

# TIPE: La répartition des nombres premiers et ses enjeux

Paul-Alexis Bon, Abel Verley

2020-2021

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Il existe une infinité de nombres premiers</b>	<b>2</b>
2.1	$\mathcal{P}$ est infini : la démonstration d'Euclide . . . . .	2
2.2	Théorème de Dirichlet : progression arithmétique . . . . .	3
2.2.1	Théorie et cas particulier . . . . .	3
2.2.2	Visualisation graphique . . . . .	4
2.3	Un algorithme pour trouver les nombres premiers . . . . .	8
2.3.1	Le crible d'Eratosthène . . . . .	8
2.3.2	Test de primalité . . . . .	10
<b>3</b>	<b>Une répartition encore méconnue</b>	<b>11</b>
3.1	Écart entre 2 nombres premiers . . . . .	12
3.2	Conjecture des nombres jumeaux . . . . .	12
3.3	Conjecture de Goldbach . . . . .	12
<b>4</b>	<b>Une solution au problème de répartition ?</b>	<b>13</b>
4.1	Approximations de $\pi(x)$ . . . . .	13
4.2	L'hypothèse de Riemann . . . . .	16
4.2.1	La fonction $\zeta$ d'Euler . . . . .	16
4.2.2	La fonction $\zeta$ de Riemann . . . . .	17
4.2.3	Brève étude de la fonction $\zeta$ . . . . .	19
<b>5</b>	<b>Application : un petit peu de cryptographie</b>	<b>20</b>
5.1	Chiffrement RSA . . . . .	20
5.1.1	L'anneau $\mathbb{Z}/n\mathbb{Z}$ . . . . .	20
5.1.2	Une brève étude de l'indicateur d'Euler $\varphi$ . . . . .	22
5.1.3	Le chiffrement RSA . . . . .	23
5.2	Le problème de la factorisation : l'algorithme $\rho$ de Pollard . . . . .	26
5.2.1	Algorithmes de recherche d'une période . . . . .	26
5.2.2	Application à la factorisation : . . . . .	28

## 1 Introduction

Depuis l'Antiquité, les nombres premiers ont fasciné les mathématiciens. On sait que l'unité est la brique élémentaire pour construire tout entier  $\geq 1$  à partir de l'addition. Ainsi les nombres premiers forment des briques élémentaires pour la multiplication. En effet on rappelle le théorème suivant :

Théorème : (Théorème fondamental de l'arithmétique) Soit  $n \in \mathbb{Z} \setminus \{-1, 0, 1\}$  alors il existe,  $\varepsilon \in \{-1, 1\}$ ,  $s \in \mathbb{N}^*$ ,  $(p_1, \dots, p_s)$  des nombres premiers tels que  $n = \varepsilon p_1 \dots p_s$ . De plus cette décomposition est unique à l'ordre des facteurs près.

Mais l'intérêt des nombres premiers ne se limite pas à ce résultat théorique et nous verrons au cours de cet exposé que les nombres premiers sont un réel enjeu sociétal sur le plan de la sécurité. Nous désignerons par  $\mathcal{P}$  l'ensemble des nombres premiers. Cet exposé vise à l'étude de  $\mathcal{P}$ , la répartition de ses éléments qui cache des propriétés toutes plus étonnantes les unes que les autres ; ainsi que discuter de leurs applications et leurs enjeux.

Beaucoup d'algorithmes seront détaillés, certains seront explicitement décrits dans ce document, mais tous sont accessibles sur une page GitHub, sur le lien suivant (<https://github.com/abelmaxv/TIPE-2020-2021>). Le lecteur est alors invité à s'appuyer sur cette ressource.

## 2 Il existe une infinité de nombres premiers

### 2.1 $\mathcal{P}$ est infini : la démonstration d'Euclide

Le premier résultat sur la répartition des nombres premiers qui est certes instinctif et élémentaire est essentiel pour la suite de cet exposé. Il s'agit d'affirmer que  $\text{card}(\mathcal{P}) = \infty$ . La démonstration a été proposée par Euclide en 300 av J.C. Cela illustre que la fascination des mathématiciens pour les nombres premiers remonte à l'antiquité.

Théorème :  $\mathcal{P}$  est infini

Démonstration : On raisonne par l'absurde. Supposons  $\mathcal{P}$  fini. On note alors  $\mathcal{P} = \{p_1, \dots, p_s\}$ . On pose

$$N = \prod_{p \in \mathcal{P}} p + 1$$

$N$  admet un diviseur premier  $p_j$ . Or  $p_j \mid \prod_{p \in \mathcal{P}} p$  donc  $p_j \mid 1$ . Ce qui est contradictoire avec  $p_j$  premier. D'où la conclusion.

## 2.2 Théorème de Dirichlet : progression arithmétique

### 2.2.1 Théorie et cas particulier

Le théorème de Dirichlet sur la progression arithmétique consiste en une généralisation la démonstration d'Euclide précédemment énoncée. Gustav Lejeune Dirichlet démontre alors son théorème en 1838. Une deuxième démonstration sera proposée en 1949 par le mathématicien Norvégien Atle Selberg.

Théorème : Pour tout  $a \in \mathbb{N}^*$  et  $b \in \mathbb{Z}$  tels que  $\text{PGCD}(a, b) = 1$  il existe une infinité de nombres premiers  $p$  de tels que  $p \equiv b[a]$

La démonstration du théorème de Dirichlet sur la progression arithmétique repose sur des résultats complexes de la théorie des groupes, des séries de Fourier ainsi que de la fonction  $\zeta$  de Riemann<sup>1</sup>. Le mathématicien allemand Jacobi affirma "En appliquant les séries de Fourier à la théorie des nombres, Dirichlet a récemment trouvé des résultats atteignant les sommets de la perspicacité humaine". Toujours est-il que cette démonstration demeure bien au-delà de notre portée. Nous allons donc porter notre attention sur des cas particuliers plus abordables.

#### Cas particulier 1 : $a = 4$ $b = 3$

Montrons qu'il existe une infinité de nombres premiers  $p$  tels que  $p \equiv 3[4]$   
On raisonne par l'absurde, supposons que  $\mathcal{A} = \{P \in \mathcal{P} | p \equiv 3[4]\}$  est fini. On pose,

$$n = \prod_{p \in \mathcal{A}} p$$

et  $m = 4n - 1$ . Supposons qu'il existe  $\alpha \in \mathcal{A}$  tel que  $\alpha|m$ , nécessairement  $\alpha|n$  donc  $\alpha|1$  ce qui est contradictoire avec  $\alpha$  premier. Par conséquent, tous les diviseurs premiers de  $m$  sont de la forme  $\alpha \equiv 1[4]$ . Il vient alors que  $m \equiv 1[4]$ . Or  $m = 4n - 1 = 4(n - 1) + 3$  donc  $m \equiv 3[4]$ . D'où la contradiction et la conclusion.

#### Cas particulier 2 : $a = 4$ $b = 1$

On rappelle le petit théorème de Fermat :

Théorème : Soit  $n \in \mathbb{N}^*$  et  $p \in \mathcal{P}$  alors  $n^p \equiv n[p]$

Démonstration : On raisonne par récurrence sur  $n$

$n = 1$  :  $n^p = n \equiv n[p]$

$n \rightarrow n + 1$  : Soit  $n$  tel que  $P(n)$ . Montrons que  $P(n + 1)$ .

---

1. cf 4.2.2

D'après le binôme de Newton,

$$(n+1)^p = \sum_{k=0}^p \binom{p}{k} n^k = 1 + \sum_{k=1}^{p-1} \binom{p}{k} n^k + n^p$$

Or pour tout  $k \in \llbracket 1, p-1 \rrbracket$ ,  $\text{PGCD}(k, p) = 1$ . Alors d'après une propriété du cours, on a que  $p \mid \binom{k}{p}$ . Donc  $p \mid \sum_{k=1}^{p-1} \binom{p}{k} n^k$ . Donc

$$(n+1)^p \equiv n^p + 1[p] \equiv n + 1[p] \text{ par HR}$$

D'où  $P(n+1)$  et la conclusion par récurrence.

Montrons alors qu'il existe une infinité de nombres premiers  $p \in \mathcal{P}$  tels que  $p \equiv 1[4]$ .

Soit  $\mathcal{S} = \{p \in \mathcal{P} \mid p \equiv 1[4]\}$

Supposons  $\text{card}(\mathcal{S}) = k \in \mathbb{N}$ . On pose

$$N = 4 \prod_{p \in \mathcal{S}} p^2 + 1$$

$$A = 2 \prod_{p \in \mathcal{S}} p$$

Supposons que  $N$  est premier alors  $N \in \mathcal{S}$  ce qui est contradictoire avec le cardinal de  $\mathcal{S}$ . Donc  $N$  possède un facteur premier  $Q$  et nécessairement,  $Q$  est de la forme  $Q = 4k + 3$ .

On applique le petit théorème de Fermat.  $A^Q \equiv A[Q]$  donc  $A^{4k+3} \equiv A[Q]$  donc  $A^{4k+2} \equiv 1[Q]$ .

Or on a aussi  $A^{4k+2} = (N-1)^{2k+1}$  et  $Q|N$  donc  $N \equiv 0[Q]$  donc  $N-1 \equiv -1[Q]$  donc  $A^{4k+2} \equiv -1[Q]$ . D'où la contradiction et la conclusion.

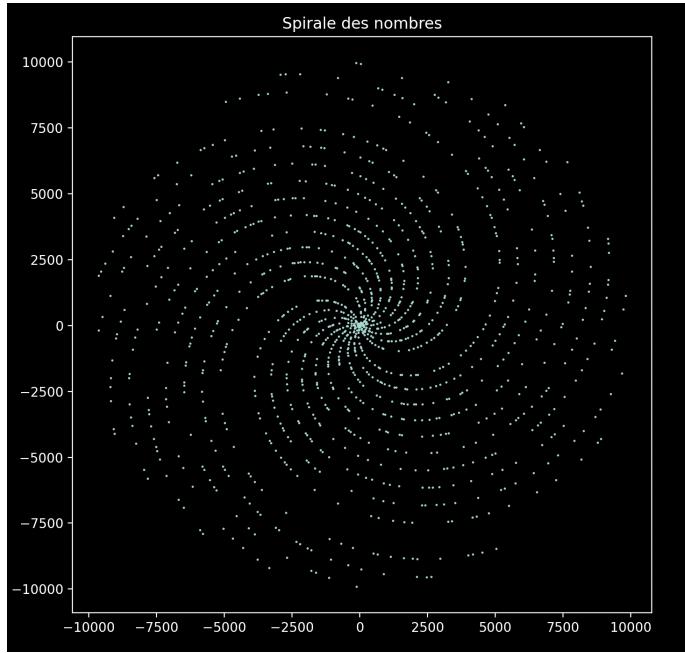
### 2.2.2 Visualisation graphique

La figure représentée ci-dessous a été construite de la façon suivante : à chaque entier  $n$ , on associe le point de module  $n$  et d'argument  $\theta = n \text{ rad}$ . On obtient alors un ensemble de spirales. On peut alors étudier l'ensemble des nombres premiers de manière graphique, en ne plaçant que les points associés aux nombres premiers.

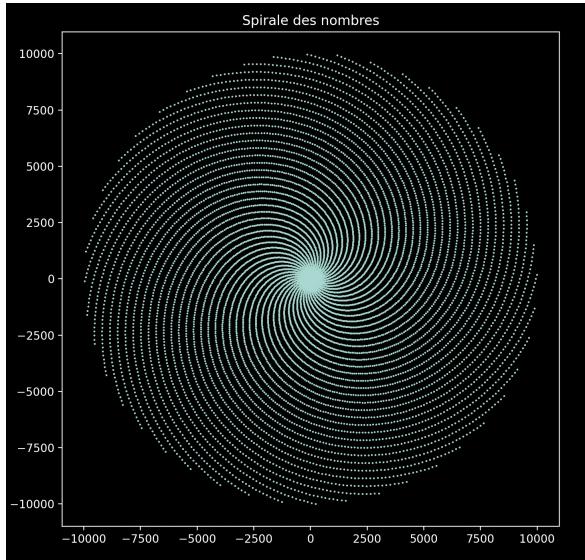
2

---

2. Voir prime\_spirals.py sur Github

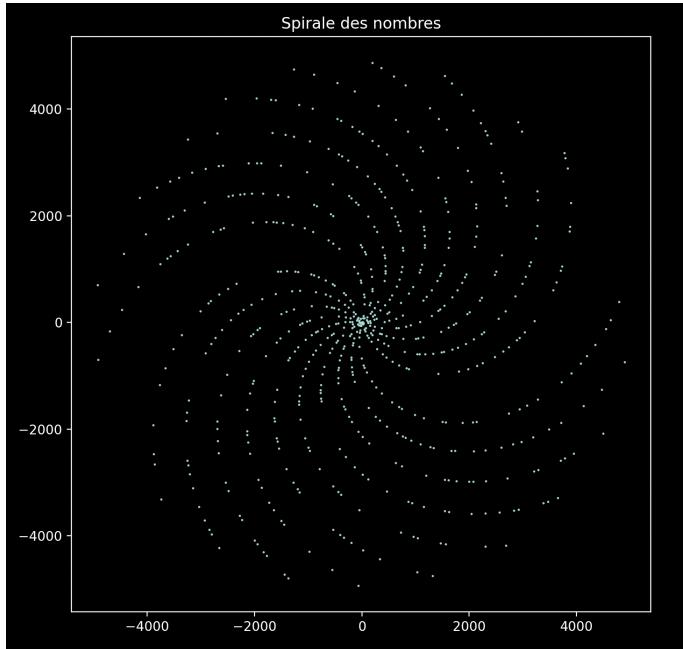


La première chose à laquelle on peut s'intéresser est la formation de spirales, il y en a exactement 20. D'où viennent-elle ? Pour répondre à cette question, on étudie dans un premier temps la figure complète, associée à l'ensemble  $\mathbb{N}$ .



On compte alors 44 spirales. Lorsqu'on regarde la position du 44<sup>eme</sup> point, on remarque que l'on a quasiment effectué 7 rotations, ce qui signifie que  $\frac{44}{7}$

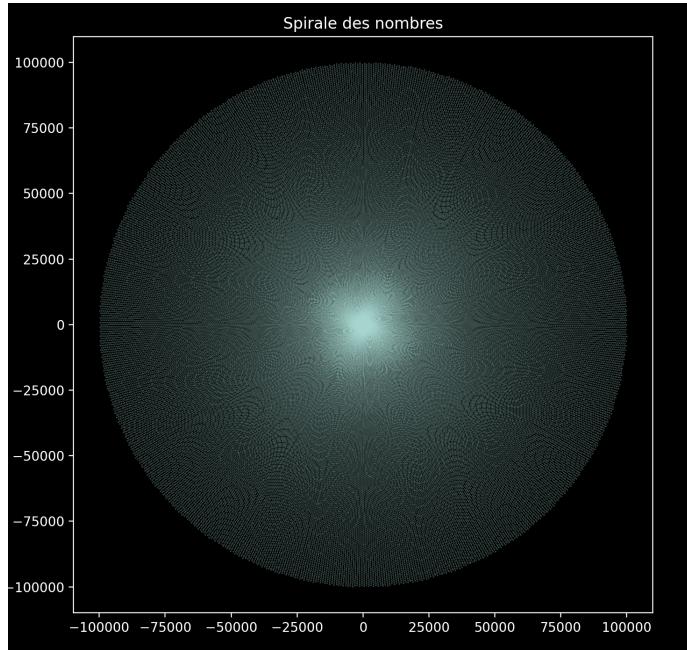
est une bonne approximation de  $2\pi$ . De ce fait, la  $i^{eme}$  spirale est composée des points associés aux entiers  $44k + i$ ,  $k \in \mathbb{N}$ , chacune représentant donc la classe de congruence  $i$  modulo 44. Maintenant, accédons aux nombres premiers. À part 2, aucun n'est pair. On supprime donc les spirales paires (soit 22 spirales). On supprime ensuite la 11<sup>eme</sup> et 33<sup>eme</sup> spirale puisque les éléments de la classe associée à chacune d'elle sont divisibles par 11. Il ne reste alors que 20 spirales, ce qui correspond bien aux spirales observées en ne plaçant que les nombres premiers.



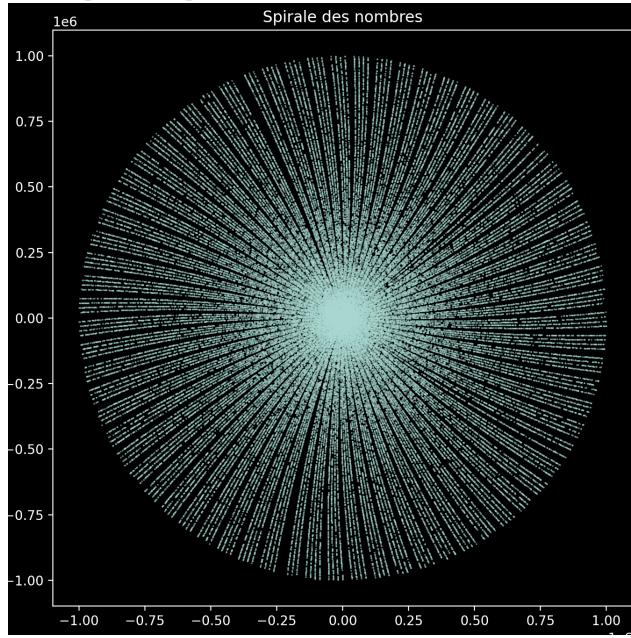
Remarque : Les spirales restantes sont exactement les  $i^{eme}$  spirales, où  $i$  désigne un entier premier avec 44.

Remarque : Il y a exactement  $\varphi(44)$  spirales, où  $\varphi$  désigne l'indicatrice d'Euler (qui sera décrite plus loin), qui correspondent aux nombres d'entiers premiers avec 44 inférieurs ou égaux à 44.

Lorsqu'on prend plus de recul sur la figure, en plaçant plus de points, on voit apparaître des lignes. Cela est dû aux mêmes raisons que les spirales.



Il y a exactement 710 lignes si l'on représente tous les entiers, contre seulement 280 pour les premiers.



En fait, en plaçant le 710<sup>eme</sup> point, on a quasiment effectué 113 rotations. Au lieu d'avoir une fois de plus des spirales, on a ici des lignes, ce qui traduit le fait que  $\frac{710}{113}$  est une meilleure approximation de  $2\pi$  que  $\frac{44}{7}$ . En supprimant les

lignes correspondant aux classes de congruence  $i$ , où  $i$  n'est pas premier avec 710 comme précédemment, on obtient finalement les 280 lignes visibles pour les nombres premiers.

Ces représentations pourraient paraître inutiles et arbitraires mais elles dissimulent en fait un intérêt conceptuel. Il faut remarquer que les nombres premiers semblent équitablement répartis parmi les différentes classes de congruence. En effet, on ne remarque presque aucune irrégularité d'une classe à l'autre et chaque classe semble être infinie. On retrouve alors le théorème de Dirichlet (énoncé précédemment), mais aussi un résultat beaucoup plus puissant qui est le suivant :

$$\lim_{n \rightarrow \infty} \left( \frac{\#\{p \in \mathbb{P} \mid p \leq n \text{ et } p \equiv a [b]\}}{\#\{p \in \mathbb{P} \mid p \leq n\}} \right) = \frac{1}{\varphi(b)}$$

Autrement dit, la densité des nombres premiers est la même dans chaque classe de congruence et vaut  $\frac{1}{\varphi(b)}$  où  $b$  désigne le modulo choisi, c'est-à-dire que les nombres premiers sont équitablement répartis dans ces classes de congruence.

## 2.3 Un algorithme pour trouver les nombres premiers

Nous savons désormais qu'il existe une infinité de nombres premiers. Cependant un nouveau problème se pose : lorsqu'on étudie des nombres suffisamment grands, il devient inenvisageable de tester sa primalité par identification de ses facteurs (à savoir : vérifier si chaque entier inférieur à sa racine le divise). On a alors besoin d'un moyen efficace pour les reconnaître.

### 2.3.1 Le crible d'Eratosthène

La première méthode que nous allons étudier remonte à plus de 2000 ans et comme son nom l'indique elle a été proposée par Eratosthène, notamment célèbre pour avoir proposé une estimation de la circonférence de la Terre.

On cherche à établir la liste des nombres premiers inférieurs à  $n \geq 2$ . Pour se faire on établit la liste des entiers entre 2 et  $n$ . Pour tout  $i \in \llbracket 2, E(\sqrt{n}) + 1 \rrbracket$  on élimine de la liste tous les multiples de  $i$ . La liste résultante est la liste des nombres premiers entre 1 et  $n$ .

On note que si un entier  $n$  est composé ( $n = n_1 n_2$ ) alors un de ses diviseur est inférieur ou égal à  $\sqrt{n}$  c'est pourquoi on arrête l'algorithme à  $E(\sqrt{n}) + 1$ .

On se propose alors d'implémenter l'algorithme en python :

```
def eratosthene(n):
    if n < 2:
        return np.array([])
    is_prime = [True] * (n+1)
    is_prime[0], is_prime[1] = False, False
```

```

for i in range(2, int(sqrt(n))+1):          #(1)
    if is_prime[i]:
        for x in range(i**2, n+1, i):      #(2)
            is_prime[x] = False

return np.array([i for i in range(n+1) if is_prime[i]])

```

Ici la fonction eratosthene prend en argument un entier  $n$  et renvoie la liste des nombres premiers inférieurs à  $n$ .

On peut alors essayer :

```

>>> eratosthene(100)
[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]

```

Intéressons nous à la complexité de ce programme. La création de la liste de booléens se fait en un temps linéaire ( $O(n)$ ) et les deux opérations qui suivent se font en un temps constant ( $O(1)$ ). Dans la boucle (2), on parcourt la liste avec un pas de  $i$  donc on fait  $\frac{n}{i}$  opérations. Or on effectue cette opération pour tout  $i$  premier inférieur à  $\sqrt{n}$  (cf. boucle (1)). On a alors,

$$\frac{n}{2} + \frac{n}{3} + \frac{n}{5} + \frac{n}{7} \dots = n \sum_{p \in \mathcal{P}} \frac{1}{p}$$

On admet alors le résultat suivant,

$$\sum_{p \in \mathcal{P}} \frac{1}{p} = \log \log n + M + o(1)$$

Avec  $M = \gamma + \sum_{p \text{ premier}} \left( \ln \left( 1 - \frac{1}{p} \right) + \frac{1}{p} \right) = 0,261497\dots$ ,  $\gamma$  étant la constante d'Euler (limite de la somme partielle de la série harmonique moins  $\ln n$ ).

Ainsi la complexité du crible d'Eratosthène se fait en  $O(n \log \log n)$ .

On implémente une fonction temps telle que temps(f, n) retourne le temps d'exécution en secondes d'une fonction  $f$  pour une entrée  $n$ . On a alors par exemple :

```

>>> temps(eratosthene, 10**8)
31.567824125289917

```

On se rend compte que pour des nombres supérieurs à l'ordre de grandeur du milliard, il devient difficile d'employer cette méthode en un temps raisonnable. Nous verrons cependant que la manipulation de très grands nombres premiers est essentielle. On étudie alors d'autres méthodes, les tests de primalité.

### 2.3.2 Test de primalité

A l'instar des algorithmes de type "Monte Carlo", les tests de primalité, reposent sur une manière probabiliste de déterminer si un nombre est premier ou non. Son issue n'est donc pas certaine mais ils ont pour avantage d'être nettement moins demandeurs en temps de calcul.

#### Test de Fermat :

On peut reformuler le petit théorème de Fermat de la manière suivante :

Théorème : Soit  $p \in \mathcal{P}$  et  $1 \leq a < p$  alors  $a^{p-1} \equiv a[p]$

Ainsi, si on choisit de manière aléatoire un entier  $a \in \llbracket 1, p - 1 \rrbracket$  et que cette égalité n'est pas respectée, on sait nécessairement que  $p$  n'est pas premier. Réciproquement, si  $p$  est composé, il est tout de même possible d'avoir  $a^{p-1} \equiv 1[p]$ . On dit alors que  $p$  est pseudo-premier de base  $a$ . Si  $p$  est un nombre pseudo-premier avec tout  $a$  inférieur à  $p$  alors on dit que  $p$  est un nombre de Carmichael. Il a alors été démontré qu'il existe une infinité de nombres de Carmichael<sup>3</sup>. Dans le cadre de notre test de primalité, on peut simplement essayer avec une nouvelle valeur de  $a$  et itérer pour augmenter la probabilité de ne pas tomber sur des nombres pseudo-premiers.

On a alors :

```
def fermat (n,k):
    premier = True
    for i in range(k):
        a = random.randrange(2,n-1)
        if expo_rap(a,n-1,n) != 1:
            premier = False
    return premier
```

Avec `expo_rap` une fonction qui calcule efficacement les puissances modulaires.

#### Test de Miller-Rabin :

Pour parer au problème des nombres de Carmichael dans le test de Fermat, le test de Miller-Rabin a été mis en place. Cette méthode se fonde sur la propriété suivante :

Soit  $p > 2 \in \mathcal{P}$ ,  $s$  et  $d$  tels que  $p - 1 = d2^s$  ( $s$  est la valuation 2-adique de  $p - 1$  et donc  $d$  est impair) alors pour tout  $a \in \llbracket 1, p \rrbracket$ ,

$$a^d \equiv 1[p] \text{ ou il existe } i \in \llbracket 0, s - 1 \rrbracket \text{ tel que } a^{d2^i} \equiv -1[p]$$

---

3. Comme 561, 1105 ou 1729 pour citer les trois premiers

Démonstration : A l'aide des méthodes de calcul et de l'intégrité, on montre facilement que dans un corps (comme  $\mathbb{Z}/p\mathbb{Z}^4$ ), l'équation,  $x^2 = 1$  admet pour uniques solutions 1 ou  $(-1)$ .

D'après le petit théorème de Fermat on a  $a^{2^s d} \equiv 1[p]$ . Supposons que pour tout  $i \in \llbracket 0, s-1 \rrbracket$ ,  $a^{d2^i} \not\equiv -1[p]$ . Alors puisqu'on a  $(a^{d2^{s-1}})^2 \equiv 1[p]$  alors  $(a^{d2^{s-1}}) \equiv 1[p]$ . Par hypothèse  $(a^{d2^{s-1}}) \equiv 1[p]$  puis on réitère jusqu'à avoir  $a^d \equiv 1[p]$ . D'où la conclusion.

Par contraposée, si  $a^d \not\equiv 1[n]$  et pour tout  $i \in \llbracket 0, s-1 \rrbracket$ ,  $a^{d2^i} \not\equiv -1[p]$  alors  $n$  est composé.  $a$  est alors un témoin de Miller et de même que précédemment on va faire le test pour des valeurs de  $a$  aléatoirement choisies. On a alors,

```
def miller_rabin(n, k):
    if n == 2:
        return True
    if n % 2 == 0:
        return False
    s, d = 0, n - 1
    while d % 2 == 0:
        s += 1
        d //= 2
    for _ in range(k):
        a = random.randrange(2, n - 1)
        x = expo_rap(a, d, n)
        if x == 1 or x == n - 1:
            continue
        for _ in range(s - 1):
            x = expo_rap(x, 2, n)
            if x == n - 1:
                break
        else:
            return False
    return True
```

On peut par ailleurs montrer en s'appuyant sur le théorème de Rabin que la probabilité d'erreur de cet algorithme est de  $(1/4)^k$ .

### 3 Une répartition encore méconnue

L'objectif de cette partie est de donner plusieurs résultats et conjectures illustrant la disparité de l'ensemble  $\mathcal{P}$ .

---

4. cf 5.1.1

### 3.1 Écart entre 2 nombres premiers

Soit  $n \in \mathbb{N}$ . Alors il est possible de trouver  $n$  nombres composés successifs, autrement dit, on peut trouver une séquence aussi grande que l'on veut de nombres consécutifs non premiers.

Démonstration : Soit  $n \in \mathbb{N}$ . On considère alors l'entier  $n!$ . On a que, pour tout  $i \in \llbracket 2; n \rrbracket$ , l'entier  $n! + i$  est composé. En effet,  $i$  est un des facteurs composant  $n!$  (puisque il décrit l'ensemble des entiers naturels inférieurs à  $n$ ). Donc  $n+2, n+3, \dots, 2n$  sont composés. On a donc créé une séquence de  $n-2$  entiers composés successifs. Puisque  $n \in \mathbb{N}$ , on peut prendre  $n$  aussi grand que l'on veut, ce qui démontre le résultat.

### 3.2 Conjecture des nombres jumeaux

Définition : Soit  $(p, p') \in \mathcal{P}^2$ . On dit que  $p$  et  $p'$  sont des nombres premiers "jumeaux" si  $p = p' + 2$  ou  $p' = p + 2$ , c'est-à-dire s'ils sont séparés exactement par un entier. Par exemple,  $(3, 5)$  sont premiers jumeaux,  $(5, 7)$  également.

La conjecture des nombres premiers affirme qu'il existe une infinité de nombres premiers jumeaux. Cela signifierait que, en considérant le résultat précédent, on peut trouver autant que l'on souhaite de nombres premiers séparés par un seul entier que de nombres premiers séparés par une grande séquence de nombres composés.

Bien qu'on ne sache pas si cette conjecture est vraie ou non, le mathématicien Brun a démontré en 1919 que la série

$$\sum \left( \frac{1}{p} + \frac{1}{p+2} \right)$$

converge, ce qui signifie que même s'il existe une infinité de nombres premiers jumeaux, ils finissent par s'éloigner les uns des autres.

### 3.3 Conjecture de Goldbach

La conjecture de Goldbach affirme l'énoncé suivant : tout entier  $n \geq 5$  peut s'écrire comme la somme de 3 nombres premiers (1), ce qui est équivalent à l'assertion suivante : tout entier pair  $2N \geq 4$  est la somme de 2 nombres premiers (2).

Démonstration : (de l'équivalence)

$\Rightarrow$  : Supposons que (1) soit vraie. Alors, pour tout  $n \geq 2$ ,  $2n+2 = p + p' + p''$ , où  $p, p'$  et  $p''$  sont premiers. Un de ces nombres est forcément pair (si aucun d'eux ne l'est, alors  $p + p' + p''$  est impair, ce qui est contradictoire avec le fait que  $2n+2$  est pair). On prend par exemple  $p'' = 2$ . Donc  $2n = p + p'$ , ce qui démontre le résultat.

$\Leftarrow$  : Réciproquement, supposons que (2) soit vraie. Alors  $2n-2 = p + p'$ , avec  $p$  et  $p'$  premiers. Ainsi,  $2n = 2 + p + p'$  et  $2n+1 = 3 + p + p'$ , ce qui démontre le résultat.

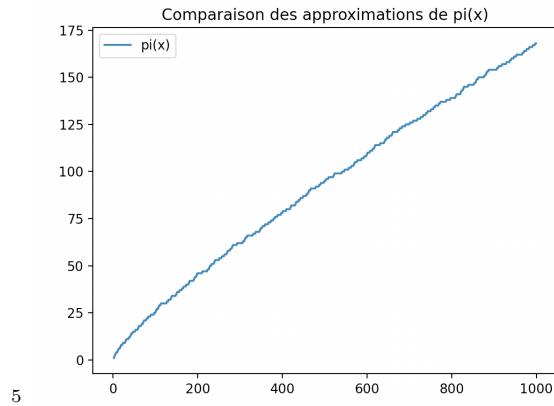
Peu de progrès ont été accomplis concernant l'étude de cette conjecture à ce jour. Seuls quelques résultats importants furent démontrés dont un théorème asymptotique démontré par Hardy et Littlewood, stipulant qu'il existe  $n_0 \in \mathbb{N}$  tel que pour tout  $n \geq n_0$ ,  $n$  est la somme de trois nombres premiers. La conjecture de Goldbach est donc vraie pour tout entier "suffisamment grand", mais cela n'a pas vraiment d'intérêt en pratique. Des calculs numériques ont pu déterminer que  $n_0 \leq 3^{3^{15}}$ .

## 4 Une solution au problème de répartition ?

### 4.1 Approximations de $\pi(x)$

L'objectif des mathématiciens est de déterminer le plus précisément possible où se trouvent les nombres premiers dans  $\mathbb{N}$ . Pour mettre en forme cette problématique, la fonction de répartition des nombres premiers a été définie ainsi :

$$\begin{aligned}\pi : \mathbb{R}^* &\longrightarrow \mathbb{N} \\ x &\longmapsto \text{Card } (\{ p \in P \text{ tel que } p \leq x\})\end{aligned}$$

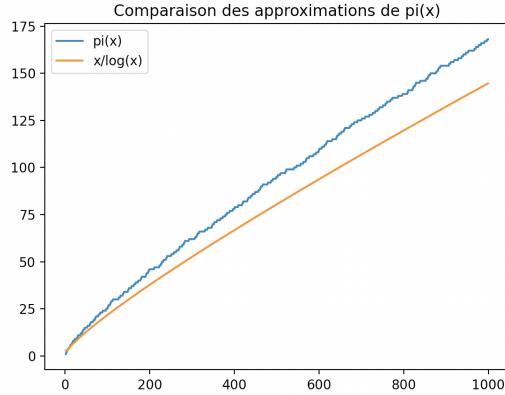


On ne peut avoir accès à la valeur de cette fonction qu'en comptant les nombres premiers un par un, ce qui se révèle peu efficace. De ce fait, pour accéder aux valeurs de cette fonction, les mathématiciens ont cherché à construire des approximations de  $\pi(x)$ . La première trouvée fut la fonction :

$$x \longmapsto \frac{x}{\ln(x)}$$

---

5. voir fonctions-approx.py sur Github

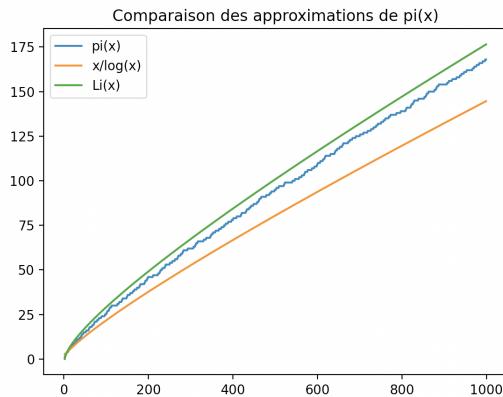


On peut en effet démontrer que cette fonction est un équivalent de  $\pi$  (démonstration que l'on admet car elle n'est pas du tout de notre niveau. Il s'agit du théorème des nombres premiers.). Cependant, bien que ces fonctions soient équivalentes, cela ne suffit pas, l'approximation reste trop vague. On a ensuite proposé la fonction Li (fonction d'écart logarithmique intégrale) définie ainsi :

$$Li : \mathbb{R} \longrightarrow \mathbb{R}$$

$$x \longmapsto \int_2^x \frac{dt}{\ln(t)}$$

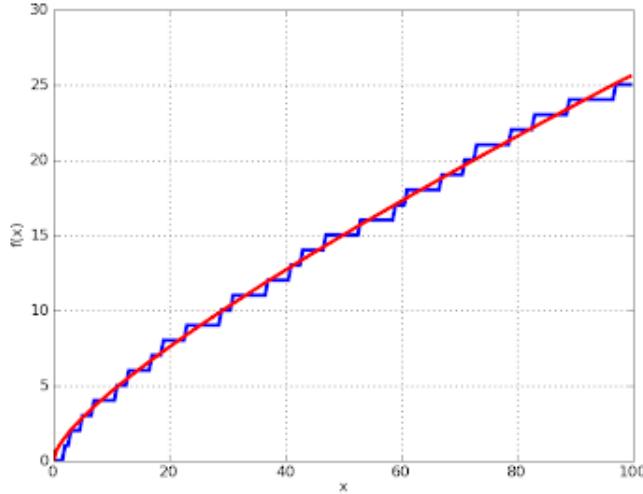
qui se révèle être une meilleure approximation.



Par la suite, le mathématicien Bernhard Riemann a proposé, comme nouvelle approximation :

$$Ri : x \longmapsto \sum_{n=1}^{+\infty} \frac{\mu(n)}{n} \operatorname{li}(x^{1/n}),$$

où  $\mu$  désigne la fonction de Möbius<sup>6</sup> et  $\operatorname{li}$  le logarithme intégral, qui contrairement à  $\operatorname{Li}$  a sa borne inférieure égale à 0.



7

Cette fonction est très proche de  $\pi$  mais n'est pas satisfaisante : elle ne présente pas de forme en escalier comme  $\pi$ . Pour remédier à ce problème Riemann a trouvé une solution : à sa fonction  $Ri$ , il apporte d'infimes modifications, qui correspondent exactement à la valeur de  $Ri$  aux points  $x^p$ , où  $p$  est un zéro non trivial de la fonction  $\zeta$ .

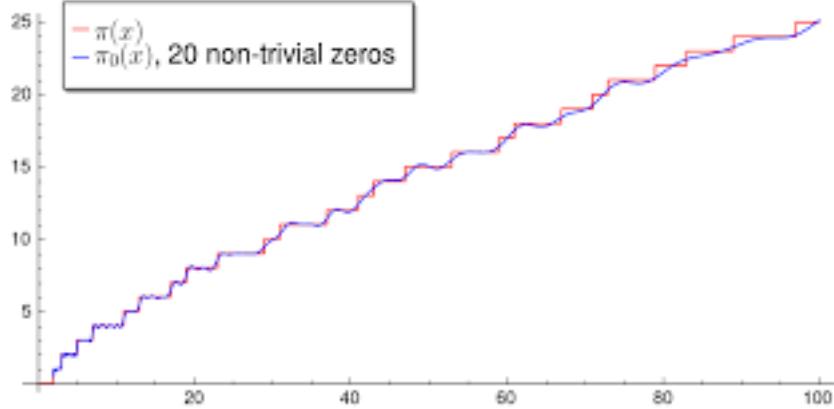
On a alors la formule suivante :

$$\pi(x) = Ri(x) - \sum_{p \text{ tel que } \zeta(p)=0} Ri(x^p)$$

---

6. Fonction de  $\mathbb{N}$  à valeur dans  $\{-1; 0; 1\}$ , et qui vaut 0 si  $n$  est divisible par un carré parfait différent de 1 ; 1 si  $n$  est le produit d'un nombre pair de nombres premiers distincts ; -1 si  $n$  est le produit d'un nombre impair de nombres premiers distincts

7. Nous n'avons malheureusement pas réussi à tracer cette figure



Grâce à Riemann, on sait maintenant comment la fonction  $\pi$  se comporte exactement, selon une fonction dont il est possible de calculer les valeurs. Il suffirait de trouver tous les zéros de la fonction  $\zeta$  afin de connaître exactement la répartition des nombres premiers. Mais se pose ici un nouveau problème : quels sont les zéros de  $\zeta$  ?

## 4.2 L'hypothèse de Riemann

### 4.2.1 La fonction $\zeta$ d'Euler

Leonhard Euler, un des plus grands mathématiciens de l'histoire notamment connu pour ses travaux en géométrie, analyse, algèbre et théorie des nombres, s'intéresse au début du XVIII-ième siècle au problème de Bâle, à savoir étudier la convergence de la série :  $\sum \frac{1}{n^2}$

Euler a alors démontré que cette série est convergente et qu'elle converge vers  $\frac{\pi^2}{6}$ .

Démonstration :

Le développement en série de Taylor de la fonction sinus au voisinage de 0 est :

$$\forall x \in \mathbb{R}, \sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

En supposant  $x$  non nul et en divisant par ce réel, on a :

$$\frac{\sin x}{x} = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \dots$$

Maintenant, les racines de  $\frac{\sin x}{x}$  apparaissent précisément pour  $x = \pm n\pi$ , avec  $n \in \mathbb{N}$ . Supposons qu'on puisse exprimer cette série infinie comme un produit

de facteurs linéaires donnés par ses racines :

$$\begin{aligned}\frac{\sin x}{x} &= \left(1 - \frac{x}{\pi}\right) \left(1 + \frac{x}{\pi}\right) \left(1 - \frac{x}{2\pi}\right) \left(1 + \frac{x}{2\pi}\right) \left(1 - \frac{x}{3\pi}\right) \left(1 + \frac{x}{3\pi}\right) \cdots \\ &= \left(1 - \frac{x^2}{\pi^2}\right) \left(1 - \frac{x^2}{4\pi^2}\right) \left(1 - \frac{x^2}{9\pi^2}\right) \cdots\end{aligned}$$

Si on effectue formellement ce produit et regroupe tous les termes en  $x^2$ , on voit que le coefficient de  $x^2$  dans  $\frac{\sin x}{x}$  est :

$$-\left(\frac{1}{\pi^2} + \frac{1}{4\pi^2} + \frac{1}{9\pi^2} + \cdots\right) = -\frac{1}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2}$$

Mais, à partir du développement de la série infinie originale de  $\frac{\sin x}{x}$ , le coefficient de  $x^2$  est :

$$-\frac{1}{3!} = -\frac{1}{6}$$

Ces deux coefficients doivent être égaux ; ainsi,

$$-\frac{1}{6} = -\frac{1}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2}$$

En multipliant les deux côtés de cette équation par  $-\pi^2$ , on obtient le résultat attendu.

Après avoir résolu le problème, il se demanda ce que cela changerait si on remplaçait l'exposant 2 du dénominateur par un réel quelconque. Il a alors posé la fonction  $\zeta$  définie par :

$$\zeta(x) = \sum_{n=1}^{+\infty} \frac{1}{n^x}$$

#### 4.2.2 La fonction $\zeta$ de Riemann

##### Quelques notions d'analyse complexe :

Dans la suite, plusieurs notions d'analyse complexe seront utilisées, il est donc utile de les détailler ici afin d'en avoir une meilleure compréhension.

**Fonction holomorphe :** Soit  $f$  une fonction à valeurs complexes définie sur  $I$ . On dit que  $f$  est holomorphe (ou analytique) si pour tout  $z_0$  de  $I$ , le taux d'accroissement de  $f$  en  $z_0$  existe, c'est-à-dire :

$$\lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

existe.

Dans ce cas, on appelle  $f'(z_0)$  cette limite, qui correspond donc à la dérivée de  $f$  en  $z_0$ . L'ensemble des fonctions holomorphes dans  $I$  est noté  $H(I)$  (c'est un anneau).

<sup>8</sup>

Remarque :  $f$  est donc une fonction de  $I$  dans  $\mathbb{R}^2$ ,  $\mathbb{R}^2$  étant ici associé au plan complexe.

Remarque : On peut démontrer que toute fonction holomorphe est de classe  $C^\infty$ .

**Fonction méromorphe :** Une fonction est dite méromorphe si elle est holomorphe dans tout le plan complexe, sauf éventuellement sur un ensemble de points isolés dont chacun est un pôle pour la fonction.

**Prolongement analytique :** En analyse complexe, la théorie du prolongement analytique détaille l'ensemble des propriétés et techniques relatives au prolongement des fonctions holomorphes (ou analytiques), c'est-à-dire qu'à une fonction  $f$  holomorphe définie sur un ensemble quelconque, on associe une nouvelle fonction (aussi notée  $f$ ), ayant le même graphe mais cette fois-ci définie sur le plan complexe.

---

Riemann s'est intéressée à la fonction  $\zeta$  d'Euler. Tout d'abord, il étend le domaine de définition de la fonction  $\zeta$  à l'ensemble  $\mathbb{C} \setminus \{1\}$ . Il pose alors une nouvelle fonction  $\zeta$ , définie comme prolongement analytique de la fonction  $\zeta$  originale. Pour faire simple, on peut montrer que  $\zeta$  peut s'écrire comme la somme de deux intégrales dont la deuxième est une fonction holomorphe.

$$\Gamma(s)\zeta(s) = \int_0^\infty \frac{t^{s-1}}{e^t - 1} dt. = \int_0^1 \frac{t^{s-1}}{e^t - 1} dt + \int_1^\infty \frac{t^{s-1}}{e^t - 1} dt.$$

où  $\Gamma$  désigne la fonction factorielle prolongée à l'ensemble des nombres complexes saufs les entiers négatifs.

Après plusieurs calculs, on parvient à l'expression de  $\zeta$  suivante :

$$\zeta(-k) = (-1)^k \frac{B_{k+1}}{k+1}$$

On remarque alors que la fonction  $\zeta$  admet bien un unique pôle en 1. (en prenant  $k = -1$ ).

La fonction  $\zeta$  est finalement définie ainsi :

$$\begin{aligned} \zeta : \mathbb{C} \setminus \{1\} &\longrightarrow \mathbb{R} \\ s &\longmapsto \sum_{n=1}^{+\infty} \frac{1}{n^s} \end{aligned}$$

---

8. (Source : Analyse réelle et complexe, W.Rudin, page 192.)

$$\begin{aligned}\zeta : \quad & \mathbb{C} \setminus \{1\} \longrightarrow \mathbb{C} \\ & x \longmapsto \sum_{n=1}^{+\infty} \frac{1}{n^s}\end{aligned}$$

#### 4.2.3 Brève étude de la fonction $\zeta$

Le lien entre la fonction  $\zeta$  et les nombres premiers avait déjà été établi par Leonhard Euler avec la formule, valable pour  $\operatorname{Re}(s) > 1$  :

$$\zeta(s) = \prod_{p \in \mathcal{P}} \frac{1}{1 - p^{-s}} = \frac{1}{(1 - \frac{1}{2^s})(1 - \frac{1}{3^s})(1 - \frac{1}{5^s}) \dots} \zeta(s)$$

où le produit infini est étendu à l'ensemble  $\mathcal{P}$  des nombres premiers. Cette relation est une conséquence de la formule pour les suites géométriques et du théorème fondamental de l'arithmétique. On appelle parfois cette formule produit eulérien.

##### Démonstration :

Soit  $s = \sigma + i\tau$ , avec  $\sigma > 1$ . On a, pour tout  $p \in \mathbb{P}$  :

$$(1 - \frac{1}{p^s})^{-1} = \sum_{n=0}^{+\infty} \frac{1}{p^{ns}}$$

En prenant le produit pour  $p$  premier,  $p \leq N$  (avec  $N \geq 2$ ) dans les deux membres, on obtient :

$$\begin{aligned}\prod_{p \leq N} (1 - \frac{1}{p^s})^{-1} &= \prod_{p \leq N} \sum_{e=0}^{+\infty} \frac{1}{p^{es}} \\ &= \prod_{i=1}^n \sum_{e=0}^{+\infty} \frac{1}{p_i^{es}} \\ &= \sum_{(e_1, \dots, e_n) \geq 0} \frac{1}{(p_1^{e_1}, \dots, p_n^{e_n})^s}\end{aligned}$$

Grâce à l'unicité de la décomposition en produit de facteurs premiers, on obtient, en notant  $P^+(n)$  le plus grand facteur premier de l'entier  $n$ ,

$$(1 - \frac{1}{p^s})^{-1} = \sum_{P^+(n) \leq N} \frac{1}{n^s}$$

Comme  $\llbracket 1 ; n \rrbracket \subset \{n \in \mathbb{N} \mid P^+(n) \leq N\}$ , on a :

$$|\zeta(s) - \prod_{p \leq N} (1 - \frac{1}{p^s})^{-1}| \leq \sum_{n \geq N} \frac{1}{n^\sigma}$$

En faisant tendre  $n$  vers  $+\infty$ , on obtient la formule d'Euler complexe.

Il est possible de démontrer que, pour tout nombre complexe différent de 0 et 1, la fonction  $\zeta$  peut se mettre sous la forme suivante, nommée "équation fonctionnelle" :

$$\zeta(s) = 2^s \pi^{s-1} \sin\left(\frac{\pi s}{2}\right) \Gamma(1-s)\zeta(1-s)$$

On remarque alors que la fonction s'annule pour tout  $s = -2k$ , avec  $k \in \mathbb{Z} \setminus \{0\}$  (le terme  $\sin\left(\frac{\pi s}{2}\right)$  s'annule pour tous les entiers pairs négatifs). Ces valeurs sont appelées "zéros triviaux" de  $\zeta$ .

Le réel enjeu de l'étude de la fonction  $\zeta$  est de trouver ses zéros non-triviaux. C'est là qu'intervient Riemann : en 1859, il conjecture que les zéros non-triviaux de  $\zeta$  se situent tous sur la droite d'équation  $\text{Re}(s) = \frac{1}{2}$  dans le plan complexe.

De nombreux mathématiciens se sont penchés sur l'étude de ces zéros mais seuls de vagues résultats furent démontrés, permettant de resserrer l'étude à la bande critique (représentant l'ensemble les nombres complexes de partie réelle comprise entre 0 et 1). On sait également qu'il existe une infinité de zéros dans cette bande. Cependant, à ce jour aucune démonstration de l'hypothèse de Riemann ne fut trouvée, bien que les mathématiciens Louis de Branges et Michael Atiyah aient tenté d'en donner une (respectivement en 2004 et 2018), rejetée par la suite par la communauté mathématique.

## 5 Application : un petit peu de cryptographie

### 5.1 Chiffrement RSA

#### 5.1.1 L'anneau $\mathbb{Z}/n\mathbb{Z}$

On note  $\mathbb{Z}/n\mathbb{Z}$  l'ensemble à  $n$  éléments des classes de congruence modulo  $n$ . On rappelle les résultats suivants sur le calcul modulaire :

Si  $x \equiv x'[n]$  et  $y \equiv y'[n]$  alors  $x + x' \equiv y + y'[n]$

Si  $x \equiv x'[n]$  et  $y \equiv y'[n]$  alors  $xx' \equiv yy'[n]$

Ainsi pour  $\alpha$  et  $\beta$  deux classes de  $\mathbb{Z}/n\mathbb{Z}$ ,  $a \in \alpha$  et  $b \in \beta$ , la classe de  $a + b$  que l'on note  $\alpha + \beta$  ne dépend que de  $\alpha$  et  $\beta$ . On définit de même  $\alpha \times \beta$ . Ainsi on peut montrer que  $(\mathbb{Z}/n\mathbb{Z}, +, \times)$  est un anneau d'élément nul  $\bar{0}$  et de l'élément unité  $\bar{1}$  appelé anneau quotient.

On remarque de plus que par définition des lois et la propriété selon laquelle  $a[\text{mod } n] = b[\text{mod } n]$  si et seulement si  $a \equiv b[n]$ , l'application  $\psi : a \mapsto a[\text{mod } n]$  est un homomorphisme d'anneaux de  $\mathbb{Z}$  et  $\mathbb{Z}/n\mathbb{Z}$  pour tout  $(a, b) \in \mathbb{Z}^2$ ,

$$\psi(a + b) = \psi(a) + \psi(b)$$

$$\psi(a \times b) = \psi(a) \times \psi(b)$$

$$\psi(1) = 1$$

On confond alors souvent  $\mathbb{Z}/n\mathbb{Z}$  et  $\{1, 2, \dots, n - 1\}$ .

Ainsi, être inversible dans signifie pour un élément  $a$  qu'il existe  $b$  tel que  $ab \equiv 1[n]$ . Or  $ax \equiv ay[n]$  n'implique pas toujours  $x \equiv y[n]$ . Il n'est donc pas évident de déterminer les éléments inversibles de  $\mathbb{Z}/n\mathbb{Z}$ .

Théorème : Les éléments inversibles de  $\mathbb{Z}/n\mathbb{Z}$  sont les éléments premiers à  $n$ .

Démonstration : Soit  $a \in \mathbb{Z}/n\mathbb{Z}$ ,  $a$  est inversible si et seulement si il existe  $b$  tel que  $ab \equiv 1[n]$  si et seulement si il existe  $k$  et  $b$  tels que  $ab - kn = 1$  si et seulement si  $a$  et  $n$  sont premiers entre eux d'après le théorème de Bézout.

On en déduit aisément les corollaires suivants :  $((\mathbb{Z}/n\mathbb{Z})^*, +, \times)$  est un corps donc en particulier  $(\mathbb{Z}/p\mathbb{Z}, +, \times)$  avec  $p$  premier est un corps. Ou encore  $((\mathbb{Z}/n\mathbb{Z})^*, \times)$  est un groupe.

On note aussi que c'est une relation de Bézout qui donne accès à l'inverse d'un élément.

On se propose d'étudier plus en détail certains résultats sur les groupes pouvant s'appliquer plus particulièrement à  $((\mathbb{Z}/n\mathbb{Z})^*, \times)$ .

Commençons par quelques définitions. Soit  $G$  un groupe d'élément neutre  $e$ ,  $g \in G$ . L'ensemble  $H = \{g^n | n \in \mathbb{Z}\}$  est un sous-groupe de  $G$ . Puisque c'est le plus petit sous-groupe contenant  $g$  on l'appelle sous-groupe engendré par  $g$ . Si  $H$  est fini, d'après le principe des tiroirs de Dirichlet, il existe  $m, m'$  tel que  $g^m = g^{m'}$  ie  $g^{m-m'} = e$ . On note  $n$  le plus petit entier tel que  $g^n = e$ . Appliquer la division euclidienne par  $n$  à n'importe quel exposant nous donne donc que  $H$  contient exactement  $n$  éléments distincts. On a aussi  $g^m = g^{m'}$  équivaut à  $m \equiv m'[n]$ . On dit alors que  $g$  est d'ordre  $n$ . Si  $G$  est fini on appelle aussi  $\text{card}G$  l'ordre de  $G$ .

On a alors le théorème suivant (Théorème de Lagrange) :

Théorème : L'ordre de tout élément d'un groupe fini  $G$  divise l'ordre de  $G$  ie pour tout  $g \in G$ ,  $g^{\text{card}(G)} = e$ .

Démonstration : On se contente d'une démonstration dans le cas  $G$  abélien qui sera suffisante pour la suite de l'exposé. Soit  $g \in G$  et  $n$  l'ordre de  $G$ . L'application  $f : x \mapsto g * x$  est une permutation de  $G$ . Si on note  $G = \{x_1, \dots, x_n\}$ , on a alors

$$f(x_1) * \dots * f(x_n) = x_1 * \dots * x_n$$

Puis par commutativité,

$$g^n * x_1 * \dots * x_n = x_1 * \dots * x_n$$

. On simplifie par  $x_1 * \dots * x_n$  pour avoir le résultat souhaité :  $g^n = e$ .

### 5.1.2 Une brève étude de l'indicateur d'Euler $\varphi$

On appelle fonction indicatrice d'Euler l'application suivante :

$$\begin{aligned}\varphi : \quad \mathbb{N}^* &\longrightarrow \mathbb{N}^* \\ n &\longmapsto \text{card}(\{a \mid 0 \leq a < n \text{ et } a \text{ premier à } n\})\end{aligned}$$

ou encore, d'après ce qui précède :

$$\begin{aligned}\varphi : \quad \mathbb{N}^* &\longrightarrow \mathbb{N}^* \\ n &\longmapsto \text{card}(\mathbb{Z}/n\mathbb{Z})^*\end{aligned}$$

Le résultat fondamental sur l'indicateur d'Euler est le suivant :

Théorème : Pour tout  $n \in \mathbb{N}^*$ , on note  $p_1, \dots, p_r$  les facteurs premiers de  $n$  alors :

$$\varphi(n) = n \prod_{i=1}^r \left(1 - \frac{1}{p_i}\right)$$

Démonstration : Il existe des démonstrations plus classiques de ce théorème mais l'approche suivante est plutôt amusante. On choisit  $x \in \llbracket 1, n \rrbracket$  aléatoirement. On note  $B$  l'événement : " $x$  et  $n$  sont premiers entre eux". Calculons  $P(B)$ . On pose aussi  $A_l$  l'événement : " $|l|x$ ".

Puisque  $p_1, \dots, p_r$  sont les diviseurs premiers de  $n$ ,  $\text{pgcd}(x, n) = 1$  si et seulement si pour tout  $i \in \llbracket 1, r \rrbracket$ ,  $p_i \nmid x$  si et seulement si, pour tout  $i \in \llbracket 1, r \rrbracket$ ,  $\overline{A_{p_i}}$ . Donc  $P(B) = P(\bigcap_{i=1}^r \overline{A_{p_i}})$ . Montrons que  $\overline{A_{p_1}} \dots \overline{A_{p_r}}$  sont mutuellement indépendants.

Pour  $i \in \llbracket 1, r \rrbracket$  il existe  $k$  tel que  $n = kp_i$  donc il y a  $k$  valeur de  $x$  satisfaisant  $A_{p_i}$  d'où  $P(A_{p_i}) = \frac{n}{np_i} = \frac{1}{p_i}$ .

De plus  $\bigcap_{i=1}^r A_i$  si et seulement si  $\prod_{i=1}^r p_i \mid x$ . Donc de même que précédemment,  $P(\bigcap_{i=1}^r A_i) = P(A_{p_1 \dots p_r}) = \frac{1}{p_1 \dots p_r} = \prod_{i=1}^r P(A_i)$ .

Donc  $A_{p_1} \dots A_{p_r}$  sont mutuellement indépendants donc  $\overline{A_{p_1}} \dots \overline{A_{p_r}}$  sont mutuellement indépendants. D'où  $P(B) = \prod_{i=1}^r P(\overline{A_i}) = \prod_{i=1}^r \left(1 - \frac{1}{p_i}\right) = \frac{\varphi(n)}{n}$  par définition de  $\varphi$ . D'où la conclusion.

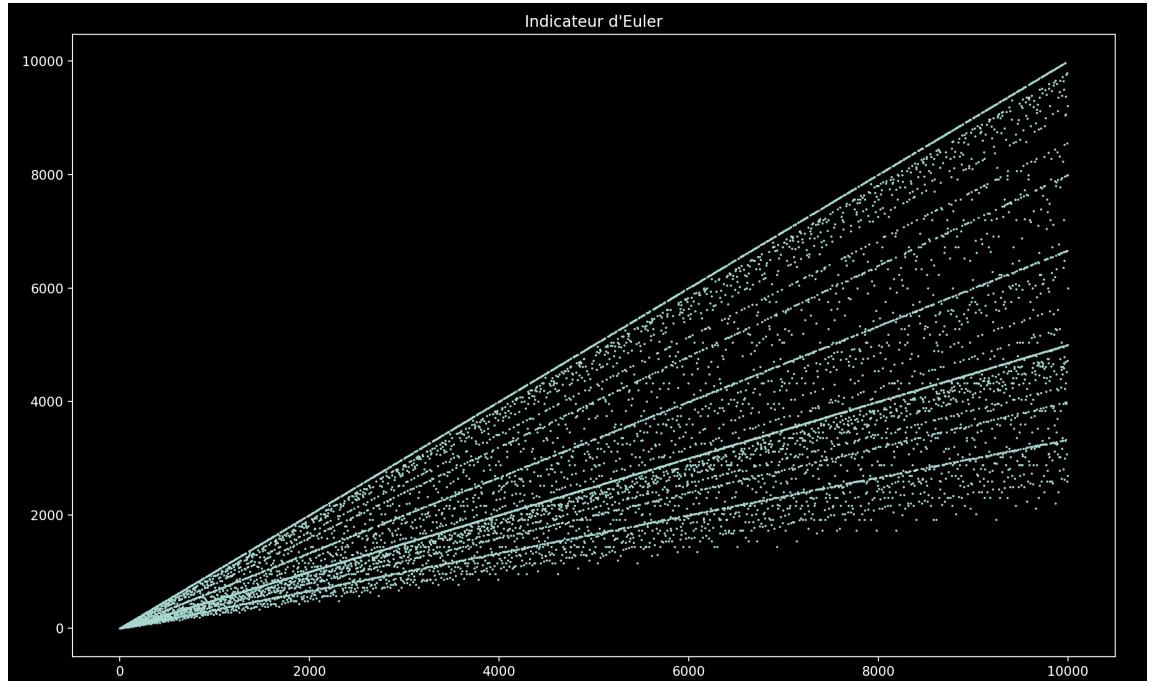
Cas particulier : Pour  $n = pq$  avec  $p$  et  $q$  premiers,  $\varphi(n) = (p-1)(q-1)$

Il est important de noter que le calcul de  $\varphi(n)$  repose sur la décomposition en facteurs premiers de  $n$  qui n'est pas chose triviale<sup>9</sup>. En appliquant le crible d'Eratosthène on est capable de calculer  $\varphi(n)$  pour des valeurs raisonnables mais il n'existe pas de méthode beaucoup plus efficace. On obtient alors la figure suivante<sup>10</sup>

---

9. cf 5.2

10. voir `indicateurEuler.py` sur Github



Un autre théorème important pour la suite de l'exposé est le théorème d'Euler qui consiste en une généralisation du petit théorème de Fermat :

Théorème : Soit  $n$  un entier  $> 0$ , et  $a$  un entier premier à  $n$ .  
Alors  $a^{\varphi(n)} \equiv 1[n]$ .

Démonstration : Il suffit d'appliquer le théorème de Lagrange à  $(\mathbb{Z}/n\mathbb{Z})^*$  en notant que  $\varphi(n) = \text{card}(\mathbb{Z}/n\mathbb{Z})^*$ .

### 5.1.3 Le chiffrement RSA

A l'aide des outils développés au cours de cette partie, nous pouvons maintenant nous intéresser à la méthode du chiffrement RSA, qui souligne l'intérêt des nombres premiers dans les enjeux sociétaux actuels et notamment la sécurité. La méthode de chiffrement RSA a été développée en 1977 par trois professeurs du MIT : Ronald Rivest, Adi Shamir, Leonard Adleman (d'où l'appellation RSA). La particularité de cette méthode est qu'elle est asymétrique. Contrairement à d'autres méthodes symétriques où obtenir la clef de chiffrement permettrait de pouvoir déchiffrer des messages interceptés, il y a ici deux clefs. Une clef dite "publique" (connue de tous) pour chiffrer et une clef "privée" pour déchiffrer le message. Nous détaillerons la relation qui lie clef publique et clé privée mais l'intérêt de cette méthode est qu'il est calculatoirement impossible de déduire la clef privée de la clef publique.

Supposons que Alice souhaite recevoir un message confidentiel de Bob<sup>11</sup> sans que le message puisse être intercepté et lu. Elle met en place la méthode RSA qui s'opère en 3 étapes :

Création des clefs : Alice commence par choisir deux nombres premiers, dans la pratique très grands,  $p$  et  $q$  qui sont en suite utilisés pour calculer  $n = pq$  le module de chiffrement. On peut alors calculer  $\varphi(n) = (p-1)(q-1)$ . On rappelle que le calcul de  $\varphi(n)$  repose sur la connaissance de  $p$  et  $q$ . Alice choisie alors l'exposant de chiffrement  $e$ , un entier inférieur et premier à  $\varphi(n)$ . Le couple  $(e, n)$  formera la clef publique.

On implémente ensuite l'algorithme d'Euclide étendu<sup>12</sup> pour obtenir une relation de Bézout et pouvoir calculer l'inverse de  $e \bmod \varphi(n)$  :

<sup>13</sup>

```
def euclide_etendu(a,b):
    r,u,v,r2,u2,v2 = a,1,0,b,0,1
    while r2 != 0:
        q = r // r2
        r, u ,v, r2 ,u2 ,v2 = r2 ,u2 ,v2 ,( r-q*r2 ),( u-q*u2 ),( v-q*v2 )
    return (r,u,v)
```

On note  $d$  cet inverse. Ainsi  $(d, n)$  forme la clef privée. Le programme suivant renvoie, une clef publique et une clef privée à partir de 2 nombres premiers donnés en entrée :

```
def creer_clefs(p,q):
    n = p*q
    phi = (p-1)*(q-1)
    e = phi
    r,d = euclide_etendu(e, phi)[0:2]
    while r !=1:
        e = random.randint(2, phi)
        r,d = euclide_etendu(e, phi)[0:2]
    return (e,n) , (d%phi,n)
```

Alice fournit  $(e, n)$  à Bob et garde  $(d, n)$  en supprimant  $p$ ,  $q$  et  $\varphi(n)$

Chiffrement du message : La deuxième étape est le chiffrement du message à partir de la clef publique fournie à Bob. Le message de Bob doit être représenté par un entier  $m$  strictement inférieur  $n$ . La conversion d'un texte à un entier peut se faire facilement avec la table ASCII par exemple. En notant  $c$  le message chiffré, le calcul de  $c$  se fait très facilement : on a

$$c \equiv m^e [n]$$

---

11. Ces noms sont en quelques sortes des conventions dans le domaine de la cryptographie  
12. Équivalent dans le cours à remonter l'algorithme d'Euclide

13. On montre par récurrence forte que  $au + bv = r$  est un invariant de boucle

En python on a tout simplement :

```
def chiffrement (m,n,e):
    return expo_rap(m,e,n)
```

Avec `exporap` une fonction pour calculer efficacement  $m^e[n]$ .

Déchiffrement : Lorsque Alice reçoit le message chiffré  $c$  de Bob, elle peut retrouver le message  $m$  original en effectuant l'opération suivante :

$$m \equiv c^d[n]$$

Ainsi la méthode de chiffrement RSA repose sur le théorème suivant :

Théorème : Soit  $d$  l'inverse de  $e$  modulo  $\varphi(n)$  avec  $n = pq$ . Si  $c \equiv m^e[n]$  alors  $m \equiv c^d[n]$

Démonstration : On a

$$\begin{aligned} c^d &\equiv (m^e)^d[n] \\ &\equiv m^{ed}[n] \end{aligned}$$

Or par hypothèse  $ed = 1 + k\varphi(n)$  et donc On a

$$\begin{aligned} c^d &\equiv m^{1+k\varphi(n)}[n] \\ &\equiv m \times (m^{\varphi(n)})^k[n] \end{aligned}$$

Si  $m$  et  $n$  sont premiers entre eux, le théorème d'Euler donne  $m^{\varphi(n)} \equiv 1[n]$ . Ce qui donne immédiatement  $c^d \equiv m[n]$

Deuxième cas : si  $\text{pgcd}(m, n) \neq 1$  alors par symétrie des rôles, on suppose que  $\text{pgcd}(m, n) = p$  et donc puisque  $m < n$ ,  $\text{pgcd}(m, q) = 1$ .

Donc  $p|m$  donc  $m \equiv 0[p]$  et  $m^{eq} \equiv 0[n]$  donc  $m^{ed} \equiv m[p]$ . D'où  $p|m^{ed} - m$

On a aussi

$$\begin{aligned} m^{ed} &\equiv m \times (m^{\varphi(n)})^k[q] \\ &\equiv m \times (m^{q-1})^{(p-1)k}[q] \end{aligned}$$

Donc  $q|m^{ed} - m$ . Ainsi  $pq = n|m^{ed} - m$  d'où  $m^{ed} \equiv m[n]$  et la conclusion dans tous les cas.

On peut implémenter une fonction déchiffrement :

```
def dechiffrement (c,d,n):
    return expo_rap(c,d,n)
```

Exemple :

Pour la création des clefs, Alice utilise les nombres de Mersennes (de la forme  $2^n - 1$ ) relativement grands :

$$p = M_{31} = 2147483647 \text{ et } q = M_{61} = 2305843009213693951$$

Les clefs privées et publiques générées sont :

```
>>> pub, priv = creer_clefs(p, q)
>>> pub
(1577853571334153130409925987, 4951760154835678088235319297)
>>> priv
(96685962259908502466291123, 4951760154835678088235319297)
```

Bob veut envoyer le message 2713 à Alice, il peut alors chiffrer son message à l'aide de la clef publique fournie par Alice :

```
>>> chiffrement(2713, pub[0], pub[1])
1176719161186747742563923737
```

Lorsque Alice reçoit le message chiffré, elle peut le déchiffrer avec sa clef privée :

```
>>> dechiffrement(1176719161186747742563923737, priv[0], priv[1])
2713
```

Exemple : On laisse au lecteur l'exercice de déchiffrer le message suivant en se servant des mêmes clefs que l'exemple précédent. On pourra se servir des fonctions du fichier RSA.py sur Github et notamment transformer l'entier déchiffré en texte grâce à la fonction int\_to\_string.

Message chiffré : 2374117310055482034146695634

## 5.2 Le problème de la factorisation : l'algorithme $\rho$ de Pollard

Dans le cadre du chiffrement RSA, on s'aperçoit, qu'il devient essentiel de pouvoir factoriser des grands nombres en leurs facteurs premiers.

Un algorithme efficace est l'algorithme  $\rho$  de Pollard.

### 5.2.1 Algorithmes de recherche d'une période

Étonnamment, la question de la factorisation d'un nombre en ses facteurs premiers renvoie dans un premier temps aux algorithmes de recherche de périodicité d'une suite.

Paradoxe des anniversaires :

Dans l'exercice 7 du TD sur le dénombrement, nous nous sommes intéressés au problème suivant : considérons un groupe de  $n$  personnes, quelle est la probabilité que 2 de ses personnes aient le même anniversaire ? Combien de personne

faut-il pour avoir une probabilité de 0,5 ?.

On montre facilement que la probabilité vaut  $p_n = 1 - \frac{A_{365}^n}{365^n} = 1 - \prod_{k=0}^{n-1} \left(1 - \frac{k}{365}\right)$ . Ainsi pour avoir  $p_n > 0,5$ , une application numérique donne  $n \geq 23$ .

Ce résultat est alors contre intuitif car on aurait tendance à penser qu'il faut au moins  $365/2 = 183$  personnes. On s'aperçoit alors que ce nombre est plus proche de  $\sqrt{365}$ .

Le résultat a alors été généralisé pour la recherche de périodicité dans  $\mathbb{N}_n$  par le résultat suivant que l'on admet :

Si on effectue des tirages avec remise dans un ensemble de cardinal  $n$  le nombre de tirages nécessaires pour espérer avoir un doublet est équivalent à  $\sqrt{\frac{\pi}{2}n}$ .

En revanche si on essaye de programmer cette tâche, on ne peut bénéficier de cette complexité améliorée, car la comparaison avec les éléments déjà trouvés (stockés dans une liste par exemple) s'effectue en une complexité quadratique. Pour bénéficier de la complexité améliorée du caractère aléatoire du paradoxe des anniversaires dans la recherche de période, on va utiliser des suites récurrentes au caractère pseudo-aléatoire.

#### Période d'une suite récurrente :

Le principe des tiroirs de Dirichlet nous assure l'assertion suivante : Pour  $E$  un ensemble fini de cardinal  $n$ ,  $a \in E$  et  $f : E \rightarrow E$ , on considère la suite  $(x_n)$  telle que  $x_0 = a$  et  $x_{n+1} = f(x_n)$ . Alors il existe  $k$  et  $l$  avec  $k \neq l$  et  $x_k = x_l$ . On dit alors que la suite est ultimement périodique. D'après le paradoxe des anniversaires on peut s'attendre à ce que la période soit proche de  $\sqrt{n}$

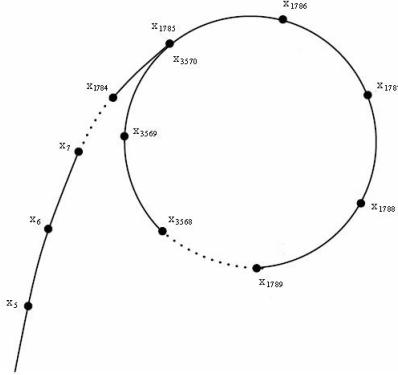
Propriété : Soit  $l$  le plus petit indice tel qu'il existe  $k$  tel que  $x_l = x_k$ . On pose alors  $p = l - k$ . Ainsi pour tout  $(i, j)$  tel que  $i \neq j$ ,  $x_i = x_j$  si et seulement si  $k \leq \min(i, j)$  et  $i \equiv j[p]$ .

Démonstration : Pour  $k \leq i$ , on a  $x_{i+p} = f^{i-k}(x_{k+p}) = f^{i-k}(x_l) = f^{i-k}(x_k) = x_i$ . Réciproquement, soit  $j > i$  avec  $x_i = x_j$ . On a alors  $k \leq j - p$  et si on applique la première implication on a  $x_{j-p} = x_j$ . Si  $j - p = i$  on peut conclure, sinon on réitère en remplaçant  $(i, j)$  par  $(i, j - p)$ .

Si on représente les valeurs prises par  $(x_n)$  on a alors une figure comme ci-dessous.<sup>14</sup>

---

14. Qui est d'ailleurs la raison de l'appellation  $\rho$  de l'algorithme.



#### Algorithme de recherche de période : Algorithme de Floyd

L'objectif est alors de trouver par un algorithme un couple  $(i, j)$  tel que  $x_i = x_j$ . L'algorithme de Floyd est alors fondé sur la constat suivant qui découle de la propriété précédente :  $x_{2i} = x_i$  si et seulement si  $i \geq k$  et  $i$  multiple de  $p$ . La méthode consiste alors à comparer les valeurs de  $i$  et  $2i$  jusqu'à obtenir un  $i$  satisfaisant ce qui se produit forcément d'après la remarque précédente. En python on peut implémenter l'algorithme de la manière suivante :

```
def floyd(f, a):
    i, x, y = 1, f(a), f(f(a))
    while x != y:
        i, x, y = i+1, f(x), f(f(y))
    return i, 2*i
```

#### 5.2.2 Application à la factorisation :

Dans l'algorithme  $\rho$  de Pollard, on applique les méthodes précédentes au service de la factorisation d'un entier  $n$ . On se place alors dans  $E = \mathbb{N}_n$  et on considère de même que précédemment une suite définie par  $x_0 = a \in E$  et  $x_{n+1} = f(x_n)$  avec  $f : E \rightarrow E$ . La période de la suite est estimée à  $\sqrt{n}$ .

Supposons l'existence d'un diviseur  $p \leq \sqrt{n}$  de  $n$ . On considère la suite des  $(x_n[p])$  qui est aussi ultimement périodique. On pourrait alors appliquer l'algorithme à la suite modulo  $p$  pour trouver  $(i, j)$  tel que  $x_i \equiv x_j [p]$  mais on ne connaît pas explicitement ce  $p$ . Pour contourner l'obstacle, on utilise l'astuce suivante : il existe  $p$  un diviseur non trivial de  $n$  tel que  $x_i \equiv x_j [p]$  si et seulement si  $PGCD(x_i - x_j, n) \neq 1$ , expression qui ne dépend plus de  $p$ . On peut alors renvoyer  $PGCD(x_i - x_j, n)$ .

On remarque cependant que si l'algorithme renvoie  $n$ , cela signifie que la période de la suite modulo  $p$  n'est pas plus courte que la période originale. Alors l'algorithme échoue. On peut recommencer en modifiant  $a$  ou  $f$ .

Pour conserver le caractère pseudo-aléatoire de la suite, il a été montré que

les  $f$  de la forme  $x \mapsto x^2 + c$  sont particulièrement efficace, on choisira alors  $f : x \mapsto x^2 + 1$ .

Voici l'algorithme  $\rho$  de Pollard implémenté en python :

```
def pollard_rho(n):
    x = 2; y = 2; d = 1
    f = lambda x: (x**2 + 1) % n
    while d == 1:
        x = f(x); y = f(f(y))
        d = pgcd(abs(x-y), n)
    if d != n: return d, n//d
    else: print('Echec')
```

La fonction  $\text{pgcd}(a, b)$  ayant été codée par l'algorithme d'Euclide.

On peut alors tester notre programme sur les exemples de la partie précédente :

$$M_{31} \times M_{61} = 2147483647 \times 2305843009213693951 = 4951760154835678088235319297$$

Alors on a :

```
>>> pollard_rho(4951760154835678088235319297)
(2147483647, 2305843009213693951)
```

## 6 Conclusion

La distribution des nombres premiers au sein de  $\mathbb{N}$  présente donc des caractéristiques à la fois aléatoires et régulières. En fait, une vision possible du problème de leur répartition serait de considérer que leur infinité est à l'origine de la véracité des résultats ou conjectures à leur égard. Étudier les "petits" nombres premiers ne pose aucun problème, mais lorsqu'on s'intéresse aux nombres plus grands, leur étude se complexifie. C'est ce qu'expriment notamment les difficultés rencontrées lors de l'exécution et de l'implémentation des algorithmes. Cela vient en partie du fait qu'on ne connaisse pas de relation explicite entre les nombres premiers, c'est comme si chaque nombre premier possédait des caractéristiques uniques qui le rendent indépendant des autres nombres premiers. De ce fait, démontrer une propriété commune à tous devient un véritable défi, on peut alors rarement faire mieux qu'énoncer une simple conjecture.

Ils présentent également des particularités fascinantes qui paraissent souvent contre-intuitives : qui aurait pu penser, 500 ans auparavant, qu'un lien direct entre l'arithmétique des nombres premiers et l'analyse complexe existait ?

L'hypothèse de Riemann joue aussi un rôle majeur dans l'étude de  $\mathcal{P}$  ; connaître exactement la répartition des nombres premiers serait une grande avancée dans le domaine de l'arithmétique. Cependant, il est possible de se demander si elle ne cache pas un problème indémontrable. En effet, il est possible

que l'hypothèse fasse partie du théorème d'incomplétude de Gödel. L'hypothèse de Riemann est-elle vraiment décidable ? Autrement dit, est-il possible de la démontrer ? Il ne serait pas absurde de penser que l'hypothèse de Riemann est indécidable, étant donné qu'elle contient, dans sa définition propre, des nombres premiers qui font partie de la branche arithmétique, elle même responsable de l'indémontrabilité de certains résultats. Il pourrait en être de même pour les autres conjectures énoncées comme celle de Goldbach ou des nombres premiers jumeaux. Le réel problème provient peut-être alors du système d'axiomes sur lequel nous nous appuyons, qui est incomplet et ne permet donc peut-être pas une étude aussi approfondie de l'arithmétique.