Une découverte sur les nombres premiers

Jean Lefort(*) d'après Eric Rowland

Les découvertes en mathématiques sont très souvent hors de portée du non spécialiste tant au niveau de l'énoncé que de la démonstration. Il arrive cependant que l'on prouve des résultats simples à énoncer (comme le théorème de Fermat-Wiles) dont la démonstration fait appel à des théories de haut niveau. Les quelques rares cas où une découverte prometteuse se démontre avec les outils du niveau lycée méritent donc une grande publicité. C'est un tel exemple datant de janvier 2008 et portant sur une suite ne fournissant que des nombres premiers ou des 1 que nous présentons aujourd'hui. La démonstration repose essentiellement sur le fait que pgcd(a, b) divise toute combinaison linéaire entière de a et b et sur le fait que p divise pgcd(pa, pb).

On peut remarquer que la quête d'une formule ne donnant que des nombres premiers est aussi longue que l'histoire des mathématiques. À côté de la formule d'Euler : $n^2 - n + 41$ et d'autres du même genre qui semblent donner beaucoup de nombres premiers mais dont le comportement à long terme n'est pas connu, la principale avancée a été faite par Matiyasevich en 1971 qui a permis la découverte d'un polynôme en 26 variables qui donne, pour toute valeur entière des variables, soit un entier négatif ou nul, soit un entier positif qui est alors premier.

Lors d'une université d'été en 2003, Matt Frank, à la tête d'un petit groupe comprenant Eric Rowland, découvrit la suite définie par la récurrence suivante :

$$f(1) = 7 \ ; f(n) = f(n-1) + \text{pgcd}(n, f(n-1))$$

Les participants remarquèrent que la quantité pgcd$(n, f(n-1))$, qui n'est autre que la suite des différences premières de la suite f [1], ne prend pour valeur que 1 ou un nombre premier. En janvier 2008, Eric Rowland publiait la démonstration de cette conjecture. C'est la partie principale de son article que nous développons ci-après[2].

Programmer les résultats

La suite des pgcd se programme très facilement et l'on obtient rapidement un grand nombre de termes. À titre d'exemple, voici, sous Maple, la programmation des 50 premiers :

(*) jlefort.apmep@wanadoo.fr

(1) On rappelle qu'une suite u_n étant donnée, on appelle suite des différences premières la suite $v_n = u_n - u_{n-1}$.

(2) On peut télécharger son article, paru sous le titre *A simple Prime-Generating Recurrence* dans Abstracts Amer. Math. Soc., 29-1-2008 sur http://arxiv.org/abs/0710.3217 en version pdf. On y trouvera outre une démonstration plus générale que celle que nous donnons ici, des éléments du type narration de recherche et des conjectures fortement étayées sur les questions que l'on peut se poser à propos de ce type de suite. L'article est évidemment en anglais, mais ne fait que dix petites pages.

```
> L:=[1]: f:=7: for n from 2 to 50 do f:=f+gcd(n,f):
  L:=[op(L),gcd(n+1,f)]: od:
> L;
```

[1, 1, 1, 5, 3, 1, 1, 1, 1, 1, 11, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 23, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 47, 3, 1, 5, 3]

Le problème c'est qu'il semble y avoir de plus en plus de 1. Il est facile d'éliminer l'écriture des 1 en modifiant légèrement le programme précédent :

```
> L:=[]: f:=7: for n from 2 to 1000 do f:=f+gcd(n,f):
  if gcd(n+1,f)>1 then L:=[op(L),gcd(n+1,f)] fi: od:
> L;
```

[5, 3, 11, 3, 23, 3, 47, 3, 5, 3, 101, 3, 7, 11, 3, 13, 233, 3, 467, 3, 5, 3, 941, 3, 7]

Mais nous voyons qu'il faut aller loin ($n = 1000$) pour n'avoir que 25 nombres premiers et qui ne sont pas tous différents.

On pourra tester ces programmes pour d'autres valeurs initiales telles que $f(1) = 3$ qui ne conduit à aucun nombre premier, $f(1) = 532$ qui contient beaucoup de 1 et de nombres premiers mais pas seulement, $f(3) = 8$ qui contient un début irrégulier, … Nous reviendrons sur ces exemples après la démonstration.

Avoir des idées en explorant la suite

Avant de faire une démonstration, il faut savoir dans quelle direction chercher. C'est ce qu'explique Rowland dans son article. Pour mieux comprendre les raisons du lemme essentiel donné plus loin, j'invite le lecteur à émettre le maximum de conjectures à partir du tableau de la page suivante où $g(n) = \text{pgcd}(n, f(n-1))$.

Un lemme

Soit un entier $n_1 \geq 2$ et la suite définie par la récurrence :

$$f(n_1) = 3n_1 \text{ et, pour } n > n_1,\ f(n) = f(n-1) + \text{pgcd}(n, f(n-1))$$

Soit $n_2 > n_1$ le plus petit entier tel que $\text{pgcd}(n_2, f(n_2 - 1)) \neq 1$. Alors :

- $\text{pgcd}(n_2, f(n_2 - 1))$ est un nombre premier p.
- p est le plus petit diviseur de $2n_1 - 1$.
- $n_2 = n_1 + \dfrac{p-1}{2}$.
- $f(n_2) = 3n_2$.

Si on a bien conjecturé à partir du tableau ci-dessous, les propositions de ce lemme sont assez naturelles : c'est évident pour la première proposition qui n'est autre que ce qu'on cherche à démontrer, à savoir que $\text{pgcd}(n, f(n-1))$ est premier quand il est différent de 1 ; c'est également évident pour la dernière puisqu'il semble bien qu'il y a équivalence entre $f(n)/n = 3$ et $\text{pgcd}(n, f(n-1))$ est un nombre premier. Les deux autres propositions sont moins immédiates mais le lecteur pourra toujours les vérifier sur le tableau précédent s'il n'a pas eu l'idée, comme Eric Rowland, de compter le nombre de 1 entre deux nombres premiers.

n	$g(n)$	$f(n)$	$f(n)/n$
1		7	7
2	1	8	4
3	1	9	3
4	1	10	2.5
5	5	15	3
6	3	18	3
7	1	19	2.71429
8	1	20	2.5
9	1	21	2.33333
10	1	22	2.2
11	11	33	3
12	3	36	3
13	1	37	2.84615
14	1	38	2.71429
15	1	39	2.6
16	1	40	2.5
17	1	41	2.41176
18	1	42	2.33333
19	1	43	2.26316
20	1	44	2.2
21	1	45	2.14286
22	1	46	2.09091
23	23	69	3
24	3	72	3
25	1	73	2.92
26	1	74	2.84615
27	1	75	2.77778
28	1	76	2.71429
29	1	77	2.65517
30	1	78	2.6
31	1	79	2.54839
32	1	80	2.5

n	$g(n)$	$f(n)$	$f(n)/n$
33	1	81	2.45455
34	1	82	2.41176
35	1	83	2.37143
36	1	84	2.33333
37	1	85	2.2973
38	1	86	2.26316
39	1	87	2.23077
40	1	88	2.2
41	1	89	2.17073
42	1	90	2.14286
43	1	91	2.11628
44	1	92	2.09091
45	1	93	2.06667
46	1	94	2.04348
47	47	141	3
48	3	144	3
49	1	145	2.95918
50	5	150	3
51	3	153	3
52	1	154	2.96154
53	1	155	2.92453
54	1	156	2.88889
.....			
.....			
99	1	201	2.0303
100	1	202	2.02
101	101	303	3
102	3	306	3
103	1	307	2.98058
104	1	308	2.96154
105	7	315	3
106	1	316	2.98113

On notera enfin que ce lemme amorce une récurrence puisqu'à partir de n_1 tel que $f(n_1) = 3n_1$, on construit n_2 satisfaisant la même équation et qu'au passage on a obtenu un nombre premier.

Si on réfléchit bien, il se pourrait que le lemme porte sur un ensemble vide pour certains choix de n_1, ce qui serait le cas si $\text{pgcd}(n, f(n-1))$ était toujours égal à 1. Il nous faudra donc démontrer l'existence de n_2 en plus des propositions du lemme.
On notera enfin que $2n_1 - 1$ étant impair, son plus petit diviseur premier p est impair (donc supérieur ou égal à 3) et par suite $\dfrac{p-1}{2}$ est un entier non nul.

• Existence de n_2

Soit p le plus petit diviseur premier de $2n_1 - 1$; posons $2n_1 - 1 = pq$ (p et q sont tous les deux impairs). Supposons que $\text{pgcd}(n_1 + i, f(n_1 + i - 1))$ soit égal à 1 pour

i variant de 1 à $\frac{p-1}{2}-1$. Si ce n'est pas le cas, alors nous sommes assurés de l'existence de n_2. Tant que ce pgcd vaut 1, $f(n)$ augmente d'une unité à chaque étape et par conséquent $f(n_1+j)=f(n_1)+j=3n_1+j$ pour j variant de 0 à $\frac{p-1}{2}-1$. Dans cette hypothèse, nous avons

$$\text{pgcd}\left(n_1+\frac{p-1}{2}, f\left(n_1+\frac{p-1}{2}-1\right)\right)=\text{pgcd}\left(n_1+\frac{p-1}{2}, 3n_1+\frac{p-1}{2}-1\right).$$

Or, d'une part $n_1+\frac{p-1}{2}=\frac{2n_1-1+p}{2}=p\frac{q+1}{2}$ qui est le produit de p par un entier puisque q est impair et, d'autre part, $3n_1+\frac{p-1}{2}-1=\frac{6n_1+p-3}{2}=p\frac{3q+1}{2}$ qui est également le produit de p par un entier. Comme p est en facteur dans les deux termes du pgcd, il divise ce pgcd et par suite celui-ci est différent de 1.

- $\boldsymbol{n_2=n_1+\frac{p-1}{2}}$

Le raisonnement précédent prouve que $n_2 \leq n_1+\frac{p-1}{2}$. Posons alors $k=n_2-n_1$.

De n_1 à $n_1+k-1=n_2-1$, $f(n)$ augmente d'une unité quand n augmente d'une unité, c'est-à-dire que :

$$\forall i \in [\![1, k-1]\!] \quad f(n_1+i)=f(n_1)+i=3n_1+i.$$

Par suite :

$$\text{pgcd}\left(n_1+k, f(n_1+k-1)\right)=\text{pgcd}\left(n_1+k, 3n_1+k-1\right)$$

et cette quantité divise toute combinaison linéaire entière des deux termes, en particulier la différence

$$(3n_1+k-1)-(n_1+k)=2n_1-1$$

et la quantité

$$3(n_1+k)-(3n_1+k-1)=2k+1.$$

Par conséquent pgcd$(n_1+k, f(n_1+k-1))$, différent de 1 par hypothèse, divise $2n_1-1$ (dont p est le plus petit facteur premier) et $2k+1$. D'où :

$$p \leq \text{pgcd}(n_2, f(n_2-1)) \leq 2k+1.$$

Ce qui montre que $p \leq 2k+1$ soit $\frac{p-1}{2} \leq k$; autrement dit $n_2 \geq n_1+\frac{p-1}{2}$.

En conclusion, on a bien $n_2=n_1+\frac{p-1}{2}$ (ou $n_2-n_1=k=\frac{p-1}{2}$).

• **pgcd$(n_2, f(n_2 - 1)) = p$.**

En reprenant l'inégalité $p \leq \text{pgcd}(n_2, f(n_2 - 1)) \leq 2k + 1$ et en remplaçant k par la valeur trouvée, il vient $p \leq \text{pgcd}(n_2, f(n_2 - 1)) \leq p$.
Ceci prouve bien que $\text{pgcd}(n_2, f(n_2 - 1)) = p$.

• **$f(n_2) = 3n_2$.**

Finalement $f(n_2) = f(n_2 - 1) + \text{pgcd}(n_2, f(n_2 - 1))$.

Or $f(n_2 - 1) = f(n_1) + \dfrac{p-1}{2} - 1 = 3n_1 - 1 + \dfrac{p-1}{2}$ et $\text{pgcd}(n_2, f(n_2 - 1)) = p$.

Donc

$$f(n_2) = 3n_1 - 1 + \frac{p-1}{2} + p = 3n_1 + \frac{3p-3}{2} = 3n_2.$$

Théorème

Si $f(1) = 7$ et si $f(n) = f(n - 1) + \text{pgcd}(n, f(n - 1))$ alors, pour tout n strictement positif, $\text{pgcd}(n, f(n - 1))$ vaut soit 1, soit un nombre premier.

En effet $f(1) = 7$, $f(2) = 8$ et $f(3) = 9$ et nous pouvons appliquer le lemme pour $n_1 = 3$ puis raisonner par récurrence ($2n_1 - 1 = 5$, $p = 5$, $\dfrac{p-1}{2} = 2$, $n_2 = 5$ et on recommence en remplaçant n_1 par n_2) et comme le résultat est également vrai pour $n = 1$ ou 2, le théorème est démontré par récurrence.

Généralisation et ouvertures

On trouvera dans l'article cité une généralisation du lemme pour $f(n_1) = 2n_1$ avec $n_1 \geq 3$, en prenant pour p le plus petit diviseur de $n_1 - 1$ et en choisissant $n_2 = n_1 + p - 1$.
Mais on y trouvera surtout toute une série de questions ouvertes tant il est vrai que toute découverte en mathématiques pose souvent plus de questions qu'elle n'en résout. Parmi les questions que soulève Rowland, je détache les trois suivantes dans la mesure où l'auteur avance quelques arguments heuristiques qui permettent de penser à une réponse affirmative.

1) Obtient-on tous les nombres premiers impairs ? C'est-à-dire, un nombre premier impair étant donné, l'obtient-on au bout d'un certain temps ? Remarquons que d'après l'étude précédente il faut au moins attendre le passage de $\dfrac{p-1}{2}$ valeurs 1. Si on regarde la liste des 5 000 premiers nombres premiers, on n'en trouve que 965 différents et 397 est le plus petit premier qui n'apparaît pas dans cette liste. Apparaît-il plus loin ?
2) En dehors de quelques cas très particuliers tels que $f(1) = 2$ ou 3, a-t-on toujours $\text{pgcd}(n, f(n - 1))$ égal à 1 ou à un nombre premier pour n assez grand ?

3) Le nombre premier 3 semble jouer un rôle particulier. Peut-on prévoir son rythme d'apparition ? (On trouvera dans l'article de Rowland quelques résultats sur ce point).

Comment éviter les 1 :

Nous avons vu au début une programmation qui évite l'affichage des 1, mais le programme calcule toutes les étapes intermédiaires. On peut éviter cela en s'appuyant sur le lemme. Il affirme que le nombre premier p qui apparaît après une suite de 1 commençant à n_1 est le plus petit facteur premier de $2n_1 - 1$ et qu'alors l'étape suivante est obtenue en $n_2 = n_1 + (p - 1)/2$. Il suffit donc d'itérer ce processus en partant de $n_1 = 3$ pour obtenir dans l'ordre les nombres premiers apparaissant dans la suite.

Voici l'algorithme écrit sous Maple pour les 50 premiers nombres premiers :

```
> n:=3: L:=[]:
> for i from 1 to 50 do
    m:=2*n-1:
    p:=op(1,op(1,op(1,ifactor(m)))):
    n:=n+(p-1)/2:
    L:=[op(L),p]:
  od:
> L;
  [5, 3, 11, 3, 23, 3, 47, 3, 5, 3, 101, 3, 7, 11, 3, 13, 233, 3, 467, 3, 5, 3, 941,
      3, 7, 1889, 3, 3779, 3, 7559, 3, 13, 15131, 3, 53, 3, 7, 30323, 3, 60647,
      3, 5, 3, 101, 3, 121403, 3, 242807, 3, 5]
```

De la façon dont l'algorithme est construit, il est normal de n'obtenir que des nombres premiers. Mais la recherche du plus petit facteur premier n'est ni simple ni rapide dès que les nombres qui interviennent sont grands (et ils vont l'être de plus en plus au fur et à mesure que l'on veut aller plus loin). C'est d'ailleurs cette difficulté de factorisation qui permet de sécuriser les transactions sur Internet à l'aide du code RSA. Cependant chaque remarque, chaque découverte sur cette suite permettra d'en mieux comprendre le devenir et de peut-être démontrer l'une ou l'autre des conjectures qu'elle suscite.

Personne ne sait encore si cette découverte va faire progresser l'étude des nombres premiers ou n'être qu'un épiphénomène. C'est bien le propre de la recherche que d'explorer tous les domaines et les applications ne viennent que dans un deuxième temps.
Bref, il y a encore du travail !