Séminaire BOURBAKI Mars 2006
58e année, 2005-2006, nO 959, p. 177 à 210

ÉCARTS ENTRE NOMBRES PREMIERS SUCCESSIFS
[d'après Goldston, Pintz, Yıldırım, ...]

par **Emmanuel KOWALSKI**

1. INTRODUCTION

Rien n'est plus commun que de remarquer l'importance de la suite infinie

$$2 = p_1 < 3 = p_2 < \cdots < p_n < p_{n+1} < \cdots$$

des nombres premiers dans l'imaginaire des mathématiciens et des arithméticiens tout particulièrement. Parmi les belles questions qui restent complètement ouvertes concernant les nombres premiers, celles qui ont trait aux *écarts* entre nombres premiers consécutifs

$$\gamma(n) = p_{n+1} - p_n \text{ pour } n \geqslant 1$$

ont exercé depuis longtemps une fascination particulière.

Le point de départ d'une analyse rigoureuse de la répartition des valeurs de $\gamma(n)$ est forcément le théorème des nombres premiers. Notons $\pi(X)$, comme d'habitude, le nombre de nombres premiers $p \leqslant X$; on a alors

$$\pi(X) \sim \frac{X}{\ln X}, \text{ quand } X \to +\infty.$$

De manière équivalente, le n-ième nombre premier p_n vérifie

$$p_n \sim n \ln n, \text{ quand } n \to +\infty,$$

et, « en moyenne », la distance $\gamma(n)$ entre nombres premiers consécutifs est de l'ordre de $\ln n$: on a

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{2\leqslant n\leqslant N} \frac{\gamma(n)}{\ln n} = 1,$$

d'où en particulier :

$$\liminf_{n\to+\infty} \frac{\gamma(n)}{\ln n} \leqslant 1.$$

L'autre source de conjectures – qui n'a pas attendu la preuve du théorème des nombres premiers – consiste à « regarder » ce qui se passe dans des tables de nombres premiers. Il avait été observé[1] qu'un nombre relativement grand de paires (p_n, p_{n+1}) vérifie $p_{n+1} - p_n = 2$. En conséquence il a été conjecturé qu'une infinité de telles paires existe. C'est la conjecture des nombres premiers jumeaux : avec la notation ci-dessus, elle affirme que

$$\liminf_{n\to+\infty} \gamma(n) = 2. \tag{1}$$

Le thème particulier de ce rapport est celui de l'existence de petits écarts entre nombres premiers (par opposition, par exemple, avec la question de borner $\gamma(n)$ pour tout n). Après des progrès sporadiques, dus à Erdös, Rankin, Bombieri-Davenport [2], Huxley, Maier, en particulier, l'étude des petits écarts entre nombres premiers a été bouleversée durant l'année 2005. Goldston, Pintz et Yıldırım [11, 12, 14] ont en effet démontré :

Théorème 1.1 (Goldston, Pintz, Yıldırım). — *On a*

$$\liminf_{n\to+\infty} \frac{\gamma(n)}{\ln n} = 0.$$

Le meilleur résultat précédemment connu était dû à Maier :

$$\liminf_{n\to+\infty} \frac{\gamma(n)}{\ln n} \leqslant 0,2484\ldots$$

De plus, en admettant certaines hypothèses usuelles concernant la répartition des nombres premiers dans les progressions arithmétiques, on a des résultats vraiment spectaculaires :

Théorème 1.2 (Goldston, Pintz, Yıldırım)

(1) *Si l'exposant de répartition θ* au sens fort *de la suite des nombres premiers vérifie $\theta > 1/2$, on a*

$$\liminf_{n\to+\infty} \gamma(n) < +\infty.$$

(2) *Si $\theta = 1$, alors*

$$\liminf_{n\to+\infty} \gamma(n) \leqslant 16.$$

L'exposant de répartition θ est défini ci-dessous (Définition 2.2). On sait que $\theta \geqslant 1/2$: c'est le célèbre théorème de Bombieri-Vinogradov.

Dans la section 7, nous présenterons des énoncés plus forts et d'autres résultats annoncés, concernant par exemple les écarts entre nombres de la forme $p_1 p_2$ avec $p_1 \neq p_2$ (c'est-à-dire, ayant exactement deux facteurs premiers).

Mais, pour commencer, nous allons expliquer la démonstration des théorèmes 1.1 et 1.2, après avoir présenté la base de la méthode telle qu'elle apparaît à l'heure actuelle.

[1] Dickson attribue cela à Polignac.

Il faut signaler que le sujet est en pleine ébullition : tous les résultats sont encore sous forme de prépublications, et différentes variantes de la méthode apparaissent dans chacun d'entre eux. Malgré la malencontreuse aventure de 2003, aucun doute ne subsiste cependant quant à la validité des deux énoncés ci-dessus.

Notations. — Rappelons quelques notations de théorie analytique des nombres.
– p désignera toujours un nombre premier, et n, m des entiers $\geqslant 1$.
– $\Lambda(n)$ désigne la fonction de von Mangoldt ; $\Lambda(n) = 0$ si n n'est pas une puissance d'un nombre premier p et $\Lambda(p^k) = \ln p$. De plus,

$$\psi(x; q, a) = \sum_{\substack{n \leqslant x \\ n \equiv a \pmod q}} \Lambda(n).$$

– $\mu(n)$ désigne la fonction de Möbius, $\tau(n)$ la fonction « nombre de diviseurs » et $\varphi(n)$ la fonction d'Euler. Pour $n = p^k$, $k \geqslant 1$, on a $\mu(p^k) = -1$ si $k = 1$, 0 sinon, $\tau(p^k) = k + 1$ et $\varphi(p^k) = p^k - p^{k-1}$.
– $f \star g$ désigne la convolution arithmétique de f et g :

$$f \star g(n) = \sum_{d|n} f(d)g(n/d) = \sum_{ab=n} f(a)g(b).$$

– $\sum^\flat$ désigne une somme restreinte aux entiers n sans facteurs carrés.
– Les notations de Landau $f = O(g)$ et de Vinogradov $f \ll g$ sont considérées comme synonymes : $f(x) = O(g(x))$ pour tout $x \in D$ signifie qu'il existe une constante « implicite » $C \geqslant 0$ (le plus souvent, fonction d'autres paramètres explicitement mentionnés) telle que $|f(x)| \leqslant Cg(x)$ pour tout $x \in D$. Cette définition *diffère* de celle, topologique, de Bourbaki [4, Chap. V]. Par contre, nous utilisons les notations $f(x) \sim g(x)$ et $f = o(g)$ dans le sens asymptotique de *loc. cit.*

Remerciements. — Je remercie D. Goldston, J. Pintz, C. Yıldırım, J. Sivak, R. de la Bretèche et É. Fouvry pour leurs remarques et corrections concernant les différents brouillons de ce texte.

2. RAPPELS DE QUELQUES NOTIONS DE CRIBLE

Cette section rappelle quelques définitions relatives aux méthodes de crible (voir par exemple [15] pour un traitement classique très complet, [18, Ch. 6], [19, Ch. 4] pour des exposés assez courts comportant les preuves des assertions de base, et [5, 20] pour deux survols dans ce séminaire).

Soit $\mathcal{A} = (a_n)$, $n \geqslant 1$, une suite de nombres réels positifs. On note

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

sa fonction sommatoire.

Soit P un entier, qui sera le plus souvent le produit de certains nombres premiers bien choisis, par exemple le produit des nombres premiers $p < z$. Les *sommes criblées* correspondantes sont

$$\mathcal{A}(x,P) = \sum_{\substack{n\leqslant x\\(n,P)=1}} a_n.$$

L'objectif général du crible est d'estimer $\mathcal{A}(x,P)$ avec la plus grande généralité possible.

Pour aborder l'étude des sommes criblées $\mathcal{A}(x,P)$, les méthodes de « petit » crible sont inspirées par la formule de Legendre[(2)] : on a

$$\mathcal{A}(x,P) = \sum_{d\mid P} \mu(d)\mathcal{A}_d(x), \qquad \text{avec} \qquad \mathcal{A}_d(x) = \sum_{\substack{n\leqslant x\\ n\equiv 0 \pmod d}} a_n.$$

Bien qu'elle ne soit pas exploitable directement en général car le nombre de termes (i.e., le nombre de diviseurs de P) est trop grand, cette formule suggère d'étudier les suites $\mathcal{A}_d = (a_{nd})$. Dans beaucoup de cas intéressants du point de vue de la théorie multiplicative des nombres, il se dégage naturellement une approximation de $\mathcal{A}_d(x)$ de la forme

$$\mathcal{A}_d(x) = g(d)X + r_d(\mathcal{A};x) \tag{2}$$

où $X \geqslant 0$ est une fonction de x, $d \mapsto g(d)$ est une fonction arithmétique que l'on suppose multiplicative[(3)] et $r_d(\mathcal{A};x)$ est un « reste ».

En pratique, cette approximation est assez précise pour des valeurs de d assez grandes, disons $d < D$, c'est-à-dire que le reste total

$$\sum_{d<D} |r_d(\mathcal{A};x)|$$

peut être estimé « suffisamment bien ». Plus précisément, la notion *d'exposant de répartition* est classiquement définie :

DÉFINITION 2.1. — *La suite $\mathcal{A}$ a un exposant de répartition* au sens faible $\geqslant \theta$ *si, pour tout $A > 0$ et tout $\varepsilon > 0$, on a*

$$\sum_{d<D} |r_d(\mathcal{A};x)| \ll \frac{X}{(\ln X)^A} \tag{3}$$

pour $x \geqslant 2$ et $D \leqslant x^{\theta-\varepsilon}$, la constante implicite dépendant au plus de A, ε et $\mathcal{A}$.

---

(2) Que l'on peut voir, au choix, comme l'expression du principe d'inclusion–exclusion, ou de la formule $\sum_{d\mid n} \mu(d) = 0$ si $n \neq 1$.

(3) Au sens où $g(nm) = g(n)g(m)$ si $(n,m) = 1$.

La méthode de Goldston, Pintz et Yıldırım requiert crucialement des informations, non seulement pour une suite $\mathcal{A} = (a_n)$, mais pour les suites « décalées » (a_{n+t}), où $t \geqslant 0$ est un entier quelconque. Cela mène à un renforcement naturel de la notion d'exposant de répartition.

Pour la définir, on généralise donc (2) en écrivant

$$\sum_{\substack{n\leqslant x\\ n\equiv t\ (d)}} a_n = g_t(d)X + r_d(\mathcal{A}; x, t) \tag{4}$$

où X est encore une fonction de x et $d \mapsto g_t(d)$ est pour tout t une fonction multiplicative dont la valeur ne dépend que de t modulo d.

Définition 2.2. — *Une suite $\mathcal{A} = (a_n)$ de réels positifs a un exposant de répartition* au sens fort $\geqslant \theta$ *si pour tout $A > 0$ et tout $\varepsilon > 0$, on a*

$$\sum_{d\leqslant D} \max_{t\,(\mathrm{mod}\ d)} \max_{y\leqslant x} |r_d(\mathcal{A}; y, t)| \ll \frac{X}{(\ln X)^A}, \tag{5}$$

pour $x \geqslant 2$ et $D \leqslant x^{\theta-\varepsilon}$, la constante implicite pouvant dépendre de A, ε et $\mathcal{A}$.

Pour le théorème 1.2, la suite concernée est la suite $(\Lambda(n))$, ou bien la fonction caractéristique des nombres premiers.

On va voir dans la preuve du théorème 1.2 qu'une notion plus faible se dégage. Comme cela pourrait fournir une voie plus accessible à la preuve inconditionnelle de ce résultat, il semble utile de l'isoler.

Définition 2.3. — *Une suite $\mathcal{A} = (a_n)$ de réels positifs a un P-*exposant de répartition $\geqslant \theta$ *si pour tout $A > 0$, tout $\varepsilon > 0$ et tout polynôme primitif $F \in \mathbf{Z}[X]$ de degré $\geqslant 1$ on a*

$$\sideset{}{^\flat}\sum_{d\leqslant D} \sum_{\substack{\nu\,(\mathrm{mod}\ d)\\ F(\nu)=0}} \max_{y\leqslant x} |r_d(\mathcal{A}; y, \nu)| \ll \frac{X}{(\ln X)^A},$$

pour $x \geqslant 2$ et $D \leqslant x^{\theta-\varepsilon}$, la constante implicite pouvant dépendre de A, ε, $\mathcal{A}$ et F.

Remarque 2.4. — Une prépublication de Pintz et Motohashi [21], parue après la première version de ce survol, introduit explicitement une variante de cette dernière notion dans le cas de la suite des nombres premiers, avec une plus grande flexibilité potentiellement intéressante.

Le cas le plus important pour notre propos est celui de la suite $a_n = \Lambda(n)$. Dans ce cas on a pour $t \geqslant 0$

$$\sum_{\substack{n\leqslant x\\ n\equiv t\ (d)}} a_n = \sum_{\substack{n\leqslant x\\ n\equiv t\ (d)}} \Lambda(n) = \psi(x; d, t),$$