TEN 2008. Il metodo ρ di Pollard.

Il metodo ρ di Pollard è un algoritmo di fattorizzazione probabilistico, efficace nella individuazione di fattori "piccoli". Questo metodo è basato sul paradosso del compleanno.

Il paradosso del compleanno.

Sia S l'insieme dei giorni dell'anno. Dunque $N = \#S = 365$. Supponiamo di estrarre k individui a caso da un gruppo di persone. La probabilità $p(k)$ che due di essi abbiamo lo stesso compleanno (qualunque esso sia) è data da

$$p(k) = 1 - \frac{N(N-1)\dots(N-k+1)}{N^k} = 1 - (1 - \frac{1}{N}) \cdot \dots \cdot (1 - \frac{k-1}{N}) = 1 - \prod_{i=1}^{k-1}(1 - \frac{i}{N}).$$

Usando la maggiorazione $1 - x < e^{-x}$, abbiamo

$$\prod_{i=1}^{k-1}(1 - \frac{i}{N}) < \prod_{i=1}^{k-1} e^{-\frac{i}{N}} = e^{-\sum_{i=1}^{k-1}\frac{i}{N}} = e^{-k(k-1)/2N},$$

da cui segue la stima

$$p(k) > 1 - e^{-k(k-1)/730}.$$

Ad esempio, se prendiamo 10 persone a caso, la probabilità che due di esse abbiano lo stesso compleanno è

$$p(10) > 1 - e^{-9/73} \sim 0.116 \sim 10\%;$$

se ne prendiamo 23, tale probabilità supera il 50%

$$p(23) > 1 - 0.50000$$

se ne prendiamo 50 supera il 96%

$$p(50) > 1 - e^{-245/73} \sim 0.965$$

e possiamo essere "praticamente" certi che due di esse avranno lo stesso compleanno se ne prendiamo 60

$$p(60) > 1 - e^{-354/73} \sim 0.9921.$$

• In generale, si verifica facilmente che pescando a caso k volte da un insieme di N elementi, la probabilità $p(k)$ che si verifichi una coincidenza, ossia di pescare due volte lo stesso elemento (qualunque esso sia), si stima con

$$p(k) > 1 - e^{-\frac{k(k-1)}{2N}}.$$

In particolare,

$$p(k) > \frac{1}{2}, \quad \text{per } k > 1,177\sqrt{N}; \qquad p(k) > 1 - e^{-5} \sim 0.9932, \quad \text{per } k > 3,16\sqrt{N}.$$

In breve, possiamo dire che pescando all'incirca $k = \sqrt{N}$ volte da un insieme di N elementi, abbiamo una probabilità positiva di trovare una concidenza.

• Viceversa, se pescando a caso k volte da un insieme di cui non conosciamo la cardinalità troviamo una coincidenza, probabilmente la cardinalità dell'insieme era al più dell'ordine di k^2.

Il metodo ρ di Pollard.

Nel metodo di fattorizzazione ρ di Pollard il paradosso del compleanno è applicato in questi termini.

Sia $\mathbf{n}$ il numero da fattorizzare e sia $\{\bar{x}_i\}_{1\leq i\leq k}$ una successione casuale in $\mathbb{Z}_n$. Se p è un fattore di n (cioè p divide n), la successione $\{\bar{x}_i\}_{1\leq i\leq k}$ si proietta su $\mathbb{Z}_p$ e definisce una succesione casuale in $\mathbb{Z}_p$. In $\mathbb{Z}_p$ si verifica una coincidenza esattamente quando

$$\bar{x}_i \equiv \bar{x}_j \mod p \quad \Leftrightarrow \quad p \mid \bar{x}_i - \bar{x}_j \qquad \text{per qualche indice } 1 \leq i < j \leq k,$$

e questa evenienza si manifesta mediante

$$p \mid \gcd(\bar{x}_i - \bar{x}_j, n) > 1.$$

Il paradosso del compleanno dice che se la successione casuale $\{\bar{x}_i\}_{1\leq i\leq k}$ è di lunghezza k e se n ha fattori primi dell'ordine di k^2, allora probabilmente

$$\gcd(\bar{x}_i - \bar{x}_j, n) > 1.$$

Calcolando

$$\gcd(\bar{x}_i - \bar{x}_j, n), \quad \textit{al variare di } 1 \leq i, j \leq k,$$

è come se noi pescassimo contemporaneamente k elementi a caso da ognuno degli $\mathbb{Z}_p$, con p fattore primo di n, senza conoscere p. Se p è un fattore primo dell'ordine di grandezza al più k^2, per il paradosso del compleanno, dentro $\mathbb{Z}_p$ probabilmente si verificherà una coincidenza e troveremo che

$$p \mid \gcd(\bar{x}_i - \bar{x}_j, n), \quad \textit{per qualche coppia } 1 \leq i < j \leq k.$$

Sempre per il paradosso del compleanno è più probabile che tale coincidenza si verifichi prima per il fattore p più piccolo.
Viceversa, se calcolando $\gcd(\bar{x}_i - \bar{x}_j, n)$, *al variare di* $1 \leq i, j \leq k$, *troviamo sempre 1, allora verosimilmente tutti i fattori di n sono maggiori di* k^2.

La successione casuale in $\mathbb{Z}_n$.

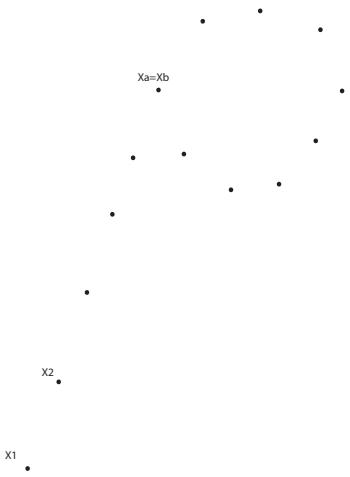
Per costruire una successione casuale in $\mathbb{Z}_n$:
- consideriamo la funzione $f\colon \mathbb{Z}_n \to \mathbb{Z}_n, \quad f(x) = x^2 + 2$;
- scegliamo a caso $\bar{x}_0 \in \mathbb{Z}_n$ e definiamo

$$\begin{aligned}
&\bar{x}_1 = f(\bar{x}_0) = \bar{x}_0^2 + 2 \mod n \\
&\bar{x}_2 = f^{(2)}(\bar{x}_0) = f \circ f(\bar{x}_0) = f(f(\bar{x}_0)) = (\bar{x}_0^2 + 2)^2 + 2 \mod n \\
&\vdots \quad \vdots \\
&\bar{x}_i = f^{(i)}(\bar{x}_0) = f \circ \ldots \circ f(\bar{x}_0))) = f(\ldots(f(\bar{x}_0))) \mod n, \qquad i > 1 \\
&\vdots \quad \vdots
\end{aligned}$$

Osservazione. Una successione di iterate a valori in un insieme finito da un certo punto in poi si ripete ciclicamente. Precisamente, $\bar{x}_a = f^{(a)}(\bar{x}_0) = f^{(b)}(\bar{x}_0) = \bar{x}_b$, allora per ogni m, compreso fra a e b, la successione è periodica di periodo $(b-a)$, cioè vale

$$\bar{x}_m = x_{m+s(b-a))}, \quad \forall s \in \mathbb{N}$$

e dunque descrive la lettera ρ che dà il nome all'algoritmo...



La ρ di Pollard.

Dim. Abbamo infatti

$$x_{m+(b-a)} = f^{(m+(b-a))}(\bar{x}_0) = f^{(m-a+b)}(\bar{x}_0) = f^{(m-a)}(f^{(b)}(\bar{x}_0)) = f^{(m-a)}(f^{(a)}(\bar{x}_0)) = f^{(m)}(x_0) = x_m$$

come richiesto.

Il prossimo Lemma contiene un'osservazione che rende molto più efficiente il metodo ρ di Pollard: se k è la lunghezza della successione delle iterate $\{\bar{x}_j\}_j$, invece che calcolare k^2 massimi comuni divisori, è sufficiente calcolarne al più k.

Lemma. *(Floyd's trick) Se esistono indici $1 \leq a, b \leq k$ tali che*

$$p \mid \gcd(\bar{x}_a - \bar{x}_b, n) \quad \Leftrightarrow \quad \bar{x}_a \equiv \bar{x}_b \mod p,$$

allora esiste anche un indice $a \leq m \leq b$ tale che

$$\gcd(\bar{x}_m - \bar{x}_{2m}, n) = p > 1 \quad \Leftrightarrow \quad x_m \equiv x_{2m} \mod p.$$

Dim. Fissato $m \geq a$, cerchiamo $s \in \mathbb{N}$ per cui valga

$$m + s(b-a) = 2m \quad \Leftrightarrow \quad m = s(b-a).$$

Volendo $m \geq a$, è evidente che basta prendere ad esempio $s = [\frac{a}{(b-a)}]+1$. Dunque per $m = ([\frac{a}{(b-a)}]+1)(b-a)$ vale

$$\bar{x}_m \equiv \bar{x}_{2m} \mod p.$$

Notiamo che vale anche $m = ([\frac{a}{(b-a)}] + 1)(b-a) \leq (\frac{a}{(b-a)} + 1)(b-a) = b$.

L'algoritmo.

- Scegliamo una funzione $f\colon \mathbb{Z}_n \to \mathbb{Z}_n$, ad esempio $f(x) = x^2 + 2 \mod n$;
- scegliamo a caso $\bar{x}_0 \in \mathbb{Z}_n$;
- calcoliamo due successioni in $\mathbb{Z}_n$ considerando rispettivamente le iterate di f e di $f^{(2)}$ in $\bar{x}_0$

$$\begin{array}{ll} \bar{x}_1 = f(\bar{x}_0) = \bar{x}_0^2 + 2 \mod n & \bar{y}_1 = \bar{x}_2 = f^{(2)}(\bar{x}_0) \mod n \\ \bar{x}_2 = f^{(2)}(\bar{x}_0) \mod n & \bar{y}_2 = \bar{x}_4 = f^{(4)}(\bar{x}_0) \mod n \\ \vdots \quad \vdots & \vdots \quad \vdots \\ \bar{x}_i = f^{(i)}(\bar{x}_0) \mod n, \qquad i > 1 & \bar{y}_i = \bar{x}_{2i} = f^{(2i)}(\bar{x}_0) \mod n, \qquad i > 1 \\ \vdots \quad \vdots & \vdots \quad \vdots \end{array}$$

- calcoliamo

$$\gcd(\bar{x}_i - \bar{y}_i, n), \qquad \text{al variare di} \quad 1 \leq i \leq k.$$

- Se per qualche i troviamo $\gcd(\bar{x}_i - \bar{y}_i, n) > 1$ abbiamo un fattore di n.

• Qual è la complessità probabilistica di questo algoritmo?

Se n è il numero da fattorizzare e p è il più piccolo fattore di n, probabilmente lo troveremo dopo aver calcolato

$$\gcd(\bar{x}_i - \bar{x}_{2i}, n), \qquad \text{al variare di} \quad 1 \leq i \leq \sqrt{p}.$$

Ogni passo richiede tre quadrature e una sottrazione modulo n per ottenere

$$\bar{x}_i - \bar{x}_{2i}$$

e successivamente un massimo comune divisore

$$\gcd(\bar{x}_i - \bar{x}_{2i}, n).$$

In totale ha complessità di un passo è data da

$$\mathcal{O}(3\log^2 n + \log n) + \mathcal{O}(\log^3 n) \sim \mathcal{O}(\log^3 n),$$

per cui la complessità di $\sqrt{p}$ passi è dell'ordine di

$$\mathcal{O}(\sqrt{p} \cdot \log^3 n).$$

Nella peggiore delle ipotesi, n ha un fattore p dell'ordine di $\sqrt{n}$. In questo caso la complessità dell'algoritmo è dell'ordine di

$$\mathcal{O}(n^{1/4} \cdot \log^3 n).$$