Autour du théorème des nombres premiers

Xavier Caruso* et David Pigeon†

Septembre 2007

Résumé

On donne une méthode générale et élémentaire pour obtenir des encadrements « à la Tchebychev » de $\pi(x)$, le nombre de nombres premiers inférieurs ou égaux à x. Numériquement, on constate que ces estimations semblent de plus en plus précises, ce qui nous amène à nous demander si la méthode converge pour donner au final le théorème des nombres premiers (*i.e.* l'équivalent $\pi(x) \sim \frac{x}{\log x}$). Nous montrons et discutons un résultat déconcertant : c'est effectivement le cas, mais sous l'hypothèse du théorème des nombres premiers.

Mots-clés : théorie analytique des nombres, théorème des nombres premiers.

De nombreux problèmes liés à la répartition des nombres premiers fascinent les mathématiciens depuis de nombreuses générations. Le plus célèbre d'entre eux est d'estimer le nombre de nombres premiers parmi les entiers compris entre 1 et x, traditionnellement noté $\pi(x)$. L'un des premiers résultats sérieux à ce sujet est l'encadrement :

$$0,92129\frac{x}{\log x} + o\left(\frac{x}{\log x}\right) \leqslant \pi(x) \leqslant 1,10556\frac{x}{\log x} + o\left(\frac{x}{\log x}\right) \tag{1}$$

dû à Tchebychev dans les années 1850. On sait depuis 1896 qu'en réalité on a l'équivalence $\pi(x) \sim \frac{x}{\log x}$, résultat connu sous l'appelation *théorème des nombres premiers*.

Dans ce papier, nous nous proposons de présenter certains raffinements des méthodes de Tchebychev qui permettent d'obtenir une panoplie d'encadrements de la même forme que (1), mais faisant intervenir des constantes numériques qui pourront être plus proches de 1. Plus précisément, on montre un théorème général d'aspect assez technique (théorème 5) à partir duquel toutes les estimations sont déduites simplement, moyennant une certaine quantité de calcul (que l'on confie à un ordinateur) qui, bien entendu, augmente avec la précision de l'approximation souhaitée.

L'article s'organise comme suit. Nous commençons par détailler dans la section 1 les objets essentiels, ainsi que les outils généraux qui seront constamment utilisés par la suite. La deuxième section est consacrée à l'énoncé et à la preuve de notre théorème principal. Finalement, dans une dernière partie, nous discutons les limites de notre résultat : essentiellement, on se demande dans quelle mesure une puissance de calcul infinie permettrait d'obtenir des constantes arbitrairement proches de 1, et donc le théorème des nombres premiers.

Dans tout l'article, $\mathcal{P}$ désigne l'ensemble des nombres premiers, et la lettre p est généralement réservée à un élément de cet ensemble. Signalons finalement que les variables muettes qui décriront dans la suite les ensembles de sommation seront implicitement astreintes à ne prendre que des valeurs entières positives, et même, sauf mention explicite du contraire, strictement positives. Ainsi, par exemple, pour un réel x, la notation $\sum_{n \leqslant x}$ est équivalente à $\sum_{n=1}^{[x]}$ (où $[x]$ est la partie entière de x).

Ce texte est essentiellement une reformulation plus moderne (et peut-être plus accessible au novice) de [1]. Le lecteur intéressé pourra se reporter à cet article et sa bibliographie pour des compléments. Les auteurs remercient Olivier Ramaré pour leur avoir signalé cette référence et avoir répondu à plusieurs de leurs questions.

*Chargé de recherche au CNRS affecté à l'université de Rennes 1
†Magistérien à l'ENS Lyon

1 Préliminaires

Nombreuses sont les démonstrations présentées dans cet article qui se dérouleront en deux temps selon le modèle suivant. On obtient tout d'abord une identité arithmétique, avant d'estimer indépendamment par des arguments de type analytique les deux membres de l'égalité. Nous expliquons dans ce paragraphe les outils principaux utilisés dans cette démarche.

1.1 L'aspect arithmétique : produit de convolution

Les identités arithmétiques évoquées précédemment s'expriment souvent à l'aide d'un produit de convolution. Si f et g sont deux fonctions de $\mathbb{N}^\star$ dans $\mathbb{R}$, on définit leur produit de convolution $f \star g$ par :

$$f \star g(n) = \sum_{ij=n} f(i)g(j) = \sum_{i|n} f(i)g\left(\frac{n}{i}\right)$$

pour tout $n \in \mathbb{N}^\star$. Il est facile de vérifier que, restreinte au sous-ensemble $\mathcal{G}$ formé des fonctions qui ne s'annulent pas en 1, cette formule définit une loi de groupe abélien. L'élément neutre en est la fonction e_1 définie par $e_1(1) = 1$ et $e_1(n) = 0$ pour tout $n > 1$.

Notons $\mathbb{1} \in \mathcal{G}$ la fonction constante égale à 1. La première fonction arithmétique qui joue un rôle essentiel est la *fonction de Möbius* μ définie comme l'inverse de $\mathbb{1}$ pour la loi $\star$. Il est remarquable que $\mu(n)$ admet une expression simple en fonction de la décomposition en facteurs premiers de n, comme le précise le lemme suivant :

Lemme 1. *La fonction μ de Möbius s'exprime comme suit :*
- $\mu(1) = 1$ *;*
- $\mu(p_1 \cdots p_k) = (-1)^k$ *si les p_i sont des nombres premiers deux à deux distincts ;*
- $\mu(n) = 0$ *si n est divisible par le carré d'un nombre premier.*

Démonstration. Si μ' est la fonction définie par les formules du lemme, il suffit de montrer que $\mathbb{1} \star \mu' = e_1$, c'est-à-dire $\sum_{d|n} \mu'(d) = e_1(n)$ pour tout entier n. Pour $n = 1$ la vérification est immédiate. Si, au contraire, $n > 1$, il existe un nombre premier p qui apparaît dans la décomposition en facteurs premiers de n avec une puissance $a \geqslant 1$. Ainsi n s'écrit $n = n'p^a$ avec n' premier avec p. Un diviseur d de n s'écrit, lui, sous la forme $d = d'p^b$ avec d' diviseur de n' et $0 \leqslant b \leqslant a$. Par ailleurs, on vérifie directement en utilisant la formule que $\mu'(d'p^b) = \mu'(d')\mu'(p^b)$. Tout ceci entraîne :

$$\sum_{d|n} \mu'(d) = \sum_{d'|n'} \sum_{b=0}^{a} \mu'(d'p^b) = \sum_{d'|n'} \mu'(d') \cdot \sum_{b=0}^{a} \mu'(p^b).$$

Or, du fait que $a \geqslant 1$, le deuxième facteur vaut $1 - 1 + 0 + \cdots + 0 = 0$ et donc le produit est bien nul, comme voulu. $\square$

Le lemme précédent a plusieurs conséquences importantes. Déjà, d'un point de vue pratique, il permet un calcul efficace, basé sur le crible d'Ératosthène, des premières valeurs prises par la fonction de Möbius. D'un point du vue plus théorique, il montre d'une part que la fonction de Möbius est *multiplicative*, dans le sens où elle vérifie $\mu(nm) = \mu(n)\mu(m)$ dès que n et m sont premiers entre eux, et d'autre part qu'elle est bornée en valeur absolue par 1, ce que nous utiliserons couramment dans la suite.

Une autre fonction qui semble intéressante pour l'estimation qui nous motive est évidemment la fonction indicatrice des nombres premiers $\mathbb{1}_\mathcal{P}$ puisqu'elle est directement reliée à la fonction π *via* la formule $\pi(x) = \sum_{n \leqslant x} \mathbb{1}_\mathcal{P}(n)$. Cependant, elle ne vérifie pas d'identité arithmétique simple et on est amené à lui substituer une fonction auxiliaire définie comme suit :

Définition 2. La *fonction de von Mangoldt* est définie par :

$$\Lambda(n) = \begin{cases} \log p & \text{si } n = p^a \text{ avec } p \text{ premier et } a \geqslant 1 \\ 0 & \text{sinon} \end{cases}$$

pour tout entier n.

De plus, on pose pour tout réel $x \geqslant 1$, $\psi(x) = \sum_{n \leqslant x} \Lambda(n)$.

Avant d'entrer plus en détails dans les propriétés de la fonction de von Mangoldt, il est bon de se persuader heuristiquement qu'elle n'est pas si éloignée que cela de $\mathbb{1}_{\mathcal{P}}$. En effet, on conçoit facilement (par exemple par le théorème des nombres premiers) que les puissances d'entiers — et donc *a fortiori* les puissances de nombres premiers — sont minoritaires par rapport aux nombres premiers eux-mêmes. Ainsi les termes parasites ne devraient pas avoir une influence importante dans le calcul de $\psi(x)$. Il reste à comprendre l'effet de la pondération en $\log p$. Mais vue « de loin » sur l'intervalle $[1, x]$, la fonction logarithme est pratiquement constante égale à $\log x$. Ces deux remarques amènent à penser que $\psi(x) \simeq \pi(x) \log x$... ce qu'effectivement nous énoncerons précisément et prouverons par la suite.

Un avantage important de la fonction Λ (par rapport à la fonction $\mathbb{1}_{\mathcal{P}}$) réside dans les propriétés énoncées dans la proposition suivante dont la démonstration (facile) est laissée au lecteur :

Proposition 3. *On a les identités* $\mathbb{1} \star \Lambda = \log$ *et* $\mu \star \Lambda = -\mu \log$.

1.2 L'aspect analytique : transformation d'Abel

Comme nous l'avons déjà signalé, la fonction de von Mangoldt a le mauvais goût de faire apparaître un facteur parasite log. Cette partie a pour but de donner les outils généraux qui vont être utilisés pour s'en débarrasser.

Plus précisément, la situation la plus fréquente qui apparaîtra au fil des démonstration sera la suivante : on a des renseignements sur la série de terme général $v_n = u_n \log n$ à partir desquels on cherche à déduire des informations sur la série $\sum u_n$. Pour ce faire, on procède à la transformation d'Abel suivante, pour $x \geqslant 2$:

$$U(x) = \sum_{n \leqslant x} u_n = u_1 + \sum_{2 \leqslant n \leqslant x} u_n \log n \, \frac{1}{\log n} = u_1 + \frac{V(x)}{\log[x]} + \sum_{2 \leqslant n \leqslant x-1} V(n) \left(\frac{1}{\log n} - \frac{1}{\log(n+1)} \right) \tag{2}$$

où $[x]$ est la partie entière de x et $V(x) = \sum_{n \leqslant x} v_n$ est la somme partielle de la série v_n dont le comportement est connu par hypothèse. Le terme $\frac{V(x)}{\log[x]}$ est donc bien maîtrisé. Il en est évidemment de même de u_1 (qui ne dépend pas de x). Le dernier terme, quant à lui, se traite en utilisant l'équivalent $\frac{1}{\log n} - \frac{1}{\log(n+1)} \sim \frac{1}{n \log^2 n}$ et donnera en général une contribution minoritaire.

Typiquement, supposons que l'on sache que $V(x) = O(x)$. Alors, le terme général de la série qui apparaît dans le membre de droite de (2) est un $O(\frac{1}{\log^2 n})$. Or :

$$\sum_{2 \leqslant n \leqslant x} \frac{1}{\log^2 n} = \sum_{2 \leqslant n \leqslant \sqrt{x}} \frac{1}{\log^2 n} + \sum_{\sqrt{x} < n \leqslant x} \frac{1}{\log^2 n} \leqslant \frac{\sqrt{x}}{\log^2 2} + \frac{4x}{\log^2 x} = O\left(\frac{x}{\log^2 x} \right)$$

d'où on déduit :

$$U(x) = u_1 + \frac{V(x)}{\log[x]} + O\left(\frac{x}{\log^2 x} \right) = \frac{V(x)}{\log x} + O\left(\frac{x}{\log^2 x} \right) \tag{3}$$

la dernière égalité provenant de :

$$\left| \frac{V(x)}{\log[x]} - \frac{V(x)}{\log x} \right| \leqslant |V(x)| \left(\frac{1}{\log(x-1)} - \frac{1}{\log x} \right) = O\left(\frac{1}{\log^2 x} \right).$$

1.3 Illustration de la méthode : premières estimations

La méthode que l'on vient d'exposer s'applique presque directement avec la suite $u_n = \mathbb{1}_{\mathcal{P}}(n)$ et permet de rendre rigoureuse l'approximation vague $\psi(x) \simeq \pi(x) \log x$ annoncée tantôt. Tout d'abord on évacue le problème des puissances de nombres premiers en utilisant les inégalités suivantes :

$$\psi(x) = \sum_{\substack{p \in \mathcal{P} \\ p \leqslant x}} \log p \left[\frac{\log x}{\log p} \right] \leqslant \sum_{\substack{p \in \mathcal{P} \\ p \leqslant x}} \log x = \pi(x) \log(x) \quad \text{et} \quad \psi(x) \geqslant V(x) \tag{4}$$

en conservant les notations de 1.2. Ceci nous ramène essentiellement à comparer les quantités $U(x)$ et $V(x)$. Pour appliquer la méthode précédente, on a besoin d'une estimation sur $V(x)$. Elle résulte directement de la proposition suivante et de l'inégalité $0 \leqslant V(x) \leqslant \psi(x)$.

Proposition 4. *On a* $\psi(x) = O(x)$.

Démonstration. On part de l'identité $\mathbb{1} \star \Lambda = \log$ et on somme sur tous les entiers $n \in [1, x]$. Cela donne :

$$A(x) = \sum_{d \leqslant x} \left[\frac{x}{d}\right] \Lambda(d) = \sum_{n \leqslant x} \log n = x \log x + O(x)$$

la dernière égalité résultant de la comparaison avec une intégrale. En appliquant cette estimation avec x et $\frac{x}{2}$, on obtient :

$$A(x) - 2A\left(\frac{x}{2}\right) = \sum_{d \leqslant x} \left(\left[\frac{x}{d}\right] - 2\left[\frac{x}{2d}\right]\right) \Lambda(d) = x \log 2 + O(x) = O(x).$$

Le facteur entre parenthèses est toujours positif et égal à 1 lorsque d est dans l'intervalle $]\frac{x}{2}, x]$. Ainsi $\psi(x) - \psi(\frac{x}{2}) = \sum_{\frac{x}{2} < d \leqslant x} \Lambda(d) = O(x)$. Appliquant ceci avec x, $\frac{x}{2}$, $\frac{x}{4}$, *etc*, sommant toutes ces égalités et utilisant finalement la convergence de la série géométrique des $\frac{1}{2^n}$, on obtient bien l'estimation de la proposition. $\square$

Fort du résultat précédent, on est en mesure d'appliquer directement l'estimation (3) qui combinée avec les inégalités (4) entraînent :

$$\frac{\psi(x)}{\log x} \leqslant \pi(x) = \frac{V(x)}{\log x} + O\left(\frac{x}{\log^2 x}\right) \leqslant \frac{\psi(x)}{\log x} + O\left(\frac{x}{\log^2 x}\right)$$

d'où on déduit :

$$\pi(x) = \frac{\psi(x)}{\log x} + O\left(\frac{x}{\log^2 x}\right). \tag{5}$$

On déduit de ceci et de la proposition 4 une première majoration encourageante, à savoir $\pi(x) = O(\frac{x}{\log x})$.

2 Théorème principal

2.1 Énoncé

Le théorème que nous allons énoncer est très pratique pour obtenir des estimations de la même veine que (1), mais il est en contrepartie très technique et demande d'introduire certaines notations préliminaires. Essentiellement, nous aurons besoin de fonctions μ tronquées, notées μ_T ($T \in \mathbb{N}^\star$) et définies comme suit :

$$\mu_T(n) = \begin{cases} \mu(n) & \text{si } n < T \\ -T \sum_{i<T} \frac{\mu(i)}{i} & \text{si } n = T \\ 0 & \text{si } n > T \end{cases}$$

La valeur attribuée à $\mu_T(T)$ permet d'avoir la relation[1] $\sum_n \frac{\mu_T(n)}{n} = 0$. On introduit également les fonctions $g_T = \mathbb{1} \star \mu_T$ et G_T définie par $G_T(x) = \sum_{n \leqslant x} g_T(n)$ pour tout réel $x \geqslant 1$. La définition du produit de convolution et la relation $1 \star \mu = e_1$ entraînent immédiatement que $g_T(n) = e_1(n)$ pour $n < T$, d'où il suit $G_T(n) = 1$ également pour $n < T$. Le théorème, à présent, s'énonce ainsi :

Théorème 5. *Soient $T \in \mathbb{N}^\star$ et $\alpha \in]0, 1]$ un nombre réel. On pose :*

$$\delta = 1 + \sum_{j \leqslant T} \frac{\log j}{j} \mu_T(j) \quad ; \quad C = 1 + \max_{T \leqslant y < \frac{T}{\alpha}} |G_T(y)| \quad \textit{et} \quad \delta' = 2\alpha + \frac{C}{T}.$$

On suppose $\delta' < 1$. Alors :

$$\limsup_{x \to \infty} \left| \frac{\pi(x) \log x}{x} - 1 \right| \leqslant \frac{|\delta| + \delta'}{1 - \delta'}.$$

Évidemment, ce théorème est intéressant car il est très facile de l'instancier pour des valeurs particulières de T et de α, obtenant ainsi sans effort de nombreux encadrements comme voulu. On remarque en particulier que le calcul des constantes C, δ et δ' ne requiert qu'un nombre fini d'opérations très faciles à automatiser,

[1] Comme on le verra par la suite, la relation analogue est aussi vérifiée pour la fonction μ (à la place de μ_T).

mais bien sûr de plus en plus longues à exécuter lorsque T grandit et α diminue. Voici par exemple les résultats obtenus pour certaines valeurs du couple (T, α) :

T	25	100	200	400	800
α	$0,005$	$0,010$	$0,010$	$0,010$	$0,002$
$\frac{\lvert\delta\rvert+\delta'}{1-\delta'}$	$0,263$	$0,127$	$0,087$	$0,063$	$0,033$

On constate qu'à partir de la troisième colonne (numérique) du tableau, les valeurs trouvées sont meilleures que celles de Tchebychev (voir formule (1)).

2.2 Preuve du théorème

Avant tout, on remarque qu'en vertu de (5), il est équivalent de montrer le théorème en remplaçant dans sa conclusion $\frac{\pi(x)\log x}{x}$ par $\frac{\psi(x)}{x}$; c'est ce que nous allons faire. À partir de maintenant, on fixe un couple (T, α) et on désigne par C, δ et δ' les constantes qui lui sont associées. La démonstration commence par l'identité $\Lambda \star g_T = \Lambda \star \mathbb{1} \star \mu_T = \log \star \mu_T$ qui fournit en sommant :

$$\sum_{n \leqslant x} \Lambda \star g_T(n) = \sum_{n \leqslant x} \log \star \mu_T(n). \tag{6}$$

La suite consiste à estimer séparément chacun des deux membres de l'égalité : précisément, nous allons montrer que le membre de gauche est proche de $\psi(x)$ alors que celui de droite est de l'ordre de x. La conclusion en découlera.

Avant de nous lancer plus avant dans les détails, remarquons que calculer la somme $\sum_{n \leqslant x} f \star g(n)$ revient à sommer les $f(i)g(j)$ sur tous les couples (i, j) au-dessous de l'hyperbole d'équation $ij = x$. Il sera bon de garder cette interprétation géométrique en tête pour mieux suivre les calculs à venir.

Le membre de droite En sommant d'abord sur les horizontales « $j =$ constante », la somme de droite s'écrit sous la forme :

$$\sum_{n \leqslant x} \log \star \mu_T(n) = \sum_{j \leqslant x} \mu_T(j) \sum_{i \leqslant \frac{x}{j}} \log i.$$

On utilise à nouveau une comparaison à une intégrale pour remplacer la somme sur i par l'estimé $\frac{x}{j} \log \frac{x}{j} - \frac{x}{j} + O(\log x)$. Réinjectant cette valeur dans l'expression, on obtient :

$$\sum_{n \leqslant x} \log \star \mu_T(n) = (x \log x - x) \sum_{j \leqslant x} \frac{\mu_T(j)}{j} - x \sum_{j \leqslant x} \frac{\log j}{j} \mu_T(j) + O(\log x) = x - \delta x + O(\log x)$$

car, si $x \geqslant T$, le premier terme de l'étape intermédiaire du calcul s'annule et le deuxième vaut $(\delta - 1)x$.

Le membre de gauche L'étude du membre de gauche, noté à présent G, est plus délicate. Nous avons à évaluer la somme des $\Lambda(i) g_T(j)$ pour (i, j) sous l'hyperbole d'équation $ij = x$. Sur la région « $j \leqslant T - 1$ », on a $g_T(j) = e_1(j)$, et donc la somme des $\Lambda(i) g_T(j)$ sur l'intersection des régions « $ij \leqslant x$ » et « $j \leqslant T - 1$ » se réduit à $\sum_{i \leqslant x} \Lambda(i) = \psi(x)$. Ainsi peut-on écrire $G = \psi(x) + G'$ où G' est la somme des $\Lambda(i) g_T(j)$ sur l'intersection des régions « $ij \leqslant x$ » et « $j \geqslant T$ » et donc s'exprime comme suit :

$$G' = \sum_{i \leqslant \frac{x}{T}} \Lambda(i) \sum_{T \leqslant j \leqslant \frac{x}{i}} g_T(j) = \sum_{i \leqslant \frac{x}{T}} \left(G_T\left(\frac{x}{i}\right) - G_T(T-1) \right) \Lambda(i) = \sum_{i \leqslant \frac{x}{T}} \left(G_T\left(\frac{x}{i}\right) - 1 \right) \Lambda(i).$$

Il nous faut maintenant majorer le terme $G_T(\frac{x}{i})$. Lorsque $\frac{x}{i} < \frac{T}{\alpha}$ ($i.e$ $i > \frac{x\alpha}{T}$), on peut utiliser la constante C du théorème. Dans les autres cas, on a besoin du lemme suivant :

Lemme 6. *La fonction G_T est majorée en valeur absolue par $2T$.*

Démonstration. Un calcul donne pour tout réel $y \geqslant 1$:

$$G_T(y) = \sum_{j \leqslant T} \left[\frac{y}{j}\right] \mu_T(j) = -\sum_{j < T} \left\{\frac{y}{j}\right\} \mu(j) + T \left\{\frac{y}{T}\right\} \sum_{j < T} \frac{\mu(j)}{j} \tag{7}$$

où $\{x\} = x - [x]$ est par définition la partie décimale de x. Le premier terme de la dernière expression est majoré en valeur absolue par $T-1$. Pour évaluer le second, on fait à nouveau une manipulation du même type en écrivant :

$$T\sum_{j<T}\frac{\mu(j)}{j} = \sum_{j<T}\left[\frac{T}{j}\right]\mu(j) + \sum_{j<T}\left\{\frac{T}{j}\right\}\mu(j) = 1-\mu(T) + \sum_{j<T}\left\{\frac{T}{j}\right\}\mu(j)$$

la deuxième égalité résultant de la définition de la fonction μ. Ainsi le dernier terme de (7) est majoré en valeur absolue par $T+1$ et le lemme en résulte. $\square$

On déduit des majorations précédentes que :

$$\begin{aligned} |G'| &\leqslant \psi\left(\frac{x}{T}\right) + \sum_{i\leqslant \frac{x\alpha}{T}}\left|G_T\left(\frac{x}{i}\right)\right|\Lambda(i) + \sum_{\frac{x\alpha}{T}<i\leqslant\frac{x}{T}}\left|G_T\left(\frac{x}{i}\right)\right|\Lambda(i) \\ &\leqslant C\cdot\psi\left(\frac{x}{T}\right) + 2T\cdot\psi\left(\frac{x\alpha}{T}\right). \end{aligned}$$

Conclusion Les estimations que l'on vient d'obtenir sur les membres de gauche et de droite de (6) impliquent :

$$|\psi(x) - x| \leqslant |\delta|x + C\cdot\psi\left(\frac{x}{T}\right) + 2T\cdot\psi\left(\frac{x\alpha}{T}\right) + O(\log x)$$

Posons $f(x) = |\frac{\psi(x)}{x} - 1|$. En utilisant $\psi(x) \leqslant xf(x) + x$, on peut transformer l'inégalité précédente pour ne faire intervenir que la fonction f. Cela donne :

$$f(x) \leqslant |\delta| + \frac{C}{T}\left(f\left(\frac{x}{T}\right)+1\right) + 2\alpha\left(f\left(\frac{x\alpha}{T}\right)+1\right) + O\left(\frac{\log x}{x}\right). \tag{8}$$

Par la proposition 4, la limite supérieure de f (quand x tend vers l'infini) est finie ; notons-la ℓ. Du fait que la limite supérieure d'une somme est inférieure ou égale à la somme des limites supérieures, la formule (8) entraîne :

$$\ell \leqslant |\delta| + \frac{C}{T}(\ell+1) + 2\alpha(\ell+1) = |\delta| + \delta' + \delta'\ell$$

d'où $\ell \leqslant \frac{|\delta|+\delta'}{1-\delta'}$: le théorème est démontré.

3 Limites du théorème

La question se pose naturellement de savoir s'il est possible de trouver des couples (T,α) pour lesquels la borne $\frac{|\delta|+\delta'}{1-\delta'}$ est arbitrairement petite. On se doute d'ores et déjà que, même si la réponse est positive, il risque d'être difficile de le prouver puisqu'elle impliquerait le théorème des nombres premiers, résultat connu comme non-élémentaire. Malgré tout, on a le résultat suivant :

Théorème 7. *On reprend les notations du théorème 5 et on admet le théorème des nombres premiers. Alors, à α fixé,*

$$\liminf_{T\to\infty}\frac{|\delta|+\delta'}{1-\delta'} = \frac{2\alpha}{1-2\alpha}.$$

En particulier, l'infimum des $\frac{|\delta|+\delta'}{1-\delta'}$ pris sur tous les couples (T,α) est nul.

L'énoncé précédent rend la situation assez cocasse. En effet, il n'implique certainement pas le théorème des nombres premiers puisque celui-ci apparaît dans les hypothèses. Toutefois, il nous dit que si l'on croit à ce théorème[2], de très bons couples (T,α) existent et par application du théorème 5 — démontré, lui — il devient possible, moyennant une certaine puissance de calcul, d'avoir des approximations aussi précises qu'on le désire du fameux théorème.

En réalité, ce scénario n'est pas si exceptionnel. Une autre illustration bien plus simple du phénomène est la suivante. Admettons que l'on dispose d'une suite (u_n) décroissante et positive, et que l'on conjecture qu'elle converge vers 0. Alors, si la conjecture est effectivement vraie, en calculant les termes de la suite suffisamment loin, on est *potentiellement* capable, pour tout $\varepsilon > 0$ fixé à l'avance, de montrer que $|u_n| \leqslant \varepsilon$ pour n suffisamment grand.

[2]Évidemment, on sait aujourd'hui qu'il est vrai, mais supposons un instant que ce ne soit pas le cas.

Concentrons-nous à présent sur la démonstration du théorème 7. De même que dans l'énoncé du théorème, on suppose que α est fixé. En outre, pour augmenter la clarté de ce qui suit, on note à partir de maintenant $\delta(T)$ (resp. $\delta'(T)$) à la place de δ (resp. δ'). Avant de nous lancer dans le vif du sujet, nous dégageons deux conséquences du théorème des nombres premiers qui vont être utiles.

3.1 Deux séries reliées à la fonction de Möbius

Le but de ce préliminaire est d'étudier les fonctions suivantes :

$$M(x) = \sum_{n \leqslant x} \mu(n) \quad \text{et} \quad m(x) = \sum_{n \leqslant x} \frac{\mu(n)}{n}.$$

On a, sans effort, $|M(x)| \leqslant x$ tandis que la relation :

$$x \sum_{j \leqslant x} \frac{\mu(j)}{j} = \sum_{j \leqslant x} \left[\frac{x}{j}\right] \mu(j) + \sum_{j \leqslant x} \left\{\frac{x}{j}\right\} \mu(j) = 1 + \sum_{j \leqslant x} \left\{\frac{x}{j}\right\} \mu(j) \tag{9}$$

implique $|m(x)| \leqslant 1 + \frac{1}{x}$. Nous allons montrer que le théorème des nombres premiers fournit des estimations bien meilleures de ces quantités.

Proposition 8. *On admet le théorème des nombres premiers. Alors $M(x) = o(x)$ et $m(x) = o(1)$.*

Démonstration. Le théorème des nombres premiers dit que $\psi(x) \sim x$, ou encore $\psi(x) - x = o(x)$. Si l'on définit $\lambda = 1 - \Lambda$, cette dernière relation s'écrit aussi sous la forme $\sum_{n \leqslant x} \lambda(n) = o(x)$. Posons pour la suite de la démonstration $\varphi(x) = \sum_{n \leqslant x} \lambda(n) = x - \psi(x)$.

La preuve de la première assertion suit alors d'assez près la stratégie donnée en 1. On part de la relation $\mu \star \Lambda = -\mu \log$ que l'on réécrit sous la forme $\mu \log = \mu \star \lambda - e_1$. Sommant cette égalité pour n variant entre 1 et x, on obtient

$$V(x) = \sum_{n \leqslant x} \mu(n) \log(n) = -1 + \sum_{ij \leqslant x} \mu(i) \lambda(j).$$

Ainsi $|V(x)| \leqslant 1 + \sum_{i \leqslant x} |\varphi(\frac{x}{i})|$. Soient $\varepsilon > 0$, et $j_0 \geqslant 1$ un réel tel que $|\varphi(j)| \leqslant \varepsilon j$ pour tout $j \geqslant j_0$. En découpant la somme $\sum_{i \leqslant x} |\varphi(\frac{x}{i})|$ en deux, selon que i soit inférieur ou non à $\frac{x}{j_0}$, on obtient $|V(x)| \leqslant \varepsilon x \log x + O(x)$. Comme ceci est vrai pour tout $\varepsilon > 0$, il vient $V(x) = o(x \log x)$. Il reste à se débarrasser du facteur $\log n$. La formule (2) s'instancie ici sous la forme :

$$M(x) = 1 + \frac{V(x)}{\log[x]} + \sum_{2 \leqslant n \leqslant x-1} V(n) \left(\frac{1}{\log n} - \frac{1}{\log(n+1)} \right).$$

Les deux premiers termes sont des $o(x)$. Le terme dans la somme est un $o(\frac{1}{\log n})$, donc un $o(1)$, ce qui assure que la somme elle-même est un $o(x)$. Il en résulte que $M(x)$ est aussi un $o(x)$ comme voulu.

Le second point est un peu plus délicat. Tout d'abord, on utilise la formule (9) qui ramène le problème à montrer que $M'(x) = \sum_{n \leqslant x} \left\{\frac{x}{n}\right\} \mu(n)$ est un $o(x)$. On a alors envie de relier $M(x)$ à $M'(x)$ en utilisant à nouveau une transformation d'Abel, mais cela demande un peu d'attention. Précisément, on commence par fixer un réel $\nu \in]0,1]$ et à couper la somme en deux :

$$M'(x) = \sum_{n \leqslant \nu x} \left\{\frac{x}{n}\right\} \mu(n) + \sum_{\nu x < n \leqslant x} \left\{\frac{x}{n}\right\} \mu(n).$$

Le premier terme est majoré en valeur absolue par νx. Il reste à estimer le second. On sépare à nouveau la somme selon la valeur de $[\frac{x}{n}]$. Sur un intervalle $]a,b]$ sur lequel $[\frac{x}{n}]$ est constant, on effectue une transformation d'Abel qui conduit, après quelques calculs, à la majoration :

$$\Big| \sum_{a < n \leqslant b} \left\{\frac{x}{n}\right\} \mu(n) \Big| \leqslant 2 \max_{a-1 < y \leqslant b} |M(y)| \leqslant 2 \max_{\nu x - 1 < y \leqslant x} |M(y)| = o(x)$$

la dernière égalité résultant de la première partie de la démonstration. Comme le nombre de valeurs que peut prendre $[\frac{x}{n}]$ lorsque n décrit l'intervalle $]\nu x, x]$ est majoré indépendamment de x (par $\frac{1}{\nu}$ en l'occurrence), on peut sommer les $o(x)$ pour obtenir au final $|M'(x)| \leqslant \nu x + o(x)$. Ceci étant vrai pour tout ν, il suit $M'(x) = o(x)$ comme annoncé. $\square$

3.2 Fin de la preuve

La démonstration du théorème 7 procède en deux étapes : on montre successivement que $\liminf |\delta(T)| = 0$, puis que $\lim \delta'(T) = 2\alpha$, formules à partir desquelles on conclut facilement.

Première étape : étude de δ(T)

Le but de cette partie est, comme nous venons de l'annoncer, de démontrer que $\liminf |\delta(T)| = 0$. Pour cela, commençons par remarquer que :

$$m_1(x) = \int_1^x m(u)\frac{du}{u} = \int_1^x \sum_{n \leqslant x} \frac{\mu(n)}{n} \mathbb{1}_{\{n \leqslant u\}} \frac{du}{u} = \sum_{n \leqslant x} \frac{\mu(n)}{n} \int_n^x \frac{du}{u} = \sum_{n \leqslant x} \frac{\mu(n)}{n} \log\left(\frac{x}{n}\right). \qquad (10)$$

Appliqué avec $x = T$, ceci fournit une relation entre $\delta(T)$ et $m_1(T)$ à savoir $\delta(T) = 1 - m_1(T)$. Il suffit donc pour conclure de montrer que la suite des $m_1(T)$ admet 1 pour valeur d'adhérence. Pour cela, on considère, pour un réel $s > 0$, l'intégrale $\int_1^\infty su^{-s-1}m_1(u)du$ qui calcule une certaine moyenne de m_1 puisque les coefficients de pondération vérifient $\int_1^\infty su^{-s-1}du = 1$. Une intégration par parties, suivie d'une transformation algébrique équivalente à celle pratiquée en (10) permet d'aboutir à :

$$\int_1^x su^{-s-1}m_1(u)du = -\frac{m_1(x)}{x^s} - \frac{m(x)}{sx^s} + \sum_{n \leqslant x} \frac{\mu(n)}{sn^{s+1}}.$$

Par les premières estimations (qui n'utilisent pas le théorème des nombres premiers) du paragraphe précédent, le second terme tend vers 0 quand x tend vers l'infini. C'est aussi le cas du premier grâce à la majoration :

$$|m_1(x)| = \Big| \sum_{n \leqslant x} \frac{\mu(n)}{n} \log\left(\frac{x}{n}\right) \Big| \leqslant \log(x) \sum_{n \leqslant x} \frac{1}{n} = O(\log^2 x).$$

Finalement, on vérifie aisément la relation $(\mu n^{-s-1}) \star (n^{-s-1}) = e_1$ et celle-ci entraîne que le troisième terme converge quand x tend vers $+\infty$ vers $\frac{1}{s\zeta(s+1)}$ où par définition $\zeta(s+1) = \sum_{n \geqslant 1} \frac{1}{n^{s+1}}$. Faisant tendre x vers l'infini, on obtient finalement :

$$\int_1^\infty su^{-s-1}m_1(u)du = \frac{1}{s\zeta(s+1)}.$$

Cette dernière quantité tend vers 1 lorsque s tend vers 0, comme on le voit en encadrant la somme des $\frac{1}{n^{s+1}}$ par des intégrales. Ceci implique que la fonction continue m_1 admet 1 pour valeur d'adhérence. (On peut le justifier en raisonnant par l'absurde et en se rappelant que $\int_1^\infty su^{-s-1}du = 1$ pour tout $s > 0$.)

Ce n'est pas fini, car il nous faut prouver que la suite des $m_1(T)$ admet 1 pour valeur d'adhérence, ce qui est plus fort que ce que l'on vient d'obtenir. Toutefois, on s'y ramène facilement après avoir noté que pour tout entier T et tout $a \in [0, 1]$:

$$|m_1(T+a) - m_1(T)| = \Big| m(T) \cdot \log\left(\frac{T+a}{T}\right) \Big| \leqslant \left(1 + \frac{1}{T}\right) \cdot \frac{a}{T} \leqslant \frac{T+1}{T^2}$$

qui tend vers 0 quand T tend vers l'infini, et ce indépendamment de a.

Remarque. Dans le raisonnement précédent, nous n'avons pas utilisé le théorème des nombres premiers. L'estimation de $m(x)$ qui nous a servi est la première que nous avons donnée en 3.1 et elle découle essentiellement de l'égalité (9).

Seconde étape : étude de δ'(T)

Pour conclure, on rappelle qu'il suffit de montrer que $\lim \delta'(T) = 2\alpha$. D'après la définition de δ' (voir théorème 5), il suffit pour cela d'établir que $C(T) = o(T)$ où par définition $C(T) = 1 + \max_{T \leqslant y \leqslant \frac{T}{\alpha}} |G_T(y)|$. On montre plus précisément que :

$$C'(T) = \max_{y \in \mathbb{R}} |G_T(y)| = o(T)$$

c'est-à-dire que $|G_T(y)|$ est un $o(T)$ indépendamment de y. Pour cela, on part de la formule (7) :

$$G_T(y) = -\sum_{j<T} \left\{\frac{y}{j}\right\} \mu(j) + T\left\{\frac{y}{T}\right\} \sum_{j<T} \frac{\mu(j)}{j}.$$

D'après la proposition 8, le dernier terme est un $o(T)$, et ce indépendamment de y puisque le coefficient $\left\{\frac{y}{T}\right\}$ est majoré par 1. Le premier terme, quant à lui, se traite par une méthode analogue à celle utilisée pour estimer $M'(x)$ dans la démonstration de la proposition 8 : on introduit un paramètre ν, on découpe la somme en deux, puis encore la deuxième somme suivant les valeurs de $[\frac{y}{T}]$ avant d'effectuer une transformation d'Abel sur chacune de ces sommes partielles puis de conclure en regroupant toutes les estimations obtenues. On laisse au lecteur le soin d'adapter ce travail un peu technique à la situation présente, puis de terminer la preuve en cours.

Références

[1] H.G. Diamond and P. Erdös, *On sharp elementary prime number estimates*, Enseign. Math., **26**, 1980, pp 313–321