

Prime vs Random Sets

Sebastien Plaasch

Maxime Rubio

Lucas Villiere

24 mai 2020

Résumé

Resume de notre projet..

Table des matières

| | |
|---|-----------|
| Introduction | 4 |
| 1 Création d'ensembles aléatoires suivant la distribution de $\pi(x)$ | 5 |
| 1.1 Approche analytique | 5 |
| 1.1.1 Méthode de création d'un ensemble aléatoire Q | 5 |
| 1.1.2 Définition de la fonction σ_q | 5 |
| 1.1.3 Écart entre $\pi(x)$ et $\sigma_q(x)$ | 6 |
| 1.1.4 Le Théorème des Nombres Premiers | 7 |
| 1.2 Approche probabiliste | 8 |
| 1.2.1 Ensembles aléatoires | 8 |
| 1.2.2 Ensembles probabilistes impairs | 10 |
| 2 Ensembles aléatoires et conjectures | 11 |
| 2.1 Les nombres premiers jumeaux | 11 |
| 2.2 Première conjecture | 15 |
| 2.2.1 Introduction | 15 |
| 2.2.2 Analyse | 15 |
| 2.2.3 Conclusion de l'analyse | 17 |

Introduction

Ici on fait l'introduction

1 Création d'ensembles aléatoires suivant la distribution de $\pi(x)$

1.1 Approche analytique

Notre première approche pour créer des ensembles aléatoires (désignés par Q) qui partagent la même distribution que celle des nombres premiers, est basée sur un théorème (théorème 1) issu de

Théorème 1. *L'hypothèse de Riemann est équivalente à l'assertion*

$$\forall n \geq 11, |p_n - \text{ali}(n)| < \frac{1}{\pi} \sqrt{n} \log^{5/2}(n)$$

où p_n représente le n -ième nombre premier.

1.1.1 Méthode de création d'un ensemble aléatoire Q

Les onze premiers éléments d'un ensemble Q sont choisis arbitrairement. Pour $n > 11$, voici la méthode de sélection de l'élément $q_n \in Q$:

- on pose $a = \max \left\{ q_{n-1}, \lceil \text{ali}(n) - \frac{1}{\pi} \sqrt{n} \log^{5/2}(n) \rceil \right\}$ (où $\lceil x \rceil$ désigne la partie entière supérieure de x) ;
- on pose $b = \lfloor \text{ali}(n) + \frac{1}{\pi} \sqrt{n} \log^{5/2}(n) \rfloor$ (où $\lfloor x \rfloor$ désigne la partie entière inférieure de x) ;
- q_n est choisi aléatoirement entre a et b .

En procédant de la sorte, le théorème 1 sera toujours vrai pour tout ensemble aléatoire Q .

Par cette méthode, nous avons créé 200 ensembles de nombres inférieurs à 10^7 , dont pour 100 d'entre-eux on a imposé la condition suivante : $\forall q_n \in Q$ tel que $n > 11$: q_n est impair. Nous testerons nos conjectures sur ces ensembles.

1.1.2 Définition de la fonction σ_q

On peut désormais définir $\sigma_q : [0, \infty[\rightarrow \mathbb{R}, x \mapsto \# \{n \in Q : n < x\}$. Nous parlerons systématiquement de la fonction σ_q alors que cette fonction n'est bien entendu pas unique, elle dépend à chaque fois de l'ensemble aléatoire Q sur lequel on travail. Cependant, grâce à la construction des ensembles Q , nous pouvons supposer que toutes les fonctions σ_q sont semblables, comme l'illustre le graphe *** en annexe.

Afin de visualiser $\sigma_q(x)$ en la comparant à $\pi(x)$ et $\frac{x}{\log(x)}$, voici leur graphe respectif :

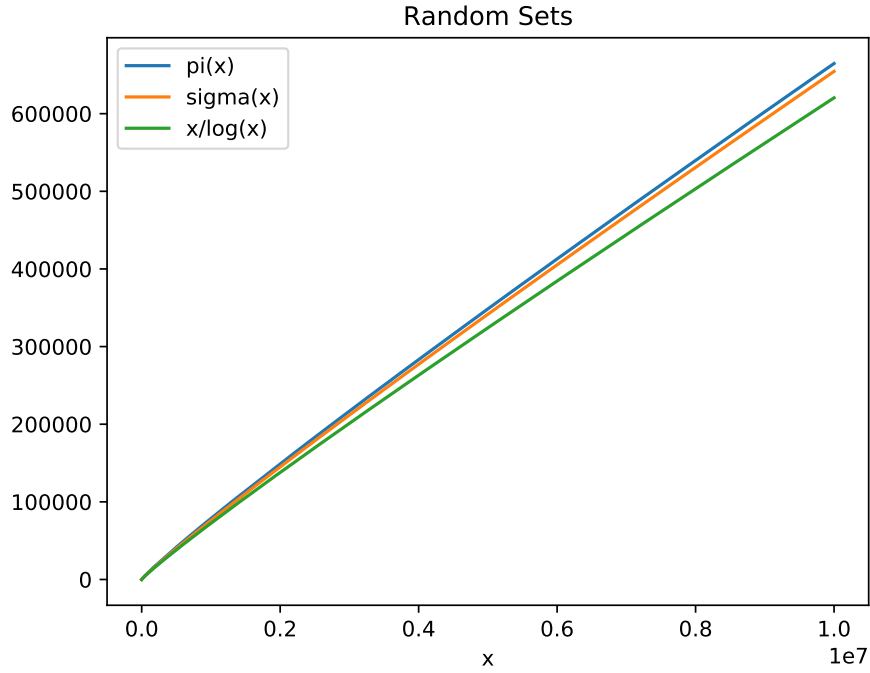


FIGURE 1 – Graphe de $\sigma_q(x)$, $\pi(x)$ et $\frac{x}{\log(x)}$

1.1.3 Écart entre $\pi(x)$ et $\sigma_q(x)$

La fonction σ_q semble suivre la même allure que $\pi(x)$. Cependant, lors de nos expérimentations, nous avons dessinés des graphes (que vous trouverez en annexe) pour des valeurs de x inférieures à celles de la figure 1. Pour des petites valeurs de x , la courbe de $\sigma_q(x)$ était presque confondue avec celle de $\frac{x}{\log(x)}$. Lorsque les valeurs de x sont de plus en plus grandes, $\sigma_q(x)$ tend vers $\pi(x)$. Pour analyser l'écart entre $\pi(x)$ et $\sigma_q(x)$, nous avons tracé le graphe de la fonction $\frac{\pi(x)}{\sigma_q(x)}$:

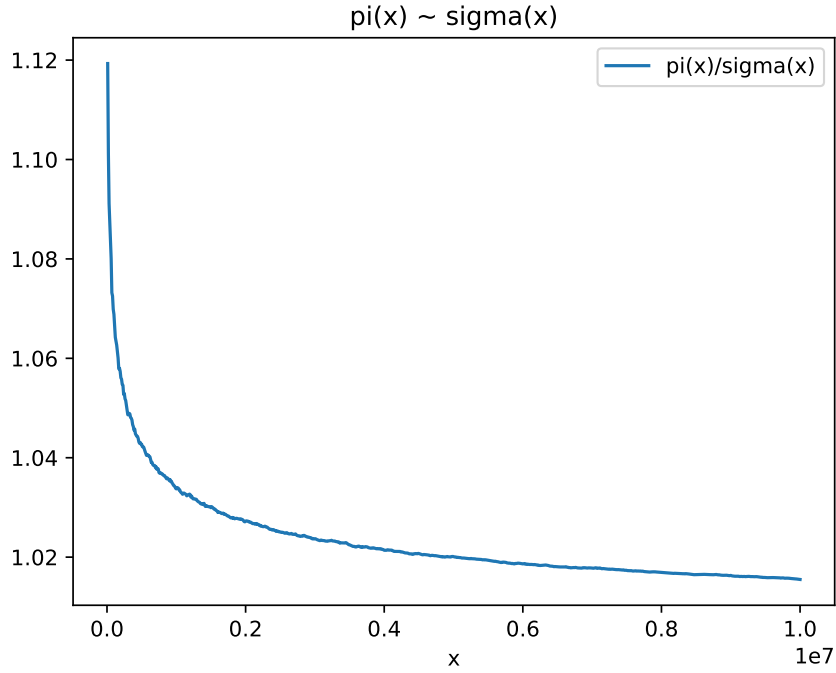


FIGURE 2 – Graphe de $\frac{\pi(x)}{\sigma_q(x)}$

1.1.4 Le Théorème des Nombres Premiers

L'objectif de cette section est de prouver la fidélité de nos ensembles aléatoires Q à la répartition des nombres premiers. Pour ce faire, nous allons vérifier si le Théorème des Nombres Premiers, cité ci-après, est vrai quand on remplace $\pi(x)$ par $\sigma_q(x)$.

Théorème 2 (Théorème des Nombres Premiers). *Quand x tend vers l'infini :*

$$\pi(x) \sim \frac{x}{\log(x)}$$

Pour démontrer le Théorème des Nombres Premiers, il a été démontré que, quand x tend vers l'infini, $\text{Li}(x) \sim \frac{x}{\log(x)}$. Nous pouvons montrer, graphiquement, que lorsque x tend vers l'infini, $\sigma_q(x) \sim \text{Li}(x)$ (voir figure 3), ce qui implique que $\sigma_q(x) \sim \frac{x}{\log(x)}$.

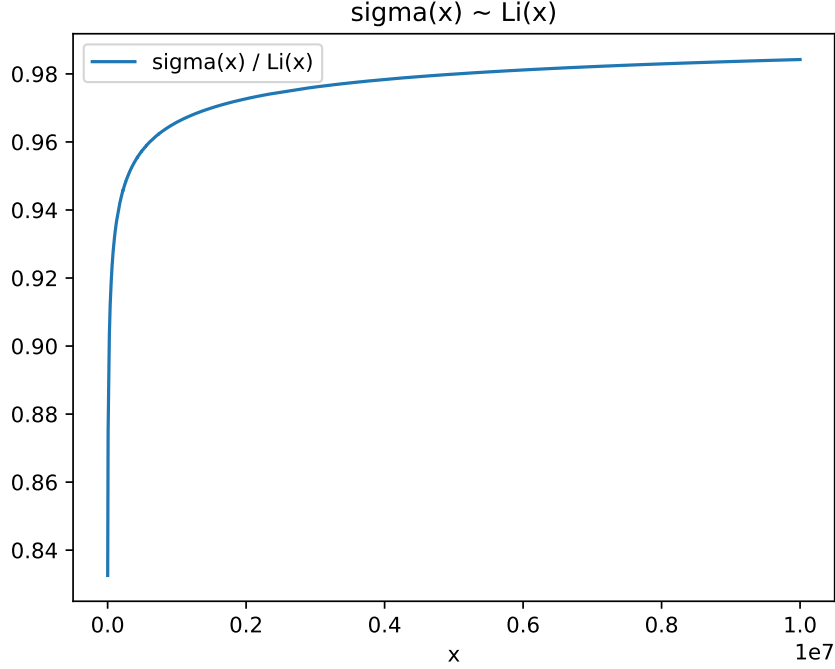


FIGURE 3 – Graphe de $\frac{\sigma_q(x)}{Li(x)}$

1.2 Approche probabiliste

Soient $E_n := \{k \in \mathbb{N}^* \mid k \leq n\}$ l'ensemble des entiers inférieurs ou égaux à n , $P_n := \{k \in E_n \mid k \text{ est premier}\}$ l'ensemble des nombres premiers inférieurs ou égaux à n , et la fonction $\pi(n) := \#P_n$, le nombre de premiers inférieurs ou égaux à n .

On a vu que la fonction $Li(x) = \int_2^x \frac{dt}{\log t}$ donne une bonne approximation de $\pi(n)$.

Cette fonction peut être approximée par la somme de Riemann de pas constant = 1 :

$$S\left(\frac{1}{\log x}\right) = \sum_{k=0}^n \frac{1}{\log(2+k)} = \frac{1}{\log 2} + \sum_{k=3}^n \frac{1}{\log k} \quad (1)$$

La fonction $\frac{1}{\log x}$ est une fonction continue, décroissante et positive sur l'intervall $[2, \infty[$. L'erreur entre $Li(n)$ et la fonction en escalier ci-dessus est donc bornée.

$$\left| S\left(\frac{1}{\log x}\right) - Li(n) \right| \leq \left| \sum_{k=2}^n \frac{1}{\log(k+1)} - \frac{1}{\log k} \right| = \frac{1}{\log(2)} - \frac{1}{\log(n+1)} < \frac{1}{\log 2} < 2$$

1.2.1 Ensembles aléatoires

Nous allons alors générer k ensembles aléatoires $R_{k_n} \subset E_n$ de sorte que :

$$\forall i \in E_n, P(i \in R_{k_n}) = \begin{cases} 0 & \text{si } i = 1 \\ 1 & \text{si } i = 2 \\ \frac{1}{\log n} & \text{si } i \geq 3 \end{cases}$$

Il est à noter deux cas particuliers :

- Le nombre 1 est exclu. En effet, $\frac{1}{\log 1}$ n'est pas défini. Par définition, 1 n'est pas un nombre premier.
- Le nombre 2 est inclu par défaut. En effet, $P(2 \in R_{k_n}) = \frac{1}{\log 2} > 1$. De plus, le nombre 2 est par définition, un nombre premier.

Nous allons aussi introduire la fonction $\sigma_k(n) := \#R_{k_n}$. Cette fonction mesure la taille de l'ensemble aléatoire.

$\sigma_k(n)$ est donc une valeur aléatoire strictement inférieure à n dont l'espérance est donnée par la formule suivante :

$$E[\sigma_k(n)] = 0 + 1 + \frac{1}{\log 3} + \dots + \frac{1}{\log n} = 1 + \sum_{k=3}^n \frac{1}{\log k} \quad (2)$$

On observe alors que l'erreur entre l'espérance de $\sigma_k(n)$ et la somme de Riemann (1) est constante, égale à $\frac{1}{\log 2} - 1 < 1$.

Les ensembles aléatoires générés de cette manière suivront donc une distribution similaire à $\text{Li}(x)$, et donc à $\pi(x)$ (voir figures 6 et 5).

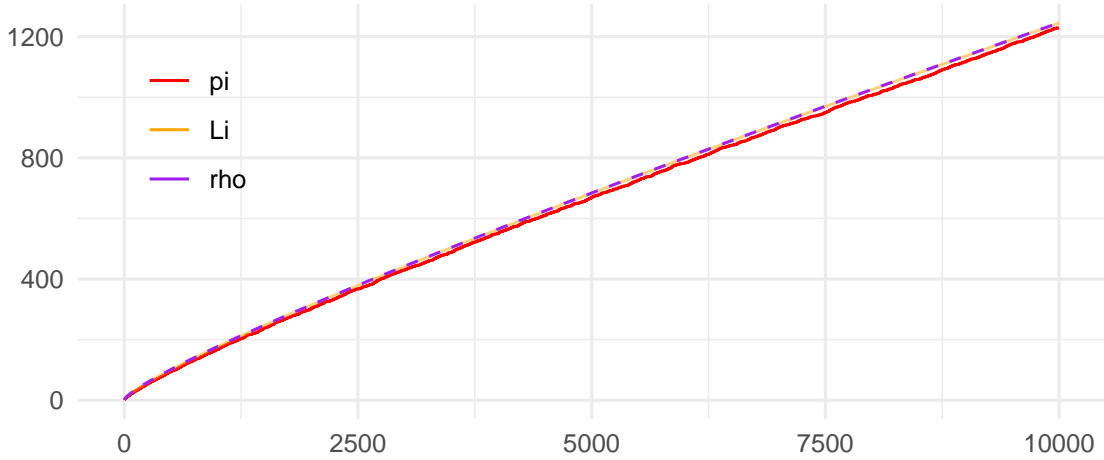


FIGURE 4 – Graphes des fonctions π , Li et σ . On observe que Li et σ sont superposées.

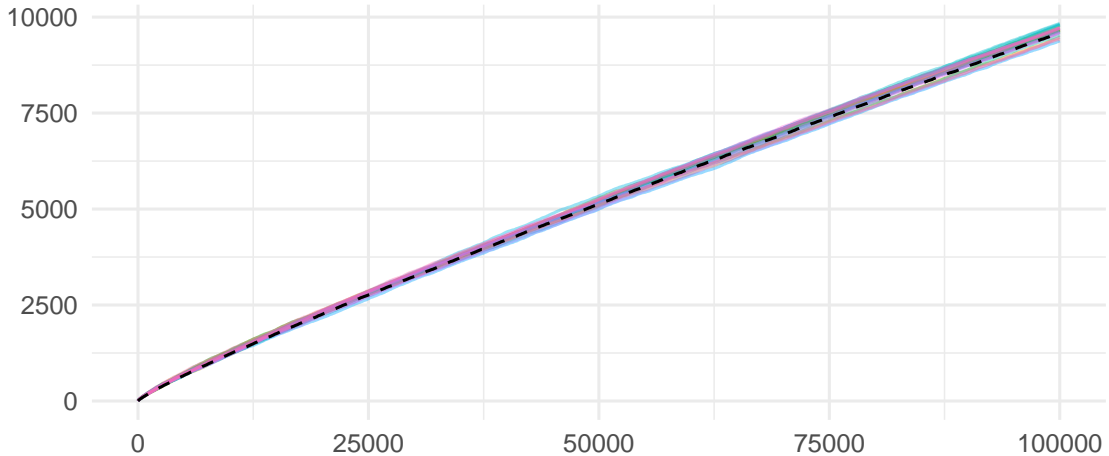


FIGURE 5 – graphes des fonctions σ_k pour $k \leq 25$ (25 premiers ensembles) et π (en pointillé)

1.2.2 Ensembles probabilistes impairs

Bien que ces ensembles aléatoires suivent la distribution de $\pi(x)$, il manque une propriété importante des nombres premiers : à l'exception de 2, tous les nombres premiers sont impairs.

Nous allons alors modifier l'algorithme mentionné précédemment afin de générer des ensembles ayant cette propriété. Soit $O_n := E_n \cap 2\mathbb{N} + 1$ l'ensemble des entiers impairs inférieurs ou égaux à n . Soit alors l'ensemble aléatoire $U_{k_n} \subset \{O_n \cup \{2\}\}$ tel que $\forall i \in \{O_n \cup \{2\}\}$, la probabilité que i soit dans l'ensemble U_{k_n} soit :

$$P(i \in U_{k_n}) = \begin{cases} 0 & \text{si } i = 1 \\ 1 & \text{si } 2 \leq i < 9 \\ \frac{2}{\log n} & \text{si } i \geq 9 \end{cases}$$

Le fait que $P(i \in U_{k_n}) = 1$ pour les entiers impairs inférieurs à 9 découle du fait que $2/\log(n) > 1$ pour ces entiers.

On observe alors que ces suites débutent avec les mêmes éléments : $(2, 3, 5, 7, \dots)$, mais contiennent des éléments aléatoires à partir du 5^e terme. Pour tout $n \geq 5$, la fonction σ est une valeur aléatoire d'espérance :

$$E[\sigma(n)] = 4 + \sum_{i=5}^{\lfloor n/2 \rfloor} \frac{2}{\log(2i-1)}$$

En réécrivant l'équation 1 de la page 8 en une somme de Riemann de pas constant = 2, on obtient que :

$$S\left(\frac{1}{\log(x)}\right) = \sum_{k=0}^n \frac{2}{\log 2 + k}$$

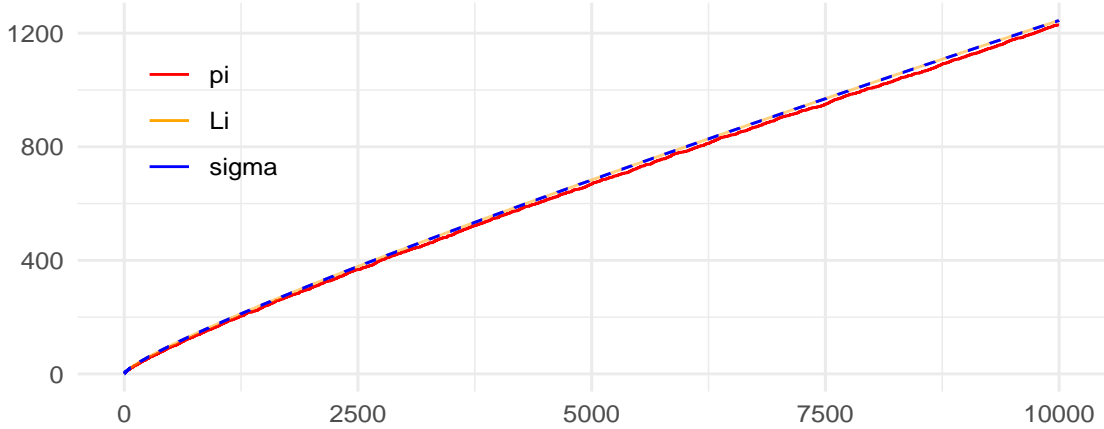


FIGURE 6 – Graphes des fonctions π et σ .

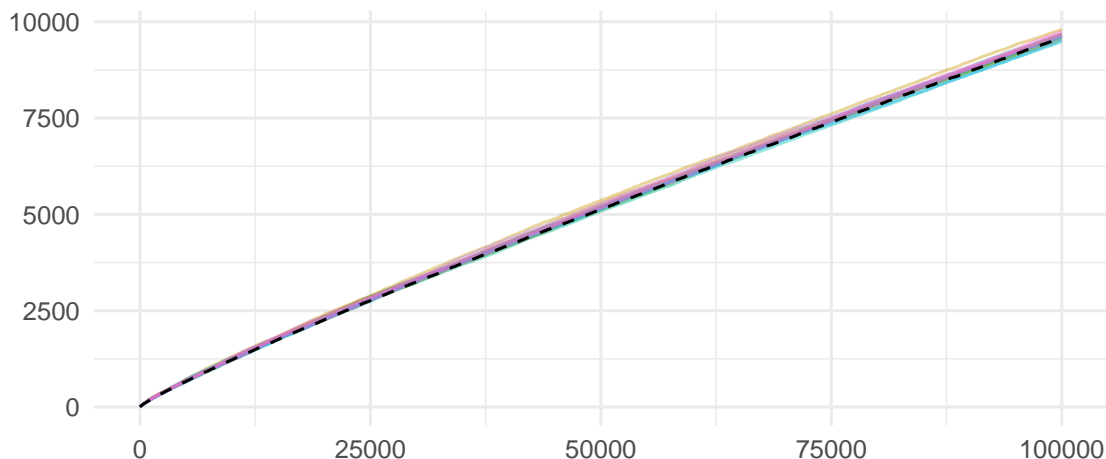


FIGURE 7 – graphes des fonctions σ_k pour $k \leq 25$ (25 premiers ensembles impairs) et π

2 Ensembles aléatoires et conjectures

Dans cette section, nous allons vérifier si des conjectures sur les nombres premiers peuvent s'appliquer aux ensembles aléatoires créés à la section 1. Ainsi, on sera en mesure d'estimer si la véracité d'une conjecture est susceptible de tenir grâce à la répartition des nombres premiers plutôt qu'à leur propriété d'être premier.

2.1 Les nombres premiers jumeaux

La première conjecture que nous allons analyser, et sans doute la plus célèbre, est la conjecture des nombres premiers jumeaux.

Définition 1. Soient $a, b \in \mathbb{N}$, $a < b$, on dit que a et b sont jumeaux si $a + 2 = b$.

Conjecture 1. *Il y a une infinité de nombres premiers jumeaux.*

Ces dernières années, il y'a eu de grosses avancées dans la démonstration de la conjecture. Ainsi, pour tout $m \geq 1$, soit $H_m := \liminf_{n \rightarrow \infty} (p_{n+m} - p_n)$, où p_n dénote le n -ième nombre premier. La conjecture des nombres premiers jumeaux est donc équivalente à $H_1 = 2$. En 2013, le mathématicien chinois Zhang Yitang est le premier à trouver une borne supérieure finie pour H_1 , il a démontré que $H_1 \leq 70\,000\,000$. Suite à la publication de Zhang Yitang, de nombreux mathématiciens se sont mis en quête de réduire la borne supérieure de H_1 . En optimisant les résultats de Zhang Yitang et grâce à d'autres méthodes ils ont pu montrer que $H_1 \leq 246$.

Afin de tester la conjecture sur les ensembles aléatoires, nous avons tracé un graphe (figure 8), où pour chaque ensembles aléatoires et l'ensemble des nombres premiers (jusqu'à 10^7) on trace une fonction qui compte le nombre de jumeaux. Nous pouvons faire les observations suivantes :

- Pour chaque type d'ensemble (selon l'approche utilisée), on constate, logiquement, que le nombre de jumeaux est plus ou moins le double pour les ensembles impairs.
- Il y'a plus de jumeaux dans les ensembles R que dans les ensembles Q ce qui peut s'expliquer par le fait qu'ils ont plus d'éléments que les ensembles Q .
- De manière générale, toutes les courbes sont croissantes, ce qui indiquerait, autant pour les nombres premiers que pour les ensembles aléatoires, que le nombre de jumeaux tend vers l'infini.

Par ailleurs, ce graphe éveille une idée intéressante. Si on désigne par f la fonction qui compte le nombre de jumeaux dans l'ensemble des nombres premiers et par g celle qui compte les jumeaux

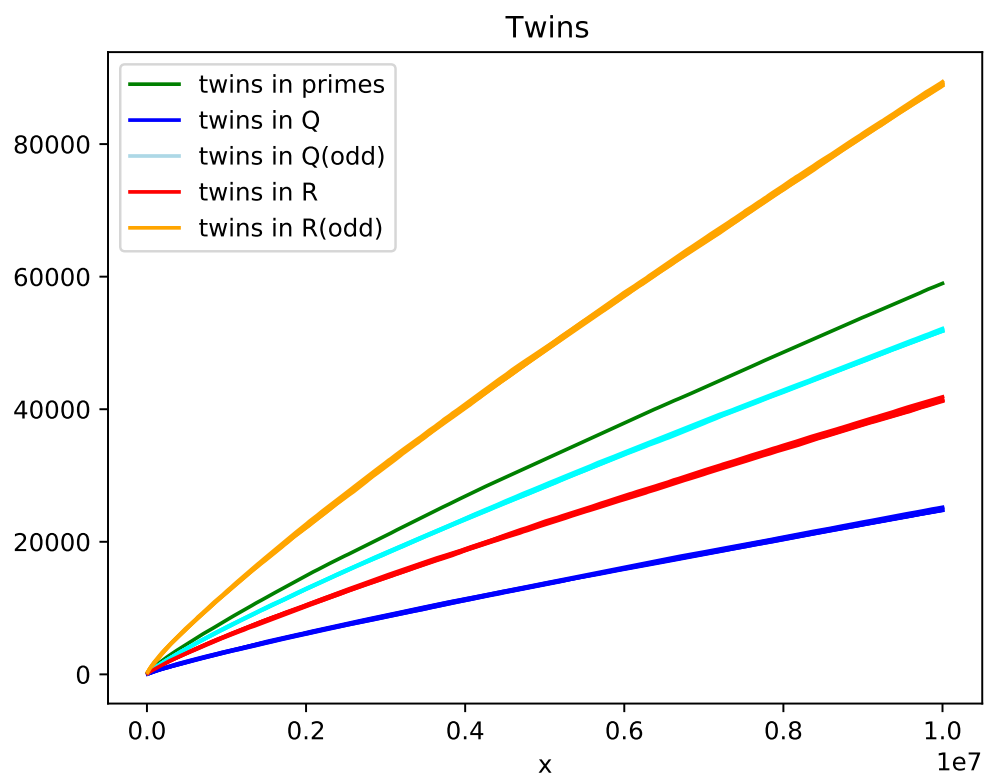


FIGURE 8 – Nombre de jumeaux dans chaque ensemble

dans un ensemble aléatoire et qu'on parvient à montrer que f et g sont semblables (sous-entendu qu'elles le sont), alors $\lim_{x \rightarrow \infty} g(x) = \infty \Rightarrow \lim_{x \rightarrow \infty} f(x) = \infty$. Ce qui démontrerait la conjecture des nombres premiers jumeaux. Pour ce faire une idée d'une éventuelle équivalence entre f et g , voici le graphe (figure 9) de $\frac{f(x)}{g(x)}$, où g est appliquée à un ensemble impair Q choisi arbitrairement.

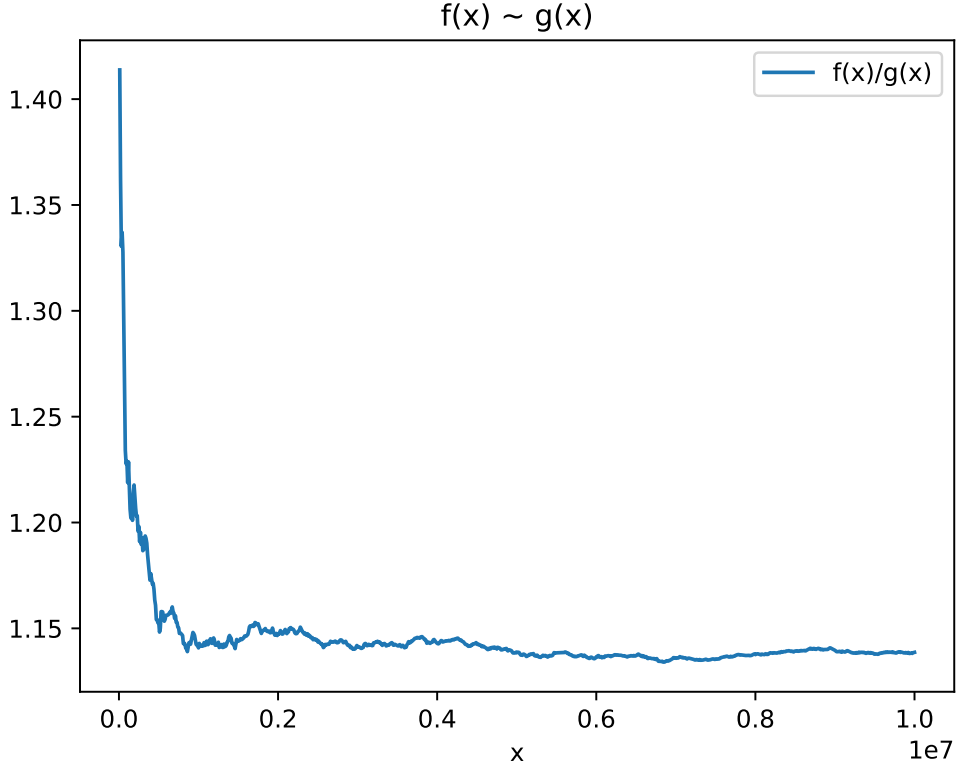


FIGURE 9 – $\frac{f(x)}{g(x)}$

Pour les ensembles aléatoires R (créés par l'approche probabiliste), on peut avoir une bonne approximation de la fonction g , qui compte le nombre de jumeaux. Soit R un tel ensemble, on définit $g : [0, \infty[\rightarrow \mathbb{R}, x \mapsto \sum_{i=2}^{\lfloor x \rfloor} \frac{1}{\log(i) \log(i+2)}$. "Elle donne une bonne approximation car c'est l'espérance mathématique." À titre comparatif, voici le graphe de $g(x)$ (figure 10), d'une fonction qui compte les jumeaux dans l'ensemble des nombres premiers et une qui les compte dans R .

On peut montrer que $\lim_{x \rightarrow \infty} g(x) = +\infty$. Donc $f \geq g \Rightarrow \lim_{x \rightarrow \infty} f(x) = +\infty$.

Pour conclure, la répartition des nombres premiers nous a permis de créer des ensembles aléatoires qui ont une infinité de nombres jumeaux.

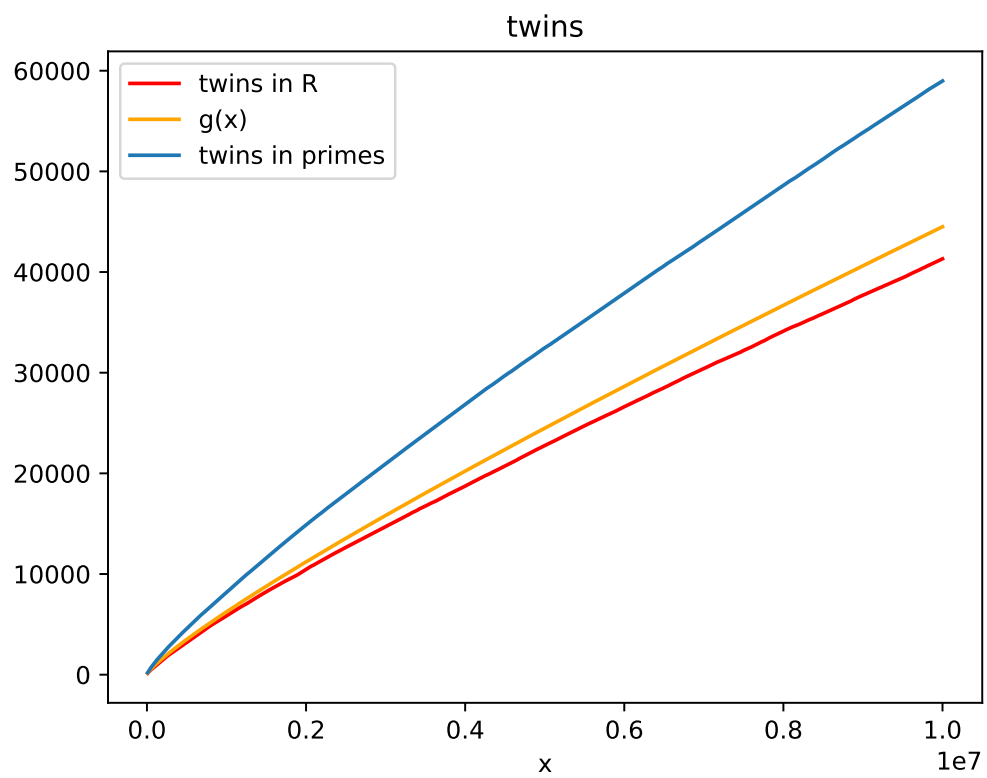


FIGURE 10 – Comparaison de $g(x)$

2.2 Première conjecture

2.2.1 Introduction

La conjoncture que nous allons tester ici est la suivante : Pour tout $n = 6, 7, \dots$, il existe un nombre premier p tel que $6n - p$ et $6n + p$ sont tous les deux premiers.¹

Nous allons d'abord vérifier que cette conjecture tient pour les nombres premiers, puis vérifier si celle-ci tient aussi pour les ensembles aléatoires suivant leur distribution. La procédure pour analyser cette conjecture est donc la suivante : pour chaque $n \in \mathbb{N}, 6 \leq n \leq 10000$, nous vérifierons si l'assertion tient. Nous collecterons alors tout n tel que $\neg P(n)$ dans un tableau de données afin d'analyser, pour chaque ensemble ou chaque groupe d'ensemble, le nombre et la distribution des erreurs.

2.2.2 Analyse

Soit $P(n)$ l'assertion "Il existe $p \in R_{k_n}, p \geq 6$ tel que $6n - p \in R_k$ et $6n + p \in R_k$ ".

En utilisant un algorithme², nous avons pu observer que l'assertion est vraie pour tout $n \leq 10^5$.

Le graphique 11 ci-dessous représente en abscisses le nombre total d'erreur (c'est à dire le nombre d'entiers n tel que $\neg P(n)$ pour un certain ensemble) et en ordonnées le plus grand entier pour laquelle l'assertion est fausse. Chaque point représente un ensemble tel que défini précédemment.

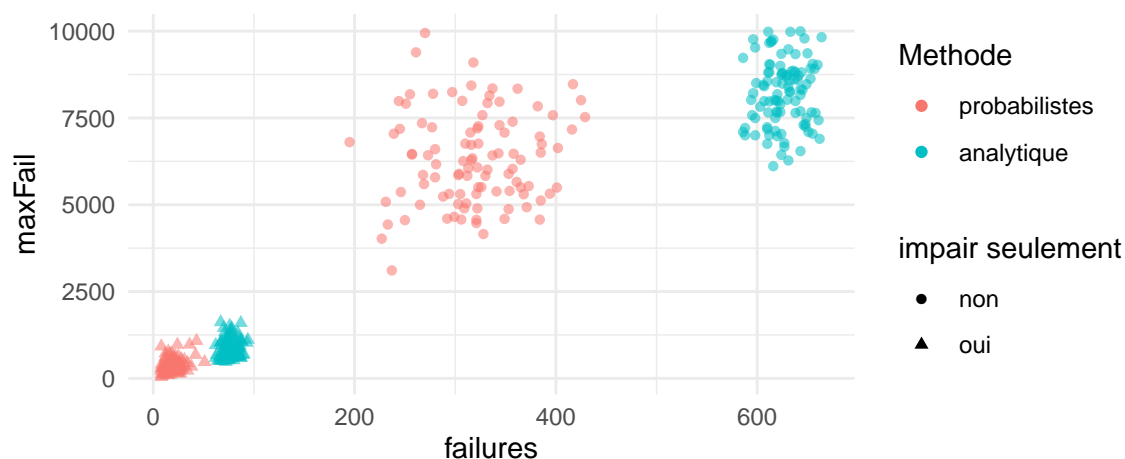


FIGURE 11 –

Nous debuterons par remarquer que les ensembles non restraints aux nombres impairs génèrent un nombre assez conséquent d'erreurs : plus de 200 erreurs pour la quasi-totalité de ces ensembles. De plus, ces erreurs persistent assez tardivement : pour la majorité de ces ensembles, il existe (au moins) un entier $n > 5000$ pour lequel l'assertion est fausse. Cependant, les ensembles composés de nombres impairs ont un nombre relativement faible d'erreurs (moins de 50 pour les ensembles générés par l'algorithme probabilistique, moins de 100 pour les ensembles générés par les algorithmes analytiques, voir figure [x]). De plus, le plus grand entier pour lequel l'assertion n'est pas vérifiée est relativement faible : cela signifie que pour tout entier $n > 2500$, l'assertion est vérifiée. Les diagrammes à boîtes 12 et 13 ci-dessus offrent un aperçu de la distribution des erreurs

1. Conjecture 2.3 de ...

2. voir /py_code/...

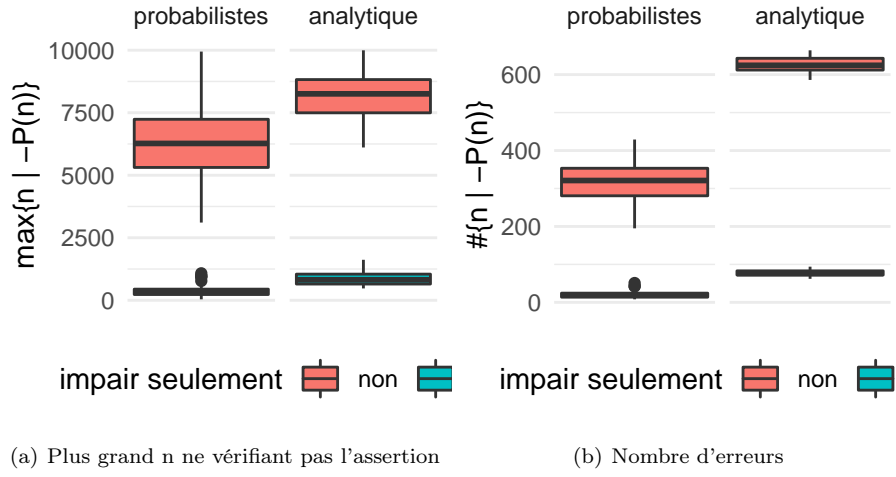


FIGURE 12 – Diagrammes en boîte

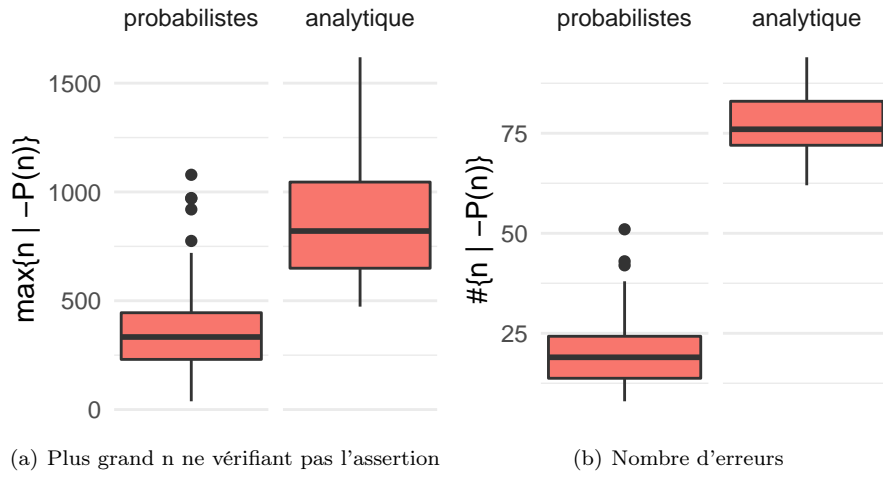


FIGURE 13 – Diagrammes en boîte - ensembles impairs.

- Il existe parmi ces ensembles 3 ensembles vérifiant l'assertion pour tout $n, 100 \leq n \leq 10000$:
- l'ensemble probabiliste impair 032 dont le plus grand entier ne vérifiant pas l'assertion est 38
 - l'ensemble probabiliste impair 091 dont le plus grand entier ne vérifiant pas l'assertion est 74
 - l'ensemble probabiliste impair 033 dont le plus grand entier ne vérifiant pas l'assertion est 97

2.2.3 Conclusion de l'analyse

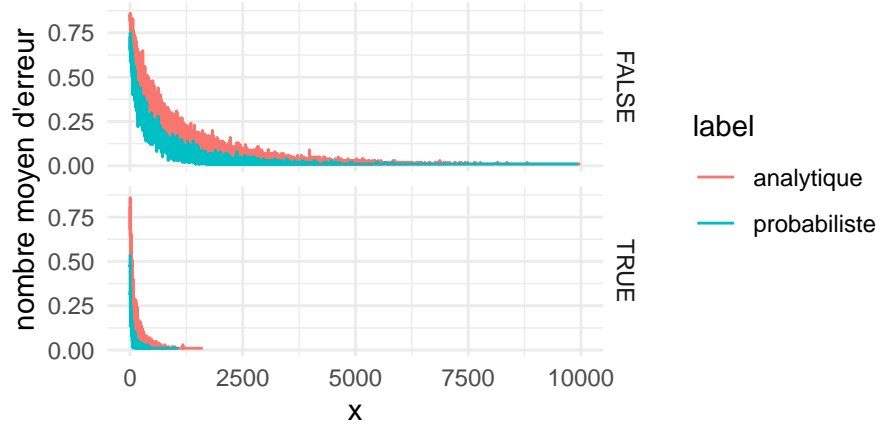


FIGURE 14 – Diagramme représentant le nombre moyen d'erreur (entier ne vérifiant pas P) sur l'intervall $(n, n + 5]$

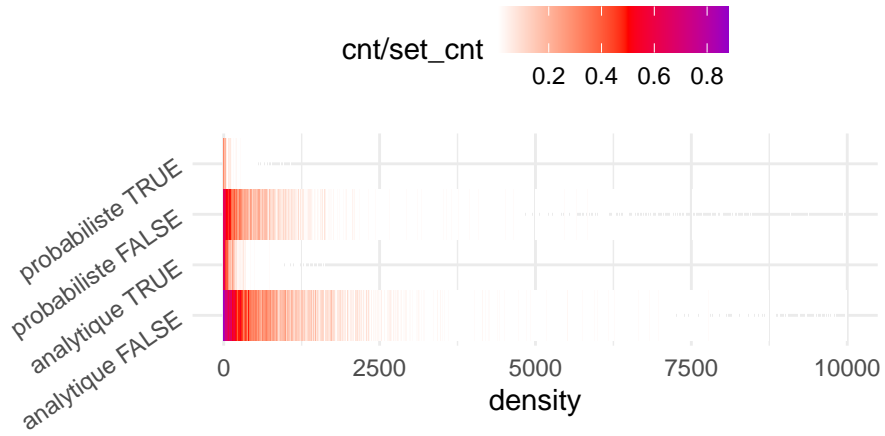


FIGURE 15 – Diagramme représentant le nombre moyen d'erreur (entier ne vérifiant pas la conjecture) sur l'intervall $(n, n + 5]$

Conclusion