Math. Proc. Cambridge Phil. Soc.
113 (2002), 133-191.

Sur la loi de répartition du k-ième facteur premier d'un entier

Jean-Marie De Koninck et Gérald Tenenbaum

Résumé. Soit $\{p_k(n)\}_{k=1}^{\omega(n)}$ la suite croissante des facteurs premiers distincts d'un entier n. Nous donnons, lorsque $k \to \infty$, une approximation uniforme de la loi de répartition limite de la fonction arithmétique $n \mapsto p_k(n)$, précisant ainsi un résultat classique d'Erdős. Deux applications en sont déduites, relatives à la médiane de cette loi et à celle de la fonction « nombre de facteurs premiers ».

Classification AMS : principal 11N25, secondaires 11K65, 11N37.

1. Introduction

Soit $P^-(n)$ (resp. $P^+(n)$) le plus petit (resp. le plus grand) facteur premier d'un entier $n \geqslant 2$, avec la convention, $P^-(1) = \infty$, $P^+(1) = 1$. Nous désignons encore par $\{p_k(n)\}_{k=1}^{\omega(n)}$ la suite croissante des facteurs premiers distincts de n, où $\omega(n)$ désigne donc le nombre de facteurs premiers de n comptés sans multiplicité. Dans tout l'article, la lettre p, avec ou sans indice, désigne un nombre premier.

Erdős a démontré en 1946 [4] que, pour presque tout entier n, la quantité $\log_2 p_k(n)$[1] est uniformément voisine de k dès que $k \to \infty$ avec n. Plus précisément, il établit que, si l'on note

$$\log_2 p_k(n) = k + \vartheta_k(n)\sqrt{2k \log_2 k} \qquad (3 \leqslant k \leqslant \omega(n)), \tag{1·1}$$

et si $\xi(n)$ est une fonction de n tendant vers l'infini assez lentement, alors

$$\sup_{\xi(n) \leqslant k \leqslant \omega(n)} \vartheta_k(n) = 1 + o(1), \qquad \inf_{\xi(n) \leqslant k \leqslant \omega(n)} \vartheta_k(n) = -1 + o(1)$$

lorsque n tend vers l'infini en restant dans une suite convenable de densité unité. Les détails d'une démonstration purement arithmétique de ce résultat

1. Ici et dans la suite nous désignons par $\log_j$ la j-ième itérée de la fonction logarithme.

sont exposés au chapitre 1 du livre de Hall & Tenenbaum [9]. Une version plus faible découle simplement de l'inégalité de Turán–Kubilius, cf. [6].

La relation (1·1) est une loi du logarithme itéré traduisant l'indépendance asymptotique, pour p assez petit, des variables aléatoires de Bernoulli ζ_p définies sur l'ensemble des n premiers entiers par

$$\zeta_p(m) = \begin{cases} 1 & \text{si } p \mid m, \\ 0 & \text{si } p \nmid m. \end{cases}$$

Cette indépendance asymptotique, dont une des conséquence majeures est le lemme fondamental du modèle de Kubilius (voir [18] pour une étude quantitative récente de l'approximation sous-jacente), fournit également, ainsi que l'a noté Erdős, une convergence vers la loi de Gauss. Désignons par $\mathbf{d}\mathcal{A}$ la densité naturelle d'une suite d'entiers générique $\mathcal{A}$ et posons

$$\lambda_k^*(y) := \mathbf{d}\{n : p_k(n) > y\}.$$

Erdős a annoncé en 1969 [5] que l'on a

$$\lambda_k^*\big(\exp\exp\big\{k + z\sqrt{k}\big\}\big) = \Phi^*(z) + o(1) \qquad (k \to \infty), \tag{1·2}$$

où $\Phi^*(z) := \int_z^\infty \mathrm{e}^{-u^2/2}\,\mathrm{d}u/\sqrt{2\pi}$ est la fonction de répartition décroissante de la loi de Gauss. Nous nous intéressons ici au terme d'erreur de cette formule asymptotique.

Pour chaque entier positif k et chaque nombre premier p, désignons par $\lambda_k(p)$ la densité de l'ensemble des entiers positifs n tels que $p_k(n) = p$, de sorte que

$$\lambda_k^*(y) = \sum_{p>y} \lambda_k(p).$$

Il découle du crible d'Ératosthène que

$$\lambda_k(p) = \frac{1}{p} \prod_{q<p} \Big(1 - \frac{1}{q}\Big) s_{k-1}(p) \qquad (k \geqslant 2), \tag{1·3}$$

où

$$s_j(p) := \sum_{\substack{P^+(m)<p \\ \omega(m)=j}} \frac{1}{m} \qquad (j \geqslant 1).^{(2)}$$

On convient, selon l'usage courant, qu'une somme vide est nulle, et qu'un produit vide vaut 1.

2. Nous étendons de manière évidente, dans la suite, la définition de $s_j(y)$ au cas où y n'est pas un nombre premier.

Il est immédiat que les $s_j(p)$ sont donnés par la formule

$$\sum_{j=0}^{\infty} s_j(p)z^j = \prod_{q<p}\Big(1+\frac{z}{q-1}\Big) \qquad (z\in\mathbb{C}). \tag{1·4}$$

Les expressions classiques des fonctions symétriques élémentaires en fonction des sommes des puissances pour les éléments d'une suite finie (cf., par exemple, [2], ex. 9, p. 167) fournissent, pour chaque p,

$$s_j = (-1)^j \sum_{\substack{m_1+2m_2+\cdots+jm_j=j\\ m_r\geqslant 0\ (1\leqslant r\leqslant j)}} \prod_{r=1}^{j} \frac{1}{m_r!}\Big(\frac{-\sigma_r}{r}\Big)^{m_r}, \tag{1·5}$$

où $s_j := s_j(p)$ et

$$\sigma_r = \sigma_r(p) := \sum_{q<p} 1/(q-1)^r \qquad (r\geqslant 1).$$

Cette expression est utile pour le calcul numérique des s_j, dans la mesure où elle comporte souvent moins de termes que la formule

$$s_j(p) = \sum_{\substack{P^+(m)<p\\ \omega(m)=j}} \frac{\mu(m)^2}{\varphi(m)}, \tag{1·6}$$

où φ désigne la fonction indicatrice d'Euler.[3] Ainsi

$$\begin{aligned}
s_1 &= \sigma_1,\\
s_2 &= \tfrac12\sigma_1^2 - \tfrac12\sigma_2,\\
s_3 &= \tfrac16\sigma_1^3 - \tfrac12\sigma_1\sigma_2 + \tfrac13\sigma_3,\\
s_4 &= \tfrac1{24}\sigma_1^4 - \tfrac14\sigma_1^2\sigma_2 + \tfrac13\sigma_1\sigma_3 + \tfrac18\sigma_2^2 - \tfrac14\sigma_4,\\
s_5 &= \tfrac1{120}\sigma_1^5 - \tfrac1{12}\sigma_1^3\sigma_2 + \tfrac16\sigma_1^2\sigma_3 - \tfrac14\sigma_1\sigma_4 + \tfrac18\sigma_1\sigma_2^2 - \tfrac16\sigma_2\sigma_3 + \tfrac15\sigma_5.
\end{aligned}$$

En 1989, Erdős & Tenenbaum [7] ont étudié les variations en p et en k de $\lambda_k(p)$. Ils ont notamment établi que, notant $\ell_1 := \log k$, $\ell_2 := \log_2 k$, on a

$$\max_p \lambda_k(p) = \exp\Big\{-k\Big(\ell_1-\ell_2-1+\frac{2\ell_2+1}{\ell_1}+\frac{2\ell_2^2-\ell_2+O(1)}{\ell_1^2}\Big)\Big\} \tag{1·7}$$

3. Si p est le n-ième nombre premier, le membre de droite de (1·6) comporte $\binom{n-1}{j}$ termes, alors que celui de (1·5) en comporte $p(j)$, où p est la fonction de partitions. On a $p(j)\leqslant\binom{n-1}{j}$ dès que $n\geqslant j+K\sqrt{j}/\log j$ avec une constante convenable K.

et que toute valeur de p réalisant ce maximum satisfait à

$$\log p = \frac{k}{\ell_1}\Big\{1 + \frac{2\ell_2}{\ell_1} + \frac{2\ell_2^2 - 3\ell_2 + O(1)}{\ell_1^2}\Big\}. \tag{1·8}$$

Nous établissons ici le résultat suivant qui constitue une version effective du résultat d'Erdős (1·2).

Théorème 1.1. *Soit*

$$A := \gamma - \sum_p \Big\{ \log\Big(\frac{1}{1-1/p}\Big) - \frac{1}{p}\Big\} \approx 0,26150. \tag{1·9}$$

On a uniformément pour $k \geqslant 1$, $z \in \mathbb{R}$,

$$\lambda_k^*\big(\exp\exp\big\{k + z\sqrt{k}\big\}\big) = \Phi^*(z) - \frac{\Phi_0(z)}{\sqrt{2\pi k}} + O\Big(\frac{1}{k}\Big), \tag{1·10}$$

avec $\Phi_0(z) = \mathrm{e}^{-z^2/2}\big\{\frac{1}{3} + A - \frac{1}{3}z^2\big\}$.

La formule (1·10) atteste des deux aspects fondamentaux de la loi de répartition de $p_k(n)$ en mettant en évidence, à côté d'un terme principal probabiliste, un terme correctif où intervient une constante arithmétique.

Nous pouvons également déduire de ces résultats d'autres estimations fines concernant la loi de répartition du k-ième facteur premier d'un entier. Un exemple intéressant est celui de la valeur médiane de cette loi, que nous définissons comme le plus petit nombre premier $p^* = p_k^*$ tel que

$$\lambda_k^*(p^*) \leqslant \tfrac{1}{2}.$$

Il est clair que $p_1^* = 2$. On verra au § 5 que $p_2^* = 37$, $p_3^* = 42719$. La notion de valeur médiane constitue un modèle heuristique quantitatif pour la taille usuelle du k-ième facteur premier d'un entier.

Nous tirons du Théorème 1.1 une formule approchée pour p_k^*.

Corollaire 1.2. *La valeur médiane p_k^* satisfait à*

$$\log_2 p_k^* = k - b + O\big(1/\sqrt{k}\,\big) \qquad (k \to \infty), \tag{1·11}$$

avec

$$b := \tfrac{1}{3} + A \approx 0,59483. \tag{1·12}$$

Comme on le verra au § 4, une étape cruciale de la démonstration du Corollaire 1.2 repose sur l'étude des sommes partielles de la série exponentielle. Ainsi le Lemme 2.1 du § 2 est-il non seulement essentiel pour l'estimation (1·11), mais il possède également d'autres applications : nous en développons une au § 3.

2. Sommes partielles de la série exponentielle

Nous commençons par évaluer asymptotiquement

$$S_n(x) := \sum_{0\leqslant j\leqslant n} \frac{\mathrm{e}^{-x}x^j}{j!},$$

uniformément pour $x > 0$ lorsque le paramètre entier n tend vers l'infini. L'étude se ramène facilement au cas $x = n$. La question correspondante a été posée et partiellement résolue par Ramanujan ([14], pp. 323-324), et Norton [13] donne un compte rendu détaillé de l'historique du problème. Cependant, comme les résultats généraux disponibles dans la littérature (voir, en particulier, Esseen [8], Delange [3], Kalinin [11]) sont d'énoncés passablement techniques et comme la formule dont nous avons besoin ne peut en être déduite sans calculs supplémentaires assez compliqués, nous présentons, pour la commodité du lecteur, une preuve autonome et relativement courte de l'évaluation nécessaire. Nous donnons ici un développement limité à l'ordre trois,(4) mais la méthode s'adapte sans difficulté à l'obtention d'un développement asymptotique.

Lemme 2.1. *On a, pour n infini et uniformément pour $x > 0$,*

$$S_n(x) = \Phi^*(z) + \frac{\Phi_1(z)}{\sqrt{n}} + \frac{\Phi_2(z)}{n} + O\Big(\frac{1}{n^{3/2}}\Big), \tag{2·1}$$

où z est défini par la relation $x = n + z\sqrt{n}$ et où l'on a posé

$$\Phi_1(z) := \frac{z^2+2}{3\sqrt{2\pi}}\mathrm{e}^{-z^2/2}, \qquad \Phi_2(z) := \int_0^z \mathrm{e}^{-u^2/2}\big\{\tfrac{1}{6} + \tfrac{1}{2}u^4 - \tfrac{1}{9}u^6\big\}\frac{\mathrm{d}u}{2\sqrt{2\pi}}.$$

Démonstration. Montrons d'abord que l'on a

$$S_n = \frac{1}{2} + \frac{2}{3\sqrt{2\pi n}} + O\Big(\frac{1}{n^{3/2}}\Big), \tag{2·2}$$

avec $S_n := S_n(n)$. Cette estimation pourrait être déduite des calculs de Ramanujan dans [14] mentionnés plus haut. Notre approche est sensiblement différente.

---

4. Un développement à un ordre plus élevé serait inutile pour notre application, compte tenu de la précision des évaluations disponibles pour les quantités arithmétiques $s_j(p)$ — voir en particulier (4·4) infra.

Posons $Q(u)=u\log u-u+1$. La formule de Stirling fournit, grâce à un calcul standard dont nous omettons les détails,

$$\frac{\mathrm{e}^{-n}n^j}{j!}=\frac{\mathrm{e}^{-nQ(j/n)}}{\sqrt{2\pi j}}\Big\{1-\frac{1}{12j}+O\Big(\frac{1}{j^2}\Big)\Big\}\qquad(1\leqslant j\leqslant n),\tag{2·3}$$

de sorte que l'on peut écrire

$$S_n=\Big\{1+O\Big(\frac{1}{n^2}\Big)\Big\}T_n+O\big(\mathrm{e}^{-cn}\big)=T_n+O\Big(\frac{1}{n^2}\Big)$$

avec

$$T_n:=\sum_{0\leqslant j\leqslant m}F_n(j),\quad m:=\big[\tfrac12 n\big],\quad c=Q(\tfrac12)>0,$$

et

$$F_n(u):=\frac{\mathrm{e}^{-nQ(1-u/n)}}{\sqrt{2\pi(n-u)}}\Big\{1-\frac{1}{12(n-u)}\Big\}.$$

Nous évaluons T_n par la formule sommatoire d'Euler-Maclaurin, soit

$$T_n=\int_0^m F_n(u)\,\mathrm{d}u+\tfrac12F_n(0)-\tfrac1{12}F_n'(0)+R_n$$

avec

$$R_n=\tfrac16\int_0^m B_3(u)F_n'''(u)\,\mathrm{d}u+O\big(\mathrm{e}^{-cn}\big),$$

où $B_3(u):=\{u\}^3-\frac32\{u\}^2+\frac12\{u\}$ est la troisième fonction de Bernoulli.

Or, on a $F_n'(u)=F_n(u)G_n(u)$ avec

$$G_n(u):=\log\Big(1-\frac{u}{n}\Big)+\frac{1}{2n-2u}-\frac{1}{(n-u)(12n-12u-1)},$$

et il est facile de vérifier que, pour $0\leqslant u\leqslant\frac12n$,

$$G_n(u)\ll\frac{u+1}{n},\qquad G_n^{(k)}(u)\ll_k\frac{1}{n^k}\qquad(k\geqslant1).$$

On en déduit que

$$F_n'(u)\ll\frac{(u+1)\mathrm{e}^{-u^2/2n}}{n^{3/2}}\ll\frac1n\qquad(0\leqslant u\leqslant m),$$

et semblablement, dans le même domaine,

$$F_n^{(k)}(u) \ll_k \frac{1}{n^{(k+1)/2}}.$$

Il suit

$$T_n = \int_0^m F_n(u)\,\mathrm{d}u + \frac{1}{2\sqrt{2\pi n}} + O\Big(\frac{1}{n^{3/2}}\Big). \tag{2·4}$$

Effectuons dans la dernière intégrale le changement de variable $u = v\sqrt{n}$. En utilisant la relation

$$Q(1-z) = \tfrac12 z^2 + \tfrac16 z^3 + \tfrac1{12} z^4 + O(z^5) \qquad (0 \leqslant z \leqslant \tfrac12),$$

nous obtenons, pour $0 \leqslant v \leqslant \frac12 n^{1/6}$,

$$\begin{aligned} F_n\big(v\sqrt{n}\big) &= \frac{\mathrm{e}^{-\frac12 v^2 - \frac16 v^3/\sqrt{n} - \frac1{12} v^4/n + O(v^5/n^{3/2})}}{\sqrt{2\pi n(1 - v/\sqrt{n})}} \Big\{1 - \frac{1}{12n} + O\Big(\frac{v}{n^{3/2}}\Big)\Big\} \\ &= \frac{\mathrm{e}^{-v^2/2}}{\sqrt{2\pi n}} \Big\{1 + \frac{A(v)}{\sqrt{n}} + \frac{B(v)}{n} + O\Big(\frac{v(1+v^6)}{n^{3/2}}\Big)\Big\} \end{aligned}$$

avec $A(v) := \frac12 v(1 - \frac13 v^2)$, $B(v) := -\frac1{12} + \frac38 v^2 - \frac16 v^4 + \frac1{72} v^6$. On vérifie sans peine que

$$\int_0^\infty A(v)\mathrm{e}^{-v^2/2}\,\mathrm{d}v = \tfrac16, \qquad \int_0^\infty B(v)\mathrm{e}^{-v^2/2}\,\mathrm{d}v = 0.$$

En utilisant l'estimation

$$F_n\big(v\sqrt{n}\big) \ll \mathrm{e}^{-v^2/2} \qquad \big(0 \leqslant v \leqslant \tfrac12\sqrt{n}\big),$$

pour évaluer la contribution du domaine $\frac12 n^{1/6} < v \leqslant \frac12\sqrt{n}$, il s'ensuit que

$$\int_0^m F_n(u)\,\mathrm{d}u = \tfrac12 + \frac{1}{6\sqrt{2\pi n}} + O\Big(\frac{1}{n^{3/2}}\Big).$$

En reportant dans (2·4), on obtient bien (2·2).

Il est maintenant facile de montrer (2·1). Nous observons d'emblée que nous pouvons, sans restreindre la généralité, supposer que $|z| \leqslant n^{1/10}$: dans le cas contraire, le résultat découle de la décroissance en x de $S_n(x)$.

Posons

$$\varphi_x(j) := \mathrm{e}^{-x} x^j / j!, \tag{2·5}$$

de sorte que $S_n(x) := \sum_{0 \leqslant j \leqslant n} \varphi_x(j)$. Comme $S_n'(x) = -\varphi_x(n)$, on peut écrire

$$S_n(x) - S_n(n) = -\int_n^x \varphi_y(n)\,\mathrm{d}y = -\int_0^z \frac{\mathrm{e}^{-n-u\sqrt{n}}(n+u\sqrt{n})^n}{n!}\sqrt{n}\,\mathrm{d}u.$$

En appliquant la formule de Stirling et un développement limité standard, il vient

$$\begin{aligned} S_n(x) - S_n(n) &= -\Big(1 - \frac{1}{12n}\Big)\int_0^z \mathrm{e}^{-\frac12 u^2 + \frac13 u^3/\sqrt{n} - \frac14 u^4/n} \frac{\mathrm{d}u}{\sqrt{2\pi}} + O\Big(\frac{1}{n^{3/2}}\Big) \\ &= -\int_0^z \Big\{1 + \frac{u^3}{3\sqrt{n}} - \frac{3 + 9u^4 - 2u^6}{36n}\Big\} \frac{\mathrm{e}^{-\frac12 u^2}\,\mathrm{d}u}{\sqrt{2\pi}} + O\Big(\frac{1}{n^{3/2}}\Big) \\ &= -\tfrac12 + \Phi^*(z) - \frac{2}{3\sqrt{2\pi n}} + \frac{\Phi_1(z)}{\sqrt{n}} + \frac{\Phi_2(z)}{n} + O\Big(\frac{1}{n^{3/2}}\Big). \end{aligned}$$

On en déduit le résultat annoncé en utilisant (2·2). □

Nous mentionnons maintenant un corollaire du Lemme 2.1 qui possède un intérêt intrinsèque, mais dont nous n'aurons pas d'usage direct dans ce travail.

Corollaire 2.2. *Pour $n \in \mathbb{N}^*$, soit x_n l'unique solution réelle positive de l'équation*

$$\sum_{0 \leqslant j \leqslant n} \frac{\mathrm{e}^{-x} x^j}{j!} = \tfrac12. \tag{2·6}$$

Alors on a

$$x_n = n + \tfrac23 + O\Big(\frac{1}{n}\Big). \tag{2·7}$$

Remarque. La méthode employée ici permet également de donner un développement asymptotique de x_n selon les puissances de $1/n$. Cela nécessite un développement de S_n selon les puissances impaires de $1/\sqrt{n}$, issu d'une application de la formule d'Euler-Maclaurin à un ordre arbitraire dans la démonstration du Lemme 2.1.

Démonstration. Posons $x_n = n + z_n\sqrt{n}$. Compte tenu de la décroissance de $x \mapsto S_n(x)$, la formule (2·1) implique $z_n = o(1)$ lorsque $n \to \infty$.[5] Comme

$$\Phi^{*\prime}(0) = -\frac{1}{\sqrt{2\pi}}, \quad \Phi^{*\prime\prime}(0) = 0, \quad \Phi_1(0) = \frac{2}{3\sqrt{2\pi}}, \quad \Phi_1'(0) = 0, \tag{2·8}$$

un développement limité permet d'en déduire

$$S_n(x_n) = \tfrac{1}{2} = \Phi^*(0) - \frac{z_n}{\sqrt{2\pi}} + O(z_n^3) + \frac{2}{3\sqrt{2\pi n}} + O\Big(\frac{z_n^2}{\sqrt{n}}\Big) + O\Big(\frac{1}{n^{3/2}}\Big),$$

d'où

$$z_n\Big\{\frac{1}{\sqrt{2\pi}} + O\Big(z_n^2 + \frac{|z_n|}{\sqrt{n}}\Big)\Big\} = \frac{2}{3\sqrt{2\pi n}} + O\Big(\frac{1}{n^{3/2}}\Big).$$

Cela implique $z_n \ll 1/\sqrt{n}$ et fournit (2·7), en reportant dans la même relation. □

3. Une application à la répartition du nombre de facteurs premiers

Les résultats du paragraphe 2 possèdent des applications évidentes à tous les problèmes impliquant une loi de Poisson, dont les exemples abondent en théorie analytique des nombres. Nous nous intéressons ici au nombre des entiers $n \leqslant x$ tels que $\Omega(n) \leqslant \log_2 x$ — ce qui constitue en fait un abord légèrement biaisé du problème de la valeur médiane de la fonction $\Omega(n)$.

Théorème 3.1. *On a*

$$\sum_{\substack{n\leqslant x\\ \Omega(n)\leqslant \log_2 x}} 1 = \tfrac{1}{2}x - x\frac{C + \{\log_2 x\}}{\sqrt{2\pi \log_2 x}} + O\Big(\frac{x}{\log_2 x}\Big), \tag{3·1}$$

avec

$$C := \gamma - \tfrac{2}{3} + \sum_p \Big\{\frac{1}{p-1} - \log\Big(\frac{1}{1-1/p}\Big)\Big\} \approx 0,36798.$$

5. Il suffit en fait d'utiliser ici la formule classique

$$\sum_{j\leqslant x+z\sqrt{x}} \mathrm{e}^{-x}x^j/j! = 1 - \Phi^*(z) + O\left(1/\sqrt{x}\right) \qquad (x \geqslant 1),$$

dont on peut donner une preuve rapide en notant que la transformée de Fourier du membre de gauche, considéré comme une fonction de z, vaut

$$\exp\Big\{x\Big(\mathrm{e}^{i\vartheta/\sqrt{x}} - 1\Big) - i\vartheta\sqrt{x}\Big\} = \begin{cases} \mathrm{e}^{-\vartheta^2/2}\Big\{1 + O\left(\dfrac{\vartheta^3}{\sqrt{x}}\right)\Big\} & (|\vartheta| \leqslant x^{1/6}),\\ O\left(\mathrm{e}^{-\vartheta^2/3}\right) & (x^{1/6} < |\vartheta| \leqslant \sqrt{x}). \end{cases}$$

et en appliquant l'inégalité de Berry–Esseen.

Démonstration. On remarque d'abord que la formule

$$\sum_{0\leqslant j\leqslant \xi} \mathrm{e}^{-\xi}\frac{\xi^j}{j!} = \frac{1}{2} + \frac{(2/3)-\{\xi\}}{\sqrt{2\pi\xi}} + O\Big(\frac{1}{\xi}\Big) \tag{3·2}$$

découle immédiatement de (2·1) et des relations (2·8) puisque la somme porte en fait sur $j \leqslant n := [\xi] = \xi - \{\xi\}$.

Soit $\pi_k(x)$ le nombre des entiers $n \leqslant x$ tels que $\Omega(n) = k$. Pour $k \leqslant \frac{3}{2}\xi$, $\xi := \log_2 x$, on a (cf., par exemple, [17], théorème II.6.4)

$$\pi_k(x) = x\mathrm{e}^{-\xi}\frac{\xi^{k-1}}{(k-1)!}\Big\{\lambda\Big(\frac{k-1}{\xi}\Big) + O\Big(\frac{1}{\xi}\Big)\Big\},$$

avec

$$\lambda(z) := \frac{1}{\Gamma(z+1)}\prod_p\Big(1+\frac{z}{p-1}\Big)\Big(1-\frac{1}{p}\Big)^{-z} \qquad (|z|<2).$$

Le membre de gauche de (3·1) vaut donc $xV_1 + xV_2 + O(x/\xi)$, avec

$$V_1 := \sum_{1\leqslant k\leqslant\xi}\mathrm{e}^{-\xi}\frac{\xi^{k-1}}{(k-1)!}, \qquad V_2 := \sum_{1\leqslant k\leqslant\xi}\mathrm{e}^{-\xi}\frac{\xi^{k-1}}{(k-1)!}\Big\{\lambda\Big(\frac{k-1}{\xi}\Big)-1\Big\}.$$

Par (3·2), on a, avec $K := [\xi]$

$$V_1 = \sum_{0\leqslant j\leqslant\xi}\mathrm{e}^{-\xi}\frac{\xi^j}{j!} - \mathrm{e}^{-\xi}\frac{\xi^K}{K!} = \frac{1}{2} - \frac{(1/3)+\{\xi\}}{\sqrt{2\pi\xi}} + O\Big(\frac{1}{\xi}\Big).$$

Ensuite, en employant l'approximation $\lambda(r) = 1+(r-1)\lambda'(1)+O((r-1)^2)$, on peut écrire

$$\begin{aligned} V_2 &= \lambda'(1)\Big\{\sum_{j\leqslant\xi-2}\mathrm{e}^{-\xi}\frac{\xi^j}{j!} - \sum_{j\leqslant\xi-1}\mathrm{e}^{-\xi}\frac{\xi^j}{j!}\Big\} + O\Big(\sum_{j\geqslant 0}\mathrm{e}^{-\xi}\frac{\xi^j}{j!}\Big(\frac{j}{\xi}-1\Big)^2\Big) \\ &= -\frac{\lambda'(1)}{\sqrt{2\pi\xi}} + O\Big(\frac{1}{\xi}\Big). \end{aligned}$$

On a

$$\begin{aligned} \lambda'(1) &= -\frac{\Gamma'(2)}{\Gamma(2)} + \sum_p\Big\{\frac{1}{p-1} - \log\Big(\frac{1}{1-1/p}\Big)\Big\} \\ &= \gamma - 1 + \sum_p\Big\{\frac{1}{p-1} - \log\Big(\frac{1}{1-1/p}\Big)\Big\} \approx 0,03465. \end{aligned}$$

Cela implique le résultat annoncé avec $C := \frac{1}{3} + \lambda'(1)$. □

4. Preuve du Théorème 1.1 et du Corollaire 1.2

Le point essentiel est l'estimation suivante de $s_j(y)$, déduite des résultats de [7] obtenus par la méthode du col. Nous posons

$$g(z) = Az - \sum_p \Big\{\frac{z}{p} - \log\Big(1 + \frac{z}{p-1}\Big)\Big\} \quad (z \in \mathbb{C} \smallsetminus [-1, -\infty[) \tag{4·1}$$

et, dans le même domaine,

$$h(z) := \mathrm{e}^{g(z)}.\text{[6]} \tag{4·2}$$

Lemme 4.1. *Soient $k \geqslant 1$ et $y > 3$ tel que $\log_2 y \asymp k$. Alors*

$$s_j(y) = h\Big(\frac{j}{\log_2 y}\Big)\frac{(\log_2 y)^j}{j!}\Big\{1 + O\Big(\frac{1}{k}\Big)\Big\} \qquad (0 \leqslant j \leqslant k). \tag{4·3}$$

Démonstration. Le corollaire 1 de [7] fournit

$$s_j(y) = F_y\Big(\frac{j}{M_j}\Big)\frac{(M_j/\mathrm{e})^j}{j!}\Big\{1 + O\Big(\frac{1}{k}\Big)\Big\} \tag{4·4}$$

uniformément pour $0 \leqslant j \leqslant k$, avec

$$F_y(z) := \prod_{p \leqslant y}\Big(1 + \frac{z}{p-1}\Big)$$

et

$$M_j := \log_2 y - \log\big(1 + \log^+(j/L_j)\big), \qquad L_j := \log_2 y - \log_2(j+1). \tag{4·5}$$

Pour $j \leqslant k$, on a clairement $j/L_j \ll 1$, d'où

$$M_j = \log_2 y + O(1) \quad (0 \leqslant j \leqslant k). \tag{4·6}$$

Maintenant, pour z réel positif borné, on peut écrire

$$\begin{aligned} F_y(z) &= \exp\Big\{z\sum_{p \leqslant y}\frac{1}{p} + \sum_p \Big\{\log\Big(1 + \frac{z}{p-1}\Big) - \frac{z}{p}\Big\} + O\Big(\frac{1}{y}\Big)\Big\} \\ &= \exp\Big\{z \log_2 y + g(z) + O(1/\log y)\Big\}, \end{aligned}$$

6. Bien entendu, la fonction $h(z)$ peut être prolongée holomorphiquement sur $\mathbb{C} \smallsetminus \{1 - p : p \text{ premier}\}$, mais nous n'utiliserons pas cette précision. En fait, nous n'emploierons que des arguments réels positifs.

où $g(z)$ est définie par (4·1). Il suit

$$s_j(y) = \Big\{1 + O\Big(\frac{1}{k}\Big)\Big\} \frac{\mathrm{e}^{-j+g(j/M_j)+j\log M_j+j(\log_2 y)/M_j}}{j!}.$$

On en déduit (4·3) par un calcul d'accroissements finis tenant compte de (4·6). □

Preuve du Théorème 1.1.

Nous allons d'abord établir que l'on a uniformément pour $k \geqslant 1$, $y > 3$,

$$\lambda_k^*(y) = \sum_{j<k} \frac{(\log_2 y)^j}{j!(\log y)} - A\frac{(\log_2 y)^{k-1}}{(k-1)!(\log y)} + O\Big(\frac{1}{k}\Big). \tag{4·7}$$

Nous verrons ensuite que cette formule est équivalente à (1·10).

Nous avons défini $\lambda_k^*(y)$ comme la densité de l'ensemble des entiers n tels que $p_k(n) > y$, c'est-à-dire ayant moins de k facteurs premiers distincts dans $[2, y]$. Il suit

$$\lambda_k^*(y) = \prod_{p\leqslant y}\Big(1-\frac{1}{p}\Big) \sum_{\substack{P^+(m)\leqslant y\\ \omega(m)<k}} \frac{1}{m} = \prod_{p\leqslant y}\Big(1-\frac{1}{p}\Big) \sum_{0\leqslant j<k} s_j(y), \tag{4·8}$$

si y n'est pas un nombre premier, ce que nous pouvons supposer sans perte de généralité.

On a $g(1) = \gamma$, $g'(1) = A$. Pour z réel positif borné, il suit donc

$$h(z) = \mathrm{e}^{\gamma}\big\{1 + A(z-1) + O\big((z-1)^2\big)\big\}.$$

Nous déduisons donc de (4·3) et (4·8) que, par exemple sous l'hypothèse $\frac{1}{10}k \leqslant \log_2 y \leqslant 10k$, on a

$$\lambda_k^*(y) = \sum_{0\leqslant j<k} \frac{(\log_2 y)^j}{j!\log y}\Big\{1+A\Big(\frac{j}{\log_2 y}-1\Big)+O\Big(\frac{1}{k}+\Big(\frac{j}{\log_2 y}-1\Big)^2\Big)\Big\}. \tag{4·9}$$

La contribution des termes d'erreur au membre de gauche de (4·9) est $O(1/k)$. On obtient donc bien (4·7) dans ce cas.

Lorsque $\log_2 y < \frac{1}{10}k$ ou $\log_2 y > 10k$, le terme principal du membre de droite de (4·7) vaut respectivement $1 + O\big(\mathrm{e}^{-k}\big)$ ou $O\big(\mathrm{e}^{-k}\big)$, ainsi qu'on peut le déduire aisément d'estimations classiques sur les sommes partielles de la série de l'exponentielle.(7) Le résultat souhaité (4·7) découle donc trivialement de la monotonie de la fonction $y \mapsto \lambda_k^*(y)$.

7. Voir par exemple [12] ou [13].

Nous sommes maintenant en mesure d'établir (1·10).

Nous pouvons supposer $|z| \leqslant k^{1/6}$: dans le cas contraire, l'estimation requise résulte de la monotonie en z de la fonction $\lambda_k^*\big(\exp\exp\{k+z\sqrt{k}\}\big)$. Posons $m := k-1$, $k+z\sqrt{k} = m+w\sqrt{m}$, de sorte que $w = z+1/\sqrt{k}+O(1/k)$. On déduit immédiatement de la formule (4·7) que

$$\lambda_k^*\big(\exp\exp\{k+z\sqrt{k}\}\big) = \sum_{j\leqslant m} \frac{(m+w\sqrt{m})^j}{j!\,\mathrm{e}^{m+w\sqrt{m}}} - A\frac{(m+w\sqrt{m})^m}{m!\,\mathrm{e}^{m+w\sqrt{m}}} + O\Big(\frac{1}{k}\Big).$$

Le résultat souhaité découle de cette estimation et de (2·1), en évaluant le facteur de A par la formule de Stirling et en remarquant que

$$\Phi^*(w) = \Phi^*(z) - \mathrm{e}^{-z^2/2}/\sqrt{2\pi k} + O(1/k).$$

□

Preuve du Corollaire 1.2.

Nous allons montrer plus précisément le résultat suivant.

Lemme 4.2. *Soit $k \geqslant 1$. Si $y > 0$ satisfait à*

$$\lambda_k^*(y) = \tfrac{1}{2} + O\big(1/k\big), \tag{4·10}$$

alors on a

$$\log_2 y = k - b + O\big(1/\sqrt{k}\big), \tag{4·11}$$

où b est défini par (1·12).

Démonstration. Posons $\log_2 y = k + z\sqrt{k}$. On déduit immédiatement de (1·10) que $z = o(1)$ lorsque $k \to \infty$. Un développement limité fournit alors, compte tenu du fait que $\Phi^*(0) = \frac{1}{2}$, $\Phi^{*\prime}(0) = 1/\sqrt{2\pi}$, $\Phi_0(0) = \frac{1}{3} + A = b$,

$$\tfrac{1}{2} - \frac{z}{\sqrt{2\pi}} + O(z^2) - \frac{b}{\sqrt{2\pi k}} + O\Big(\frac{z}{\sqrt{k}}\Big) = \tfrac{1}{2} + O\Big(\frac{1}{k}\Big),$$

d'où

$$z\Big\{1 + O\Big(|z| + \frac{1}{\sqrt{k}}\Big)\Big\} = -\frac{b}{\sqrt{k}} + O\Big(\frac{1}{k}\Big).$$

On en déduit successivement $z \ll 1/\sqrt{k}$ et $z\sqrt{k} = -b + O\big(1/\sqrt{k}\big)$. Cela achève la démonstration. □

k	q_k	$\log q_k$	valeur prédite
5	23	$3,13549$	$3,77484$
6	47	$3,85015$	$4,41316$
7	113	$4,72739$	$5,00341$
8	199	$5,29330$	$5,55575$
9	283	$5,64545$	$6,07892$
10	467	$6,14633$	$6,57916$
11	887	$6,78784$	$7,06090$
12	1627	$7,39449$	$7,52737$
13	2083	$7,64156$	$7,98094$
14	4297	$8,36567$	$8,42344$
15	6397	$8,76358$	$8,85628$
16	10343	$9,24407$	$9,28061$
17	16111	$9,68726$	$9,69733$
18	24251	$10,0962$	$10,1072$
19	37717	$10,5379$	$10,5109$

Tableau 1.— Valeurs de q_k telles que $\lambda_k(q_k) = \max_p \lambda_k(p)$.

5. Calculs numériques

Désignons par q_k le plus petit nombre premier pour lequel la fonction $\lambda_k(p)$ atteint sa valeur maximale. Il est facile d'obtenir à partir de la formule (1·3) que $q_1 = 2$, $q_2 = 3$, $q_3 = 5$ et 7, $q_4 = 13$. Pour des valeurs plus grandes de k, un logiciel de calcul est nécessaire. Le Tableau 1 donne les valeurs de q_k et $\log q_k$ pour $5 \leqslant k \leqslant 19$ ainsi que les valeurs de $\log q_k$ prédites par le terme principal de la formule (1·8) d'Erdős et Tenenbaum.

Ainsi que l'a remarqué Balazard (cf. [7]) la suite $\{\lambda_k(p)\}_{k=1}^{\infty}$ est unimodale pour chaque p, comme l'est en fait la suite des fonctions symétriques élémentaires de toute suite finie de nombres positifs. Il n'en va pas de même de la suite $\{\lambda_k(p_j\}_{j=1}^{\infty}$, ainsi que l'atteste la courbe de la Figure 1, représentant les variations de $f(j) = \lambda_k(p_j)$ pour $k = 9$ et $1 \leqslant j \leqslant 120$.

Il est facile de constater que

$$\lambda_k(p_j) > \lambda_k(p_{j-1}) \iff p_j - p_{j-1} + 1 < s_{k-2}(p_{j-1})/s_{k-1}(p_{j-1}).$$

D'après le corollaire 2 de [7], on a pour $j \geqslant 2$, $p_j > k^2$,

$$s_{k-2}(p_{j-1})/s_{k-1}(p_{j-1}) \asymp (\log_2 p_{j-1})/k,$$

et il est facile de vérifier(8) que le membre de gauche est $\ll 1/\sqrt{k}$ si $p_k \leqslant p_j \leqslant k^2$. En employant l'encadrement

$$3 \leqslant p_j - p_{j-1} + 1 \ll p_j^{107/200},$$

où la majoration découle d'un résultat récent de Baker et Harman [1], on voit donc qu'il existe des constantes positives c_1 et c_2 telles que $j \mapsto \lambda_k(p_j)$ soit croissante pour $p_j \leqslant c_1(k/\log k)^{200/107}$ et décroissante pour $p_j > \exp\exp(c_2 k)$. Sous la conjecture de Cramér, le domaine de croissance peut être étendu à $p_j \leqslant \exp\{c_1\sqrt{k/\log k}\}$. En utilisant le fait que $p_j - p_{j-1}$ vaut en moyenne $\log p_j$, on voit que $j \mapsto \lambda_k(p_j)$ est en moyenne croissante pour $p_j \leqslant \exp\{c_1 k/\log k\}$ et en moyenne décroissante pour $p_j > \exp\{c_2 k/\log k\}$, ce qui est en accord qualitatif avec l'estimation asymptotique (1·8) pour l'argument q_k du maximum de λ_k.

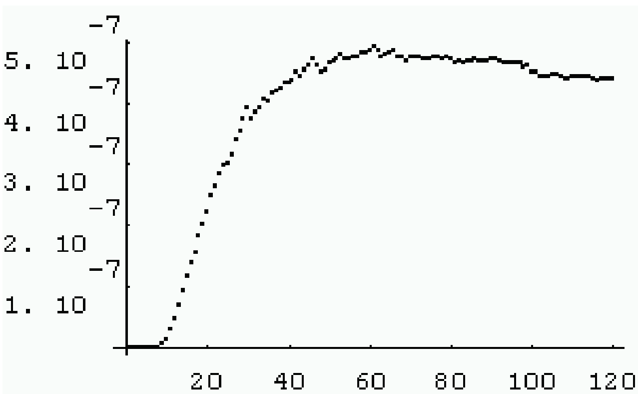

FIGURE 1.— $\lambda_9(p_j)$ pour $1 \leqslant j \leqslant 120$.

8. En utilisant, par exemple, la croissance de la suite $\nu \mapsto s_{\nu-1}(p_{j-1})/s_\nu(p_{j-1})$, qui découle d'une propriété générale des fonctions symétriques — cf. [10], *theorem* 53.

On peut constater sur la Figure 1 que la fonction $\lambda_9(p_j)$ atteint son maximum lorsque $j = 61$, ce qui correspond à $p_j = p_{61} = 283$ et que la fonction $\lambda_9(p)$ est croissante pour $p \leqslant p_{30} = 113$. Cependant, $\lambda_9(p_{31}) = \lambda_9(127) < \lambda_9(p_{30}) = \lambda_9(113)$. De plus, la courbe révèle des maximums locaux à $p_{34} = 139 < p_{35} = 149$, $p_{42} = 181 < p_{43} = 191$, $p_{46} = 199 < p_{47} = 211$ et à $p_{53} = 241 < p_{54} = 251$. On peut donc constater, en concordance avec la discussion théorique ébauchée plus haut, que, dans une certaine zone critique, les nombres premiers qui sont suivis de beaucoup de nombres composés sont d'excellents candidats pour être des maximums locaux de $\lambda_k(p)$.

En utilisant (1·3), on établit aisément que $\sum_{3\leqslant p\leqslant 31} \lambda_2(p) \approx 0,490611$ et que $\sum_{3\leqslant p\leqslant 37} \lambda_2(p) \approx 0,500248$. On en déduit que la valeur médiane du deuxième facteur premier d'un entier est 37. Nous avons ainsi établi que, essentiellement une fois sur deux, le deuxième facteur premier d'un entier choisi au hasard (dans l'ensemble des naturels) est $\leqslant 37$.

La Figure 2 représente la courbe de $\lambda_2(p_j)$ pour $j \geqslant 2$. Ainsi le point $j = 12$, correspondant au nombre premier $p = 37$, sépare l'aire sous la courbe en deux parties « presque » égales.

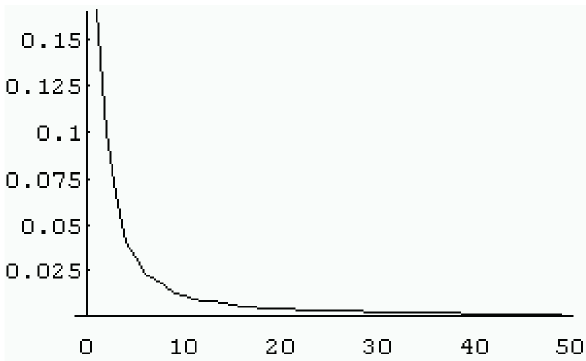

Figure 2. — $\lambda_2(p_j)$ pour $j \geqslant 2$.

La valeur médiane p_3^* du troisième facteur premier est le plus petit nombre premier tel que

$$\prod_{p<p_3^*}\Big(1-\frac{1}{p}\Big)\sum_{0\leqslant j\leqslant 2} s_j(p_3^*) \leqslant \tfrac{1}{2}.$$

Le logiciel Mathematica fournit la valeur $p_3^* = p_{4467} = 42\,719$.

Si l'on tente, par la même méthode, de déterminer la valeur médiane du quatrième facteur premier, on constate que le temps de calcul est beaucoup trop long, et en fait prendrait plusieurs mois, même avec un ordinateur très puissant. Il est toutefois possible d'obtenir une valeur approchée de p_4^* en faisant appel à des encadrements précis des quantités

$$\prod_{p\leqslant x}\Big(1-\frac{1}{p}\Big) \qquad \text{et} \qquad \sum_{p\leqslant x}\frac{1}{p}, \tag{5·1}$$

analogues à ceux de Rosser & Schoenfeld [15] et Schoenfeld [16]. On obtient ainsi

$$5,7144\cdot 10^{12} < p_4^* < 5,7164\cdot 10^{12}.$$

Les encadrements plus fins des quantités (5·1) obtenus par Schoenfeld [16] sous l'hypothèse de Riemann permettent l'estimation conjecturale

$$p_5^* \approx 7,887\cdot 10^{34}.$$

La même méthode fournit $p_5^* \approx 7\cdot 10^{34}$ inconditionnellement.

Bibliographie

[1] R. C. Baker & G. Harman, The difference between consecutive primes, *Proc. London Math. Soc.* (3) **72**, no. 2 (1996), 261-280.

[2] L. Comtet, *Analyse combinatoire*, Tome premier, Presses universitaires de France, 1970.

[3] H. Delange, Sur le nombre des diviseurs premiers de n, *Acta Arith.* **7** (1962), 191–215.

[4] P. Erdős, On the distribution function of additive functions, *Ann. of Math.* **47**, 1–20.

[5] P. Erdős, On the distribution of prime divisors, *Aequationes Math.* **2**, (1969), 177–183.

[6] P. Erdős, Some unconventional problems in number theory, Astérisque **61** (1979), 73-82.

[7] P. Erdős & G. Tenenbaum, Sur les densités de certaines suites d'entiers, *Proc. London Math. Soc.* (3) **59** (1989), 417-438.

[8] C. G. Esseen, Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law, *Acta Math. (Uppsala)* **77** (1945), 1–125.

[9] R.R. Hall & G. Tenenbaum, *Divisors*, Cambridge University Press, 1988.

[10] G. H. Hardy, J. E. Littlewood & G. Pólya, *Inequalities*, Cambridge University Press (1952).

[11] V. M. Kalinin, Special functions and the limit properties of probability distributions, I, *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **13** (1969), 5–137 [en russe].

[12] K. K. Norton, On the number of restricted prime factors of an integer I, *Illinois J. Math.* **20** (1976), 681–705.

[13] K. K. Norton, Estimates for partial sums of the exponential series, *J. Math. Analysis and Appl.* **63**, no. 1 (1978), 265–296.

[14] S. Ramanujan, *Collected papers*, Cambridge University Press, 1927 ; réimpression par Chelsea, New York, 1962.

[15] J.B. Rosser & L. Schoenfeld, Approximate formulas for some functions of prime numbers, *Illinois J. Math.* **6** (1962), 64-94.

[16] L. Schoenfeld, Sharper bounds for the Chebyshev functions $\vartheta(x)$ and $\psi(x)$ II, *Math. Comp.* **30** (1976), 337-360.

[17] G. Tenenbaum, *Introduction à la théorie analytique et probabiliste des nombres*, Cours spécialisés, no. 1, Société Mathématique de France (1995), xv + 457 pp.

[18] G. Tenenbaum, Crible d'Ératosthène et modèle de Kubilius, in : K. Győry, H. Iwaniec, J. Urbanowicz (eds.), *Number Theory in Progress*, Proceedings of the conference in honor of Andrzej Schinzel, Zakopane, Poland 1997, 1099–1129, Walter de Gruyter, Berlin, New York, 1999.

Jean-Marie De Koninck
Département de Mathématiques
et de Statistique
Faculté des Sciences et de Génie
Université Laval
Québec G1K 7P4, Canada

jmdk@mat.ulaval.ca

Gérald Tenenbaum
Institut Élie Cartan
Université de Nancy 1
BP 239
54506 Vandœuvre Cedex
France

tenenb@ciril.fr