Arithmétique et Tests de Primalité

Pierre Rouchon

Centre Automatique et Systèmes
Mines ParisTech, PSL Research University
pierre.rouchon@mines-paristech.fr

Octobre 2018

Plan

- On note $\mathbb{P}$ l'ensemble des nombres premiers et la fonction de comptage $\pi(x)$:
$$\pi(x) := \#\{p \in \mathbb{P} \mid p \leq x\}$$
Gauss et Legendre avaient déjà suggéré au cours des dernières années du *XVIII^e^* siècle que $\pi(x) \sim x/\log(x)$ pour x grand.
- Ce n'est qu'en 1896 que Hadamard et de la Vallée-Poussin ont montré ce résultat en utilisant une fonction de la variable complexe "qui code les nombres premiers", la fonction ζ de Riemann définie par la série de Diriclet
$$\zeta(s) = \sum_{n \geq 1} \frac{1}{n^s}$$
absolument convergente pour $\Re(s) > 1$.
- Ainsi si on tire au hazard un nombre de 1024 bits, il est premier avec une probabilité de $1/\log\left(2^{1024}\right) \approx 1/710$.

Exo: Supposons que nous ayons classé les nombres premiers par ordre croissant p_n avec $n \in \mathbb{N}^*$, $p_n < p_{n+1}$. Alors montrer que $\pi(x) \sim x/\log(x)$ pour x grand, équivaut à $p_n \sim n\log(n)$ pour n grand.

On utilise le fait que pour $x \in [p_n, p_{n+1}[$, $\pi(x) = n$.

- $\pi(x) \sim x/\log x \quad \Rightarrow \quad p_n \sim n \log n$. Puisque $\pi(p_n) = n$, on a $\frac{p_n}{\log p_n} = n(1 + \epsilon_n)$ avec $\lim_{n \mapsto +\infty} \epsilon_n = 0$. Donc $\log p_n \left(1 - \frac{\log \log p_n}{\log p_n}\right) = \log n \left(1 + \frac{\log(1+\epsilon_n)}{\log n}\right)$. Comme $p_n \mapsto +\infty$, on en déduit $\log p_n = \log n \, (1 + \eta_n)$ avec $\lim_{n \mapsto +\infty} \eta_n = 0$. Ainsi en revenant à $\frac{p_n}{\log p_n} = n(1 + \epsilon_n)$, on a $p_n = n \log n \, (1 + \mu_n)$ avec $\lim_{n \mapsto +\infty} \mu_n = 0$.
- $p_n \sim n \log n \quad \Rightarrow \quad \pi(x) \sim x/\log x$. Pour tout entier x il existe un unique n_x tel que $p_{n_x} \leq x < p_{n_x+1}$. Comme $p_n/p_{n+1} \sim 1$ on en déduit que $x \sim p_{n_x}$ soit $x \sim n_x \log n_x$. Mais $n_x = \pi(x)$. Donc $x \sim \pi(x) \log \pi(x)$ d'où facilement $\log \pi(x) \sim \log x$ et donc $\pi(x) \sim x/\log x$.

- Si l'on tire au hasard un nombre de 1024 bits, on sait, en utilisant $\pi(x) \equiv x/\log(x)$, que l'on a une chance sur $\log(2^{1024}) \approx 710$ de tomber sur un nombre premier.
- Le moyen le plus simple pour obtenir un grand nombre premier est de prendre au hasard un grand entier et de tester s'il est premier (test de Miller-Rabin : un algorithme très efficace qui garantie la primalité avec une probabilité aussi petite que l'on veut).

Algorithme AKS

- Prime is in **P** : en 2002, Manindra Agrawal de l'Indian Institute of Technology à Kanpur et deux de ses étudiants Neeraj Kayal et Nitin Saxena ont trouvé un algorithme simple et polynômial en au plus $0(\log^{12}(n))$ qui teste la primalité de n
- Leur algorithme s'appuie de façon ingénieuse sur le petit théorème de Fermat qui dit que pour tout entier premier n et tout entier $a \neq 0 \mod (n)$ (i.e. a premier avec n) alors $a^{n-1} = 1 \mod (n)$. Nous ne décrirons pas ici cet algorithme. Nous renvoyons le lecteur à "google" avec comme mots clés "AKS" et "Prime" pour avoir les informations les plus récentes sur cet algorithme.
- Enfin, l'algorithme AKS répond par "oui" ou "non" à la question :" n, est-il premier ?". Si la réponse est "non", l'algorithme ne donne pas de diviseur non trivial de n. Ainsi, la difficulté de la factorisation, difficulté sur laquelle repose le système RSA reste entière.

- Le petit théorème de Fermat : si n est premier alors pour tout $a \neq 0 \mod (n)$, $a^{n-1} = 1 \mod (n)$.
- Ainsi, si après avoir tiré au hasard un nombre n, on trouve un $1 \leq a < n$ tel que $a^{n-1} \neq 1 \mod (n)$, on sait que n n'est pas premier. Aussi, il est tentant de faire le dépistage heuristique suivant appelé test de Fermat : prendre $a < n$ au hasard et calculer $a^{n-1} \mod (n)$. Si $a^{n-1} = 1 \mod (n)$ on dit que n est premier en base a.

- Etudier l'efficacité de ce test revient à étudier la densité des nombres composés et premiers en base a par rapport celle des nombres premiers : on parle alors d'entiers pseudo-premiers en base a. Si on note $\pi_a(x)$ le nombre d'entiers n composés, premiers en base a et $\leq x$, alors on sait que $\pi_a(x)/\pi(x)$ tend vers zéros quand x tend vers l'infini.
- Par exemple, pour $a = 2$, Pomerance (1981) a montré que

$$\exp\left(\log(x)^{5/14}\right) \leq \pi_2(x) \leq x \exp\left(-\frac{\log(x) \log\log\log(x)}{2 \log\log(x)}\right)$$

Ainsi pour $x = 2^{512}$ on a

$$\pi_2(x)/\pi(x) \leq 5.3\ 10^{-32}.$$

- On peut dire que pour n grand choisi au hasard, il faut être très malchanceux pour que n soit pseudo-premier en base a pour un grand nombre de a.
- Cependant, cela ne veut pas dire qu'il n'existe pas de nombre n qui soit pseudo-premier pour toutes bases a entre 2 et $n-1$ avec a premier avec n.
- En fait, il en existe une infinité. Ce sont les nombres de Carmichael qui sont la plaie du test de Fermat. On a montré en 1994 que pour x assez grand il existe au moins $x^{2/7}$ nombres de Carmichael inférieurs à x

- Pour n premier et $2 \leq a \leq n-1$, on sait que $a^{n-1} = 1 \mod (n)$. Mais $n-1$ est pair donc $b = a^{\frac{n-1}{2}}$ vérifie $b^2 = 1 \mod (n)$. Donc b est racine du polynôme $y^2 - 1 = 0$ dans $\mathbb{Z}/n\mathbb{Z}$ qui est un corps pour l'addition et la multiplication modulo-n car n est premier.
- Dans un corps, le nombre de racines d'un polynôme est au plus égal à son degré. Donc on a nécessairement $b = 1 \mod (n)$ ou $b = -1 \mod (n)$, car 1 et -1 sont deux racines distinctes de $y^2 - 1$. Ainsi, si n est premier, $a^{\frac{n-1}{2}} = \pm 1 \mod (n)$ pour tout $0 < a < n$. Si $(n-1)/2$ est encore pair et si $a^{\frac{n-1}{2}} = 1 \mod (n)$, alors un raisonnement identique montre que nécessairement $a^{\frac{n-1}{4}} = \pm 1 \mod (n)$.
- En continuant jusqu'à l'entier s pour lequel $\frac{n-1}{2^s}$ soit impair on obtient le test de Miller Rabin

- Pour un entier impair n que l'on écrit $n = 1 + 2^s t$ avec $s \geq 1$ et t impair : on prend au hasard a entier entre 2 et $n - 1$ et on calcule les nombres $r_i = a^{2^i t} \mod (n)$.
- L'entier n passe le test dans les deux cas suivants : soit $r_0 = r_1 = ... = r_s = 1$; soit il existe i entre 0 et $s - 1$ tel que $r_i = -1$.
- Un nombre n qui passe le test de Miller-Rabin pour un certain a est dit fortement premier en base a.
- Pas d'analogue des nombres de Carmichael car si n est composé, n est fortement premier en base a pour au plus $1/4$ des entiers a entre 2 et $n - 1$ (résultat admit).

Exo: Donner une version optimisée sur le plan algorithmique du test de Miller-Rabin. Evaluer sa complexité en terme de multiplication modulo-n .

- On choisit a_1 entre 2 et $n-1$ au hasard. Si n n'est pas fortement premier en base a_1, n n'est pas premier on s'arrête. Sinon, on choisit toujours au hasard un nouveau a_2 différent de a_1 entre 2 et $n-1$.
- Si n est fortement premier en base a_2, on choisit un troisième nombre a_3 différent des deux précédents.
- Et ainsi de suite jusqu'à avoir choisi au plus k nombres différents a_i entre 2 et $n-1$.

Supposons maintenant que le nombre n soit fortement premier pour les k bases $(a_1, \ldots, a_k)$ tirées au hazard. Quelle est la probabilité pour qu'il soit premier ?
Si n est composé, un simple comptage montre que, la proportion des k-uples $(a_1, ..., a_k) \in \{2, ..., n-1\}^k$ tels que n soit fortement premier pour chaque a_i est plus petite que $\frac{1}{4^k}$. Ainsi il est tentant dire qu'un tel n est premier avec une probabilité d'au moins $1 - \frac{1}{4^k}$.

Soit N un grand nombre entier. On prend au hasard $n \leq N$. On suppose que pour un entier k assez grand, n ait passé avec succès k tests de Miller-Rabin.

- La probabilité que n soit premier est de $1/\log(N)$.
- Un raisonnement un peu rapide nous dirait que n est premier avec une probabilité supérieure à $1 - 1/4^k$.
- Cela est faux, car il faut aussi considérer le fait que n est pris au hasard parmi les nombres plus petits que N, avec une probabilité de $1/\log(N)$ de tomber, lors de ce premier tirage, sur un nombre premier.

Pour le bon calcul, il suffit de minorer $a_N/(a_N + b_N^k)$ avec a_N le cardinal des nombres premiers plus petits que N et b_N^k, le cardinal des nombres composés plus petits que N et qui passent k tests :

$$a_N \approx \frac{N}{\log N}, \quad b_N^k \leq \frac{N - \frac{N}{\log N}}{4^k}.$$

Soit un grand entier N, et k assez grand pour que $4^k \gg \log(N)$: on tire au hasard $n \leq N$ pour lequel on fait k tests de Miller-Rabin.

- Loi de Bayes : pour toutes partitions $(A_\mu)_\mu$ de l'ensemble des possibles $P(A_\mu/B) = \frac{P(B/A_\mu)P(A_\mu)}{\sum_\nu P(B/A_\nu)P(A_\nu)}$. Prendre ici $B \equiv$ test passé k fois, $\nu \in \{1, 2\}$, $A_1 \equiv n$ premier, $A_2 \equiv n$ composé.
- Alors la probabilité pour que n, choisi au hasard $\leq N$ et passant le test k fois, soit effectivement premier vérifie :

$$\text{Prob}(n \text{ premier} \mid \text{test passé } k \text{ fois}) \geq 1 - \log(N)/4^k.$$

- On donne $\varepsilon \ll 1$ et un ordre de grandeur N pour n, on en déduit le nombre minimal k de tests à faire par la formule

$$k = \frac{\log(\log(N)/\varepsilon)}{\log(4)}.$$

- Par exemple, pour un nombre de 512-bits pris au hasard, il faut prendre $k = 26$ (resp. $k = 42$) si on veut avoir une probabilité d'erreur de moins de 10^{-5} (resp. 10^{-10}) de se tromper après k tests de Miller-Rabin positifs.

- Il s'agit d'un algorithme déterministe et polynômial qui repose sur une conjecture plausible en théorie des nombres : l'hypothèse de Riemann généralisée.
- Hypothèse de Riemann pour la fonction $\zeta(s)$:

$$\zeta(s) = \sum_{n \geq 1} \frac{1}{n^s}$$

ne semble définir $\zeta(s)$ que pour $\Re(s) > 1$, on peut prolonger ζ sur tout le plan complexe sauf en $s = 1$ qui est un pôle simple : $\zeta(s) - 1/(s-1)$ est une fonction entière comme le sont les fonctions $\cos s$ ou $\sin s/s$ par exemple (le développement en séries en $s = 0$ admet un rayon infini de convergence).

- $\zeta(-2n) = 0$ pour n entier > 0. On sait aussi depuis longtemps que les autres zéros, appelés zéros non triviaux, sont dans la bande verticale $0 < \Re(s) < 1$. Dans son fameux article de 1859, Riemann émet l'hypothèse que les zéros non triviaux de ζ sont sur la droite verticale $\Re(s) = 1/2$.
- L'hypothèse de Riemann généralisée porte sur des fonctions similaires à ζ du type

$$\sum_{n \geq 1} \frac{\chi(n)}{n^s}$$

où les fonctions $\chi : \mathbb{N} \mapsto \mathbb{S}^1$ sont des fonctions périodiques et multiplicatives ($\chi(n_1 n_2) = \chi(n_1)\chi(n_2)$) particulières, les caractères de Dirichlet modulo-N, N étant la période de χ. L'hypothèse est alors que les zéros non triviaux de ces fonctions sont tous sur la droite $\Re(s) = 1/2$.

- En 1985, Miller et Bach ont prouvé, en supposant vraie l'hypothèse de Riemann généralisée, que si l'entier impair n est fortement premier pour toute base a entre 2 et $2\log^2(n)$ alors il est premier.
- Ainsi comme le test de Miller-Rabin est polynômial, il est facile de voir que les $E(2\log^2(n)) - 1$ tests qui constituent l'algorithme de Miller-Bach impliquent une complexité polynômiale.

Exo: Donner la complexité du test de Miller-Bach en reprenant les estimations obtenues pour celle du test de Miller-Rabin.

- On note $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ l'ensemble des classes modulo n. Il y en a n ($\#\mathbb{Z}_n = n$) et on identifie $\mathbb{Z}_n$ à l'ensemble $\{0, 1, ..., n-1\}$. $\mathbb{Z}_n$ est muni d'une structure naturelle d'anneau pour l'addition et la multiplication.
- Si $k \in \mathbb{Z}_n$ est inversible, on calcule son inverse via l'algorithme d'Euclide et l'identité de Bezout. On note $\mathbb{Z}_n^*$ l'ensemble des $k \in \mathbb{Z}_n$ inversibles.
- $\mathbb{Z}_n$ est un corps ssi n est premier et alors $\mathbb{Z}_n^* = \mathbb{Z}_n/\{0\}$ dans ce cas.

- Soient donc deux entiers strictement positifs $k < n$:

$$
\begin{aligned}
n &= kq_0 + r_0, \quad r_0 < k \\
k &= r_0 q_1 + r_1, \quad r_1 < r_0 \\
r_0 &= r_1 q_2 + r_2, \quad r_2 < r_1 \\
&\vdots \\
r_{m-2} &= r_{m-1} q_m + r_m, \quad r_m < r_{m-1} \\
r_{m-1} &= r_m q_{m+1} + r_{m+1}, \quad 0 = r_{m+1} < r_m
\end{aligned}
$$

  où la suite $(n, k, r_0, r_1, ..., r_m, r_{m+1})$ est strictement décroissante et arrive à zéro avec $r_{m+1} = 0$.
- Le pgcd est r_m et on a u et v tels que

$$un + vk = \text{pgcd}\,(n, k) = r_m \quad \text{Identité de Bezout}$$

- Matrices uni-modulaires : matrices à coefficients entiers et dont l'inverse est aussi à coefficients entiers (système linéaire d'inconnue $(r_0, \dots, r_m)$).

- Evaluons le nombre D de divisions de l'algorithme d'Euclide en fonction de la taille de n. L'algorithme est le plus long lorsque chaque quotient q_i vaut 1.
- $r_{m+1} = 0$, $r_m = 1$ et

$$r_i = r_{i+1} + r_{i+2}, \quad i = m-1, m-2, ..., 0 \quad \text{avec } k = r_0 + r_1, \; n = k + r_0$$

- Avec $j = m + 1 - i$ on a $n = F_{m+3}$ où F_j est la suite de Fibonacci avec le nombre d'or $\vartheta = (1 + \sqrt{5})/2$:

$$F_j = F_{j-1} + F_{j-2}, \quad j \in \{2, \ldots, m+3\}$$

avec comme départ de la récurrence, $F_0 = 0$ et $F_1 = 1$.
- Comme $F_j = (\vartheta^j - (1 - \vartheta)^j)/\sqrt{5}$ et $F_{m+3} = n$ on a au plus $m + 2 = D \leq \log_\vartheta(n)$ divisions avec reste à faire (Lamé (1845)).

- Pour tout entier $n > 1$, on note $\varphi(n)$ le nombre d'entiers entre 1 et $n-1$ premiers avec n :

$$\varphi(n) = \#\mathbb{Z}_n^*.$$

- Théorème de Fermat-Euler : Si a est premier avec n alors $a^{\varphi(n)} = 1 \mod (n)$ (preuve via $\prod_{x\in\mathbb{Z}_n^*} x = \prod_{x\in\mathbb{Z}_n^*} ax \mod (n)$).
- Lorsque n est premier $\varphi(n) = n-1$ on a le petit théorème de Fermat : si n est premier et si a entier entre 1 et $n-1$, alors $a^{n-1} = 1 \mod (n)$.

- Théorème d'Euler :

$$\sum_{d|n} \varphi(n/d) = \sum_{d|n} \varphi(d) = n$$

- Preuve : pour $1 \leq d \leq n$, on pose

$$\psi(d,n) = \#\{x \in \{1,\dots,n\} \mid \text{pgcd}\,(x,n) = d\}.$$

Si d divise n alors $\psi(d,n) \neq 0$ sinon $\psi(d,n) = 0$. Donc $n = \sum_{d=1}^{n} \psi(d,n) = \sum_{d|n} \psi(d,n)$. Mais, $\psi(d,n) = \varphi(n/d)$ si d divise n. Il suffit de diviser par d, pour mettre en bijection les nombres x tels que pgcd $(x,n) = d$ et les nombres y tels que pgcd $(y, n/d) = 1$. Ainsi on a

$$n = \sum_{d|n} \varphi(n/d) = \sum_{d/n} \varphi(d).$$

- Théorème Chinois : soient deux entiers p et $q \geq 1$ premiers entre eux. Alors les anneaux $\mathbb{Z}_p \times \mathbb{Z}_q$ et $\mathbb{Z}_{pq}$ sont isomorphes.
- Preuve : l'application $\pi : \mathbb{Z} \mapsto \mathbb{Z}_p \times \mathbb{Z}_q$ qui à $x \in \mathbb{Z}$ associe le couple $(x \mod (p), x \mod (q))$ est un homomorphisme d'anneau, surjectif et de noyau $pq\mathbb{Z}$. Il suffit de quotienter par le noyau pour avoir un isomorphisme entre $\mathbb{Z}/pq\mathbb{Z} = \mathbb{Z}_{pq}$ et $\mathbb{Z}_p \times \mathbb{Z}_q$.
- Corollaire : si p et q sont premiers entre eux, alors $\varphi(pq) = \varphi(p)\varphi(q)$
- Corollaire : Si n admet comme décomposition en facteurs premiers $n = p_1^{\alpha_1} ... p_k^{\alpha_k}$ alors

$$\varphi(n) = \prod_{i=1}^{k} (p_i^{\alpha_i} - p_i^{\alpha_i - 1})$$

Déchiffrement RSA

- On part d'un grand nombre n (d'au moins 1024 bits) qui est le produit de deux grands nombres premiers p et q : clé publique n et e inversible modulo $\varphi(n) = (p-1)(q-1)$, clé secrète (p, q).
- Chiffrement d'un message en clair défini par un entier $M \mod (n)$ se fait par la transformation suivante :

$$M \mapsto M^e \mod (n).$$

- Déchiffrement : Bob reçoit via un canal public $A = M^e$. Pour déchiffrer, il lui suffit d'élever A à la puissance d où d est l'inverse de e modulo $\varphi(n) = (p-1)(q-1)$.

$$A^d = M \mod (n).$$

Pourquoi a-t-on $M^{ed} \equiv M \mod (n)$?

- La base du déchiffrement RSA via la clé secrète d du message chiffré M^e. :

$$\forall M \in \{0, 1, ..., n-1\}, \quad M^{ed} = M \mod (n)$$

dès que $n = pq$ où p et q sont deux nombres premiers, $e \in \{1, ..., \varphi(n) - 1\}$ inversible modulo $\varphi(n)$ et d'inverse $d \in \{1, ..., \varphi(n) - 1\}$: $ed = 1 + k\varphi(n)$ pour un certain entier k.
- Si M et n sont premiers entre eux, alors par le théorème d'Euler-Fermat $M^{\varphi(n)} = 1 \mod (n)$. Ainsi

$$M^{ed} = M^{1+k\varphi(n)} = M \mod (n).$$

- Si M et n non premiers entre eux, utiliser l'isomorphisme de la preuve du théorème chinois.

- Théorème de l'élément primitif : si p est premier alors, le groupe $(\mathbb{Z}_p^*, \times)$ est cyclique, i.e., il est de la forme

$$\mathbb{Z}_p^* = \{1, a, a^2, ..., a^{p-2}\}$$

où $a \in \mathbb{Z}_p^*$ est appelé élément primitif (non nécessairement unique).

- Preuve : le nombre d'éléments primitifs a est $\varphi(p-1)$ ($\sum_{d=1}^{p-1} N_d = p-1$ où N_d est le nombre d'éléments d'ordre d. N_d vaut soit 0 ou est supérieur à $\varphi(d)$. Conclure par théorème d'Euler en montrant, lorsque d divise $p-1$, que $N_d = \varphi(d)$.

- Théorème de Lucas : le nombre n est premier, si et seulement si, il existe $\alpha \in \mathbb{Z}_n^*$ tel que $\alpha^{n-1} = 1 \mod (n)$ et $\alpha^{\frac{n-1}{p}} \neq 1 \mod (n)$ pour tout diviseur premier p de $n-1$.
- Preuve : Si n est premier alors il suffit de prendre pour α un élément primitif modulo n. Inversement, si un tel élément α existe alors il est forcément d'ordre $n-1$ dans $\mathbb{Z}_n^*$. Or tout élément de $\mathbb{Z}_n^*$ est d'ordre un diviseur de $\varphi(n)$ (reprendre des bouts de la preuve du théorème sur l'élément primitif). Comme $\varphi(n) \leq n-1$ on voit que nécessairement $\varphi(n) = n-1$ mais cela signifie n premier.

Théorème (Lucas)

Le nombre n est premier, si et seulement si, il existe $2 \leq \alpha \leq n-1$, *tel que*

$$\alpha^{n-1} = 1 \mod (n) \quad \textit{et} \quad \alpha^{\frac{n-1}{p}} \neq 1 \mod (n)$$

pour tout diviseur premier p de $n-1$.

Un α qui vérifie les conditions ci-dessus est alors nécessairement primitif modulo n.

Si on connaît la décomposition en facteur premier de $n-1$, i.e., si on sait que $n = 1 + p_1^{\nu_1} ... p_k^{\nu_k}$, il est facile d'avoir un test qui garantie la primalité de n : on choisit un α au hasard et on calcule

$$\alpha^{n-1} \mod (n), \quad \alpha^{\frac{n-1}{p_1}} \mod (n) \quad .... \quad \alpha^{\frac{n-1}{p_k}} \mod (n)$$

- Si le test est positif on est sûr que n est premier,
- s'il est négatif alors on change de α.
- Si au bout de plusieurs essais avec des α différents les tests sont toujours négatifs, on a de fortes craintes que n ne soit pas premier et on change de n en jouant sur les exposants ν_i.

- Obtenir un grand nombre premier p avec un élément primitif α.
- On construit récursivement des nombres premiers de plus en plus grands par cette méthode en posant $n = 1 + p_1^{\nu_1} ... p_k^{\nu_k}$ où l'on sait que les p_i sont premiers.
- Une fois que l'on a trouvé des exposants ν_i tels que n soit premier alors on rajoute n dans la liste des p_i et on continue l'opération avec $k+1$ nombres premiers cette fois.

Si l'on souhaite utiliser une exponentielle modulaire, il faut disposer d'un nombre premier p et d'un élément primitif α modulo-p. Comme p et α sont publics, la fabrication par le théorème de Lucas est possible puisqu'elle donne en même temps de grands nombres premiers avec éléments primitifs. Il faut cependant veiller à ce que $n-1$ ait un diviseur premier grand de façon à rendre inefficace la méthode de Pohlig-Hellman pour calculer le logarithme.