions arithmétiques, $\chi : \mathbb{N} \to \mathbb{C}$, désignées sous le nom de *caractères*, qui sont périodiques de période q, complètement multiplicatives,[3] et dont une combinaison linéaire convenable est la fonction indicatrice de la progression $n \equiv a \pmod q$.

La construction des caractères de Dirichlet est subordonnée au fait que la structure de l'ensemble $(\mathbb{Z}/q\mathbb{Z})^*$ est cyclique : il

1. Ainsi l'ensemble $4\mathbb{N}$ contient une infinité de nombles de la forme $p+1$ et une infinité de nombres de la forme $p-1$. Par des techniques différentes et beaucoup plus sophistiquées, Sárközy d'une part, Kamae & Mendès France d'autre part, ont établi en 1978 que la propriété persiste lorsque l'on remplace $4\mathbb{N}$ par un ensemble-différence $D := \{a_j - a_k : j \geqslant 0, k \geqslant 0\}$, où $\{a_n\}_{n=0}^{\infty}$ est une suite strictement croissante d'entiers vérifiant $a_n \ll n$.

2. À strictement parler, la preuve de Dirichlet n'était complète que dans le cas d'une raison q égale à un nombre premier. Le cas général était subordonné à la « formule des classes » que Dirichlet ne devait établir qu'en 1839/40.

3. C'est-à-dire qu'elles vérifient identiquement $\chi(mn) = \chi(m)\chi(n)$ lorsque m, n parcourent $\mathbb{N}$.

existe au moins un élément g, appelé *racine primitive*, dont les puissances g^j décrivent $(\mathbb{Z}/q\mathbb{Z})^*$ tout entier. Étant donnée une racine primitive g modulo q, fixée une fois pour toutes, on peut donc associer à chaque entier n premier à q un unique entier $\varrho = \varrho(n)$ de $\{1, \dots, q-1\}$ tel que $n \equiv g^\varrho \pmod q$. Pour chaque b $(0 \leqslant b \leqslant q-2)$, on pose alors

$$\chi_b(n) = \mathrm{e}^{2\pi i b\varrho(n)/(q-1)},$$

et l'on étend la définition de χ_b en posant $\chi_b(n) = 0$ si $q \mid n$. On construit donc ainsi $\varphi(q) = q-1$ caractères distincts.[1]

Le caractère χ_0, égal à la fonction indicatrice de l'ensemble des entiers premiers à q, joue un rôle particulier. On le désigne sous le nom de *caractère principal*.

La multiplicativité des caractères découle immédiatement du fait que

$$n_1 n_2 \equiv g^{\varrho(n_1)} g^{\varrho(n_2)} \equiv g^{\varrho(n_1)+\varrho(n_2)} \pmod q$$

si $(n_1 n_2, q) = 1$, d'où $\chi(n_1 n_2) = \chi(n_1)\chi(n_2)$ — une propriété qui persiste trivialement si $q \mid n_1 n_2$.

Il est facile de montrer que l'on a pour tout entier a premier à q

$$\sum_{0 \leqslant b \leqslant q-2} \overline{\chi_b(a)}\chi_b(n) = \begin{cases} q-1 & \text{si } n \equiv a \pmod q, \\ 0 & \text{si } n \not\equiv a \pmod q. \end{cases}$$

C'est clair si $n \equiv a \pmod q$; dans le cas contraire, il suffit d'observer que le membre de gauche est la somme des termes d'une progression géométrique. Cette relation, communément désignée *formule d'orthogonalité des caractères*, fournit la combinaison linéaire mentionnée plus haut qui permet de repérer les entiers d'une progression $a \pmod q$.

Cela étant, Dirichlet introduit les séries

$$L(s, \chi) := \sum_{n \geqslant 1} \chi(n) n^{-s}$$

1. Il est clair que c'est le nombre maximal.

associées aux caractères $\chi = \chi_b$ définis plus haut. On montre facilement que ces séries convergent absolument dans le demi-plan $\sigma > 1$, où elles vérifient l'analogue de la formule d'Euler[1]

$$L(s,\chi) = \prod_p (1 - \chi(p)p^{-s})^{-1}.$$

En particulier,

$$L(s,\chi_0) = \prod_{p \nmid q} (1 - p^{-s})^{-1} = (1 - q^{-s})\zeta(s).$$

Par dérivation logarithmique, il suit, toujours lorsque $\sigma > 1$,

$$\begin{aligned} \frac{-L'}{L}(s,\chi) &= \sum_p \frac{\chi(p)\ln p}{p^s - \chi(p)} \\ &= \sum_p \sum_{\nu \geqslant 1} \frac{\chi(p^\nu)\ln p}{p^{\nu s}} = \sum_{n \geqslant 1} \frac{\chi(n)\Lambda(n)}{n^s}. \end{aligned}$$

En multipliant cette relation par $\overline{\chi(a)}$ et en sommant sur $\chi = \chi_b$, nous obtenons la formule fondamentale[2]

$$\sum_{n \equiv a \,(\mathrm{mod}\, q)} \frac{\Lambda(n)}{n^s} = \frac{1}{q-1} \sum_{0 \leqslant b \leqslant q-2} \overline{\chi_b(a)}\, \frac{-L'(s,\chi_b)}{L(s,\chi_b)}\,. \tag{3.1}$$

Lorsque $s \to 1+$ par valeurs réelles, le membre de gauche vaut

$$\sum_{p \equiv a \,(\mathrm{mod}\, q)} \frac{\ln p}{p^s} + O(1),$$

et le travail de Dirichlet a essentiellement consisté à montrer que, dans cette circonstance, le membre de droite tend vers l'infini.

1. Voir le Théorème 3.1 (p. 120) pour une démonstration complète.
2. Pour la commodité du lecteur, rappelons que Λ est la fonction de von Mangoldt — § 1.9, p. 27.

Le terme correspondant à $b = 0$ dans la somme en b de (3.1) vaut

$$-\frac{\zeta'(s)}{\zeta(s)} - \frac{\ln q}{q^s - 1} = \frac{1}{s-1} + O(1),$$

donc il suffit d'établir que chacun des autres termes tend vers une limite finie lorsque $s \to 1+$. Il est commode de remarquer, à ce stade, que, lorsque $b \neq 0$, les séries $L(\sigma, \chi_b)$, absolument convergentes pour $\sigma > 1$, sont convergentes pour $\sigma > 0$. Cela provient du fait que l'application $n \mapsto bQ(n)$, induit trivialement une injection de $(\mathbb{Z}/q\mathbb{Z})^*$ dans lui-même, et donc, puisque l'ensemble est fini, une bijection. On en déduit en particulier que

$$\sum_{1 \leqslant n \leqslant q} \chi_b(n) = \sum_{0 \leqslant r \leqslant q-2} \mathrm{e}^{2\pi i r/(q-1)} = 0 \qquad (1 \leqslant b \leqslant q-2),$$

et, au vu de la décroissance de $n \mapsto n^{-\sigma}$, la convergence de $L(\sigma, \chi_b)$ découle du classique critère d'Abel.[1]

Nous pouvons donc finalement énoncer que *l'hypothèse* $L(1, \chi_b) \neq 0$ $(1 \leqslant b \leqslant q-2)$ *implique que toute progression arithmétique* $a + q\mathbb{N}$ *avec* $(a, q) = 1$ *contient une infinité de nombres premiers.* Plus précisément, il découle de ce qui précède que, si les nombres $L(1, \chi)$ sont non nuls, on a

$$\sum_{p \equiv a \,(\mathrm{mod}\, q)} \frac{\ln p}{p^\sigma} = \frac{1}{(q-1)(\sigma-1)} + O(1) \qquad (\sigma \to 1+).$$

Par intégration en σ sur $[\alpha, 2]$ avec $\alpha > 1$, on en déduit encore que

$$\sum_{p \equiv a \,(\mathrm{mod}\, q)} \frac{1}{p^\alpha} = \frac{1}{q-1} \ln\left(\frac{1}{\alpha - 1}\right) + O(1) \quad (\alpha \to 1+).$$

1. Le critère d'Abel est rappelé au § 4.2, p. 117. Un argument analytique un peu plus sophistiqué permet de montrer en fait la convergence de $L(s, \chi_b)$ dans le demi-plan complexe $\sigma > 0$.

Cette relation implique à son tour la divergence de la série

$$\sum_{p\equiv a \,(\mathrm{mod}\, q)} 1/p.$$

On voit ainsi que l'approche de Dirichlet pour les nombres premiers en progressions arithmétiques est quantitativement de même nature que celle d'Euler pour l'ensemble de tous les nombres premiers.

Il reste à établir que $L(1,\chi_b) \neq 0$ pour $1 \leqslant b \leqslant q-2$.

L'analyse du problème montre rapidement que le cas où χ_b est réel (i.e. $b = \frac{1}{2}(q-1)$) est plus profond et doit être considéré séparément.

Lorsque $b \neq \frac{1}{2}(q-1)$, on a $\chi_b^2 \neq \chi_0$ et l'astuce de Mertens, consistant à utiliser la positivité du polynôme trigonométrique $3+4\cos\vartheta+\cos 2\vartheta$ pour montrer que $\zeta(1+i\tau) \neq 0$, fonctionne encore. En effet, en écrivant $\chi_b(p) = \mathrm{e}^{2\pi i\vartheta_p}$ et en notant que le développement eulérien de $L(\sigma,\chi_b)$ fournit la formule

$$\ln L(\sigma,\chi_b) = \sum_{p,\,\nu\geqslant 1} \chi_b(p)^\nu p^{-\sigma\nu}/\nu \qquad (\sigma > 1),$$

on obtient que $3\ln\zeta(\sigma)+4\Re e\,\ln L(\sigma,\chi_b)+\Re e\,\ln L(\sigma,\chi_b^2) \geqslant 0$, d'où

$$\zeta(\sigma)^3|L(\sigma,\chi_b)|^4|L(\sigma,\chi_b^2)| \geqslant 1.$$

Puisque $\chi_b^2 = \chi_{2b} \neq \chi_0$, la série $L(s,\chi_b^2)$ est convergente en $s=1$. Si $L(1,\chi_b)=0$, on en déduit que le membre de gauche est $O(\zeta(\sigma)^3(\sigma-1)^4) = O(\sigma-1)$ lorsque $\sigma \to 1+$, ce qui contredit l'inégalité pour σ assez proche de 1.

Le cas $b = \frac{1}{2}(q-1)$ est notablement plus délicat. On a alors, avec la notation du symbole de Legendre introduite au § 1.6 (p. 20),

$$\chi_b(n) = \chi(n) = \left(\frac{n}{q}\right).$$

La démonstration originale de Dirichlet est fondée sur le calcul des sommes de Gauss

$$G(n) := \sum_{1 \leqslant m \leqslant q-1} \left(\frac{m}{q}\right) \mathrm{e}^{2\pi i m n/q}.$$

En effectuant le changement de variable consistant à sommer selon les valeurs de $\ell = mn$ modulo q, on obtient en effet

$$G(n) = G(1)\chi(n) \qquad (n \geqslant 1).^{(1)}$$

Cela permet d'exprimer $L(1, \chi)$ comme une combinaison linéaire finie des nombres $\chi(n)$ pour $1 \leqslant n \leqslant q$. Dirichlet obtenait ensuite la non nullité de $L(1, \chi)$ en faisant appel, d'une part, au calcul de $G(1)$ et, d'autre part, à des formules établies par Gauss dans son travail sur les polynômes cyclotomiques.

Nous n'emprunterons pas cette voie historique, et procéderons de manière élémentaire et purement analytique en observant que l'on a

$$f(n) := \sum_{d|n} \chi(d) = \prod_{p^\nu \| n} \sum_{0 \leqslant j \leqslant \nu} \chi(p)^j = \prod_{p^\nu \| n} \frac{1 - \chi(p)^{\nu+1}}{1 - \chi(p)} \geqslant 0,$$

ainsi que l'on peut le déduire du simple fait que $\chi(p)$ ne prend que les valeurs 0 et ± 1. On a de plus $f(n^2) \geqslant 1$ pour tout entier $n \geqslant 1$. Cela implique, pour tout σ avec $0 < \sigma \leqslant 1$,

$$F(\sigma) := \sum_{n \geqslant 1} f(n)\mathrm{e}^{-\sigma n} \geqslant \sum_{1 \leqslant m \leqslant 1/\sqrt{\sigma}} \mathrm{e}^{-1} \geqslant \frac{1}{6\sqrt{\sigma}}.$$

Par ailleurs, revenant à la définition de f, on peut aussi écrire

$$F(\sigma) = \sum_{d \geqslant 1} \chi(d) \sum_{m \geqslant 1} \mathrm{e}^{-\sigma m d} = \sum_{d \geqslant 1} \frac{\chi(d)}{\mathrm{e}^{\sigma d} - 1}.$$

1. La formule n'est initialement valable que lorsque $(n, q) = 1$. Elle persiste dans le cas contraire puisqu'alors $G(n) = 0$.

En invoquant une fois encore la convergence de la série $L(1,\chi)$ et en introduisant la fonction décroissante

$$\vartheta_\sigma(d) := \frac{1}{\sigma d} - \frac{1}{\mathrm{e}^{\sigma d}-1},$$

il suit

$$F(\sigma) = \frac{L(1,\chi)}{\sigma} - \sum_{d\geqslant 1} \vartheta_\sigma(d)\chi(d).$$

Par la règle de convergence d'Abel,[1] la série en d est $\ll \vartheta_\sigma(1) \ll 1$ lorsque $\sigma \to 0+$. On a donc obtenu, pour une constante convenable $A > 1$,

$$L(1,\chi) \geqslant \sigma F(\sigma) - A\sigma \geqslant \tfrac{1}{6}\sqrt{\sigma} - A\sigma.$$

En choisissant convenablement σ (e.g. $\sigma = \frac{1}{144}A^{-2}$), on obtient bien que $L(1,\chi) > 0$.

Cette méthode a été étendue par Dirichlet au cas d'un module q quelconque. Lorsque $q = p^\nu$ et p est un nombre premier impair, le groupe $(\mathbb{Z}/q\mathbb{Z})^*$ est encore cyclique et l'on définit simplement les $\varphi(q)$ caractères modulo q par

$$\chi_b(n) := \mathrm{e}^{2\pi i b\varrho(n)/\varphi(q)} \qquad (0 \leqslant b \leqslant \varphi(q)-1).$$

Le cas où $q = 2^\nu$ est un peu plus compliqué : on pose

$$\chi_{b_1,b_2}(n) = \mathrm{e}^{2\pi i\{b_1\varrho_1(n)/2 + b_2\varrho_2(n)/2^{\nu-2}\}},$$

où $\varrho_1(n)$ et $\varrho_2(n)$ sont définis par la relation

$$n \equiv (-1)^{\varrho_1(n)} 5^{\varrho_2(n)} \pmod{2^\nu},$$

avec $b_1 = 0$ ou 1, et $0 \leqslant b_2 < 2^{\nu-2}$. Dans la situation générique d'un module q qui est le produit de puissances de nombres premiers distincts, on forme simplement les produits des caractères relatifs à

1. Rappelée au § 4.2, p. 117.

ces différentes puissances. L'analyse précédente peut être étendue sans difficulté majeure et conduit à l'estimation quantitative

$$\sum_{\substack{p\leqslant x\\ p\equiv a\,(\mathrm{mod}\,q)}} \frac{1}{p} \sim \frac{\ln_2 x}{\varphi(q)} \qquad (x\to\infty).$$

La méthode de Dirichlet peut sans encombre être insérée dans celle de Riemann, menée à terme par Hadamard et La Vallée-Poussin. L'argument crucial d'absence de zéro sur la droite $\sigma = 1$ doit bien entendu être étendu à toutes les fonctions $L(s,\chi)$, ce qui peut être établi avec des moyens analogues à ceux que nous avons décrits plus haut pour traiter le cas du seul point $s = 1$. Cela conduit naturellement à la formule asymptotique annoncée au début de ce paragraphe pour la fonction $\pi(x;a,q)$.

Les développements plus récents ont surtout visé à préciser l'uniformité en q d'un tel résultat. Siegel a montré en 1936 que, pour chaque $\varepsilon > 0$, il existe une constante $C_\varepsilon > 0$ telle que $L(1,\chi) \geqslant C_\varepsilon/q^\varepsilon$ pour tout caractère χ non principal modulo q. On en déduit le *théorème de Siegel–Walfisz : pour toute constante $A > 0$, il existe une constante $c = c(A) > 0$ telle que l'on ait uniformément pour $x \geqslant 3$, $(a,q) = 1$, $1 \leqslant q \leqslant (\ln x)^A$,*

$$\pi(x;a,q) = \frac{\pi(x)}{\varphi(q)} + O\Big(x\mathrm{e}^{-c\sqrt{\ln x}}\Big) = \frac{\mathrm{li}(x)}{\varphi(q)} + O\Big(x\mathrm{e}^{-c\sqrt{\ln x}}\Big).$$

Malheureusement la constante C_ε du théorème de Siegel n'est pas « effective », autrement dit pas effectivement calculable si on donne à ε une valeur trop petite.[1]

Cela induit une semblable ineffectivité pour la constante c du théorème de Siegel–Walfisz, qui n'est pas calculable si $A \geqslant 1$. Les meilleurs résultats effectifs connus reposent sur la théorie des formes quadratiques réelles et la formule du nombre des

1. En l'état actuel des connaissances, on ne sait pas majorer numériquement C_ε si $\varepsilon < \frac{1}{2}$.

classes, due à Dirichlet. Des estimations de qualité très légèrement inférieure peuvent être obtenues, beaucoup plus simplement, par voie analytique, grâce à l'inégalité de Pólya–Vinogradov (1918), stipulant que l'on a, pour tout caractère χ non principal modulo q,

$$\max_{x\geqslant 1}\left|\sum_{n\leqslant x}\chi(n)\right| \leqslant 2\sqrt{q}\,\ln q.$$

Les deux approches ne fournissent une formule asymptotique pour $\pi(x;q,a)$ que lorsque $q\leqslant(\ln x)^A$ avec $A<2$.

Ces questions d'ineffectivité sont pour le moins insolites. Nous nous proposons maintenant de lever un coin du voile. L'explication repose sur l'existence hypothétique d'un spectre ricanant nommé « zéro de Siegel » qui hante et épouvante cette partie de la théorie analytique des nombres.

L'affaire mérite qu'on s'y attarde un instant. Les régions sans zéro connues pour les fonctions $L(s,\chi)$ sont essentiellement du type

$$\mathcal{D}_q=\{s:\sigma\geqslant 1-c/\ln(3+q|\tau|)\}$$

sauf pour au plus un caractère de module q, disons χ_1, qui est réel et tel que $L(s,\chi_1)$ possède au plus un unique zéro dans $\mathcal{D}_q$, également réel, que nous désignerons par β_1. Lorsque l'on évalue les sommes

$$\psi(x;\chi):=\sum_{n\leqslant x}\chi(n)\Lambda(n)$$

par intégration complexe, l'éventuel zéro de Siegel induit une contribution supplémentaire égale à $-x^{\beta_1}/\beta_1$, de sorte qu'on trouve finalement

$$\sum_{\substack{n\leqslant x\\ n\equiv a\,(\mathrm{mod}\,q)}}\Lambda(n)=\frac{x}{\varphi(q)}-\frac{\chi_1(a)x^{\beta_1}}{\beta_1\varphi(q)}+O\Big(x\mathrm{e}^{-c\sqrt{\ln x}}\Big)$$

uniformément pour, disons, $q\leqslant \mathrm{e}^{\sqrt{\ln x}}$, et où la constante positive c est effectivement calculable. L'estimation (ineffective) de Siegel

$$L(1,\chi_1)\geqslant C_\varepsilon/q^\varepsilon$$

permet, par un simple calcul d'accroissements finis, de « repousser » le zéro au delà du point $1 - C_{2\varepsilon}/q^{2\varepsilon}$ (par exemple), et partant de traiter cette contribution comme un terme d'erreur lorsque $q \leqslant (\ln x)^A$, quitte à choisir $\varepsilon < 1/2A$.

On conjecture bien entendu que le zéro de Siegel n'existe pas et qu'en fait les fonctions $L(s, \chi)$ ont, à l'instar de la fonction $\zeta(s)$, tous leurs zéros non triviaux sur la droite critique $\sigma = \frac{1}{2}$. C'est *l'hypothèse de Riemann généralisée.* Le cas échéant, elle fournirait l'estimation

$$\pi(x; a, q) = \frac{\mathrm{li}(x)}{\varphi(q)} + O\big(\sqrt{x}\,\ln x\big)$$

uniformément pour $x \geqslant 3$, $q \geqslant 1$, $(a, q) = 1$.

Cette évaluation, bien que reposant sur une hypothèse très forte, ne fournit qu'un piètre résultat lorsque $q \geqslant \sqrt{x}$. On peut espérer qu'elle est améliorable en moyenne, voire pour « presque tous » les modules q — en un sens à préciser. Dans cette voie de recherche, le meilleur résultat connu actuellement est le *théorème de Bombieri & Vinogradov* (1965) qui énonce que, pour toute constante $A > 0$, on a uniformément pour $x \geqslant 3$, $Q \geqslant 1$,

$$\sum_{q \leqslant Q} \max_{(a,q)=1} E(x; a, q) \ll \frac{x}{(\ln x)^A} + \sqrt{x}\, Q(\ln Qx)^4,$$

où l'on a posé $E(x; a, q) := \max_{3 \leqslant y \leqslant x} |\pi(y; a, q) - \mathrm{li}(y)/\varphi(q)|$.

Ce résultat permet de « contourner » l'hypothèse de Riemann généralisée dans nombre d'applications. Si l'on pouvait remplacer le terme $Q(\ln Qx)^4$ par $\sqrt{Q}\,(Qx)^{\varepsilon}$,[1] cela fournirait un succédané supérieur à l'original dans bien des circonstances.

1. C'est la conjecture d'Elliott–Halberstam.

4. Le théorème de Green et Tao

Parmi toutes les conjectures relatives aux nombres premiers, l'une des plus simples stipule que cette suite contient des progressions arithmétiques arbitrairement longues. Comme indiqué dans l'avant-propos, la suite de 10 termes

199, 409, 619, 829, 1039, 1249, 1459, 1669, 1879, 2089

est un exemple de progression arithmétique (de raison 210) entièrement composée de nombres premiers.

Un raisonnement probabiliste élémentaire vient à l'appui de cette hypothèse : si un nombre entier de taille N est premier avec probabilité approximativement $1/\ln N$, la probabilité que

$$n,\ n+q,\ \ldots,\ n+(k-1)q$$

soient simultanément premiers est proche de $1/(\ln N)^k$ lorsque, disons, $N < q \leqslant 2N$ et $(n,q) = 1$. Ainsi, le raisonnement statistique suggère l'existence de $\approx N^2/(\ln N)^k$ progressions arithmétiques de longueur k dont tous les termes sont des nombres premiers de taille N.

Le théorème de Green et Tao, qui fournit également une borne effective (très grande) pour la valeur minimale $N = N_k$ en deçà de laquelle nous sommes certains de trouver k nombres premiers en progression arithmétique, répond à une question ancienne, pour laquelle très peu de résultats partiels étaient connus : seul un théorème de van der Corput (1939) correspondant à $k = 3$ était précédemment disponible dans cette direction.

Deux autres conjectures toujours ouvertes impliquent le résultat : celle des nombres premiers jumeaux généralisée de Hardy et Littlewood (cf. chap. 5, p. 148) et celle d'Erdős et Turán (1936) selon laquelle tout ensemble d'entiers dont la série des inverses diverge contient des progressions arithmétiques arbitrairement longues.

Étant donnée une fonction $f : (\mathbb{N}^*)^r \to \mathbb{C}$, notons

$$E_N(f) := \frac{1}{N^r} \sum_{1 \leqslant n_1, \ldots, n_r \leqslant N} f(n_1, \ldots, n_r)$$

la moyenne de ses valeurs sur les N^r premiers r-uples de nombres entiers. Posons encore

$$f_k(m,n) := \prod_{0 \leqslant j < k} f(n+mj).$$

Un célèbre théorème de Szemerédi (1975) s'énonce, dans sa version finie, sous la forme suivante : *si $\delta > 0$ et $k \in \mathbb{N}^*$ sont donnés, si N est un nombre entier assez grand, et si $f = \mathbf{1}_A$ est la fonction indicatrice d'un ensemble d'entiers tel que $E_N(f) > \delta$, alors $E_N(f_k) > c(\delta,k)$ pour une constante convenable $c(\delta,k) > 0$, ne dépendant ni de f, ni de N.*[1]

Cela signifie en particulier qu'un ensemble d'entiers assez dense contient nécessairement des progressions arithmétiques arbitrairement longues.

Le premier travail de Green et Tao a consisté à étendre le champ de cet énoncé à des fonctions f plus générales. Une étape significative avait préalablement été franchie par Gowers (2001), grâce à une approche reposant sur la démonstration ergodique[2] de Furstenberg (1982) du théorème de Szemerédi : le résultat reste valable pour toute fonction f à valeurs dans $[0,1]$.

Green et Tao prouvent ensuite que l'énoncé de type Szemerédi persiste si, pour chaque N, on a

$$0 \leqslant f(n) \leqslant \nu_N(n) \qquad (1 \leqslant n \leqslant N)$$

où ν_N est une fonction de poids dite «pseudo-aléatoire» en ce sens qu'elle vérifie

$$E_N(\nu_N) = 1 + o(1) \qquad (N \to \infty)$$

et d'autres conditions de bon comportement statistique.[3]

1. Il est à noter que $E_N(f)$ est évalué avec $r = 1$, alors que l'on choisit évidement $r = 2$ pour calculer $E_N(f_k)$.

2. La théorie ergodique est une théorie mathématique née dans les années 1930 et reposant sur une hypothèse fondamentale de nature statistique.

3. Ces conditions portent sur les produits de formes linéaires et sur les corrélations. Donner les définitions précises nous entraînerait largement hors des limites de cet opuscule.

C'est la démonstration de cette généralisation qui occupe la plus grande partie de l'article de Green et Tao.

Les deux auteurs établissent ensuite l'existence d'un poids pseudo-aléatoire et d'une fonction f satisfaisant aux conditions précédentes et tels que f soit nulle hors de l'ensemble $\mathcal{P}$ des nombres premiers.

Cette seconde partie, qui participe d'une approche plus classique de la théorie analytique des nombres, repose sur des techniques de crible voisines de celles qui sont exposées au § 6 *infra*. Nous n'en dirons pas plus.

5. Le modèle de Cramér

S'il est souvent difficile de répondre aux questions les plus naturelles concernant les nombres premiers, il est parfois malaisé de se faire même une idée de ce que devrait être la réponse exacte. Ayant étudié la répartition des nombres premiers depuis les années 1920, Cramér a proposé à la fin des années 1930 une méthode simple et fascinante pour forger des conjectures sur les nombres premiers : *la suite des nombres premiers se comporte comme une suite aléatoire soumise à la même contrainte de croissance.* Formalisons un peu. Le théorème des nombres premiers nous indique que la probabilité pour qu'un entier de taille n soit premier est proche de $1/\ln n$.[1] Désignant par $\{X_n\}_{n=2}^{\infty}$ une suite de variables aléatoires indépendantes prenant les valeurs 0 et 1 avec

$$\mathbb{P}(X_n = 1) = 1/\ln n \qquad (n \geqslant 3),^{(2)}$$

la suite aléatoire

$$S := \{n \geqslant 2 : X_n = 1\} \tag{3.2}$$

1. La quantité $1/\ln n$ apparaît naturellement ici comme la dérivée en $x = n$ de $\mathrm{li}(x)$.

2. On peut poser, par exemple, $\mathbb{P}(X_2 = 1) = 1$ pour fixer les idées, cette valeur n'ayant en fait pas d'influence notable. Ici et dans la suite, nous utilisons la lettre $\mathbb{P}$ pour désigner la probabilité.

constitue donc pour Cramér un modèle stochastique de la suite des nombres premiers.

En termes plus intuitifs, on peut considérer le dispositif suivant. On dispose d'une suite infinie (ou d'un très grand nombre) d'urnes numérotées, disons $U_3,\ U_4$, etc. L'urne U_n contient une unique boule blanche et environ $\ln n$ boules noires. On tire au hasard une boule de chaque urne et l'on affecte l'entier n à la suite S si la boule issue de U_n est blanche. La suite S modélise les nombres premiers. Bien entendu, pour chaque tirage particulier la suite S possédera des propriétés spécifiques qui la différenciera considérablement de la suite $\mathcal{P}$ de tous les nombres premiers. Cependant, si une propriété s'avère suffisamment insistante pour être réalisée *presque sûrement* lorsque l'on multiplie les tirages, on conjecturera avec Cramér qu'elle est également partagée par $\mathcal{P}$.

Étant donné un ensemble d'entiers A, désignons par $\pi_A(x)$ le cardinal de $A \cap [1, x]$. Lorsque S est défini par (3.2), $\pi_S(x)$ est donc la variable aléatoire égale au nombre des entiers n n'excédant pas x et tels que $X_n = 1$. Ainsi $\pi_S(x)$ est un modèle aléatoire de $\pi(x) = \pi_{\mathcal{P}}(x)$, et il est facile de montrer que l'on a presque sûrement (i.e. avec probabilité 1)

$$\pi_S(x) = \mathrm{li}(x) + \vartheta_S(x)\sqrt{2x\frac{\ln_2 x}{\ln x}}$$

où $\vartheta_S(x)$ oscille asymptotiquement entre -1 et 1.

C'est une première confirmation, éclatante, de la pertinence du modèle de Cramér : Littlewood a montré[1] que l'on a

$$\vartheta_{\mathcal{P}}(x) = \Omega_{\pm}\big(\ln_3 x/\sqrt{\ln x \ln_2 x}\big)$$

inconditionnellement, et l'hypothèse de Riemann implique

$$\vartheta_{\mathcal{P}}(x) = O\big((\ln x)^{3/2}/(\ln_2 x)^{1/2}\big).$$

Une autre conséquence du modèle de Cramér concerne la différence entre les nombres premiers consécutifs. Posons

1. Voir le § 2.7, p. 69.

$d_n := p_{n+1} - p_n$, où p_n désigne le n-ième nombre premier. On remarque incidemment que la suite $\{d_n\}_{n=1}^{\infty}$ est non bornée puisque les nombres $k! + j$ $(2 \leqslant j \leqslant k)$ sont tous composés. On a

$$\sum_{n<N} d_n = p_N - 2 \sim N \ln N \quad (N \to \infty),$$

d'après le théorème des nombres premiers, donc

$$\liminf_{n\to\infty} d_n / \ln n \leqslant 1 \leqslant \limsup_{n\to\infty} d_n / \ln n.$$

Maintenant, puisque la probabilité pour qu'un entier m de $]p_n, 2p_n]$ soit premier est proche de $1/\ln p_n$, la probabilité $P_{n,k}$ de l'événement $d_n \geqslant k$ satisfait, sous réserve d'indépendance, à

$$P_{n,k} \approx (1 - 1/\ln p_n)^k \approx \mathrm{e}^{-k/\ln n}.$$

Lorsqu'on choisit $k = k(n) = \delta(\ln n)^2$, on obtient donc $P_{n,k(n)} \approx n^{-\delta}$, de sorte que la série $\sum_n P_{n,k(n)}$ diverge pour $\delta < 1$ et converge pour $\delta > 1$. Un résultat classique de la théorie des probabilités, le *lemme de Borel–Cantelli,* permet d'en déduire que les nombres premiers aléatoires du modèle de Cramér satisfont à

$$\limsup_{n\to\infty} d_n/(\ln p_n)^2 = 1.$$

L'assertion légèrement plus faible

$$d_n \ll (\ln p_n)^2$$

est célèbre sous le nom de *conjecture de Cramér.*

On peut encore utiliser le modèle de Cramér pour prévoir le comportement de la fonction $\pi(x)$ dans les petits intervalles.

En utilisant le fait bien connu que la loi binomiale converge vers celle de Poisson,[1] on obtient facilement que, posant $y = \lambda \ln x$,

1. Si les X_n $(1 \leqslant n \leqslant N)$ sont des variables aléatoires de Bernoulli, indépendantes et de même loi $\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0) = \lambda/N$, et si l'on pose $S_N := \sum_{1\leqslant n\leqslant N} X_n$, alors on a, lorsque $N \to \infty$ et λ reste fixé, $\lim \mathbb{P}(S_N = k) = \mathrm{e}^{-\lambda}\lambda^k/k!$.

on a avec probabilité 1

$$|\{x \leqslant \xi : \pi_S(x+y) - \pi_S(x) = k\}| \sim \xi e^{-\lambda} \frac{\lambda^k}{k!} \qquad (\xi \to \infty).$$

Gallagher a montré en 1966 que la relation précédente est effectivement réalisée pour $S = \mathcal{P}$ sous réserve de la validité d'une forme forte de la conjecture des nombres premiers jumeaux généralisée.[1]

Ainsi le modèle de Cramér semble être en cohérence à la fois avec l'hypothèse de Riemann (cas des intervalles longs) et avec la conjecture des nombres premiers jumeaux (cas des intervalles très courts : $y = \lambda \ln x$). Il ne faut cependant pas trop demander : le modèle de Cramér prévoit que p et $p+2$ (voire p et $p+1$!) sont simultanément premiers avec probabilité $1/(\ln x)^2$, et n'intègre donc pas la conjecture de Hardy & Littlewood mentionnée au Chapitre 1 (§ 1.11, p. 41).

Il n'est pas difficile de voir que le modèle de Cramér implique que l'on a presque sûrement

$$\pi_S(x+y) - \pi_S(x) = \mathrm{li}(x+y) - \mathrm{li}(x) + O(\sqrt{y}) \sim y/\ln x$$

dès que $y/(\ln x)^2 \to \infty$, avec $y \leqslant x$. En 1943, Selberg a partiellement confirmé la validité du modèle sur ce point en montrant, sous l'hypothèse de Riemann, qu'avec de telles valeurs de y on a, pour presque tous les entiers x,

$$\pi(x+y) - \pi(x) \sim y/\ln x.$$

Cependant, à la surprise générale, Maier a prouvé en 1985 que si $y = (\ln x)^\alpha$ avec $\alpha > 2$, cette relation *n'est pas satisfaite lorsque* $x \to \infty$. Il a en fait établi l'existence d'une constante $\delta = \delta(\alpha) > 0$ telle que chacune des inégalités

$$\pi(x+y) - \pi(x) \gtrless (1 \pm \delta)\frac{y}{\ln x} \quad \big(y = (\ln x)^\alpha\big),$$

soit vérifiée pour une infinité de valeurs entières de x.

1. Voir le Chapitre 5.

Ce résultat, qui représente la première réfutation du modèle de Cramér, mérite quelques développements. Il est fondé sur la disparité, déjà constatée au Chapitre 1, entre le terme principal du crible d'Ératosthène et la taille réelle de la fonction $\pi(x)$.

Entrons un peu dans les détails. Le langage des probabilités, et notamment celui des probabilités conditionnelles[1] sera particulièrement bien adapté à notre propos. Rappelons la notation $P^-(n)$ pour le plus petit facteur premier d'un entier n, avec la convention $P^-(1) = \infty$. Nous avons vu au Chapitre 1, comme conséquence du crible de Brun, que l'on a (cf. § 1.11, p. 39)

$$\Phi(x,z) := \sum_{n\leqslant x,\, P^-(n)>z} 1 \sim x\prod_{p\leqslant z}(1-1/p)$$

si $x \to \infty$ et $z = x^{o(1)}$. Lorsque $z \to \infty$ sous cette dernière condition, on peut donc énoncer que l'on a, pour un entier n de taille x,

$$\begin{aligned}\mathbb{P}\big(n\in\mathcal{P} \mid P^-(n)>z\big) &\sim \frac{\mathbb{P}(n\in\mathcal{P})}{\mathbb{P}(P^-(n)>z)}\\ &\sim \frac{1/\ln x}{\prod_{p\leqslant z}(1-1/p)} \sim \frac{\mathrm{e}^\gamma \ln z}{\ln x},\end{aligned}$$

où la dernière équivalence résulte de la formule de Mertens (§ 1.10, p. 33).

En sommant cette relation sur les $\Phi(x+y,z)-\Phi(x,z)$ entiers $n\in]x,x+y]$ tels que $P^-(n)>z$, il est donc naturel de supputer que

$$\pi(x+y)-\pi(x) \sim \frac{\mathrm{e}^\gamma \ln z}{\ln x}\{\Phi(x+y,z)-\Phi(x,z)\}.$$

Cela étant, posons $M := \prod_{p\leqslant z} p$. Comme la condition $P^-(n) > z$ est déterminée par la classe de congruence de n

1. Nous noterons $\mathbb{P}(A \mid B)$ la probabilité de l'événement A sachant que B est réalisé.

modulo M, on a, si $x \equiv 0 \pmod{M}$,

$$\Phi(x+y,z) - \Phi(x,z) = \Phi(y,z).$$

À ce stade, Maier utilise un résultat asymptotique classique[1] concernant la fonction $\Phi(y,z)$ et qui précise le « changement de phase » entre le comportement de type crible (où les différentes conditions de divisibilité modulo les nombres premiers $p \leqslant z$ sont asymptotiquement indépendantes) et le comportement de type $\pi(x)$ (où un facteur correctif $\frac{1}{2}e^{\gamma}$ doit être introduit pour tenir compte de la dépendance) : on a, uniformément pour $y^{\varepsilon} \leqslant z \leqslant \sqrt{y}$, $y = z^u$,

$$\Phi(y,z) \sim \frac{y\omega(u)}{\ln z},$$

où ω désigne la *fonction de Buchstab*, définie comme l'unique solution continue sur $[2,\infty[$ de l'équation différentielle aux différences

$$(u\omega(u))' = \omega(u-1) \qquad (u > 2)$$

avec la condition initiale $u\omega(u) = 1$ $(1 \leqslant u \leqslant 2)$.

On a clairement $\lim_{u\to\infty} \omega(u) = e^{-\gamma}$, puisque la formule précédente doit recouvrir celle du crible de Brun lorsque z est « petit ». Il s'avère, ainsi que l'a établi Iwaniec, que $\omega(u)$ oscille indéfiniment autour de sa limite : en fait $\omega(u) - e^{-\gamma}$ change de signe au moins une fois dans chaque intervalle de longueur 1.

En rassemblant nos estimations, nous obtenons donc la conjecture suivante, dont les raisonnements heuristiques qui précèdent constituent en fait une ébauche de démonstration,

$$\pi(x+y) - \pi(x) \sim e^{\gamma}\omega(u)\frac{y}{\ln x} \quad \big(x \equiv 0 \pmod{M},\ y = z^u\big).$$

Il reste, formellement, à choisir $z = \ln x$ et u égal à un extremum de ω pour obtenir le théorème de Maier.

1. Dû au mathématicien soviétique A.A. Buchstab, dans les années trente.

Pour établir rigoureusement son résultat, Maier introduit un élégant argument de moyennes croisées qu'il nous entraînerait trop loin de décrire dans cet opuscule. Nous nous contentons d'indiquer qu'il repose de manière cruciale sur la régularité de la répartition des nombres premiers dans les progressions arithmétiques.[1]

Le théorème de Maier, qui a été étendu de diverses manières,[2] nous incite à amender le modèle de Cramér, en remplaçant l'heuristique $\mathbb{P}(n \in \mathcal{P}) = 1/\ln n$ par

$$\mathbb{P}(n \in \mathcal{P} \mid P^-(n) > z) = \frac{1}{\ln n} \prod_{p \leqslant z} (1 - 1/p)^{-1} \quad (z \approx \ln n).$$

Un tel modèle modifié se trouve ainsi en cohérence à la fois avec la conjecture des nombres premiers jumeaux généralisée et l'hypothèse de Riemann. Il laisse supposer que la conjecture de Cramér forte

$$\limsup_{n \to \infty} d_n/(\ln p_n)^2 = 1$$

est fausse et doit être remplacée par

$$\limsup_{n \to \infty} d_n/(\ln p_n)^2 = 2\mathrm{e}^{-\gamma}.$$

Les meilleurs résultats qualitatifs connus concernant les variations de d_n sont celui de Rankin (1938),

$$\limsup_{n \to \infty} \frac{d_n}{\ln p_n \ln_2 p_n \ln_4 p_n/(\ln_3 p_n)^2} > 0,$$

et celui de Goldston, Pintz et Yıldırım (2005),

$$\liminf_{n \to \infty} d_n/\ln p_n = 0,$$

améliorant le résultat historique d'Erdős (1940), stipulant que la limite inférieure ci-dessus est < 1.

1. Voir le § 3 *supra*.

2. Notamment, par Friedlander, Granville, Hildebrand et Maier, pour fixer des limites inattendues aux domaines de validité des théorèmes de type Siegel–Walfisz et Bombieri–Vinogradov.

6. Le théorème de Goldston, Pintz et Yıldırım

Nous nous proposons ici de revenir sur la démonstration du résultat de Goldston, Pintz et Yıldırım concernant les petits écarts entre nombres premiers. Comme annoncé au paragraphe précédent, ces trois auteurs ont établi en 2005 que, pour tout $\varepsilon > 0$, l'inégalité

$$d_n < \varepsilon \ln n \tag{3.3}$$

a lieu pour une infinité d'entiers n. Cette avancée majeure dans notre compréhension des nombres premiers a ensuite été raffinée : dès 2007, les trois coauteurs ont établi la relation

$$\liminf_{n\to\infty} \frac{d_n}{\sqrt{\ln n}(\ln_2 n)^2} < \infty.$$

Nous allons à présent décrire le principe de la preuve de la relation (3.3), sans toutefois entrer dans les détails calculatoires.

Pour $N \in \mathbb{N}$, posons $I_N :=]N, 2N]$. L'idée initiale consiste à chercher une fonction arithmétique positive ou nulle F telle que

$$S_N := \sum_{n\in I_N} R(n)F(n) > Q_N := \sum_{n\in I_N} F(n) \quad (N \to \infty) \tag{3.4}$$

avec

$$R(n) := \sum_{1\leqslant h\leqslant H} \mathbf{1}_{\mathcal{P}}(n+h).$$

Dan Goldston, János Pintz et Cem Yıldırım (2005)

En effet, (3.4) implique l'existence d'un $p \in]N, 2N + H]$ tel que $p^{\#} - p \leqslant H$, où $p^{\#}$ désigne le nombre premier suivant immédiatement p.

Il est tentant de choisir $F(n) := R(n)$, ou $F(n) := \mathbf{1}_{\{R(n)\geqslant k\}}(n)$ pour un entier k bien choisi. Cependant, dans le premier cas, l'évaluation du membre de gauche de (3.4) reviendrait à résoudre, au moins en moyenne, la conjecture des nombres premiers jumeaux généralisée, alors que, dans le second cas, on a affaire à une sous-somme de la précédente — un problème a priori plus difficile.

Il est donc nécessaire de rechercher une fonction F moins restrictive. Pour atteindre notre but, il est néanmoins naturel d'imposer

$$F(n) \geqslant \mathbf{1}_{\{R(n)\geqslant k\}}(n),$$

par exemple sous la forme

$$F(n) \geqslant g(n;\mathcal{H}) := \prod_{1\leqslant j\leqslant k} \mathbf{1}_{\mathcal{P}}(n + h_j) \tag{3.5}$$

lorsque $\mathcal{H} := \{h_1, \dots, h_k\}$ est un sous-ensemble de $[1, H] \cap \mathbb{N}$.[1]

Or, le crible de Selberg, dont la première version date de 1947, fournit une méthode pour minimiser la quantité Q_N lorsque F appartient à une certaine classe et sous la contrainte (3.5). L'idée de Selberg, très simple et très efficace, consiste à chercher F sous la forme

$$F(n;\mathcal{H}) := \Big(\sum_{d|P(n;\mathcal{H})} \lambda_d \Big)^2$$

où l'on a posé $P(n;\mathcal{H}) := \prod_{h\in\mathcal{H}}(n + h) = \prod_{1\leqslant j\leqslant k}(n + h_j)$, et où la fonction $d \mapsto \lambda_d$ vérifie

$$\lambda_1 = 1, \quad \lambda_d = 0 \quad (d > R).$$

1. Nous verrons qu'il s'avérera en fait nécessaire d'introduire plusieurs tels ensembles $\mathcal{H}$ afin de tirer avantage d'un effet de moyenne.

Si $R \leqslant N$, on voit que seul $d = 1$ intervient dans la somme lorsque $g(n;\mathcal{H}) = 1$, d'où (3.5), puisque $F(n;\mathcal{H}) \geqslant 0$ pour tout entier n.

Avec ces notations, $Q_N = Q_N(\mathcal{H})$ est une forme quadratique des variables λ_d $(1 \leqslant d \leqslant R)$. Le calcul de minimisation effectué par Selberg montre que le choix

$$\lambda_d := \mu(d)\left(\frac{\ln^+ R/d}{\ln R}\right)^k, \tag{3.6}$$

où l'on a posé $\ln^+ x = \max(0, \ln x)$, est proche de l'optimum.

On obtient ainsi, grâce à un calcul classique faisant intervenir les sommes de fonctions multiplicatives,

$$Q_N(\mathcal{H}) \sim \mathfrak{S}(\mathcal{H})\frac{k!N}{(\ln R)^k}, \tag{3.7}$$

avec la notation traditionnelle

$$\mathfrak{S}(\mathcal{H}) := \prod_p \Big(1 - \frac{\nu_{\mathcal{H}}(p)}{p}\Big)\Big(1 - \frac{1}{p}\Big)^{-k} \tag{3.8}$$

où $\nu_{\mathcal{H}}(p)$ désigne le nombre de résidus distincts occupés par les éléments de $\mathcal{H}$ modulo p. On note que $\nu_{\mathcal{H}}(p) = k$ dès que $p > H$, ce qui assure la convergence du produit.

Il reste à évaluer $S_N = S_N(\mathcal{H})$. En ignorant les termes d'erreur issus du théorème des nombres premiers en progressions arithmétiques, nous pouvons écrire

$$\begin{aligned} T_N(\mathcal{H}) &:= \sum_{n\in I_N} \mathbf{1}_{\mathcal{P}}(n+h)F(n;\mathcal{H}) \\ &\approx \sum_{1\leqslant h\leqslant H}\ \sum_{1\leqslant d,t\leqslant R} \lambda_d\lambda_t \frac{\varrho([d,t];h,\mathcal{H})N}{\varphi([d,t])\ln N} \end{aligned}$$

où $\varrho(m;h,\mathcal{H})$ désigne le nombre de classes résiduelles a modulo m telles que $(a+h,m) = 1$ (pour assurer la compatibilité de la

condition $n+h \in \mathcal{P}$) et $-a \in \mathcal{H} \pmod m$, et où le facteur $\ln N$ du dénominateur prend en compte le fait que la sommation porte sur des nombres premiers translatés ($n+h \in \mathcal{P}$). On a donc

$$\varrho(p;h,\mathcal{H}) = \nu_{\mathcal{H}\cup\{h\}}(p) - 1 \qquad (p \leqslant R)$$

et

$$\varrho(m;h,\mathcal{H}) = \prod_{p\mid m} \varrho(p;h,\mathcal{H})$$

lorsque m est un entier sans facteur carré.

Cette technique prometteuse doit en fait être légèrement modifiée pour conduire à la conclusion souhaitée $S_N > Q_N$. En effet, le choix (3.6) pour les λ_d minimise bien $Q_N(\mathcal{H})$ et donc aussi $\sum_{\mathcal{H}} Q_N(\mathcal{H})$, mais notre problème consiste en réalité à maximiser le rapport

$$\Big(\sum_{\mathcal{H}} S_N(\mathcal{H})\Big) \Big/ \Big(\sum_{\mathcal{H}} Q_N(\mathcal{H})\Big).$$

Le remède est cependant à portée de main. En posant

$$\lambda_d := \mu(d)\Big(\frac{\ln^+ R/d}{\ln R}\Big)^{k+r}$$

où r est un paramètre entier positif, et en utilisant le résultat de Gallagher (1976)

$$\sum_{\substack{\mathcal{H}\subset[1,H]\\ |\mathcal{H}|=k}} \mathfrak{S}(\mathcal{H}) \sim \binom{H}{k},$$

nous obtenons d'une part, par un raisonnement analogue à celui qui conduit à (3.7),

$$\sum_{\substack{\mathcal{H}\subset[1,H]\\ |\mathcal{H}|=k}} Q_N(\mathcal{H}) \sim \frac{H^k(k+2r)!N}{k!(\ln R)^{k+2r}}$$

et d'autre part

$$\sum_{\substack{\mathcal{H}\subset[1,H]\\|\mathcal{H}|=k}} S_N(\mathcal{H}) \sim \frac{H^k(k+2r)!N}{k!(\ln R)^{k+2r}\ln N}\Big(H+\frac{2k(2r+1)\ln R}{(r+1)(k+2r+1)}\Big).$$

De plus, ces évaluations sont uniformes, pour tout k fixé, dès que $H \ll \ln N$ et $R \leqslant N^{1/4-\varepsilon}$, cette dernière restriction provenant essentiellement des conditions de validité du théorème de Bombieri–Vinogradov énoncé p. 88.

En choisissant $H := \eta \ln N$, où η est un paramètre positif arbitrairement petit, ε assez petit, $R := \lfloor N^{1/4-\varepsilon} \rfloor$, k assez grand et $r := \lfloor \sqrt{k} \rfloor$, nous obtenons bien

$$\sum_{\substack{\mathcal{H}\subset[1,H]\\|\mathcal{H}|=k}} S_N(\mathcal{H}) > \sum_{\substack{\mathcal{H}\subset[1,H]\\|\mathcal{H}|=k}} Q_N(\mathcal{H})$$

pour N assez grand.

7. Équirépartition modulo un

L'étude de l'adéquation de la répartition des nombres premiers avec le modèle stochastique de Cramér peut être poursuivie et généralisée en analysant la distribution des nombres premiers relativement à d'autres types de mesures. Le domaine de *l'équirépartition modulo un*, qui constitue une branche à part entière de la théorie des nombres, fournit à cet égard un vaste champ d'investigations.

Une suite $\{u_n\}_{n=1}^{\infty}$ à valeurs dans $[0,1[$[(1)] est dite *équirépartie modulo 1* si chaque sous-intervalle de $[0,1[$ reçoit asymptotiquement une quote-part équitable des valeurs de u_n, en d'autres termes si l'on a pour tous α, β avec $0 \leqslant \alpha < \beta < 1$,

$$\lim_{N\to\infty} (1/N)|\{n \leqslant N : \alpha < u_n \leqslant \beta\}| = \beta - \alpha.$$

(Une telle suite est donc dense dans $[0,1[$.) Plus généralement, on dit qu'une suite $\{u_n\}_{n=1}^{\infty}$ à valeurs réelles est équirépartie modulo 1

1. Il serait plus exact de parler ici du tore $\mathbb{R}/\mathbb{Z}$, mais nous n'aurons pas l'usage de cette subtilité conceptuelle.

si la suite $\langle u_1\rangle, \langle u_2\rangle, \ldots$ de ses parties fractionnaires satisfait à la condition précédente.

Le résultat de base de la théorie est le critère d'équirépartition d'Hermann Weyl (1916) qui énonce que la suite réelle $\{u_n\}_{n=1}^{\infty}$ est équirépartie modulo 1 si, et seulement si, l'une des deux conditions équivalentes suivantes est réalisée :

$$\text{(i)}\quad \lim_{N\to\infty}\frac{1}{N}\sum_{n\leqslant N} f(\langle u_n\rangle)=\int_0^1 f(x)\,\mathrm{d}x \qquad (\forall f\in\mathcal{R}[0,1[),$$

$$\text{(ii)}\quad \sum_{n\leqslant N} \mathrm{e}^{2\pi i\nu u_n} = o(N) \quad (N\to\infty,\ \nu=1,2,\ldots),$$

où, dans (i), nous désignons par $\mathcal{R}[0,1[$ l'espace des fonctions intégrables au sens de Riemann sur $[0,1[$.

On peut donner rapidement une idée de la démonstration en observant d'abord que la définition de l'équirépartition modulo 1 correspond à la condition (i) pour les fonctions indicatrices d'intervalles, dont les combinaisons linéaires permettent d'approcher convenablement les éléments de $\mathcal{R}[0,1[$. Ensuite, la condition (ii), qui est clairement un cas particulier de (i), contient en fait le cas général puisque les combinaisons linéaires d'exponentielles $x\mapsto \mathrm{e}^{2\pi i\nu x}$ $(\nu\in\mathbb{Z})$ permettent d'approcher les fonctions périodiques continues, et donc aussi les fonctions Riemann-intégrables.

Il est immédiat de déduire du critère de Weyl que la suite $n\mapsto\langle n\alpha\rangle$ est équirépartie modulo 1 pour tout nombre α irrationnel : les sommes d'exponentielles de (ii), qu'on peut calculer explicitement puisqu'elles correspondent à des progressions géométriques, sont bornées.

Considérons la suite $\{\ln n\}_{n=1}^{\infty}$. La somme trigonométrique associée s'écrit

$$\begin{aligned} S_N &= \sum_{n\leqslant N} n^{2\pi i\nu} = N^{2\pi i\nu}\sum_{n\leqslant N}\left(\frac{n}{N}\right)^{2\pi i\nu} \\ &\sim N^{1+2\pi i\nu}\int_0^1 x^{2\pi i\nu}\,\mathrm{d}x = \frac{N^{1+2\pi i\nu}}{1+2\pi i\nu} \qquad (N\to\infty), \end{aligned}$$

d'après la théorie de l'intégrale de Riemann. Ainsi S_N/N ne tend pas vers 0, et la suite $\{\ln n\}_{n=1}^{\infty}$ n'est donc pas équirépartie modulo 1. Nous observons cependant qu'elle est dense modulo 1 puisque $\ln n \to \infty$ et $\ln(n+1) - \ln n \to 0$ lorsque $n \to \infty$.

On peut montrer, grâce à une méthode générale d'estimation de sommes trigonométriques mise au point par van der Corput dans les années vingt, que la suite $\{\alpha n \ln n\}_{n=1}^{\infty}$ est équirépartie modulo 1 pour tout $\alpha \neq 0$. Cela suggère, au vu du théorème des nombres premiers, qu'il en va de même de la suite $\{\alpha p_n\}_{n=1}^{\infty}$ lorsque $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Le critère de Weyl nous enseigne que le comportement des sommes d'exponentielles

$$\sum_{n \leqslant N} \mathrm{e}^{2\pi i \nu u_n}$$

reflète la nature de la répartition de $\{u_n\}_{n=1}^{\infty}$ modulo 1. Mais peut-on également y lire le caractère aléatoire de la suite ?

Il est acquis qu'une suite aléatoire est presque sûrement équirépartie modulo 1. Si $\{X_n\}_{n=1}^{\infty}$ est une suite de variables aléatoires indépendantes et uniformément réparties sur $[0,1[$,[1] la loi du logarithme itéré des probabilités fournit même l'estimation presque sûre

$$\limsup_{N \to \infty} \frac{\left| \sum_{n \leqslant N} \mathrm{e}^{2\pi i \nu X_n} \right|}{\sqrt{2N \ln_2 N}} = 1 \qquad (\nu \neq 0).$$

Cela suggère que le comportement de sommes d'exponentielles portant sur les nombres premiers, et en particulier celui de

$$S_N(\alpha) := \sum_{n \leqslant N} \mathrm{e}^{2\pi i \alpha p_n},$$

contient une information capitale sur la nature stochastique de la répartition des nombres premiers.

1. En d'autres termes, $\mathbb{P}(X_n \leqslant z) = z$ $(0 \leqslant z \leqslant 1)$.

Non seulement la validité de l'estimation $S_N(\nu\alpha) = o(N)$ pour tout entier $\nu \neq 0$ équivaut à l'équirépartition de $\{\alpha p_n\}_{n=1}^{\infty}$ modulo 1, mais encore la qualité de l'évaluation effective nous permet, le cas échéant, de préciser si ce résultat d'équirépartition doit être interprété comme un signe de régularité déterministe ou d'irrégularité aléatoire pour la suite $\{p_n\}_{n=1}^{\infty}$.

En effet, une suite $\{u_n\}_{n=1}^{\infty}$ peut être équirépartie modulo 1 pour plusieurs raisons fondamentalement distinctes. Nous avons vu plus haut que le caractère aléatoire en est une. Considérons, à l'opposé, le cas de la suite $\{n\alpha\}_{n=1}^{\infty}$ avec α irrationnel. Les sommes d'exponentielles sont alors aussi petites que possible, et l'équirépartition provient de la parfaite régularité de la suite $n \mapsto n$. D'une manière générale, si le nombre irrationnel α est approché par un rationnel réduit a/q, la partie fractionnaire $\langle u_n\alpha\rangle$, lorsque $u_n \in \mathbb{Z}$, est proche de r_n/q où r_n est le reste de au_n modulo q. Dans le cas de la suite $u_n = n$, ces restes décrivent uniformément les entiers de 0 à $q-1$, et cela explique l'équirépartition.

Bien entendu, une suite peut aussi être très régulière sans pour autant être équirépartie modulo 1 : la suite nulle est un exemple trivial et nous avons mentionné plus haut le cas moins évident de la suite $\{\ln n\}_{n=1}^{\infty}$.[1]

L'argument d'approximation des irrationnels par des rationnels présenté plus haut pour justifier l'équirépartition modulo 1 de la suite $\{n\alpha\}_{n=1}^{\infty}$ ($\alpha \in \mathbb{R} \setminus \mathbb{Q}$) peut à l'évidence être transposé à la suite $\{\alpha p_n\}_{n=1}^{\infty}$: on fait maintenant appel au théorème de Dirichlet qui assure, pour tout q, une bonne répartition des p_n dans les classes de congruences admissibles modulo q.[2] L'équirépartition de la suite $\{\alpha p_n\}_{n=1}^{\infty}$ a été essentiellement établie par Vinogradov (1937) dans sa démonstration que tout entier impair assez grand

1. Le défaut d'équirépartition tient alors à la lenteur de croissance, qui imprime à la suite un caractère « stationnaire », même modulo 1.

2. Cet argument heuristique n'est valable que dans les cas « non pathologiques » et doit être utilisé avec précaution : Meijer & Sattler (1972) ont construit une suite d'entiers $\{u_n\}_{n=1}^{\infty}$ qui est bien répartie modulo q pour tout q mais telle que $\{u_n\alpha\}_{n=1}^{\infty}$ ne soit équirépartie modulo 1 pour presque aucun α.

est somme d'au plus trois nombres premiers — ce qui constitue, encore aujourd'hui, le pas le plus significatif vers la conjecture de Goldbach.[1] La preuve, dont l'objectif est donc d'établir la relation asymptotique

$$S_N(\alpha) = o(N) \qquad (N \to \infty)$$

pour tout irrationnel α fixé, se scinde naturellement en deux selon la taille du dénominateur $q \leqslant Q_N$ d'une approximation a/q de α telle que $(a,q) = 1$, $|\alpha - a/q| \leqslant 1/qQ_N$,[2] où $Q_N = N/(\ln N)^A$ pour une constante A arbitrairement choisie.

Si $q \leqslant (\ln N)^A$, on peut faire appel au théorème de Siegel–Walfisz pour estimer

$$S_N(a/q) = \sum_{1\leqslant b\leqslant q} \mathrm{e}^{2\pi iab/q}\pi(p_N; b, q).$$

Grâce à l'identité de Ramanujan $\sum_{\substack{1\leqslant b\leqslant q\\(b,q)=1}} \mathrm{e}^{2\pi iab/q} = \mu(q)$, il suit

$$S_N(a/q) = \frac{N\mu(q)}{\varphi(q)} + O\Big(N\mathrm{e}^{-c\sqrt{\ln N}}\Big).$$

Posant $\beta := \alpha - a/q$ et $R_n := S_n(a/q) - n\mu(q)/\varphi(q)$, on en déduit

$$\begin{aligned} S_N(\alpha) &= \sum_{n\leqslant N} \mathrm{e}^{2\pi i\beta p_n}\{S_n(a/q) - S_{n-1}(a/q)\} \\ &= \sum_{n\leqslant N} O\Big(\frac{1}{\varphi(q)}\Big) + \sum_{n\leqslant N} \mathrm{e}^{2\pi i\beta p_n}\{R_n - R_{n-1}\}. \end{aligned}$$

1. Voir le Chapitre 5, p. 152.

2. L'existence de a/q est assurée par un théorème classique de Dirichlet établi grâce au principe des tiroirs.

La première somme en n est clairement $\ll N/\varphi(q)$. La seconde vaut

$$\sum_{n<N} R_n\{\mathrm{e}^{2\pi i\beta p_n} - \mathrm{e}^{2\pi i\beta p_{n+1}}\} + R_N\mathrm{e}^{2\pi i\beta p_N}$$
$$\ll \sum_{n<N} |R_n|\,|\beta|(p_{n+1} - p_n) + |R_N|$$
$$\ll N\mathrm{e}^{-c\sqrt{\ln N}}\Big(\frac{p_N}{qQ_N} + 1\Big) \ll \frac{N}{q},$$

de sorte que, finalement,

$$S_N(\alpha) \ll N/\varphi(q).$$

Cette majoration est pleinement acceptable puisque l'irrationalité de α implique que q, donc aussi $\varphi(q)$, tend vers l'infini avec N.

Si $q > (\ln N)^A$, la technique précédente est rendue caduque par notre ignorance de la répartition des nombres premiers dans les progressions arithmétiques, et il faut recourir à un argument indirect global. L'idée essentielle consiste à introduire une identité combinatoire (analogue à celle du crible d'Ératosthène) permettant d'exprimer $S_N(\alpha)$ au moyen de sommes du type

$$\sum_m \sum_n a_m b_n \mathrm{e}^{2\pi imn\alpha},$$

où m et n sont des entiers ordinaires, astreints à parcourir certains intervalles, et a_m, b_n sont des fonctions arithmétiques de tailles contrôlées.

Il existe à ce jour trois familles d'identités fournissant des décompositions utilisables dans ce contexte, dont les versions initiales sont dues respectivement à Vinogradov (1937), Vaughan (1977) et Daboussi (1996).

Il nous entraînerait trop loin de décrire ici les formules correspondantes. Nous nous contenterons d'indiquer que les deux dernières ont permis une considérable simplification de la méthode de Vinogradov, ainsi du reste qu'une amélioration des bornes

obtenues. La méthode de Vaughan, par exemple, fournit, dès que $|\alpha - a/q| \leqslant 1/q^2$, $(a, q) = 1$,

$$S_N(\alpha) \ll (\ln N)^4 \{N/\sqrt{q} + N^{4/5} + \sqrt{Nq}\}.$$

Choisissant $A = 2B + 8$, et donc $Q_N = N/(\ln N)^{2B+8}$, nous pouvons ainsi énoncer que *l'on a, pour toute constante* $B > 0$,

$$S_N(\alpha) \ll N/\varphi(q) + N/(\ln N)^B,$$

où a/q *est une approximation rationnelle de* α *telle que* $q \leqslant Q_N$, $(a, q) = 1$, $|\alpha - a/q| \leqslant 1/qQ_N$.

Bien qu'à notre connaissance aucun résultat de minoration de $|S_N(\alpha)|$ ne soit connu, on conjecture généralement que la suite $\{\alpha p_n\}_{n=1}^{\infty}$ est plus aléatoire que déterministe et partant que l'ordre de grandeur de $S_N(\alpha)$ est usuellement comparable à $\sqrt{N}$. La majoration issue de la méthode de Vaughan fournit certainement, pour tout $\varepsilon > 0$, une borne $O_\varepsilon(N^{4/5+\varepsilon})$ lorsque les « bonnes » approximations rationnelles de α ont de « grands » dénominateurs — ce qui est le cas de presque tous les nombres réels, et en particulier des nombres réels algébriques.[1]

8. Vision géométrique

En l'absence de résultats complets sur la répartition modulo 1 des nombres premiers, on peut soutenir l'intuition par des constructions géométriques. Étant donnée une suite $u := \{u_n\}_{n=1}^{\infty}$ de nombres réels, on construit une ligne polygonale $\mathcal{L}(u)$ dont les sommets successifs sont les points z_N $(N \geqslant 0)$ définis dans le plan complexe par

$$z_N := \sum_{1 \leqslant n \leqslant N} e^{2\pi i u_n} \qquad (N \geqslant 0).$$

On note que chaque côté de $\mathcal{L}(u)$ est de longueur unité.

1. Rappelons qu'un nombre réel est dit algébrique s'il est racine d'un polynôme à coefficient entier. Un nombre qui n'est pas algébrique est dit *transcendant*. Ainsi $\sqrt{2}$ est algébrique alors que π et e sont transcendants.

D'après le critère de Weyl, si la suite u est équirépartie modulo 1, on a $z_N = o(N)$, donc la courbe $\mathcal{L}(u)$ ne s'éloigne pas trop rapidement de l'origine. La visualisation de $\mathcal{L}(u)$ illustre en quelque sorte la qualité de la répartition modulo 1, et pourra mettre en évidence des phénomènes cachés ou difficiles à établir directement par le calcul.

Considérons la suite $\{n\alpha\}_{n=1}^{\infty}$ avec $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Nous savons qu'il y a équirépartition modulo 1. Les premières valeurs de la suite réduite sont les suivantes (pour $\alpha = \sqrt{2}$).

n	1	2	3	4	5	6	7	8	9
$10^3 \langle n\sqrt{2} \rangle$	414	828	242	555	71	485	899	313	727

Les points u_n semblent remplir l'intervalle $[0, 1[$ de façon dense. De fait, l'allure de $\mathcal{L}(u)$ est extrêmement structurée, ce qui conforte notre analyse du paragraphe précédent selon laquelle l'équirépartition de $\{n\alpha\}_{n=1}^{\infty}$ est de nature déterministe.

Ainsi qu'on peut le montrer aisément, $\mathcal{L}(u)$ est une « étoile » à une infinité de branches. Elle est dense dans la couronne centrée au point $\frac{1}{2}(-1 + i/\tan \pi\alpha)$ dont les cercles limites ont pour rayons respectifs $1/|2 \tan \pi\alpha|$ et $1/|2 \sin \pi\alpha|$ — voir Figure 1. L'allure générale ne dépend que faiblement de la valeur de α.

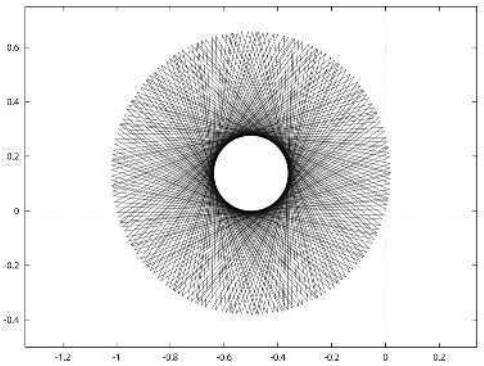


Figure 1
$\sum_{n=1}^{200} \mathrm{e}^{2i\pi n\sqrt{2}}$

Dans le cas de la suite $\{\alpha n \ln n\}_{n=1}^{\infty}$, la lecture des valeurs prises donne, plus encore que pour la suite précédente, l'impression d'un engendrement aléatoire. Cependant la courbe $\mathcal{L}(n \mapsto \alpha n \ln n)$ est remarquablement bien structurée, ainsi que l'atteste la Figure 2, sur laquelle on a tracé, pour $\alpha = \sqrt{2}$, les 10^4 premiers côtés.

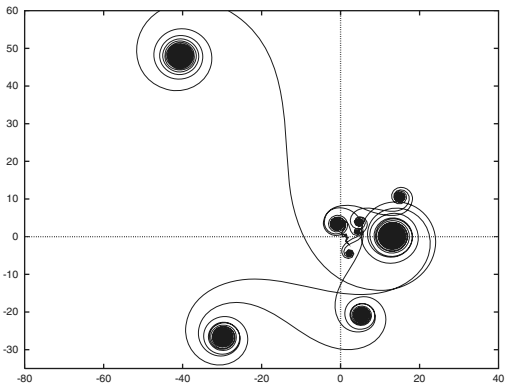


Figure 2
$\sum_{n=1}^{10000} \mathrm{e}^{2i\pi\sqrt{2}n\ln n}$

L'aspect spiralé est facile à interpréter : à cause de la faible croissance de $\ln n$, la courbe se comporte localement comme la courbe associée à $\{n\alpha_N\}_{n=1}^{\infty}$ où $\alpha_N \approx \alpha \ln N \pmod 1$ pour n de taille N. La courbe se présente donc comme une succession de couronnes, liées par des portions quasi rectilignes, correspondant aux valeurs de N telles que $\alpha_N \approx 0 \pmod 1$.

La courbe $\mathcal{L}(n \mapsto \alpha \lfloor n \ln n \rfloor)$, qui pourrait *a priori* ressembler plus encore à $\mathcal{L}(n \mapsto \alpha p_n)$, présente le même type de structure que la précédente : voir la Figure 3.

L'expérimentation fait immédiatement apparaître un phénomène de nature profondément différente pour la courbe $\mathcal{L}(n \mapsto \alpha p_n)$ — cf. Figure 4.

Les spirales ont disparu et la comparaison avec une suite aléatoire (Figure 5) est assez frappante. On retrouve ainsi, malheureusement sans pouvoir le démontrer, que les nombres premiers semblent occuper tout le hasard qui leur est imparti.

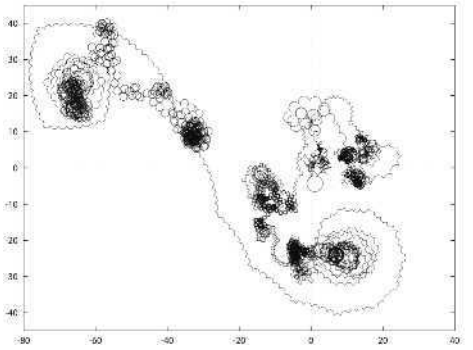


Figure 3
$\sum_{n=1}^{10000} \mathbf{e}^{2i\pi\sqrt{2}\lfloor n \ln n \rfloor}$

Les théories du chaos nous enseignent que certains phénomènes sont instables. Une faible variation dans les données initiales provoque de formidables différences de comportement final. La situation est analogue ici. Les paramètres de la répartition globale des nombres premiers ne permettent pas d'en décrire les irrégularités locales.

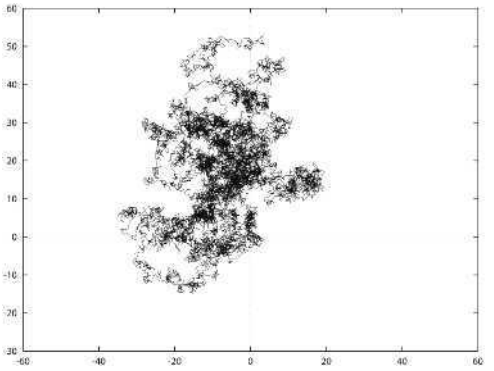


Figure 4
$\sum_{n=1}^{10000} \mathbf{e}^{2i\pi\sqrt{2}p_n}$

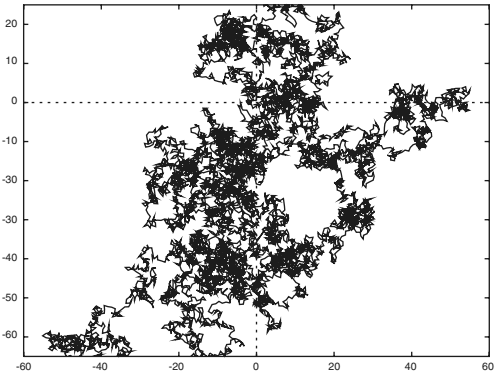


Figure 5

$\sum_{n=1}^{10000} \mathrm{e}^{2i\pi\sqrt{2}\,\mathsf{aléa}(n)}$

Chapitre 4

Une preuve élémentaire du théorème des nombres premiers

1. Introduction

Après les preuves du théorème des nombres premiers, en 1896, par Hadamard et La Vallée-Poussin, les mathématiciens ont cherché à élucider les rapports entre la théorie des fonctions de variable complexe et celle de la répartition des nombres premiers, qui fait *a priori* partie de l'arithmétique élémentaire.

Ainsi que nous l'avons vu au Chapitre 2, la preuve analytique repose de manière essentielle sur l'absence de zéro de la fonction zêta de Riemann sur la droite $\sigma = 1$. En outre, les *théorèmes taubériens*[1]

1. Ainsi nommés par référence au mathématicien autrichien A. Tauber qui a établi en 1897 une réciproque du théorème d'Abel sur la continuité radiale de la somme d'une série entière en tout point de convergence : si $\lim_n na_n = 0$ et si

$$\lim_{x \to 1-} \sum_{n \geqslant 0} a_n x^n = \ell,$$

alors $\sum_{n \geqslant 0} a_n$ converge et est de somme ℓ. Hardy & Littlewood ont montré en 1913 que la condition $na_n > -K$ $(K > 0)$ est suffisante.

développés par Wiener au début des années 1930, et notamment le célèbre théorème d'Ikehara (1931), fournissent un cadre général dans lequel le théorème des nombres premiers et l'assertion sur les zéros de $\zeta(s)$ apparaissent de manière éclatante comme des propositions équivalentes. Cela a induit l'idée qu'en un certain sens la théorie des fonctions analytiques était plus « profonde » que l'analyse réelle et qu'il était hautement improbable, pour ne pas dire impossible, de parvenir au théorème des nombres premiers[1] par de simples manipulations d'inégalités.

Ainsi, le théorème des nombres premiers apparaissait comme irréductible, pour sa démonstration, au cadre naturel dans lequel on pouvait l'énoncer — une frustration qui engendrait bien des spéculations méta-mathématiques !

Ce fut donc un grand choc lorsqu'en 1949 Erdős et Selberg produisirent une preuve élémentaire,[2] mais certainement astucieuse et nullement facile, du théorème des nombres premiers — rendant *ipso facto* caduques toutes les considérations sur les profondeurs relatives des méthodes de l'analyse complexe et de l'analyse réelle.

L'outil essentiel est ici la formule asymptotique

$$\sum_{p\leqslant x}(\ln p)^2+\sum_{pq\leqslant x}\ln p\ln q=2x\ln x+O(x),$$

aujourd'hui célèbre sous le nom d'*identité de Selberg*.

Cette relation peut être vue comme une sorte de formule de Stirling de degré supérieur : au lieu d'employer la fonction de von Mangoldt $\Lambda(n)$ qui vérifie

$$\sum_{d|n}\Lambda(d)=\ln n\qquad(n\geqslant 1),$$

1. Sous forme « minimale » : $p_n\sim n\ln n$ ou $\pi(x)\sim x/\ln x$, où p_n désigne le n-ième nombre premier et $\pi(x)$ le nombre des indices n tels que $p_n\leqslant x$.

2. Au sens où elle n'utilise pratiquement aucun outil sophistiqué d'analyse à l'exception des propriétés du logarithme.

Atle Selberg (1917–2007)

on introduit une nouvelle fonction $\Lambda_2(n)$ satisfaisant à

$$\sum_{d|n} \Lambda_2(d) = (\ln n)^2 \qquad (n \geqslant 1)$$

et l'on évalue sa fonction sommatoire à partir d'un calcul approché élémentaire (et facile !) de $\sum_{n\leqslant x}(\ln n)^2$. La procédure est semblable à celle employée au § 1.9 pour estimer la valeur moyenne de $\Lambda(n)$ — autrement dit, obtenir l'encadrement de Tchébychev, cf. p. 29 — à partir d'une forme faible de la formule de Stirling, i.e. l'évaluation de la valeur moyenne de $\ln n$. Le gain (capital) dû à l'introduction de l'exposant 2 est que la technique produit ici directement une formule asymptotique.

En utilisant l'identité de Selberg, Erdős a prouvé élémentairement que le rapport p_{n+1}/p_n de deux nombres premiers consécutifs tend vers 1. Il a même établi que, pour tout $\delta > 0$ et $x > x_0(\delta)$ il existe entre x et $x + \delta x$ au moins $c(\delta)x/\ln x$ nombres premiers, où $c(\delta)$ est une constante positive ne dépendant que de δ.

Cela fournissait le pendant local de la régularité globale mise en évidence par la formule de Selberg, et il n'est guère surprenant, du moins *a posteriori*, qu'une preuve élémentaire du théorème

Paul Erdős (1913–1996)

des nombres premiers ait pu résulter de la conjonction de ces deux informations : Selberg effectua la liaison finale deux jours seulement après qu'Erdős lui ait communiqué sa preuve des deux résultats cités plus haut. Quelque temps plus tard, les deux mathématiciens simplifièrent ensemble l'argument. La nouvelle preuve n'utilisait plus directement le résultat d'Erdős mais reposait fondamentalement sur les mêmes idées.

Ces idées sont présentes, sous une forme ou une autre, dans toutes les démonstrations élémentaires du théorème des nombres premiers, et, sur ce plan, la preuve que nous présentons dans ce chapitre ne fait pas exception. On peut dire *grosso modo* qu'il s'agit de mettre en évidence des équations ou des inéquations fonctionnelles. La signification profonde d'une équation/inéquation fonctionnelle est celle d'une tendance lourde à la régularité : l'espace analytique est soumis à un champ structurant, qui impose un comportement régulier à toute solution située dans son domaine d'attraction. La technique nécessaire consiste alors à montrer que des estimations préliminaires (souvent grossières) sont suffisantes pour imposer aux quantités étudiées d'appartenir au domaine d'attraction requis.

Hédi Daboussi

Nous nous proposons ici de fournir au lecteur une démonstration élémentaire complète et autonome du théorème des nombres premiers. Nous avons choisi de présenter la preuve de Daboussi (1984), qui n'est ni la première historiquement, ni la plus courte, ni celle qui conduit au meilleur résultat effectif, ni même celle qui fournit le plus vaste champ de généralisations.[1]

1. Notons cependant que, comme l'a montré Daboussi lui-même, la méthode s'étend sans difficulté majeure au cas des progressions arithmétiques.

Il reste que nous avons, pour ce parti pris, des motivations d'ordre esthétique, initiatique et heuristique.

Esthétique d'abord, parce que la preuve conjugue avec raffinement idées raisonnables et actes de foi — au chapitre desquels nous rangeons, bien entendu, le principe de départ, qui n'est pleinement justifiable qu'*a posteriori*.

Initiatique ensuite, car cette preuve fournit un merveilleux prétexte à une exploration discursive du paysage actuel de la théorie analytique des nombres : convolutions de fonctions arithmétiques, crible, solutions d'équations différentielles aux différences — la plupart des outils du chercheur moderne de la discipline sont présents sous une forme rudimentaire qui les rend accessibles au néophyte.

Heuristique enfin, car l'idée sous-jacente (à savoir que l'on peut construire un modèle multiplicativement pertinent de l'ensemble des nombres entiers en omettant tous les nombres premiers assez grands) est riche d'enseignements, et de promesses, aptes à faire vibrer notre petite cosmogonie personnelle de la méta-philosophie arithmétique.

2. Intégration par parties

La *sommation d'Abel* est la manipulation élémentaire qui consiste à exprimer une somme finie de produits, disons $S = \sum_{M<n\leqslant N} a_n b_n$, en fonction des sommes partielles de l'un des termes, par exemple $A_n = \sum_{M<k\leqslant n} a_k$. De l'identité $a_n = A_n - A_{n-1}$, on déduit en effet que

$$S = \sum_{M<n\leqslant N-1} A_n(b_n - b_{n+1}) + A_N b_N.$$

Ce procédé permet d'étudier la convergence des séries oscillantes. On en déduit en particulier le *critère d'Abel* : si la suite $\{a_n\}_{n\geqslant 1}$ est à sommes partielles bornées et si la suite $\{b_n\}_{n\geqslant 1}$ décroît vers 0, alors la série $\sum_n a_n b_n$ converge et son reste de rang N est $O(b_N)$.

Il est très utile d'avoir à disposition une version effective de la sommation d'Abel adaptée à la situation, fréquente dans les applications, où $\{a_n\}_{n\geqslant 1}$ est une suite à caractère arithmétique dont la fonction sommatoire est connue de manière approchée et où $\{b_n\}_{n\geqslant 1}$ est une suite de nature analytique, par exemple $b_n = b(n)$ où b est une fonction régulière.

Lemme 2.1 *Soient $w, x \in \mathbb{R}$ et $\{a_n\}_{n\geqslant 1}$ une suite complexe satisfaisant à*

$$\sum_{w<n\leqslant t} a_n = E(t) + O(R(t)) \quad (w \leqslant t \leqslant x)$$

où $E \in \mathcal{C}^1[w,x]$ et $R : [w,x] \to \mathbb{R}^+$ est une fonction monotone. Pour toute fonction $b : [w,x] \to \mathbb{C}$ continue et de classe $\mathcal{C}^1$ par morceaux, on a

$$\sum_{w<n\leqslant x} a_n b(n) = \int_w^x E'(t)b(t)\,\mathrm{d}t + O(R_1(x)) \tag{4.1}$$

avec

$$R_1(x) = |E(w)b(w)| + |R(x)b(x)| + \int_w^x |R(t)b'(t)|\,\mathrm{d}t.$$

Démonstration. Posons $A(t) = \sum_{w<n\leqslant t} a_n$. Le membre de gauche de (4.1) vaut

$$b(x)A(x) - \sum_{w<n\leqslant x} a_n \int_n^x b'(t)\,\mathrm{d}t = b(x)A(x) - \int_w^x A(t)b'(t)\,\mathrm{d}t.$$

Le résultat indiqué en découle en remplaçant dans la dernière intégrale $A(t)$ par $E(t) + O(R(t))$ et en intégrant par parties la contribution du terme principal $E(t)$. □

3. Convolution des fonctions arithmétiques

Une *fonction arithmétique* est tout simplement une suite à valeurs complexes $f : \mathbb{N}^* \to \mathbb{C}$. La convolution[1] des fonctions arithmétiques est l'opération qui au couple (f, g) permet d'associer la nouvelle fonction $h = f * g$ définie par

$$f * g(n) = \sum_{d|n} f(d)g(n/d).$$

Cette opération correspond au produit ordinaire des séries de série de Dirichlet : on a formellement

$$\sum_{n\geqslant 1} \frac{f * g(n)}{n^s} = \sum_{n\geqslant 1} \frac{f(n)}{n^s} \sum_{n\geqslant 1} \frac{g(n)}{n^s}$$

et il est facile de vérifier que cette égalité formelle possède un sens analytique dès que les trois séries sont absolument convergentes.[2]

La fonction zêta de Riemann $\zeta(s)$ est la série de série de Dirichlet associée à la fonction arithmétique $\mathbf{1}$, définie par $\mathbf{1}(n) = 1$ $(n \geqslant 1)$, et son carré $\zeta(s)^2$ est associé à la fonction nombre de diviseurs

$$\tau(n) = \mathbf{1} * \mathbf{1}(n) = \sum_{d|n} 1.$$

Il est facile de vérifier que la fonction

$$\delta(n) := \begin{cases} 1 & \text{si } n = 1, \\ 0 & \text{si } n > 1, \end{cases}$$

1. Également appelée *convolution de Dirichlet*.

2. Cette condition peut être affaiblie de diverses manières. Il est, par exemple, suffisant que l'une des deux séries associées à f ou g soit absolument convergente, l'autre étant semi-convergente, pour que la série associée à $f * g$ converge et ait pour somme le produit des deux séries initiales.

est un élément neutre pour la convolution et que l'ensemble des fonctions arithmétiques muni de l'addition ordinaire et du produit de convolution est un anneau commutatif unitaire intègre.[1]

Les *fonctions multiplicatives* sont les fonctions arithmétiques qui respectent la structure multiplicative de $\mathbb{N}$, i.e. celles qui vérifient

$$f(1) = 1, \quad f(n) = \prod_{p^\nu \| n} f(p^\nu) \quad (n > 1).$$

Une définition équivalente consiste à imposer

$$f(1) \neq 0, \text{ et } f(mn) = f(m)f(n) \text{ dès que } (m, n) = 1.$$

Les fonctions multiplicatives constituent un sous-groupe du groupe des éléments inversibles pour la convolution. Ainsi la fonction nombre de diviseurs $\tau = \mathbf{1} * \mathbf{1}$ est multiplicative. Comme une puissance de nombre premier p^ν possède $\nu + 1$ diviseurs, on en déduit que

$$\tau(n) = \prod_{p^\nu \| n} (\nu + 1).$$

L'importance cruciale des fonctions multiplicatives en théorie analytique des nombres tient essentiellement au théorème suivant.

Théorème 3.1 *Soit f une fonction multiplicative. Alors la série de série de Dirichlet $F(s) = \sum_{n \geqslant 1} f(n)/n^s$ est absolument convergente si et seulement s'il en va de même de la série $\sum_p \sum_{\nu \geqslant 1} f(p^\nu)/p^{\nu s}$. Lorsqu'il en est ainsi, on a le développement en produit eulérien convergent*

$$F(s) = \prod_p \sum_{\nu \geqslant 0} \frac{f(p^\nu)}{p^{\nu s}}.$$

1. On peut aussi montrer qu'il est *factoriel*, autrement dit toute fonction arithmétique peut se décomposer en produit d'un élément inversible et d'éléments premiers, avec décomposition unique au facteur inversible près. Une fonction arithmétique f est inversible si, et seulement si, l'on a $f(1) \neq 0$.

Démonstration. Observons tout d'abord que la convergence absolue de la somme double, qui est une conséquence triviale de celle de $F(s)$, implique la convergence du produit infini

$$M := \prod_p \left(1 + \sum_{\nu \geqslant 1} \left|\frac{f(p^\nu)}{p^{\nu s}}\right|\right).$$

Réciproquement, sous l'hypothèse que la série double converge absolument, nous pouvons écrire pour $x \geqslant 1$

$$\sum_{n \leqslant x} \left|\frac{f(n)}{n^s}\right| \leqslant \sum_{P^+(n) \leqslant x} \left|\frac{f(n)}{n^s}\right| = \prod_{p \leqslant x} \left(1 + \sum_{\nu \geqslant 1} \left|\frac{f(p^\nu)}{p^{\nu s}}\right|\right) \leqslant M.$$

Cela montre bien que la série $F(s)$ est absolument convergente. La majoration

$$\left|\sum_{n \geqslant 1} \frac{f(n)}{n^s} - \prod_{p \leqslant x} \sum_{\nu \geqslant 0} \frac{f(p^\nu)}{p^{\nu s}}\right| = \left|\sum_{P^+(n) > x} \frac{f(n)}{n^s}\right| \leqslant \sum_{n > x} \left|\frac{f(n)}{n^s}\right|$$

implique alors, en faisant tendre x vers l'infini, la formule du développement eulérien. □

Lorsqu'on applique le théorème précédent à la fonction $f = \mathbf{1}$, on retrouve la formule d'Euler, déjà rencontrée aux Chapitres 1 et 2.

Le *principe de l'hyperbole*, dû à Dirichlet, est un procédé commode pour évaluer la fonction sommatoire d'un produit de convolution.

L'utilisation éponyme est liée à la fonction nombre de diviseurs $\tau = \mathbf{1} * \mathbf{1}$, dont la fonction sommatoire

$$\sum_{n \leqslant x} \tau(n) = \sum_{uv \leqslant x} 1$$

dénombre exactement le nombre de points à coordonnées entières situés dans le premier quadrant ouvert et sous l'hyperbole équilatère $uv = x$.

Théorème 3.2 (Principe de l'hyperbole) *Soient f, g deux fonctions arithmétiques, de fonctions sommatoires respectives*

$$F(x) := \sum_{n\leqslant x} f(n), \qquad G(x) := \sum_{n\leqslant x} g(n).$$

On a pour $1 \leqslant y \leqslant x$

$$\sum_{n\leqslant x} f*g(n) = \sum_{n\leqslant y} g(n)F\Big(\frac{x}{n}\Big) + \sum_{m\leqslant x/y} f(m)G\Big(\frac{x}{m}\Big) - F\Big(\frac{x}{y}\Big)G(y).$$

Démonstration. Le membre de gauche s'écrit encore

$$\begin{aligned}\sum_{md\leqslant x} f(m)g(d) &= \sum_{md\leqslant x,\, d\leqslant y} f(m)g(d) + \sum_{md\leqslant x,\, d>y} f(m)g(d)\\ &= \sum_{d\leqslant y} g(d)F(x/d) + \sum_{m\leqslant x/y} f(m)\left\{G(x/m) - G(y)\right\}.\end{aligned}$$

Cela implique immédiatement le résultat annoncé, en développant le dernier terme. □

L'application historique de cette méthode est la célèbre formule de Dirichlet

$$\sum_{n\leqslant x} \tau(n) = x\{\ln x + 2\gamma - 1\} + O(\sqrt{x}).$$

Le théorème précédent appliqué avec $y = \sqrt{x}$ et $f = g = \mathbf{1}$ (donc $F(x) = G(x) = \lfloor x\rfloor$), montre en effet que le membre de gauche vaut

$$2\sum_{m\leqslant\sqrt{x}} \lfloor x/m\rfloor - \lfloor\sqrt{x}\rfloor^2 = 2\sum_{m\leqslant\sqrt{x}} x/m - x + O(\sqrt{x}).$$

On conclut en appliquant le résultat classique

$$\sum_{m\leqslant z}\frac{1}{m}=\ln z+\gamma+O\Big(\frac{1}{z}\Big)$$

avec $z=\sqrt{x}$.

4. La fonction de Möbius

Nous avons déjà rencontré la fonction de Möbius, notée μ, dans la formule du crible d'Ératosthène (§ 1.8). Nous la redéfinissons ici (afin de garantir l'autonomie de ce chapitre) comme l'inverse de convolution de la fonction $\mathbf{1}$, soit

$$\mathbf{1}*\mu=\delta \quad \text{ou encore} \quad \sum_{d|n}\mu(d)=\begin{cases}1 & \text{si } n=1,\\ 0 & \text{si } n>1.\end{cases}$$

L'existence de l'inverse est garantie par le fait que les fonctions multiplicatives forment un sous-groupe de l'ensemble des fonctions arithmétiques inversibles. La multiplicativité nous permet, pour le calcul de cet inverse, de nous restreindre aux puissances de nombres premiers.[1] La relation de convolution $\mathbf{1}*\mu=\delta$ fournit alors, de proche en proche, $\mu(p)=-1$ et $\mu(p^\nu)=0$ pour $\nu\geqslant 2$. On a donc, pour un entier générique n de décomposition canonique $n=\prod_{j=1}^{k}p_j^{\nu_j}$,

$$\mu(n)=\begin{cases}(-1)^k & \text{si } \nu_j=1\ (1\leqslant j\leqslant k),\\ 0 & \text{si } \max_j \nu_j\geqslant 2.\end{cases}$$

Comme on peut le vérifier sans peine, la fonction Λ de von Mangoldt rencontrée au Chapitre 1 (p. 27) satisfait à la relation de convolution $\Lambda*\mathbf{1}=\ln$. En convolant par μ, on obtient

$$\Lambda=\mu*\ln\,.$$

On peut aussi vérifier que $\Lambda=-(\mu\ln)*\mathbf{1}$.

1. On a $\mu(1)=1$ par définition puisque l'inverse de convolution d'une fonction multiplicative est encore multiplicative.

Plus généralement, on peut rassembler les propriétés algébriques de la fonction de Möbius dans l'énoncé suivant.

Théorème 4.1 (Inversion de Möbius)

(a) *Soient f et g des fonctions arithmétiques. Les deux propriétés suivantes sont équivalentes*

$$\text{(i)} \qquad g(n) = \sum_{d|n} f(d) \qquad (n \geqslant 1)$$

$$\text{(ii)} \qquad f(n) = \sum_{d|n} g(d)\,\mu(n/d) \qquad (n \geqslant 1).$$

(b) *Soient F et G des fonctions complexes définies sur $[1, +\infty[$. Les deux conditions suivantes sont équivalentes*

$$\text{(iii)} \qquad F(x) = \sum_{n\leqslant x} G(x/n) \qquad (x \geqslant 1)$$

$$\text{(iv)} \qquad G(x) = \sum_{n\leqslant x} \mu(n)F(x/n) \qquad (x \geqslant 1).$$

Démonstration. Le point (a) exprime directement l'équivalence des relations $g = f * \mathbf{1}$ et $f = g * \mu$. Établissons par exemple l'implication (iii) $\Rightarrow$ (iv), la réciproque étant analogue. Pour $x \geqslant 1$, nous avons

$$\begin{aligned}\sum_{n\leqslant x} \mu(n)F(x/n) &= \sum_{n\leqslant x} \mu(n) \sum_{m\leqslant x/n} G(x/mn) \\ &= \sum_{k\leqslant x} G(x/k) \sum_{mn=k} \mu(n).\end{aligned}$$

Par définition de μ, la somme intérieure vaut $\delta(k)$. Cela implique bien (iv). □

En appliquant le Théorème 4.1 pour $G(x) \equiv 1$, on obtient

$$\sum_{n\leqslant x} \mu(n)\,\lfloor x/n \rfloor = 1 \qquad (x \geqslant 1).$$

Cela montre que

$$\sup_{x\in\mathbb{R}}\left|\sum_{n\leqslant x}\frac{\mu(n)}{n}\right|\leqslant 1$$

et suggère que $\lim_{x\to\infty}\sum_{n\leqslant x}\mu(n)/n=0$. Nous verrons au paragraphe suivant que cette relation est élémentairement équivalente au théorème des nombres premiers.

Posons

$$M(x):=\sum_{n\leqslant x}\mu(n).$$

Par interversion de sommations, on a pour $a>0,\ b>0$,

$$\int_a^b M(x)\frac{\mathrm{d}x}{x^2}=\frac{M(a)}{a}-\frac{M(b)}{b}+\sum_{a<n\leqslant b}\frac{\mu(n)}{n}$$

de sorte que

$$\sup_{a,b\in\mathbb{R}^+}\left|\int_a^b M(x)\frac{\mathrm{d}x}{x^2}\right|\leqslant 4. \tag{4.2}$$

Nous ferons un usage crucial de cette majoration au cours de la démonstration élémentaire du théorème des nombres premiers.

Remarque. La définition du produit de convolution en termes de produit ordinaire de séries de série de Dirichlet permet de retrouver immédiatement que l'on a $\sum_{n\geqslant 1}\mu(n)/n^s=1/\zeta(s)$ pour $\sigma>1$ et en particulier,

$$\sum_{n\geqslant 1}\mu(n)/n^2=6/\pi^2,$$

puisque $\zeta(2)=\pi^2/6$ (cf. § 2.4, p. 57). Le théorème suivant permet d'interpréter ce nombre comme la probabilité que deux entiers choisis au hasard soient premiers entre eux.

Théorème 4.2 *Désignons par $G(x,y)$ le nombre des couples d'entiers (m,n) tels que $1\leqslant m\leqslant x$, $1\leqslant n\leqslant y$, $(m,n)=1$. Alors*

la relation asymptotique $G(x,y) \sim (6/\pi^2)xy$ *a lieu lorsque* x *et* y *tendent vers l'infini. Plus précisément, on a en posant* $z := \min(x,y)$

$$G(x,y) = xy\Big\{\frac{6}{\pi^2} + O\Big(\frac{\ln z}{z}\Big)\Big\} \qquad (x,y \geqslant 2).$$

Démonstration. Grâce à la relation de convolution $\delta = \mathbf{1} * \mu$, on peut écrire

$$\begin{aligned} G(x,y) &= \sum_{m\leqslant x,\, n\leqslant y} \delta\left((m,n)\right) = \sum_{d\leqslant z} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor \left\lfloor \frac{y}{d} \right\rfloor \\ &= xy \sum_{d\leqslant z} \frac{\mu(d)}{d^2} + O\Big((x+y)\sum_{d\leqslant z}\frac{1}{d}\Big) \\ &= xy\Big\{\frac{6}{\pi^2} + O\Big(\frac{1}{z} + \Big(\frac{1}{x}+\frac{1}{y}\Big)\ln z\Big)\Big\}. \end{aligned}$$

□

5. Valeur moyenne de la fonction de Möbius et théorème des nombres premiers

Nous avons vu au Chapitre 1 que la fonction de von Mangoldt $\Lambda = \mu * \ln$ est fortement liée à la fonction caractéristique des nombres premiers et possède une fonction sommatoire

$$\psi(x) = \sum_{n\leqslant x} \Lambda(n),$$

qui, comme il a été établi p. 29, satisfait à

$$\psi(x) \sim \pi(x)\ln x \qquad (x \to \infty).$$

Il est donc naturel de se demander si la valeur moyenne de $\mu(n)$ possède une interprétation simple dans le cadre de l'étude asymptotique des fonctions de Tchébychev $\pi(x)$, $\psi(x)$, c'est-à-dire dans le domaine de la répartition des nombres premiers. Le théorème suivant, dû à Landau (1909), répond complètement à cette question.

Théorème 5.1 *Les assertions suivantes sont élémentairement équivalentes :*

(i) $\psi(x) \sim x \qquad (x \to \infty)$,

(ii) $M(x) := \sum_{n\leqslant x} \mu(n) = o(x) \qquad (x \to \infty)$,

(iii) $\sum_{n\geqslant 1} \mu(n)/n = 0$.

Remarques. Seule l'implication (ii) $\Rightarrow$ (i) sera utilisée dans la preuve du théorème des nombres premiers. Ainsi que le lecteur pourra le constater aisément, ce n'est pas la partie la plus difficile de la démonstration. L'assertion (iii) signifie bien entendu que la série figurant au membre de gauche converge et est de somme nulle.

Démonstration. L'implication (iii) $\Rightarrow$ (ii) résulte d'une simple intégration par parties. Supposons, en effet, que

$$m(x) := \sum_{n\leqslant x} \mu(n)/n = o(1) \quad (x \to \infty).$$

En appliquant le Lemme 2.1 avec $w = 1$, $E(t) = 0$ et

$$R(t) = \sup_{y\geqslant t} |m(y)|,$$

nous obtenons,

$$M(x) \ll xR(x) + \int_1^x R(t)\,\mathrm{d}t = o(x).$$

Pour prouver l'implication (ii) $\Rightarrow$ (i), nous établissons d'abord une identité de convolution pour la fonction $\Lambda - \mathbf{1}$, dont nous devons montrer qu'elle possède une fonction sommatoire $o(x)$. On a

$$\begin{aligned}\Lambda - \mathbf{1} = (\ln - \tau) * \mu &= (\ln - \tau + 2\gamma\mathbf{1}) * \mu - 2\gamma\delta \\ &= f * \mu - 2\gamma\delta,\end{aligned}$$

où la fonction $f := \ln - \tau + 2\gamma\mathbf{1}$ satisfait à

$$F(x) := \sum_{n\leqslant x} f(n) = O(\sqrt{x}),$$

d'après la formule de Dirichlet établie au § 3 et l'évaluation classique

$$\sum_{n\leqslant x} \ln n = x\ln x - x + O(1+\ln x) \quad (x \geqslant 1).$$

Nous allons montrer que

$$H(x) := \sum_{n\leqslant x} f*\mu(n) = o(x)$$

en utilisant l'estimation précédente pour $F(x)$ et en faisant appel au principe de l'hyperbole (Théorème 3.2). Pour chaque $y > 2$ fixé, on peut en effet écrire

$$H(x) = \sum_{n\leqslant x/y} \mu(n)F\Big(\frac{x}{n}\Big) + \sum_{m\leqslant y} f(m)M\Big(\frac{x}{m}\Big) - F(y)M\Big(\frac{x}{y}\Big).$$

Sous l'hypothèse (ii), il suit donc, pour chaque y fixé,

$$\begin{aligned}\limsup_{x\to\infty}\Big|\frac{H(x)}{x}\Big| &\leqslant \limsup_{x\to\infty}\frac{1}{x}\sum_{n\leqslant x/y}\Big|F\Big(\frac{x}{n}\Big)\Big| \\ &\ll \limsup_{x\to\infty}\frac{1}{x}\sum_{n\leqslant x/y}\sqrt{\frac{x}{n}} \ll \frac{1}{\sqrt{y}}.\end{aligned}$$

Comme y peut être choisi arbitrairement grand, cela implique bien que $H(x) = o(x)$.

Il reste à établir l'implication (i) $\Rightarrow$ (iii). Nous utiliserons la majoration

$$|m(x)| \leqslant 1$$

prouvée au paragraphe précédent. Nous remarquons d'abord que la formule d'inversion de Möbius $\mu * \mathbf{1} = \delta$ implique

$$\sum_{md=n} \frac{\mu(m)}{md} = \frac{\delta(n)}{n} = \delta(n) \qquad (n \geqslant 1)$$

d'où, par sommation pour $n \leqslant x$,

$$\sum_{m\leqslant x} \frac{\mu(m)}{m} \sum_{d\leqslant x/m} \frac{1}{d} = 1 \qquad (x \geqslant 1).$$

En évaluant la somme intérieure, il vient

$$\begin{aligned} 1 &= \sum_{m\leqslant x} \frac{\mu(m)}{m} \Big\{ \ln\Big(\frac{x}{m}\Big) + \gamma + O\Big(\frac{m}{x}\Big) \Big\} \\ &= m(x)(\ln x + \gamma) - G(x) + O(1), \end{aligned}$$

où l'on a posé

$$G(x) := \sum_{m\leqslant x} \frac{\mu(m) \ln m}{m}.$$

Il nous suffit donc d'établir que l'on a

$$G(x) = o(\ln x) \qquad (x \to \infty).$$

À cette fin, nous utilisons la relation de convolution

$$\mu \ln = -\Lambda * \mu = (\mathbf{1} - \Lambda) * \mu - \delta,$$

dont nous omettons la vérification, qui est facile. Posant

$$P(x) := \lfloor x \rfloor - \psi(x) = \sum_{n\leqslant x} (1 - \Lambda(n)) = o(x),$$

on peut donc écrire

$$\begin{aligned} G(x) + 1 &= \sum_{j\leqslant x} \Big(\frac{P(j) - P(j-1)}{j} \Big) m\Big(\frac{x}{j}\Big) \\ &= \sum_{j\leqslant x} P(j) \Big(\frac{m(x/j)}{j} - \frac{m\,(x/(j+1))}{j+1} \Big) + o(1) \\ &= \sum_{j\leqslant x} \frac{P(j)}{j} \Big\{ m_j(x) + \frac{1}{j+1} m\Big(\frac{x}{j+1}\Big) \Big\} + o(1), \end{aligned}$$

avec $m_j(x) := m(x/j) - m\,(x/(j+1))$. Donnons-nous alors, pour chaque $\varepsilon > 0$, un nombre entier $j_0(\varepsilon)$ tel que $|P(j)| \leqslant \varepsilon j$ pour $j > j_0(\varepsilon)$. En utilisant la majoration $\sup_t |m(t)| \leqslant 1$ et l'inégalité triviale

$$|m_j(x)| \leqslant \sum_{x/(j+1)<n\leqslant x/j} \frac{1}{n},$$

il suit

$$|G(x)+1| \leqslant 3 \sum_{j\leqslant j_0(\varepsilon)} \frac{|P(j)|}{j} + 2\varepsilon \sum_{n\leqslant x} \frac{1}{n} + o(1),$$

d'où $\limsup_{x\to\infty} |G(x)|/\ln x \leqslant 2\varepsilon$. Comme ε est arbitrairement petit, cela montre bien que

$$G(x) = o(\ln x)$$

et achève la démonstration du théorème. □

On peut obtenir d'autres formulations équivalentes à celles du Théorème 5, par exemple

(iv) $$\sum_{n\leqslant x} \Lambda(n)/n = \ln x - \gamma + o(1) \qquad (x \to \infty).$$

L'implication (iv) ⇒ (i) découle directement du Lemme 2.1, et nous omettons les détails. Pour montrer la réciproque, on utilise à nouveau la fonction

$$f = \ln - \tau + 2\gamma \mathbf{1}.$$

On a

$$\begin{aligned}\sum_{n\leqslant x} \frac{\Lambda(n)-1}{n} &= \sum_{kd\leqslant x} \frac{f(k)}{k}\,\frac{\mu(d)}{d} - 2\gamma \\ &= \sum_{d\leqslant x/y} \frac{\mu(d)}{d} L\left(\frac{x}{d}\right) + \sum_{k\leqslant y} \frac{f(k)}{k} m\left(\frac{x}{k}\right) - L(y)m\left(\frac{x}{y}\right) - 2\gamma,\end{aligned}$$

avec

$$L(z) := \sum_{k\leqslant z} \frac{f(k)}{k} = C + O\Big(\frac{1}{\sqrt{z}}\Big) \qquad (z \geqslant 1).$$

Cette dernière relation est établie, pour une constante convenable C, par sommation d'Abel en utilisant l'estimation

$$F(z) \ll \sqrt{z}.$$

En faisant tendre x puis y vers l'infini, on obtient, grâce à (iii),

$$\sum_{n\leqslant x} \frac{\Lambda(n)-1}{n} = -2\gamma + o(1),$$

d'où la conclusion souhaitée.

6. Entiers sans grand ou sans petit facteur premier

Dans la théorie analytique moderne, les ensembles d'entiers

$$S(x,y) := \{n \leqslant x : P^+(n) \leqslant y\},$$
$$T(x,y) := \{n \leqslant x : P^-(n) > y\},$$

jouent des rôles de plus en plus importants. La source de ce phénomène réside principalement dans le fait que, pour y donné, chaque entier $n \leqslant x$ peut être décomposé de manière unique sous la forme d'un produit $n = ab$ avec $a \in S(x,y)$, $b \in T(x,y)$. On exploite ensuite cette factorisation avec l'idée générale que le nombre a se comporte essentiellement comme un entier ordinaire, alors que b possède les caractéristiques d'un nombre premier.

En d'autres termes, $S(x,y)$ et $T(x,y)$ constituent respectivement des modèles simplifiés de l'ensemble des entiers et de celui des nombres premiers n'excédant pas x.

La démonstration élémentaire du théorème des nombres premiers que nous allons présenter fait appel aux propriétés de base de ces ensembles. Nous notons

$$\Psi(x,y) := |S(x,y)|, \qquad \Phi(x,y) := |T(x,y)|.$$

La plupart des estimations qui suivent s'expriment de manière naturelle en fonction de la quantité

$$u := (\ln x)/\ln y.$$

Nous adopterons systématiquement cette notation.

Nous ferons dans la suite un usage intensif (et parfois implicite) de la propriété suivante : *pour tout $y \geqslant 2$ la série*

$$\sum_{P^+(n)\leqslant y} \frac{1}{n^\sigma} = \prod_{p\leqslant y} \Big(1 - \frac{1}{p^\sigma}\Big)^{-1}$$

est convergente dès que $\sigma > 0$. Cela découle immédiatement du Théorème 3.1, puisque la somme double

$$\sum_{p\leqslant y} \sum_{\nu\geqslant 1} \frac{1}{p^{\nu\sigma}} = \sum_{p\leqslant y} \frac{1}{p^\sigma - 1}$$

est finie.

Théorème 6.1 *On a*

$$\Psi(x,y) \ll x\mathrm{e}^{-u/2} \qquad (x \geqslant 1,\ y \geqslant 2).$$

Démonstration. Nous pouvons supposer $y \geqslant 11$ puisque dans le cas contraire $S(x,y)$ est réduit aux nombres de la forme $2^\alpha 3^\beta 5^\gamma 7^\delta$ n'excédant pas x, d'où

$$\Psi(x,y) \ll (\ln x)^4 \quad (y < 11).$$

Maintenant, donnons-nous un paramètre positif α et considérons la fonction multiplicative f_α définie par $f_\alpha(p^\nu) = p^{\alpha\nu}$ si $p \leqslant y$ et $f_\alpha(p^\nu) = 0$ dans le cas contraire. Si $n > x^{3/4}$ et $n \in S(x,y)$, on a $f_\alpha(n) > x^{3\alpha/4}$, donc

$$\Psi(x,y) \leqslant x^{3/4} + x^{-3\alpha/4} \sum_{n\leqslant x} f_\alpha(n).^{(1)}$$

1. Le principe très simple de cette inégalité, consistant à majorer une fonction indicatrice par un multiple constant d'une fonction multiplicative a été employé par Rankin dans ce contexte précis. Les arithémticiens y réfèrent souvent sous le nom de «méthode de Rankin».

Choisissons $\alpha = 2/(3\ln y)$, de sorte que

$$x^{-3\alpha/4} = \mathrm{e}^{-u/2} \quad \text{et} \quad x^{3/4} \leqslant x\mathrm{e}^{-u/2}.$$

Nous allons montrer que la somme en n est $O(x)$, ce qui suffit pleinement à fournir l'estimation annoncée.

L'idée essentielle consiste à introduire la fonction multiplicative positive ou nulle g_α définie par $g_\alpha(p^\nu) = p^{\nu\alpha}(1-p^{-\alpha})$ si $p \leqslant y$ et $g_\alpha(p^\nu) = 0$ si $p > y$, et à remarquer que $f_\alpha \leqslant g_\alpha * \mathbf{1}$. En fait $(g_\alpha * \mathbf{1})(n) = n^\alpha = f_\alpha(n)$ si $P^+(n) \leqslant y$ et, dans le cas contraire, $(g_\alpha * \mathbf{1})(n) \geqslant 1 > f_\alpha(n) = 0$. On a donc

$$\sum_{n\leqslant x} f_\alpha(n) \leqslant \sum_{n\leqslant x}\sum_{d\mid n} g_\alpha(d) = \sum_{d\leqslant x} g_\alpha(d)\left\lfloor \frac{x}{d}\right\rfloor,$$

après interversion de sommations. Il en résulte que la somme à majorer n'excède pas

$$\sum_{P^+(d)\leqslant y} g_\alpha(d)\frac{x}{d} = x\prod_{p\leqslant y}\sum_{\nu\geqslant 0}\frac{g_\alpha(p^\nu)}{p^\nu}.$$

On a $g_\alpha(p^\nu)/p^\nu = p^{\nu\alpha}(1-p^{-\alpha})/p^\nu \leqslant \alpha(\mathrm{e}^{2/3}/p)^\nu \ln p$ et $\mathrm{e}^{2/3} < 2$. La somme en ν vaut donc $1 + O(\alpha\ln p/p)$. Le premier théorème de Mertens (cf. p. 32) implique alors que le produit en p est borné, ce qui achève la démonstration. □

Théorème 6.2 *On a pour $x \geqslant 1$, $y \geqslant 2$,*

$$\Phi(x,y) = x\prod_{p\leqslant y}\Big(1-\frac{1}{p}\Big) + O\big(x\mathrm{e}^{-u/2}\ln y\big).$$

Démonstration. Par la formule d'inversion de Möbius, la quantité $\sum_{d\mid n,\, P^+(d)\leqslant y}\mu(d)$ vaut 1 ou 0 selon que l'entier n appartient ou non à $T(x,y)$. En sommant pour $n \leqslant x$ et en intervertissant les sommations, on obtient

$$\Phi(x,y) = \sum_{d\leqslant x,\, P^+(d)\leqslant y}\mu(d)\left\lfloor x/d\right\rfloor.$$

La formule annoncée en découle en remplaçant $\lfloor x/d \rfloor$ par $(x/d) + O(1)$ pour $d \leqslant x$ et en notant que le Théorème 6.1 permet, par intégration par parties (cf. le Lemme 2.1), d'obtenir la majoration

$$\sum_{d>x,\, P^+(d)\leqslant y} \mu(d)/d \ll \mathrm{e}^{-u/2} \ln y.$$

□

Bien entendu, l'estimation du Théorème 6.2 n'est utile que pour des valeurs de y sensiblement plus petites que x : il faut que

$$y \leqslant x^{\{1/2+o(1)\}/\ln_2 x}$$

pour que l'on puisse déduire du Théorème 6.2 une majoration meilleure que l'inégalité triviale $\Phi(x,y) \leqslant x$. Cette limitation ne sera toutefois nullement un handicap dans la perspective où nous nous plaçons. Nous n'utiliserons en fait que des valeurs *bornées* de y.

Soit

$$M(x,y) := \sum_{n\in S(x,y)} \mu(n).$$

Conformément au principe exposé plus haut de modéliser $\{n : n \leqslant x\}$ par $S(x,y)$, la méthode de Daboussi a pour point de départ un résultat liant le comportement de $M(x)$ à celui de $M(x,y)$.

Proposition 6.3 *Pour chaque $y \geqslant 2$ fixé, on a :*

$$\limsup_{x\to\infty} \frac{|M(x)|}{x} \leqslant \prod_{p\leqslant y} \Big(1 - \frac{1}{p}\Big) \int_1^\infty \frac{|M(x,y)|}{x^2}\,\mathrm{d}x.$$

Démonstration. Le Théorème 6.1 implique la convergence de l'intégrale. Décomposons canoniquement chaque entier n n'excédant

pas x sous la forme $n = ab$ avec $a \in S(x,y)$, $b \in T(x,y)$. On a alors nécessairement $(a,b) = 1$ donc $\mu(n) = \mu(a)\mu(b)$. Il suit

$$M(x) = \sum_{b \in T(x,y)} \mu(b) M(x/b, y),$$

et donc

$$\begin{aligned} |M(x)| &\leqslant \sum_{b \in T(x,y)} |M(x/b,y)| \\ &= \sum_{n \leqslant x} |M(n,y)| \Big\{ \Phi\Big(\frac{x}{n}, y\Big) - \Phi\Big(\frac{x}{n+1}, y\Big) \Big\}. \end{aligned}$$

Utilisons maintenant l'approximation du Théorème 6.2 pour évaluer l'expression entre accolades, en notant Π_y le produit en p. La contribution du terme principal est

$$\sum_{n \leqslant x} |M(n,y)| x \int_n^{n+1} \Pi_y \frac{\mathrm{d}t}{t^2} \leqslant x \Pi_y \int_1^\infty \frac{|M(t,y)|}{t^2}\, \mathrm{d}t.$$

Cette majoration est bien compatible avec le résultat annoncé et il reste seulement à montrer que la contribution des termes d'erreur est $o(x)$. Cette contribution vaut exactement $R := \sum_{n \leqslant x} |M(n,y)| R_n$ avec

$$R_n := \Phi\Big(\frac{x}{n}, y\Big) - \Phi\Big(\frac{x}{n+1}, y\Big) - x \Pi_y \int_n^{n+1} \frac{\mathrm{d}t}{t^2},$$

de sorte que le Théorème 6.2 implique, pour tout entier $N \leqslant x$,

$$\sum_{N < n \leqslant x} R_n \ll (x/N)^{1-\varepsilon(y)} \ln y,$$

où l'on a posé $\varepsilon(y) = 1/(2 \ln y)$. En observant que

$$|M(n,y)| - |M(n-1,y)|$$

est majoré en valeur absolue par la fonction indicatrice de $S(x,y)$, une simple sommation d'Abel fournit alors

$$R \ll (\ln y) \sum_{n\in S(x,y)} (x/n)^{1-\varepsilon(y)} \ll (\ln y)^2 x^{1-\varepsilon(y)}.$$

Cela termine la démonstration de la Proposition 6.3. □

Il s'agit maintenant de montrer que la borne supérieure obtenue pour $\limsup |M(x)|/x$ tend vers 0 quand $y \to \infty$. Cette estimation, qui constitue le cœur de la méthode, nécessite les développements qui suivent.

7. La fonction de Dickman

On appelle fonction de Dickman l'unique fonction continue $\varrho : [0,\infty[\to \mathbb{R}$, dérivable sur $]1,\infty[$, telle que

$$\begin{cases} \varrho(v) = 1 & (0 \leqslant v \leqslant 1), \\ v\varrho'(v) + \varrho(v-1) = 0 & (v > 1). \end{cases}$$

Cela définit bien ϱ de manière récursive : pour tout entier $k \geqslant 1$ et $v \in [k, k+1[$, on a

$$\varrho(v) = \varrho(k) - \int_k^v \varrho(t-1)\frac{\mathrm{d}t}{t}\cdot$$

Les principales propriétés de la fonction de Dickman sont rassemblées dans l'énoncé suivant. Nous rappelons (cf. § 2.4) la définition de la fonction Γ d'Euler et l'équation fonctionnelle

$$v\Gamma(v) = \Gamma(v+1),$$

de sorte que $\Gamma(n+1) = n!$ pour tout entier $n \geqslant 0$. On note en particulier la relation asymptotique

$$\ln\Gamma(v) \sim v\ln v \qquad (v \to \infty).$$

Théorème 7.1 *On a :*

(i)	$v\varrho(v) = \int_{v-1}^{v} \varrho(w)\,\mathrm{d}w$	$(v \geqslant 1)$
(ii)	$\varrho(v) > 0$	$(v \geqslant 0)$
(iii)	$\varrho'(v) < 0$	$(v > 1)$
(iv)	$\varrho(v) \leqslant 1/\Gamma(v+1)$	$(v \geqslant 0)$.

Démonstration. Le point (i) est une conséquence immédiate de l'équation différentielle aux différences et des conditions initiales de la définition de ϱ. En effet, les deux membres de (i) ont même dérivée pour $v > 1$ et même valeur en $v = 1$.

Montrons (ii). Soit $\tau := \inf\{v : \varrho(v) = 0\}$. Si τ est fini, alors $\tau > 1$ puisque ϱ est continue et satisfait à $\varrho(v) = 1$ pour $0 \leqslant v \leqslant 1$. Par (i), on peut donc écrire

$$0 = \tau\varrho(\tau) = \int_{\tau-1}^{\tau} \varrho(w)\,\mathrm{d}w.$$

La continuité de ϱ implique alors que le membre de droite de cette égalité est strictement positif — d'après la définition même de τ. Cela montre que τ n'est pas fini, d'où (ii).

Le point (iii) découle immédiatement de (ii) et de l'équation fonctionnelle satisfaite par ϱ.

Nous établissons (iv) par récurrence sur $k := \lfloor v \rfloor$. La propriété est satisfaite pour $k = 0$ puisque $\varrho(v) = 1$ $(0 \leqslant v \leqslant 1)$. Si $k \geqslant 1$, on déduit de (i), (ii), (iii) et de l'hypothèse de récurrence que l'on a

$$\varrho(v) = \frac{1}{v}\int_{v-1}^{v} \varrho(w)\,\mathrm{d}w \leqslant \frac{\varrho(v-1)}{v} \leqslant \frac{1}{v\Gamma(v)} = \frac{1}{\Gamma(v+1)}.$$

□

Les solutions d'équations différentielles aux différences, comme la fonction de Dickman, interviennent dans les problèmes de théorie des nombres parce qu'elles sont des versions continues de quantités arithmétiques qui vérifient des équations analogues

discrètes. La fonction de Dickman est ainsi liée à la fonction $\Psi(x,y)$ que nous avons rencontrée au paragraphe précédent. En classant les entiers n de $S(x,y)$ selon la taille de leur plus grand facteur premier (i.e. en écrivant $n = mp$ avec $P^+(m) \leqslant p \leqslant y$) on obtient immédiatement *l'équation de Buchstab* :

$$\Psi(x,y) = 1 + \sum_{p \leqslant y} \Psi(x/p, p) \quad (x \geqslant 1,\ y \geqslant 1).$$

L'analogie entre l'équation de Buchstab et l'équation différentielle aux différences définissant la fonction de Dickman fournit le principe du théorème suivant.

Théorème 7.2 *Soit $\varepsilon > 0$. On a uniformément pour $1 \leqslant x^\varepsilon \leqslant y$, $u = (\ln x)/\ln y$,*

$$\Psi(x,y) = x\varrho(u)\Big\{1 + O\Big(\frac{1}{\ln(2x)}\Big)\Big\}.$$

Démonstration. Introduisons, pour chaque entier $k \geqslant 1$, la quantité

$$\Delta_k(x) := \sup_{y \geqslant x^{1/k}} |\Psi(x,y) - x\varrho(u)|.$$

Nous devons montrer que $\Delta_k(x) \ll x/\ln(2x)$ pour tout k fixé : le résultat annoncé en découlera pour le choix $k := \lfloor 1/\varepsilon \rfloor + 1$.

On a trivialement $\Delta_1(x) \ll 1$ puisque $\Psi(x,y) = \lfloor x \rfloor$ pour $y \geqslant x$ et $\varrho(u) = 1$ pour $0 \leqslant u \leqslant 1$. On raisonne ensuite par récurrence sur k. L'équation de Buchstab fournit pour chaque entier $k \geqslant 1$, lorsque $x^{1/(k+1)} \leqslant y < x^{1/k}$,

$$\Psi\big(x, x^{1/k}\big) - \Psi(x,y) = \sum_{y < p \leqslant x^{1/k}} \Psi(x/p, p).$$

En remarquant que $x/p \leqslant p^k$ pour tout $p > x^{1/(k+1)}$, on en déduit que

$$\Delta_{k+1}(x) \leqslant \Delta_k(x) + xR_k(x) + \sum_{x^{1/(k+1)} < p \leqslant x^{1/k}} \Delta_k(x/p)$$

avec

$$R_k(x) := \sup_{k\leqslant u\leqslant k+1}\Big|\varrho(k)-\varrho(u)-\sum_{y<p\leqslant x^{1/k}}\frac{1}{p}\varrho\Big(\frac{\ln x}{\ln p}-1\Big)\Big|.$$

La dernière somme en p relève du Lemme 2.1 d'intégration par parties. Le second théorème de Mertens (cf. p. 34) nous permet en effet d'appliquer ce résultat avec

$$E(t)=\ln_2 t-\ln_2 y,\quad R(t)=1/\ln(2t).$$

Il suit, après changement de variable $t=x^{1/v}$,

$$\sum_{y<p\leqslant x^{1/k}}\frac{1}{p}\varrho\Big(\frac{\ln x}{\ln p}-1\Big)=\int_k^u\frac{\varrho(v-1)}{v}\,\mathrm{d}v+O_k\Big(\frac{1}{\ln(2x)}\Big),$$

d'où, compte tenu de l'équation fonctionnelle satisfaite par ϱ,

$$R_k(x)\ll_k 1/\ln(2x).$$

Cette estimation fournit immédiatement l'évaluation souhaitée pour $\Delta_{k+1}(x)$ en reportant dans l'inégalité de récurrence la majoration

$$\Delta_k(x)\ll_k\begin{cases}1 & (k=1)\\ x/\ln(2x) & (k\geqslant 2).\end{cases}$$

□

Une conséquence inattendue du théorème précédent consiste en une preuve arithmétique de la formule

$$\int_0^\infty\varrho(v)\,\mathrm{d}v=\mathrm{e}^\gamma,$$

où γ désigne la constante d'Euler. Le point de départ est l'identité

$$S(y):=\prod_{p\leqslant y}\Big(1-\frac{1}{p}\Big)^{-1}=\sum_{P^+(n)\leqslant y}\frac{1}{n}=\int_1^\infty\frac{\Psi(x,y)}{x^2}\,\mathrm{d}x.$$

La formule de Mertens (cf. p. 33) fournit d'une part l'évaluation

$$S(y) = \mathrm{e}^{\gamma} \ln y + O(1) \quad (y \to \infty).$$

D'autre part, le Théorème 7.2 permet d'écrire pour tout entier $k \geqslant 1$ fixé

$$\int_1^{y^k} \frac{\Psi(x,y)}{x^2}\,\mathrm{d}x = (\ln y) \int_0^k \varrho(u)\,\mathrm{d}u + O_k(1)$$

alors que le Théorème 6.1 procure la majoration

$$\int_{y^k}^{\infty} \frac{\Psi(x,y)}{x^2}\,\mathrm{d}x \ll \int_{y^k}^{\infty} \mathrm{e}^{-u/2} \frac{\mathrm{d}x}{x} \ll \mathrm{e}^{-k/2} \ln y.$$

Il suit

$$\mathrm{e}^{\gamma} = \int_0^k \varrho(u)\,\mathrm{d}u + O_k(1/\ln y) + O\left(\mathrm{e}^{-k/2}\right).$$

La formule annoncée en découle en faisant tendre y puis k vers l'infini.

8. La preuve de Daboussi, revisitée

Soit

$$\alpha := \limsup_{x\to\infty} |M(x)|/x.$$

On a $0 \leqslant \alpha \leqslant 1$ et nous devons établir que $\alpha = 0$. D'après la Proposition 6.3, on a

$$\alpha \leqslant \prod_{p\leqslant y} \left(1 - \frac{1}{p}\right)\Big\{I_1(y) + I_2(y)\Big\}$$

avec

$$I_1(y) := \int_1^y \frac{|M(x)|}{x^2}\,\mathrm{d}x, \quad I_2(y) := \int_y^{\infty} \frac{|M(x,y)|}{x^2}\,\mathrm{d}x.$$

Nous allons montrer que pour tout $\beta > \alpha$, on a

$$\limsup_{y\to\infty} I_1(y)/\ln y \leqslant \beta\delta,$$
$$\limsup_{y\to\infty} I_2(y)/\ln y \leqslant \beta(\mathrm{e}^{\gamma}-1),$$

où $\delta = \delta(\alpha)$ est une quantité < 1 si $\alpha > 0$. En reportant dans notre inégalité pour α et en faisant tendre y vers l'infini, cela implique que $\alpha \leqslant \mathrm{e}^{-\gamma}\{\beta\delta + \beta(\mathrm{e}^{\gamma}-1)\}$. Comme β peut être choisi arbitrairement proche de α, il suit $\alpha \leqslant \alpha\{1-\mathrm{e}^{-\gamma}(1-\delta)\}$, et finalement $\alpha = 0$.

Michel Mendès France
& Hédi Daboussi vers 1990

L'inégalité annoncée pour $I_1(y)$ résulte simplement de l'estimation

$$\sup_{1\leqslant a\leqslant b}\left|\int_a^b M(x)\frac{\mathrm{d}x}{x^2}\right| \leqslant 4$$

obtenue au § 4. En effet, nous allons voir que l'estimation requise a lieu pour $\delta := 1-\frac{1}{16}\alpha^2 \geqslant \frac{15}{16}$. Considérons à cette fin un nombre réel $x_0 = x_0(\beta)$ tel que $|M(x)| \leqslant \beta x$ pour tout $x \geqslant x_0$. Pour chaque $y \geqslant x_0$, on peut scinder l'intervalle $[x_0, y[$ en un nombre fini de sous-intervalles à bornes entières $[x_j, x_{j+1}[$ $(0 \leqslant j \leqslant J)$ à l'intérieur desquels $M(x)$ est de signe constant. Comme $M(x)$ est à valeurs entières et varie par sauts de ± 1, on a certainement, pour $0 < j \leqslant J$,

$$M(x_j) = 0,$$
$$T_j := \int_{x_j}^{x_{j+1}} \frac{|M(x)|}{x^2}\,\mathrm{d}x = \left|\int_{x_j}^{x_{j+1}} \frac{M(x)}{x^2}\,\mathrm{d}x\right| \leqslant 4.$$

Maintenant, remarquons que, pour $x_j \leqslant x \leqslant x_{j+1}$, on a

$$|M(x)| = |M(x) - M(x_j)| \leqslant x - x_j.$$

Posant $\lambda_j = \ln(x_{j+1}/x_j)$, et effectuant le changement de variable $x = tx_j$ dans l'intégrale, il suit

$$T_j \leqslant T_j^* := \min\Big(4, \int_1^{\exp\lambda_j} \min(\beta, 1 - 1/t)\frac{\mathrm{d}t}{t}\Big).$$

Nous allons prouver que $T_j^* \leqslant \beta\delta\lambda_j$, ce qui suffit pleinement à établir le résultat souhaité concernant $I_1(y)$. Si $\beta\lambda_j \geqslant 8$, on a

$$T_j^* \leqslant 4 \leqslant \tfrac{1}{2}\beta\lambda_j \leqslant \beta\delta\lambda_j.$$

On peut donc supposer $\beta\lambda_j < 8$. Grâce à l'inégalité $1 - 1/t \leqslant \ln t$, on obtient alors

$$T_j^* \leqslant \begin{cases} \displaystyle\int_1^{\exp\lambda_j} \ln t\,\frac{\mathrm{d}t}{t} = \tfrac{1}{2}\lambda_j^2 \leqslant \tfrac{1}{2}\beta\lambda_j, & \text{si } \lambda_j \leqslant \beta, \\ \beta(\lambda_j - \tfrac{1}{2}\beta) \leqslant (1 - \tfrac{1}{16}\beta^2)\beta\lambda_j, & \text{si } \lambda_j > \beta, \end{cases}$$

la seconde majoration étant obtenue en scindant l'intégrale à $t = \mathrm{e}^{\beta}$. Nous avons donc $T_j^* \leqslant \beta\delta\lambda_j$ en toute circonstance, et cela fournit la majoration annoncée pour $I_1(y)$.

L'inégalité relative à $I_2(y)$ est une conséquence facile du résultat suivant.

Lemme 8.1 *Soient $\varepsilon > 0$, $\beta > \alpha$. On a uniformément pour $1 \leqslant x^{\varepsilon} \leqslant y$, $u = (\ln x)/\ln y$,*

$$|M(x,y)| \leqslant \beta x\varrho(u)\Big\{1 + O\Big(\frac{1}{\ln(2x)}\Big)\Big\}.$$

Admettons momentanément cet énoncé et déduisons-en l'estimation requise. Grâce à la majoration triviale $|M(x,y)| \leqslant \Psi(x,y)$ et

au Théorème 6.1, il suit en effet, pour tout entier $k \geqslant 1$,

$$\int_y^{y^k} \frac{|M(x,y)|}{x^2}\,\mathrm{d}x \leqslant \beta\{\ln y + O_k(1)\}\int_1^k \varrho(u)\,\mathrm{d}u,$$
$$\int_{y^k}^{\infty} \frac{|M(x,y)|}{x^2}\,\mathrm{d}x \ll (\ln y)\int_k^{\infty} \mathrm{e}^{-u/2}\,\mathrm{d}u \ll \mathrm{e}^{-k/2}\ln y.$$

En faisant tendre y puis k vers l'infini, cela implique immédiatement

$$\limsup_{y\to\infty} I_2(y)/\ln y \leqslant \beta(\mathrm{e}^{\gamma}-1),$$

puisque

$$\int_1^{\infty} \varrho(u)\,\mathrm{d}u = \int_0^{\infty} \varrho(u)\,\mathrm{d}u - 1 = \mathrm{e}^{\gamma}-1.$$

Démonstration du 8.1. Nous employons une technique analogue à celle de la preuve du Théorème 7.2 en observant que $M(x,y)$ satisfait aussi une équation fonctionnelle. L'idée consiste à exploiter cette relation sous la forme d'une inéquation qui permet, au prix d'un affaiblissement du terme d'erreur, de propager la majoration

$$|M(x,y)| \leqslant \beta\varrho(u)x + O(1),$$

initialement valable pour $y \geqslant x$, à tout domaine $y \geqslant x^{\varepsilon}$. Toutefois, on ne peut utiliser directement l'équation de Buchstab[1]

$$M(x,y) = 1 - \sum_{p\leqslant y} M(x/p, p-1)$$

car la majoration de récurrence qui en résulte, à savoir

$$|M(x,y)| \leqslant |M(x,x^{1/k})| + \sum_{y<p\leqslant x^{1/k}} |M(x/p,p-1)|,$$

n'a pas $x\varrho(u)$ pour solution asymptotique.

1. La preuve de cette identité est semblable à celle de l'équation de Buchstab pour $\Psi(x,y)$ établie p. 138 : il suffit ici de classer les entiers selon leur plus petit facteur premier.

On contourne cette difficulté technique en évaluant la somme pondérée

$$\begin{aligned}\sum_{n\in S(x,y)}\mu(n)\ln n &= \sum_{p\leqslant y}\sum_{\substack{m\in S(x/p,y)\\ p\nmid m}}\mu(mp)\ln p\\ &= -\sum_{p\leqslant y}\{M(x/p,y)+O(x/p^2)\}\ln p\\ &= -\sum_{p\leqslant y}M(x/p,y)\ln p+O(x).\end{aligned}$$

Observant par ailleurs que

$$\Big|\sum_{n\in S(x,y)}\mu(n)\ln(x/n)\Big|\leqslant\sum_{n\leqslant x}\ln(x/n)\ll x,$$

et donc $\sum_{n\in S(x,y)}\mu(n)\ln n = M(x,y)\ln x+O(x)$, on déduit de ce qui précède que

$$M(x,y)\ln x = -\sum_{p\leqslant y}M(x/p,y)\ln p+O(x).$$

En particulier, il existe une constante absolue C telle que l'on ait pour $x\geqslant 1$, $y\geqslant 1$

$$|M(x,y)|\ln x\leqslant\sum_{p\leqslant y}|M(x/p,y)|\ln p+Cx.$$

Nous allons utiliser cette inéquation fonctionnelle pour déduire de la majoration

$$|M(x,y)| = |M(x)|\leqslant\beta x\quad(y\geqslant x\geqslant x_0(\beta))$$

une estimation valable pour $y\geqslant x^{\varepsilon}$. À cette fin, nous introduisons les quantités

$$h(v) := |M(y^v,y)|/\beta\varrho(v)y^v,\quad h^*(u) := \sup_{u_0\leqslant v\leqslant u}h(v),$$

où $u_0 = u_0(\beta) = \ln x_0(\beta)/\ln y$. Notre hypothèse peut donc encore s'exprimer sous la forme simple

$$h^*(1) \leqslant 1,$$

alors que l'énoncé du lemme équivaut à

$$h^*(u) \leqslant 1 + O_k(1/\ln y) \quad (1 \leqslant u \leqslant k).^{(1)}$$

Considérons des nombres réels $u \in [1,k]$ et $v \in [u_0,u]$. Appliquée avec $x = y^v$, notre inéquation fonctionnelle pour $|M(x,y)|$ fournit, après division par $\beta\varrho(v)y^v \ln y^v$,

$$h(v) \leqslant \sum_{p\leqslant y,\, y^v/p>x_0} \frac{h^*(u-v_p)\varrho(v-v_p)\ln p}{p\varrho(v)\ln y^v} + \sum_{p\leqslant y,\, y^v/p\leqslant x_0} \frac{|M(y^v/p,y)|\ln p}{\beta\varrho(v)y^v\ln y^v} + \frac{C}{\beta\varrho(v)\ln y^v},$$

où l'on a posé $v_p := (\ln p)/\ln y$. Notons que $M(y^v/p,y) = 0$ si $p > y^v$. Le premier théorème de Mertens permet de majorer la seconde somme en p par

$$\sum_{y^v/x_0\leqslant p\leqslant y^v} \frac{\ln p}{\beta p\varrho(v)\ln y^v} \ll \frac{1}{v\varrho(v)\ln y}.$$

Nous avons donc montré l'existence d'une constante $K(\beta)$ telle que

$$h(v) \leqslant \sum_{p\leqslant y} \frac{h^*(u-v_p)\varrho(v-v_p)\ln p}{pv\varrho(v)\ln y} + \frac{K(\beta)}{v\varrho(v)\ln y}.$$

À ce stade, nous utilisons le fait que $h^*(u)$ est une fonction croissante de u. Posant

$$S_\vartheta := \sum_{p\leqslant y^\vartheta} \frac{\varrho(v-v_p)\ln p}{v\varrho(v)p\ln y} \quad (0\leqslant\vartheta\leqslant 1),$$

1. Le résultat souhaité découle formellement de cette évaluation avec $k := \lfloor 1/\varepsilon\rfloor + 1$.

nous déduisons de l'inégalité précédente que

$$h(v) \leqslant h^*(u)S_{\frac{1}{2}} + h^*(u-\tfrac{1}{2})\{S_1 - S_{\frac{1}{2}}\} + \frac{K(\beta)}{v\varrho(v)\ln y}\cdot$$

Nous pouvons évaluer S_ϑ grâce au Lemme 2.1 en posant

$$a_p = (\ln p)/p\ln y, \;\; b(t) = \varrho(v - \ln t/\ln y)/v\varrho(v).$$

Le premier théorème de Mertens fournit la formule asymptotique $\sum_{p\leqslant y^t} a_p = t + O(1/\ln y)$, et l'on obtient

$$S_\vartheta = r(\vartheta) + O_k(1/\ln y),$$

avec

$$r(\vartheta) := \frac{1}{v\varrho(v)}\int_0^\vartheta \varrho(v-t)\,\mathrm{d}t.$$

On déduit du Théorème 7.1(i) que $r(1) = 1$ et la décroissance de la fonction ϱ implique $r(\frac{1}{2}) \leqslant 1 - r(\frac{1}{2})$, soit $r(\frac{1}{2}) \leqslant \frac{1}{2}$. En reportant dans l'inégalité liant $h(v)$, $h^*(u)$ et $h^*(u-\frac{1}{2})$ et en prenant le supremum en v, il suit pour $k \geqslant 1$ et $1 \leqslant u \leqslant k$

$$h^*(u) \leqslant h^*(u-\tfrac{1}{2}) + O_k(1/\ln y).$$

Une application inductive de cette inégalité fournit finalement

$$h^*(u) \leqslant h^*(1) + O_k(1/\ln y) \leqslant 1 + O_k(1/\ln y),$$

ce qu'il fallait démontrer. □

Chapitre 5
Les grandes conjectures

Si la théorie des nombres premiers doit une large part de son pouvoir de fascination aux mystères dont elle demeure parsemée, elle avance surtout grâce aux conjectures qui, au fil des siècles, ont forgé notre philosophie de l'objet et précisé les contours des buts à atteindre.

À tout seigneur tout honneur. Nous avons déjà mentionné, au Chapitre 2, l'hypothèse de Riemann, dont une forme élémentaire équivalente est

$$(\forall\, \varepsilon > 0) \quad \pi(x) = \mathrm{li}(x) + O(x^{1/2+\varepsilon}).$$

On sait qu'un tel résultat, s'il est vérifié, serait optimal et que la différence

$$\Delta(x) = \pi(x) - \mathrm{li}(x)$$

change de signe infiniment souvent et possède des oscillations « comparables » à $\sqrt{x}$.[1] Le premier changement de signe est certainement supérieur à 10^{10} et, ainsi que l'a montré te Riele (1986), inférieur à 7×10^{370}.

1. Voir le 2.7, p. 69.

On peut cependant se demander quelle est la fréquence de ces oscillations. Un raisonnement heuristique sur la base d'un modèle de type Cramér laisse supposer que, pour tout X assez grand, il y a au moins un changement de signe dans chaque intervalle $]X, 2X]$. Le meilleur résultat connu dans cette direction, dû à Kaczorowski (1985), établit cette hypothèse en moyenne : pour une constante convenable $c > 0$, il y a au moins $c \log X$ changements de signe dans l'intervalle $[2, X]$.

À l'appui de la philosophie du « tout ce qui est possible se réalise », nous avons déjà mentionné, au § 1.11 (p. 41) la conjecture des nombres premiers jumeaux. En 1923, Hardy & Littlewood ont proposé la généralisation suivante : pour tous k-uples d'entiers positifs ou nuls $\{a_j\}_{j=1}^k$, $\{b_j\}_{j=1}^k$ ayant la propriété que le polynôme

$$P(\xi) = \prod_{1 \leqslant j \leqslant k} (a_j \xi + b_j)$$

ne s'annule identiquement modulo p pour aucun nombre premier p, il existe une infinité de nombres entiers n tels que les nombres $a_j n + b_j$ $(1 \leqslant j \leqslant k)$ soient simultanément premiers. C'est la *conjecture des nombres premiers jumeaux généralisée.*[1]

On peut facilement étendre le calcul heuristique présenté au § 1.11 pour fournir un équivalent asymptotique de la forme $Cx/(\log x)^k$ pour le nombre des nombres premiers du type indiqué n'excédant pas x. Ainsi, cette conjecture contient, sous une forme quantitative,[2] à la fois le théorème de Dirichlet sur les progressions arithmétiques et la conjecture des nombres premiers jumeaux.

Cette extension a été placée par Schinzel & Sierpiski en 1958 dans un cadre plus général encore qui lui confère un statut quasi définitif. Leur conjecture, connue sous le nom d'*hypothèse H* s'énonce comme suit.

1. La conjecture classique des nombres premiers jumeaux correspond donc à $k = 2$, avec $(a_1, b_1) = (1, 0)$, $(a_2, b_2) = (1, 2)$.

2. La forme qualitative est due à Dickson (1904).

Hypothèse *H*. *Soit $\{Q_j(\xi)\}_{j=1}^k$ une suite finie de polynômes irréductibles à coefficients entiers. On note $\varrho(p)$ le nombre de racines modulo p du polynôme $Q(\xi) := \prod_{j=1}^k Q_j(\xi)$. Alors le nombre $\pi_Q(x)$ d'entiers $n \leqslant x$ tels que $|Q_j(n)|$ soit premier pour $1 \leqslant j \leqslant k$ satisfait à la relation asymptotique*

$$\pi_Q(x) = \{C_Q + o(1)\}\, x \prod_{1 \leqslant j \leqslant k} \frac{1}{\log |Q_j(x)|} \qquad (x \to \infty),$$

où l'on a posé

$$C_Q := \prod_p \Big(1 - \frac{\varrho(p)}{p}\Big)\Big(1 - \frac{1}{p}\Big)^{-k}.$$

On peut montrer grâce à une généralisation adéquate du théorème des nombres premiers dans les extensions algébriques de $\mathbb{Q}$, que le produit définissant C_Q converge.

Le cas trivial où $Q(n)$ possède des facteurs fixes (i.e. où $Q(n)$ est divisible par un nombre premier fixe pour tout n) est couvert par l'énoncé puisqu'alors $C_Q = 0$.

Lorsque $Q(\xi) = \prod_{h \in \mathcal{H}}(\xi + h)$ où $\mathcal{H}$ est un ensemble fini de nombres entiers, on reconnaît

$$C_Q = \mathfrak{S}(\mathcal{H}),$$

tel que défini en (3.8) p. 100.

La théorie moderne du crible[1] fournit pour $\pi_Q(x)$ des majorations du « bon » ordre de grandeur, c'est-à-dire une expression du type conjecturé, mais avec une constante éventuellement plus grande que C_Q.[2] Pour les bornes inférieures, il faut se contenter

1. Voir en particulier l'ouvrage de Halberstam & Richert *Sieve methods*, Academic Press 1974.

2. C'est par exemple le cas pour la quantité (3.7) p. 100, qui constitue une majoration du nombre des k-uples de nombres premiers de la forme $(n + h_1, \ldots, n + h_k)$ lorsque $\mathcal{H} := \{h_1, \ldots, h_k\}$ est un système admissible. Compte tenu de la limitation en R imposée par le théorème de Bombieri–Vinogradov, on constate que la majoration excède l'équivalent asymptotique conjecturé d'un facteur $2^k k! + o(1)$.

de remplacer $\pi_Q(x)$ par le nombre des entiers $n \leqslant x$ pour lesquels $Q(n)$ ne possède (au plus) qu'un nombre borné $K > k$ de facteurs premiers. Dans le cas de $k = 4$ polynômes linéaires, par exemple, on peut choisir $K = 15$.

Il est particulièrement agaçant de ne pas savoir s'il existe ou non une infinité de nombres premiers de la forme $n^2 + 1$ — une conséquence très affaiblie du cas le plus simple de l'hypothèse H après celui des polynômes de degré 1.

Plusieurs voies ont été explorées pour aborder ce problème par une variante moins difficile.

On a cherché des minorations pour le plus grand facteur premier $P^+(Q(n))$ de valeurs polynomiales. Ainsi Deshouillers & Iwaniec (1982), améliorant un résultat de Hooley, ont montré que l'on a $P^+(n^2 + 1) > n^{6/5}$ infiniment souvent. Heath-Brown (1999) a obtenu un résultat qualitativement comparable pour $P^+(n^3 + 2)$. Dans le cas d'un polynôme irréductible $Q(n)$ de degré arbitraire, la meilleure estimation connue est celle de Tenenbaum (1990), améliorant un résultat d'Erdős et Schinzel et fournissant, pour une infinité de valeurs de n, la minoration

$$P^+(Q(n)) > n \exp\{(\log n)^\alpha\}$$

dès que $\alpha < 2 - \log 4 \approx 0{,}6137$.

Paul Erdős & Gérald Tenenbaum vers 1990

La principale difficulté inhérente au problème de trouver des nombres premiers dans les suites polynomiales est due à la vitesse de croissance : il n'y a que $O(\sqrt{x})$ valeurs d'un polynôme quadratique dans les entiers $\leqslant x$, et ce nombre est trop faible pour que les techniques actuellement disponibles permettent d'exhiber un nombre premier parmi ces valeurs.[1]

Piatetski-Shapiro (1953) a proposé de rechercher des nombres premiers dans la suite $\{\lfloor n^c \rfloor\}_{n=1}^{\infty}$ avec $c > 1$, $c \notin \mathbb{N}$, ce qui fournit un champ d'investigation continu entre le cas des polynômes linéaires (Dirichlet) et celui des polynômes quadratiques (hypothèse H). Il a obtenu une formule asymptotique pour $c < 12/11$. Le record actuel pour l'existence d'une infinité de nombres premiers $p = \lfloor n^c \rfloor$ est dû à Rivat et Wu (2001), avec $c = 243/205$.

Une voie analogue consiste à rechercher des nombres premiers dans les valeurs de polynômes à plusieurs variables, disons du type

$$Q(n_1, \ldots, n_k) = \sum_{1 \leqslant j \leqslant k} a_j n_j^{d_j}.$$

Il s'agit alors de rendre la quantité $b = \sum_{1 \leqslant j \leqslant k} 1/d_j$, qui est comparable à $1/c$ dans le problème de Piatetski-Shapiro, aussi petite que possible. Friedlander & Iwaniec (1996) sont parvenus à une réponse positive pour le polynôme $n_1^2 + n_2^4$, ce qui représente une avancée tout à fait remarquable dans ce domaine. Heath-Brown (1999) a pu traiter le cas $Q(n_1, n_2) := n_1^3 + 2n_2^3$, avec donc $b = 2/3$. Dans les deux situations, les auteurs fournissent même une formule asymptotique pour le cardinal des nombres premiers de la forme indiquée n'excédant pas x.

Une vieille conjecture concernant la répartition des nombres premiers affirme que la fonction $x \mapsto \pi(x)$ est *sous-additive*, en d'autres termes que

$$\pi(x + y) \leqslant \pi(x) + \pi(y) \qquad (x \geqslant 2, y \geqslant 2).$$

1. La difficulté est comparable à celle du problème des nombres premiers $\leqslant x$ dans une progression arithmétique de raison $q \geqslant \sqrt{x}$. Il est d'ailleurs intéressant de noter que c'est précisément la limite issue de l'hypothèse de Riemann généralisée.

Cette inégalité est asymptotiquement raisonnable au vu du théorème des nombres premiers, et Montgomery & Vaughan ont effectivement établi en 1973 que l'on a

$$\pi(x+y) \leqslant \pi(x) + 2\pi(y) \qquad (x \geqslant 2, y \geqslant 2).$$

Cependant la sous-linéarité est probablement fausse : Hensley & Richards ont montré, également en 1973, qu'elle est incompatible avec la conjecture des nombres premiers jumeaux généralisée, et l'opinion générale est largement en faveur de cette dernière.

Une autre conjecture ultra-célèbre, émise par Goldbach en 1742 dans deux lettres adressées à Euler, peut être vue comme duale de la conjecture des nombres premiers jumeaux. Au lieu de considérer les différences $p - q$ de nombres premiers (et de conjecturer que la plus petite valeur pour laquelle il n'y a pas d'obstruction évidente est atteinte infiniment souvent) on forme les sommes $p + q$ et l'on conjecture que, puisque la seule contrainte manifeste est que ces nombres soient pairs (lorsque p et q sont des nombres premiers > 2), tous les nombres pairs sont effectivement représentables ainsi. La *conjecture de Goldbach* est donc que *tout nombre pair* > 2 *est somme de deux nombres premiers*, ce qui implique trivialement que tout nombre impair > 5 est somme de trois nombres premiers.

Le nombre $R_k(n)$ de représentations d'un entier naturel n sous la forme

$$n = \sum_{1 \leqslant j \leqslant k} p_j$$

Hugh L. Montgomery

Robert C. Vaughan

satisfait évidemment à

$$R_k(n) = \int_0^1 \Big(\sum_{p \leqslant n} \mathrm{e}^{2\pi i \alpha p} \Big)^k \mathrm{e}^{-2\pi i \alpha n} \, \mathrm{d}\alpha.$$

On voit ici apparaître la somme trigonométrique $S_N(\alpha)$ étudiée au § 3.7 avec $N = \pi(n)$. Nous savons que $S_N(\alpha)$ est pratiquement de l'ordre de N lorsque α possède une « bonne » approximation rationnelle avec un « petit » dénominateur, et que $S_N(\alpha) = o(N)$ dans le cas contraire. Ces observations ont conduit Hardy & Littlewood, dans les années 1920, à évaluer asymptotiquement l'intégrale en mettant en évidence une contribution dominante issue des voisinages des nombres rationnels à petits dénominateurs. C'est la fameuse *méthode du cercle*, dont les applications en arithmétique ne se comptent plus.[1] I.M. Vinogradov a ainsi montré en 1937, comme conséquence de ses estimations sur $S_N(\alpha)$, que tout nombre impair assez grand est effectivement somme de trois nombres premiers et que presque tout nombre pair est somme de deux nombres premiers : le nombre d'exceptions n'excédant pas x est $o(x)$, et Montgomery& Vaughan (1975) ont amélioré cette estimation en $O(x^{1-\delta})$ pour un $\delta > 0$ convenable.

En utilisant une méthode de crible,[2] le mathématicien chinois Chen Jing-Run a montré en 1973 que tout entier pair assez grand est représentable sous la forme $p + P_2$ où p est premier et P_2 est un entier ayant au plus deux facteurs premiers.

Tant par le résultat de Chen que celui de Vinogradov, on pourrait croire proche la solution complète de la conjecture de Goldbach. Mais il y a loin de la coupe aux lèvres : la présence de trois variables au lieu de deux est, dans chacune des deux méthodes, une contrainte essentielle, et les spécialistes s'accordent pour penser que le plus dur reste à faire.

1. Voir le livre de R.C. Vaughan, *The Hardy–Littlewood method*, Cambridge University Press, 1981.

2. En fait une version pondérée du crible de Selberg, cf. l'ouvrage cité plus haut de Halberstam & Richert.

La répartition dans les progressions arithmétiques fait évidemment partie des grandes questions qui restent à élucider sur les nombres premiers. Nous avons déjà mentionné (§ 3.3, p. 88), l'hypothèse de Riemann généralisée aux fonctions $L(s,\chi)$ et la conjecture d'Elliott–Halberstam qui impliquerait, en moyenne, une plus grande régularité encore.

L'estimation du plus petit nombre premier, disons $P(a,q)$, dans la progression $p \equiv a \pmod q$ est une question fine et difficile participant de la même problématique. Un argument probabiliste suggère que la quantité

$$P^*(q) := \max_{(a,q)=1} P(a,q)$$

est de l'ordre de $\varphi(q)(\log q)^2$ alors que l'hypothèse de Riemann généralisée fournit $P^*(q) \ll q^{2+\varepsilon}$ pour tout $\varepsilon > 0$. Linnik (1944) a établi l'existence d'une constante L telle que $P^*(q) \ll q^L$ et Heath-Brown (1992) a montré que l'on peut choisir $L \leqslant 11/2$.

Au chapitre des problèmes d'irrégularité de répartition, il faut mentionner une curieuse conjecture de Tchébychev affirmant qu'en un certain sens il y a plus de nombres premiers de la forme $4m+3$ que de la forme $4m+1$, ce qui est effectivement fortement suggéré par les tables.

Cette affirmation ne peut être prise en un sens trop strict puisqu'il est possible d'établir, avec des moyens analogues à ceux employés par Littlewood pour étudier les oscillations de $\pi(x)-\mathrm{li}(x)$, que la différence $\pi(x;1,4)-\pi(x;3,4)$ change de signe infiniment souvent. Cependant, Tchébychev a émis l'hypothèse que, si l'on note

$$T(\sigma,\chi) := \sum_{p>2} (-1)^{(p-1)/2} \mathrm{e}^{-\sigma p}$$

alors

$$\lim_{\sigma\to 0+} T(\sigma,\chi) = -\infty.$$

Un tel résultat confirmerait effectivement la prépondérance, en un sens assez subtil, des nombres premiers du type $4m + 3$. Hardy & Littlewood ont montré en 1918 que cela découlerait de l'hypothèse de Riemann pour la fonction L associée à l'unique caractère non trivial modulo 4, soit

$$L(s, \chi) = \sum_{n \geqslant 0} \frac{(-1)^n}{(2n+1)^s}.$$

Landau avait remarqué dès 1918 que, plus généralement, la formule explicite pour $\pi(x; a, q)$ fait apparaître un premier terme résiduel positif occasionnellement prépondérant si a est non-résidu quadratique modulo q. Sous l'hypothèse de Riemann généralisée, la présence de ce terme suffit à impliquer la conjecture de Tchébychev.

Landau (1918) a également établi la réciproque : si $L(s, \chi)$ possède un zéro dans le demi-plan $\sigma > \frac{1}{2}$, alors

$$\limsup_{\sigma \to 0+} T(\sigma, \chi) = +\infty \quad \text{et} \quad \liminf_{\sigma \to 0+} T(\sigma, \chi) = -\infty.$$

Les résultats modernes sur ce problème vont tous dans la direction sinon d'une réfutation de l'hypothèse de Tchébychev, du moins dans celle de l'existence de limitations à la prépondérance de certaines classes de résidus sur d'autres. Par exemple, Kaczorowski a montré en 1991 que, si $N(X; a, q)$ désigne, pour $(a, q) = 1$, $a \neq 1$, le nombre des entiers $x \leqslant X$ tels que $\pi(x; a, q) \geqslant \pi(x; 1, q)$, alors on a, sous l'hypothèse de Riemann généralisée pour les fonctions L de module q,

$$c_1 X \leqslant N(X; a, q) \leqslant (1 - c_2)X \qquad (X \geqslant X_0),$$

avec des constantes positives convenables c_1 et c_2.

La plupart des autres problèmes profonds et non résolus relatifs aux nombres premiers concernent leur présence dans les suites rares.

Nous avons déjà mentionné le cas des valeurs polynomiales, mais les conjectures sur le comportement statistique des nombres

premiers conduisent à des hypothèses beaucoup plus hardies encore.

Les nombres de la forme $2^m + 1$ ont, depuis longtemps, intrigué les arithméticiens, amateurs et professionnels. L'idée sous-jacente est qu'en translatant un entier, tel que 2^m, de structure multiplicative très éloignée de celle d'un nombre premier, on devrait obtenir un entier sinon premier du moins d'une structure voisine. Il est facile de constater que $2^m + 1$ ne peut être premier que si m est lui-même une puissance de 2,[1] et Fermat a conjecturé que tous les nombres $F_n = 2^{2^n} + 1$ sont premiers. Cette hypothèse, fondée sur le raisonnement heuristique précédent qui tend à « accroître » la probabilité que F_n soit premier, est contraire au raisonnement statistique puisque l'espérance naïve du nombre de valeurs de $n \leqslant N$ telles que F_n soit premier est $\sum_{n \leqslant N} 1/\log F_n \ll 1$.

Euler a observé en 1732 que F_5 est divisible par 641 et l'on sait aujourd'hui que F_n est composé pour $5 \leqslant n \leqslant 21$. On conjecture que tous les F_n sont composés à partir d'un certain rang, mais on ne sait pas montrer qu'une infinité d'entre eux sont composés.

Semblablement, $2^m - 1$ ne peut être premier que si m est lui-même premier. Les nombres de la forme

$$M_p := 2^p - 1$$

sont appelés nombres de Mersenne, d'après le nom de celui qui a donné en 1644 une liste (partiellement inexacte) des valeurs de $p \leqslant 257$ pour lesquelles M_p est premier. Lucas a trouvé une méthode pour tester la primalité des nombres de Mersenne qui sert encore à produire de grands nombres premiers — comme celui cité dans l'avant–propos.

On ignore s'il existe une infinité de nombres premiers de Mersenne, mais l'argument probabiliste suggère une réponse positive. Hooley (1976) a fourni une preuve conditionnelle que, si b est un entier impair avec $|b| > 1$, le nombre $\pi_b(x)$ de valeurs de

1. Si $m = kp$ avec $p > 2$, on a $2^m + 1 = (2^k + 1) \sum_{0 \leqslant j \leqslant p-1} (-1)^j 2^{kj}$.

$n \leqslant x$ telle que $2^n + b \in \mathcal{P}$ satisfait $\pi_b(x) = o(x)$. Il a également établi, inconditionnellement, un résultat analogue concernant les *nombres premiers de Cullen*, i.e. de la forme $K_n := n2^n + 1$: on a l'estimation $|\{n \leqslant x : K_n \in \mathcal{P}\}| = o(x)$.

La liste de conjectures présentée ici est bien entendu loin d'être exhaustive. Certains problèmes mentionnés précédemment dans le texte, comme la conjecture de Cramér, n'ont pas été rediscutés dans ce chapitre. D'autres questions, comme celles d'origine algébrique, ne sont pas abordées. On se consolera en observant que l'entreprise est par nature illimitée : les nombres premiers, le lecteur l'aura compris, sont énigmes avant toute autre chose.

Lectures complémentaires

Z.I. Borevitch & I.R. Chafarevitch, 1967. *Théorie des nombres*, Gauthier–Villars, Paris.

H. Cohen, 1993. *A course in computational algebraic number theory*, Graduate Texts in Mathematics 138, Springer, New York, Heidelberg, Berlin.

H. Davenport, 1980. *Multiplicative number theory*, seconde édition révisée par H. L. Montgomery, Springer, New York, Heidelberg, Berlin.

J.-P. Delahaye, 2000. *Merveilleux nombres premiers*, coll. Pour la science, Belin, Paris.

H.M. Edwards, 2003. *Riemann's Zeta Function*, Dover Publications Inc.

P.D.T.A. Elliott, 1979. *Probabilistic number theory : mean value theorems*, Grundlehren der Math. Wiss. 239, Springer-Verlag, New York, Berlin, Heidelberg.
1980. *Probabilistic number theory : central limit theorems*, Grundlehren der Math. Wiss. 240, Springer-Verlag, New York, Berlin, Heidelberg.

H. Halberstam & H.-E. Richert, 1974. *Sieve methods*, London Mathematical Society Monographs 4, Academic Press, London, New York, San Francisco.

R.R. Hall & G. Tenenbaum, 1988. *Divisors*, Cambridge Tracts in Mathematics 90, Cambridge University Press.

G.H. Hardy & E.M. Wright, 2006. *Introduction à la théorie des nombres*, préface de Catherine Goldstein, traduction de François Sauvageot, Vuibert.

A.E. Ingham, 1990. *The distribution of prime numbers*, préface de R.C. Vaughan, Cambridge University Pres.

A. Ivić, 2003. *The Riemann Zeta-Function : Theory and Applications*, Dover Publications Inc.

M. Kac, 1972. *Statistical independence in probability, analysis and number theory*, Carus Mathematical Monographs 12, Mathematical Association of America, John Wiley & Sons.

N. Koblitz, 2000. *Algebraic aspects of cryptography*, 2e édition, Springer-Verlag, Berlin, Heidelberg.

A.J. Menezes, P.C. van Oorschot & S.A. Vanstone, 2001. *Handbook of Applied Cryptography*, 5e édition, CRC Press, 2001.

H.L. Montgomery & R.C. Vaughan, 2007. *Multiplicative number theory, I. Classical theory*, Cambridge Studies in Advanced Mathematics 97, Cambridge University Press, Cambridge.

I. Niven, H. S. Zuckerman, H. L. Montgomery, 1991. *An introduction to the theory of numbers*, John Wiley New York, Chichester, Brisbane, Toronto, Singapore, xiii+527 pp.

P. Ribenboim, 2000. *Nombres premiers : mystères et records*, Presses Universitaires de France.

G. Robin, 1991. *Algorithmique et cryptographie*, SMAI, coll. Ellipses.

P. Samuel, 1967. *Théorie algébrique des nombres*, Hermann, Paris.

J.-P. Serre, 1968. *Corps locaux*, Hermann, Paris.

G. Tenenbaum, 2008. *Introduction à la théorie analytique et probabiliste des nombres*, troisième édition, coll. Échelles, Belin, 592 pp.

E.C. Titchmarsh, 1951. *The theory of the Riemann zeta-function*, (seconde édition, révisée par D. R. Heath-Brown en 1986) Oxford University Press.

R.C. Vaughan, 1997. *The Hardy-Littlewood Method*, Cambridge Tracts in Mathematics 125, Cambridge University Press.

Index