This is the last preprint. The final paper appeared in "Matematica e Cultura 2000", Venice 1999, ed. by M.Emmer, Springer Verlag Italia, 2000, 227-233.

NUMERI PRIMI E CRITTOGRAFIA

A. LANGUASCO A. PERELLI

In questo intervento tratteggeremo dapprima alcuni elementi dello sviluppo della teoria dei numeri, ponendo particolare attenzione al problema della distribuzione dei numeri primi ed argomenti collegati. Descriveremo poi un'applicazione al problema della sicurezza della trasmissione dei dati, ossia alla crittografia.

Numeri primi

Ricordiamo per prima cosa che un intero $n > 1$ si dice *primo* se è divisibile solamente per 1 e per se stesso. Ad esempio, i numeri 2, 3, 5, 7, 11, 13, 17 e 19 sono primi.

Già gli antichi Greci si interessarono alle proprietà dei numeri primi. Una tecnica sviluppata in quel periodo è il ben noto *crivello di Eratostene*, che consente di calcolare tutti i numeri primi tra 2 e x, dove x è un qualunque numero reale positivo fissato. Essenzialmente, dopo aver scritto tutti gli interi compresi tra 2 ed x, si fissa il numero 2 e si cancellano tutti i suoi multipli; si considera poi il primo numero più grande di 2 che non è stato cancellato (cioè il 3) e si cancellano tutti i suoi multipli. Si procede in tal modo (al passo successivo si considera il 5 e si cancellano tutti i suoi multipli) fino al massimo intero più piccolo di $\sqrt{x}$. Poiché al termine di tale procedura abbiamo cancellato tutti gli interi che hanno divisori propri più piccoli di $\sqrt{x}$, gli interi non cancellati sono tutti e soli i primi nell'intervallo $[2, x]$.

Un altro problema a cui si interessarono i Greci è: *i numeri primi sono infiniti* ? La risposta è affermativa, e sono conosciute dimostrazioni di varia natura; presentiamo qui quella di Euclide, di natura *aritmetica*.

Teorema. (Euclide) *Esistono infiniti numeri primi.*

Dimostrazione. Supponiamo per assurdo che esista solamente un numero finito di numeri primi, siano essi $p_1 < p_2 < \ldots < p_k$. Consideriamo il numero

$$N = p_1 p_2 \cdots p_k + 1$$

che, chiaramente, non può essere primo in quanto $N > p_k$; d'altronde N non è divisibile per alcun p_j e quindi N è primo, assurdo. ▪

A questo punto ci si può chiedere: *perché i numeri primi sono interessanti* ? La risposta più immediata a tale domanda è fornita dal fatto che, in un certo senso, i numeri primi sono i "mattoni" fondamentali con cui costruire tutti gli altri interi. Formalmente tale affermazione si esprime mediante il ben noto

Teorema Fondamentale dell'Aritmetica. *Ogni intero positivo si fattorizza in modo unico come prodotto di numeri primi.*

Osservazioni. (1) La dimostrazione del Teorema Fondamentale dell'Aritmetica è essenzialmente divisa in due parti:
a) esistenza della fattorizzazione (diretta conseguenza della definizione di numero primo)
b) unicità della fattorizzazione (semplice ma non del tutto banale).
Ricordiamo, per quanto riguarda b), l'esempio di *Hilbert* che mostra come si possano costruire semplici "sistemi numerici" in cui *non vale la fattorizzazione unica*. Consideriamo gli interi della forma $4k+1$, $k = 0, 1, \ldots$, che costituiscono un sistema chiuso rispetto alla moltiplicazione. É facile verificare che

$$693 = 9 \cdot 77 = 21 \cdot 33$$

fornisce due distinte fattorizzazioni come prodotto di "primi" in tale sistema; infatti 9, 77, 21 e 33 non ammettono una fattorizzazione non banale come prodotto di interi della forma $4k+1$.

(2) Dal Teorema Fondamentale dell'Aritmetica segue facilmente il

Corollario. $\sqrt{2}$ *è irrazionale.*

Dimostrazione. Ancora per assurdo: supponiamo che $\sqrt{2} = \frac{m}{n}$. Allora $n\sqrt{2} = m$, da cui

$$2n^2 = m^2.$$

L'assurdo si ottiene osservando che il fattore 2 ha esponente dispari nel termine di sinistra dell'ultima equazione, mentre ha esponente pari nel termine di destra, in contraddizione con il Teorema Fondamentale dell'Aritmetica. ▪

Grosso modo, possiamo suddividere le *problematiche* relative ai numeri primi in due tipologie principali:
- di tipo *algebrico*, riguardanti principalmente il comportamento dei primi nelle estensioni algebriche dei numeri razionali;
- di tipo *analitico*, riguardanti principalmente la distribuzione dei primi tra i numeri interi.

Nel seguito tratteremo solamente le problematiche di tipo analitico.

È naturale chiedersi: *quanti sono i numeri primi* ? Sappiamo già che sono infiniti, ma quello che ci chiediamo è quale sia l'*ordine di grandezza* della quantità

$$\pi(x) = \text{ numero dei primi tra 1 e } x.$$

Il primo tentativo di risolvere tale problema fu fatto da *Gauss* verso la fine del '700. Basandosi sulle tavole di numeri primi da lui stesso calcolate, Gauss *congetturò* che l'andamento asintotico di $\pi(x)$ dovesse essere

$$\frac{\pi(x)}{x/\log x} \to 1 \quad \text{per } x \to \infty.$$

Come vedremo in seguito, la congettura di Gauss si rivelò esatta, ed è oggi nota come *Teorema dei Numeri Primi* (brevemente *TNP*).

I primi risultati nella direzione della congettura di Gauss furono provati da *Chebyshev* verso la metà del 1800.

Teorema. (Chebyshev) *Esistono due costanti* $0 < c < 1 < C$ *tali che per* x *sufficientemente grande*

$$c\frac{x}{\log x} \leq \pi(x) \leq C\frac{x}{\log x}.$$

La dimostrazione del teorema di Chebyshev si basa su una tecnica elementare ma ingegnosa che coinvolge alcune proprietà dei coefficienti binomiali.

Il passo decisivo nella direzione del *TNP* fu fatto da *Riemann*, pochi anni dopo i risultati di Chebyshev, ponendo le basi per la dimostrazione del *TNP*. La novità fondamentale del metodo di Riemann fu quella di studiare la funzione $\pi(x)$ con metodi di *analisi complessa* (da cui l'aggettivo "analitico" che prendono le ricerche di questo tipo).

Riemann introdusse la funzione della variabile *complessa* s

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \qquad \Re(s) > 1,$$

oggi nota come *funzione zeta di Riemann*. La funzione zeta di Riemann è collegata ai numeri primi per mezzo dell'*identità di Eulero*

$$\zeta(s) = \prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1} \qquad \Re(s) > 1,$$

dove il prodotto è esteso a tutti i numeri primi; l'identità di Eulero è una semplice conseguenza del Teorema Fondamentale dell'Aritmetica, ed anzi è considerata l'*equivalente analitico* della fattorizzazione unica degli interi.

Il punto cruciale dell'identità di Eulero è che al lato destro appaiono esplicitamente i numeri primi, mentre il lato sinistro è definito in modo indipendente da essi. Il metodo di Riemann apre quindi la strada alla possibilità di ottenere informazioni sui numeri primi mediante lo studio delle *proprietà analitiche* della funzione $\zeta(s)$. Ad esempio, sfruttando il fatto che

$$\lim_{s \to 1^+} \zeta(s) = +\infty$$

si può ottenere facilmente una dimostrazione di natura *analitica* del teorema di Euclide sull'infinità dei numeri primi.

Osserviamo che tale dimostrazione analitica è dovuta a Eulero, che già considerò $\zeta(s)$ come funzione della variabile *reale* s. Riemann mostrò che la funzione $\zeta(s)$ è prolungabile su tutto il piano complesso, e che la distribuzione dei numeri primi è strettamente collegata alla *distribuzione degli zeri* della funzione $\zeta(s)$; tale collegamento ha dato origine ad alcuni tra i più profondi problemi della matematica.

Il *TNP* fu dimostrato indipendentemente da *Hadamard* e *de la Vallée Poussin* nel 1896. Tale dimostrazione, basata sul metodo di Riemann, rappresenta il culmine di una serie di ricerche sulla teoria delle funzioni di una variabile complessa, condotte prevalentemente da Hadamard. Il punto cruciale della dimostrazione consiste nel mostrare che $\zeta(1+it) \neq 0$ per ogni numero reale t; è oggi noto che il non annullamento della funzione zeta di Riemann sulla retta $\Re s = 1$ è in realtà *equivalente* al *TNP*.

Per quasi tutta la prima metà del '900 si riteneva (soprattutto a causa all'influenza dei matematici inglesi *Hardy* e *Littlewood*, cui sono dovuti contributi fondamentali alla teoria analitica dei numeri) che non fosse possibile ottenere una dimostrazione del *TNP* senza far uso di tecniche di analisi complessa. Tale convinzione si rivelò però errata quando, alla metà del '900, *Selberg* e *Erdös* diedero una *dimostrazione elementare* del *TNP*, ovvero una dimostrazione basata su tecniche essenzialmente aritmetiche. Precisiamo comunque che "elementare" non sta in questo caso a significare "facile"; infatti la dimostrazione di Selberg-Erdös è concettualmente più complicata di quella analitica.

Una volta noto il *TNP*, il passo successivo fu quello di capire "quanto buona" fosse l'*approssimazione* di $\pi(x)$ mediante la funzione $\frac{x}{\log x}$ o, più precisamente, mediante la funzione *logaritmo integrale*

$$li(x) = \int_2^x \frac{dt}{\log t}$$

(di cui $\frac{x}{\log x}$ è il primo termine dello sviluppo asintotico). Attualmente non è nota una risposta definitiva a tale problema; ricordiamo però che la famosa *Ipotesi di Riemann*

$$\zeta(s) \neq 0 \ \text{ per } \ \Re s > \tfrac{1}{2}$$

gioca un ruolo fondamentale in questo ambito. Si può infatti provare che l'Ipotesi di Riemann è *equivalente* all'approssimazione

$$\pi(x) = li(x) + O(\sqrt{x}\log x)$$

ossia, essenzialmente, al fatto che l'errore che si commette approssimando $\pi(x)$ con $li(x)$ è, in valore assoluto, più piccolo di $\sqrt{x}\log x$. Osserviamo infine che tale approssimazione, se vera, è ottimale, che sono noti svariati *argomenti euristici* a favore dell'Ipotesi di Riemann e che computazioni su larga scala ne confermano la validità.

Al momento attuale molti risultati sono stati provati, ma molti altri rimangono *problemi aperti* per la ricerca sulla distribuzione dei numeri primi; oltre all'Ipotesi di Riemann, tra questi vogliamo ricordare alcuni *problemi classici*:

1) (*primi rappresentati dai polinomi*) esistono infiniti interi n per cui n^2+1 è un numero primo ? (o, più in generale, $P(n)$ è un numero primo per infiniti n, con $P(x)$ polinomio irriducibile senza divisori fissi ?)

2) (*distanza tra numeri primi consecutivi*) esiste sempre un numero primo tra due quadrati perfetti consecutivi ?

3) (*primi gemelli*) esistono infiniti numeri primi p tali che $p+2$ è ancora primo ?

4) (*congettura di Goldbach*) ogni intero pari maggiore di 2 può essere scritto come somma di due numeri primi ?

Concludiamo il paragrafo osservando che in tali problemi insorgono difficoltà di varia natura; ad esempio, la difficoltà fondamentale dei problemi 3) e 4) risiede nel fatto che i numeri primi sono definiti mediante proprietà *moltiplicative*, mentre i problemi in questione coinvolgono proprietà *additive*.

Crittografia

Con il termine *crittografia* intendiamo lo studio dei metodi che consentono la trasmissione *sicura* dell'informazione. Usualmente si distingue tra due diversi tipi di crittografia:

a) *a chiave segreta*: è il metodo classico (usato già dagli antichi Romani) ed è utile solamente nel caso vi siano pochi utenti poiché, per funzionare, è necessario che ogni utente preventivamente concordi e scambi la propria chiave segreta con ogni altro utente;

b) *a chiave pubblica*: è il metodo moderno e consente una trasmissione sicura anche nel caso di molti utenti poiché non necessita di uno scambio preventivo delle chiavi segrete. È stato proposto per la prima volta da *Diffie* e *Hellman* nel 1976.

A prima vista la crittografia a chiave pubblica sembra impossibile. Per convincersi del contrario proponiamo l'esempio classico del *doppio lucchetto*. Supponiamo di avere due utenti A e B e che A voglia spedire un messaggio segreto a B:

1) A mette il messaggio in una scatola che chiude con il suo lucchetto L_A (di cui lui solo ha la chiave) e che poi spedisce a B;

2) B riceve la scatola chiusa con L_A, aggiunge il suo lucchetto L_B (di cui lui solo ha la chiave) e rispedisce il tutto ad A;

3) A, ricevuta la scatola con il doppio lucchetto, toglie il lucchetto L_A e rispedisce la scatola a B;

4) a questo punto, ricevuta la scatola, B può togliere il lucchetto L_B e leggere il messaggio di A.

La sicurezza di questo schema risiede nel fatto che le chiavi per aprire i due lucchetti sono

conosciute solamente ai rispettivi proprietari, che non le hanno preventivamente concordate e scambiate.

Una delle "versioni matematiche" di tale idea è il *metodo crittografico a chiave pubblica R.S.A.*, proposto da *Rivest, Shamir* ed *Adleman* nel 1978. Vediamo schematicamente come A può mandare un messaggio segreto a B usando il metodo R.S.A.:

B sceglie in modo *casuale*

- due primi p, q *grandi* (di 200-300 cifre in base 10), e *calcola* $N = pq$ e $\varphi(N) = (p-1)(q-1)$

- un intero e *coprimo* con $\varphi(N)$ tale che $e < \varphi(N)$, e *calcola* l'intero $d < \varphi(N)$ tale che $de \equiv 1 \pmod{\varphi(N)}$

e poi *rende pubblici* i numeri N ed e.

A, per mandare un messaggio a B, compie le seguenti operazioni:

1) codifica il messaggio in un modo standard usando i numeri $\leq N$;

2) spedisce a B ogni numero M di tale codifica sotto forma di

$$M^e \pmod{N}.$$

Per *decodificare* il messaggio, B calcola semplicemente

$$(M^e)^d \pmod{N};$$

quello che ottiene è proprio M grazie al teorema di *Fermat-Eulero* che, in *questa situazione*, afferma che $M^{ed} \equiv M \pmod{N}$.

Il punto fondamentale è ora: *in cosa consiste la sicurezza del metodo* ? Da quanto appena visto, per decodificare il messaggio è *necessario* conoscere d; noto e, per calcolare d è *necessario* conoscere $\varphi(N)$; ma, noto N, *calcolare* $\varphi(N)$ *è computazionalmente equivalente a fattorizzare* N.

Quindi, in definitiva, la *sicurezza del metodo R.S.A.* dipende dai seguenti fatti:

- per *codificare* il messaggio bisogna saper costruire dei *numeri primi grandi*, e tale operazione è *computazionalmente "veloce"*; si può infatti dimostrare che la complessità computazionale di opportuni *test di primalità* per stabilire se un numero n è primo è della forma

$$(\log n)^{c \log\log\log n},$$

ovvero è "quasi-polinomiale" in $\log n$ (si noti che $\log n$ è, essenzialmente, il numero di cifre di n)

- per *violare* il sistema bisogna saper fattorizzare *interi grandi* ottenuti come prodotto di due primi; tale problema è *computazionalmente "lento"*; si *congettura* infatti che la complessità computazionale sia della forma

$$e^{c \sqrt[3]{\log n (\log\log n)^2}},$$

ovvero è "sub-esponenziale" in $\log n$.

Tale marcata differenza tra la velocità di esecuzione delle operazioni di costruzione di numeri primi grandi e di fattorizzazione di interi grandi garantisce la sicurezza del metodo, almeno per un tempo sufficientemente lungo.

Ad esempio, con la tecnologia attuale l'operazione di fattorizzazione di un intero di 140 cifre in base 10, prodotto di due primi casuali calcolati in pochi secondi su un computer disponibile in commercio, richiede, utilizzando vari supercomputers operanti parallelamente, circa un mese ! Incrementando il numero di cifre si aumenta la sicurezza del sistema; attualmente si raccomanda di utilizzare interi con almeno 220 cifre in base 10.

Riferimenti bibliografici

Consigliamo i testi classici di Ingham [4] e Davenport [2] per un'esposizione chiara dei risultati fondamentali sulla distribuzione dei numeri primi. Segnaliamo inoltre l'eccellente introduzione alla teoria elementare dei numeri di Davenport [3] (in lingua italiana), che contiene anche un capitolo sulla crittografia.

Per maggiori dettagli sugli aspetti storici dello sviluppo della crittografia si consiglia il libro di Kahn [5] mentre, per una modellizzazione matematica della crittografia a chiave pubblica più rigorosa di quella qui esposta, i testi di Koblitz [6] e [7] forniscono una esauriente presentazione. Per quanto concerne gli algoritmi di fattorizzazione e i test di primalità, si consigliano i libri di Koblitz [6], di Cohen [1] e di Riesel [8].

[1] H. Cohen - *A Course in Computational Algebraic Number Theory* - Springer 1994.

[2] H. Davenport - *Multiplicative Number Theory* - Springer 1981.

[3] H. Davenport - *Aritmetica Superiore* - Zanichelli 1994.

[4] A. E. Ingham - *The Distribution of Prime Numbers* - Cambridge U. P. 1932.

[5] D. Kahn - *The Codebreakers, the Story of Secret Writing* - Macmillan 1967.

[6] N. Koblitz - *A Course in Number Theory and Cryptography* - Springer 1987.

[7] N. Koblitz - *Algebraic Aspects of Cryptography* - Springer 1998.

[8] H. Riesel - *Prime Numbers and Computer Method for factorization* - Birkhäuser 1994.

Alessandro Languasco
Dipartimento di Matematica Pura e Applicata
Università di Padova
Via Belzoni 7
35131 Padova, Italy
e-mail: languasco@math.unipd.it

Alberto Perelli
Dipartimento di Matematica
Università di Genova
Via Dodecaneso 35
16146 Genova, Italy
e-mail: perelli@dima.unige.it