Matematica, Cultura e Società

Rivista dell'Unione Matematica Italiana

Alessandro Zaccagnini

Macchine che producono numeri primi

Matematica, Cultura e Società. Rivista dell'Unione Matematica Italiana, Serie 1, Vol. **1** (2016), n.1, p. 5–20.

Unione Matematica Italiana

<http://www.bdim.eu/item?id=RUMI_2016_1_1_1_5_0>

Matematica, Cultura e Società. Rivista dell'Unione Matematica Italiana, Unione Matematica Italiana, 2016.

Matematica, Cultura e Società – Rivista dell'Unione Matematica Italiana
Serie I, Vol. 1, N. 1, Aprile 2016, 5-20

Macchine che producono numeri primi

ALESSANDRO ZACCAGNINI
Università di Parma
E-mail: alessandro.zaccagnini@unipr.it

Sommario: *In questo articolo ci occupiamo della possibilità di produrre "meccanicamente" i numeri primi. In particolare, trattiamo criticamente il Crivello di Eratostene, la macchina di Conway e la formula di Gandhi, che permettono di determinare tutti i numeri primi.*

Abstract: *The present paper is devoted to the study of "mechanical" means to produce prime numbers. In particular, we deal critically with the Eratosthenes sieve, Conway's machine and Gandhi's formula, all of which yield the sequence of all prime numbers.*

1. – Metodi meccanici per produrre numeri primi

In questo articolo analizziamo il problema di generare "meccanicamente" i numeri primi. Non è più una curiosità meramente matematica, ma è diventato un problema concreto con applicazioni alla crittografia, e dunque alla sicurezza e alla riservatezza delle comunicazioni digitali. Il punto di partenza sarà la descrizione del crivello di Eratostene, per poi passare alla "macchina di Conway" che produce la lista di tutti i numeri primi in ordine crescente. Vedremo la formula di Gandhi che, in linea di principio, permette di calcolare un numero primo conoscendo tutti i precedenti; parleremo di polinomi che assumono valori primi e concluderemo con una versione "moderna" e quantitativa del Crivello di Eratostene, e cioè la formula di Legendre. Fino al XVIII secolo le formule per i numeri primi erano considerate un po' come la pietra filosofale degli antichi alchimisti. È solo con Gauss e i suoi contemporanei che si diffonde l'idea di studiare la distribuzione dei numeri primi dal punto di vista macroscopico, ignorando il dettaglio, come può essere una formula per l'n-esimo numero primo. Questo è stato l'inizio degli studi più proficui, cominciati intorno alla fine del XVIII secolo e non ancora compiuti.

Pur non essendo uno degli obiettivi principali di questo articolo, vogliamo enfatizzare il ruolo dei numeri primi come mattoni moltiplicativi degli interi: infatti, la struttura moltiplicativa è molto più complicata e interessante di quella additiva. I numeri primi compaiono esattamente 2 volte nella "tavola pitagorica" infinita, e questa è la loro più semplice caratterizzazione, e spiega perché si tende a non considerare 1 come un numero primo, dal momento che compare una volta sola. Piú seriamente, il numero 1 ha un reciproco nei numeri naturali, e quindi il concetto di primalità non ha molto significato in questo caso.

2. – Il crivello di Eratostene

Tutti sanno che il matematico alessandrino Eratostene trovò un procedimento, detto crivello o setaccio, per determinare i numeri primi nell'intervallo tra 2 ed un certo limite N. Ricordiamo brevemente come funziona: si scrivono in ordine gli interi da 2 ad N (in termini moderni, si allocano $N-1$ "bit" di memoria) e si eliminano tutti i multipli di 2 a partire da $2^2 = 4$. Poi si ricomincia, cercando il più piccolo numero non ancora cancellato, e cioè 3, ed eliminando i suoi multipli a partire

Accettato: l'1 dicembre 2015.

da $3^2 = 9$. Si va avanti così, eliminando successivamente i multipli di 5 a partire da $5^2 = 25$, di 7 a partire da $7^2 = 49$, ... Ci si può fermare quando il più piccolo numero da cancellare è maggiore di N, cioè quando si raggiunge un numero primo che supera $N^{1/2}$. Ciò che resta nell'elenco sono i numeri primi fino ad N. Tra i pregi del Crivello, è importante notare la relativa velocità di esecuzione, una necessità di memoria non troppo esosa e, come vedremo sotto, una certa flessibilità e adattabilità anche ad altri contesti oltre all'estrema semplicità della programmazione.

Un significativo risultato della seconda metà del XIX secolo, dovuto a F. Mertens, afferma che

$$\sum_{p \le N} \frac{1}{p} \sim \log(\log(N)),$$

dove la notazione implica che la somma è fatta solo sui numeri primi. Da questo è facile dedurre che il numero di operazioni elementari (quali addizioni, moltiplicazioni, accessi o scritture in memoria) necessarie per eseguire il Crivello di Eratostene sui primi N numeri interi non supera approssimativamente $N \log(\log(N))$. In altre parole, ogni intero in $[2, N]$ richiede, in media, circa $\log(\log(N))$ operazioni: se vogliamo, il "costo unitario" del Crivello è molto basso.

Naturalmente non è sensato cercare di utilizzare il Crivello, così com'è, per determinare numeri primi "grandi" come quelli necessari al giorno d'oggi per le applicazioni crittografiche, con qualche centinaio di cifre decimali. Allo stesso tempo, non vogliamo rinunciare ad usare uno strumento così efficiente: mostriamo dunque che il Crivello è anche flessibile e lo si può usare efficacemente anche per altri scopi. Si congettura che vi sia almeno un numero primo in ogni intervallo del tipo $[N, N + C(\log(N))^2]$, dove $C > 0$ è una costante sufficientemente grande. Se dunque vogliamo trovare almeno un numero primo nell'intervallo $I = [N, N + M]$ dove M è molto più piccolo di N, possiamo usare il crivello per eliminare dall'intervallo I tutti i multipli dei numeri primi $p \le L$, dove L è un parametro arbitrario, che conviene scegliere non troppo grande. Gli interi residui non sono necessariamente primi poiché è sensato scegliere L molto più piccolo di $N^{1/2}$. È possibile stimare il numero di interi sopravvissuti a questa eliminazione, cioè gli interi che appartengono all'intervallo $[N, N + M]$ e che non hanno fattori primi $\le L$. Ci serve una variante del crivello e un'altra formula di Mertens: se $L \to +\infty$ allora

$$\prod_{p \le L} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log(L)},$$

dove $\gamma \approx 0.577216...$ è la costante di Eulero. In questo modo si eliminano tutti i numeri che hanno fattori primi molto piccoli in modo molto efficiente. Questi, evidentemente, non hanno alcuna speranza di essere primi. Poi si sottopongono i numeri rimasti, che sono relativamente pochi, ad un criterio di primalità, che è più oneroso dal punto di vista computazionale. Oggi esistono criteri di primalità o di pseudo-primalità molto efficienti: fra i più noti citiamo quello di Miller & Rabin, e naturalmente quello di Agrawal, Kayal & Saxena, che nel 2002 ha messo la parola fine all'annosa questione della esistenza di un algoritmo polinomiale per la primalità. In altre parole, oggi sappiamo che è possibile dimostrare rigorosamente che un certo intero n è primo eseguendo non più di $C(\log_2 n)^6$ operazioni elementari, dove C è una certa costante positiva; la funzione $C(\log_2 n)^6$ è un polinomio nella variabile $\log_2 n$ che è, approssimativamente, il numero di cifre binarie di n e dunque la misura della grandezza dell'*input* dell'algoritmo.

Per fare un esempio concreto, prendiamo $N \approx 10^{100}$ ed $M \approx 60\,000 > (\log(N))^2$, che sono ordini di grandezza ragionevoli per le applicazioni crittografiche, ed $L = 10^8$. In questo caso specifico, si può stimare il numero di interi residui in $[N, N + M]$ con $\approx Me^{-\gamma}/\log(L) \approx 1829$.

Alla fine del XVIII secolo, il matematico francese A. M. Legendre osservò che è possibile ottenere informazioni teoriche sul *numero* dei numeri primi che non superano N utilizzando il crivello, e cioè contando accuratamente quanti interi sono eliminati ad ogni iterazione: ne parliamo in dettaglio nel §7.

3. – Criteri di primalità

Resta aperta la domanda: come fare a decidere se un dato intero n è primo oppure no? Una delle prime risposte è data dal Teorema di Wilson, che fornisce una condizione *necessaria e sufficiente*.

TEOREMA 3.1 [Wilson]. – *Il numero intero $n \ge 2$ è un numero primo se e solo se $(n-1)! \equiv -1 \bmod n$.*

Diamo solo l'idea della dimostrazione: se $n \geq 6$ è composto, allora non è difficile vedere che $(n-2)! \equiv 0 \bmod n$; bisogna scrivere $n = ab$ con $1 < a \leq b < n$ e distinguere il caso speciale in cui $n = p^2$ con $p \geq 3$, in cui saremmo costretti a prendere $a = b$ e si deve procedere diversamente; da questo segue immediatamente che $(n-1)! \equiv 0 \bmod n$. Il caso interessante del Teorema di Wilson è ovviamente quello in cui n è un numero primo p: se $a \in \{2, \ldots, p-2\}$ allora a è invertibile modulo p ed inoltre $a^{-1} \not\equiv a \bmod p$. Quest'ultima affermazione dipende dal fatto che l'equazione $x^2 \equiv 1 \bmod p$ ha solo due soluzioni $\bmod p$, che sono 1 e $p-1$: è qui che si sfrutta l'ipotesi di primalità, la quale implica, in definitiva, che $\mathbb{Z}/p\mathbb{Z}$ è un campo. Dunque $(p-2)! \equiv 1 \bmod p$ e il risultato voluto segue, moltiplicando ambo i membri di quest'ultima congruenza per $p-1$.

Per esempio, quando $p = 17$ abbiamo

$$\begin{aligned} 15! &= 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 \cdot 11 \cdot 12 \cdot 13 \cdot 14 \cdot 15 \\ &= 1 \cdot (2 \cdot 9) \cdot (3 \cdot 6) \cdot (4 \cdot 13) \cdot (5 \cdot 7) \cdot (8 \cdot 15) \cdot (10 \cdot 12) \cdot (11 \cdot 14) \\ &\equiv 1 \bmod 17, \end{aligned}$$

poiché abbiamo associato ogni elemento in $\{2, \ldots, 15\}$ con il suo reciproco modulo 17 e, come già detto, tutti questi elementi sono distinti dal proprio reciproco.

Un esame anche superficiale di questo risultato mostra che si tratta, in realtà, di una caratterizzazione essenzialmente inutilizzabile in pratica. Infatti, per dimostrare che 101 è un numero primo è necessario calcolare *esattamente* $100! \approx 9.33 \cdot 10^{157}$: anche utilizzando qualche scorciatoia (in fondo, ci serve solo il risultato modulo 101 e quindi possiamo calcolare 100! iterativamente, inserendo dopo ogni moltiplicazione la riduzione modulo 101) il numero di moltiplicazioni necessarie è, in generale, dell'ordine del numero n di cui ci si chiede se sia primo o composto. Inoltre, non è evidentemente possibile utilizzare formule approssimate per $n!$, per quanto precise, come quella dovuta a Stirling. Dunque, questo procedimento risulta di gran lunga più oneroso dell'algoritmo "ingenuo" detto della divisione per tentativi: per verificare che n è primo, è infatti sufficiente controllare che non abbia divisori nell'intervallo $[2, n^{1/2}]$: nel caso in questione bastano 9 divisioni invece delle 99 moltiplicazioni di cui sopra. Il numero di divisioni necessarie può essere ulteriormente ridotto avendo a disposizione l'elenco di numeri primi fino ad $n^{1/2}$, o utilizzando semplici accortezze di buon senso. Qui, a differenza del Crivello di Eratostene, in cui cerchiamo *tutti* i numeri primi in un certo intervallo, vogliamo determinare gli eventuali fattori primi di uno *specifico* intero n.

Concludiamo enunciando esplicitamente una conseguenza dell'osservazione fatta qui sopra a proposito del valore di $(n-2)! \bmod n$: questa permette di affermare che

$$f(n) = n\left\{\frac{(n-2)!}{n}\right\}$$

per $n \geq 5$ è la *funzione caratteristica* dei numeri primi, e cioè vale 1 se n è primo e 0 altrimenti. Qui $\{x\} \in [0,1)$ indica la *parte frazionaria* del numero reale x, e cioè $x - \lfloor x \rfloor$, mentre $\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\} \in \mathbb{Z}$ indica la *parte intera*. Questa funzione f può essere usata, per esempio, per scrivere una formula per contare i numeri primi fino ad un certo numero reale x, come faremo di nuovo, più seriamente, nel §7. Per la precisione, posto $\pi(x) = |\{n \leq x : n \text{ è primo}\}|$, per $x \geq 5$ si ha

$$\pi(x) = 2 + \sum_{5 \leq n \leq x} f(n).$$

Questa formula soffre degli stessi difetti del Teorema di Wilson; la enunciamo solamente per mettere in guardia i lettori: ciò che è possibile in linea di principio (scrivere una formula per trovare o per contare i numeri primi) non è necessariamente utile o efficiente. La limitazione $x \geq 5$ è resa necessaria dal fatto che $f(4) = 2$, mentre per tutti i numeri composti $n \geq 6$ abbiamo $f(n) = 0$, come già osservato.

4. – La macchina di Conway

Torniamo alla domanda iniziale: esistono metodi meccanici per produrre numeri primi? Anche interpretando alla lettera questa richiesta, la risposta è positiva. La macchina di Conway consiste in quattordici frazioni:

$\frac{17}{91}$	$\frac{78}{85}$	$\frac{19}{51}$	$\frac{23}{38}$	$\frac{29}{33}$	$\frac{77}{29}$	$\frac{95}{23}$	$\frac{77}{19}$	$\frac{1}{17}$	$\frac{11}{13}$	$\frac{13}{11}$	$\frac{15}{14}$	$\frac{15}{2}$	$\frac{55}{1}$
A	B	C	D	E	F	G	H	I	J	K	L	M	N

Il meccanismo di funzionamento è semplice: si parte dal numero $n_0 = 2$ come *input* e si scorre la lista delle frazioni elencate qui sopra da sinistra verso destra, fino a trovare la prima frazione che, moltiplicata per n_0 dà un numero intero. In questo caso la frazione è M. Si rimpiazza $n_0 = 2$ con $Mn_0 = 15$ (cioè si pone $n_1 = 15$) e si verifica se quest'ultimo valore è una potenza di 2. In caso affermativo, l'esponente è il più piccolo numero primo. In caso negativo, si ricomincia con la stessa regola: partendo da $n_1 = 15$, la prima frazione che ha la proprietà richiesta è l'ultima della fila, e cioè N. Si rimpiazza $n_1 = 15$ con $Nn_1 = 825$, che non è una potenza di 2, cioè si pone $n_2 = 825$. Dunque si deve iterare per la terza volta: in questo caso si trova che $825E = 725$ è intero e non è una potenza di 2, e così via. Ogni tanto compare una potenza di 2, per esempio dopo i primi 19 passi illustrati nella Figura 1. Gli esponenti di queste potenze sono precisamente i numeri primi 2, 3, 5, 7, ..., in ordine di grandezza.

$$2 \overset{\times M}{\to} 15 \overset{\times N}{\to} 825 \overset{\times E}{\to} 725 \overset{\times F}{\to} 1925 \overset{\times K}{\to} 2275 \overset{\times A}{\to} 425$$

$$\overset{\times B}{\to} 390 \overset{\times J}{\to} 330 \overset{\times E}{\to} 290 \overset{\times F}{\to} 770 \overset{\times K}{\to} 910 \overset{\times A}{\to} 170$$

$$\overset{\times B}{\to} 156 \overset{\times J}{\to} 132 \overset{\times E}{\to} 116 \overset{\times F}{\to} 308 \overset{\times K}{\to} 364 \overset{\times A}{\to} 68 \overset{\times I}{\to} 4 = 2^2$$

FIGURA 1 – La lettera sopra la freccia indica quale frazione stiamo usando: i primi diciannove passi danno 2^2 e quindi 2 è il primo numero primo.

Qui le frazioni *nascondono* una versione del procedimento di "divisione per tentativi," il più semplice algoritmo di fattorizzazione noto, descritto qui sopra. Non si tratta di un vero e proprio crivello come quello di Eratostene perché la macchina ha un numero finito di registri di memoria, come vedremo sotto, e il crivello richiede almeno un *bit* per ogni intero da controllare: piuttosto, la macchina di Conway esegue una sequenza di divisioni.

La macchina di Conway maschera un algoritmo adattabile ad un gran numero di casi simili: è una specie di macchina di Turing. Il risultato è un po' meno sorprendente se si adotta un punto di vista più astratto: i vari passaggi corrispondono a cambiamenti di stato di alcuni registri che possono assumere solo valori interi non negativi, che traducono le operazioni aritmetiche di base. Dietro le quinte, la macchina sta verificando la divisibilità di un certo intero per *tutti* gli interi precedenti: per questo motivo, in definitiva la macchina di Conway è tutt'altro che efficiente. In pratica, Conway è riuscito a "travestire" un algoritmo fatto di addizioni, moltiplicazioni e le altre operazioni elementari, ma anche la semplice divisione fra interi in realtà è eseguita con il metodo delle "sottrazioni ripetute", fatto questo che fa aumentare a dismisura il numero di iterazioni necessarie.

La tabella nella Figura 2 contiene i dati sul numero di iterazioni necessarie a calcolare i primi 9 numeri primi per dare un'idea concreta dell'efficienza (o, piuttosto, dell'inefficienza) della macchina di Conway: nell'ultima riga riportiamo anche il numero di *test* che la macchina deve effettuare. Questi dati non hanno bisogno di particolari commenti: aggiungiamo solo che per determinare $p_{100} = 541$ sono necessarie 213 898 044 iterazioni e 895 927 135 test.

Rimandiamo i lettori all'Appendice A per una spiegazione un po' più esauriente sul funzionamento della macchina di Conway vera e propria e per un esempio rudimentale.

5. – Formule per i numeri primi: la formula di Gandhi

L'equivalente matematico di una macchina è una formula. Esistono formule per i numeri primi? Precisiamo la domanda: esiste una combinazione suffi-

n	1	2	3	4	5	6	7	8	9
p_n	2	3	5	7	11	13	17	19	23
Iterazioni	19	69	280	707	2363	3876	8068	11 319	19 201
Test	129	425	1563	3735	11 674	18 811	38 010	52 854	88 134

FIGURA 2 – Il numero di iterazioni e di test della macchina di Conway.

cientemente complicata di simboli matematici con questa proprietà? Sì, ma le formule note sono tutte molto deludenti e di fatto inutilizzabili. La maggior parte di queste si basa sul Teorema di Wilson 3.1, che è il prototipo di tutte le caratterizzazioni "inutili" dei numeri primi. Qui scegliamo di descrivere una formula di natura piuttosto diversa, dovuta ad un omonimo del più noto Mahatma.

TEOREMA 5.1 [Formula di Gandhi]. – *Sia* p_n *l'n-esimo numero primo e poniamo* $P(0)=1$ *e* $P(n)=p_1\cdot p_2\cdots p_n$ *per* $n\geq 1$. *Inoltre sia*

$$S(n)=\sum_{d|n} a_d \qquad \textit{dove} \qquad a_d=\frac{\mu(d)}{2^d-1}.$$

Allora per $n\geq 0$ *si ha*

$$p_{n+1}=\left\lfloor 1-\log_2\Big(S(P(n))-\frac{1}{2}\Big)\right\rfloor.$$

Cominciamo spiegando la notazione: $\sum_{d|n}$ indica una somma fatta su tutti i divisori d interi positivi di n, mentre $\mu(d)$ indica la *funzione di Möbius*. Poniamo $\mu(1)=1$ e $\mu(n)=0$ se n è divisibile per il quadrato di qualche numero primo. Se $n>1$ è il prodotto di k numeri primi *distinti*, allora poniamo $\mu(n)=(-1)^k$.

Per chiarire le cose, vediamo subito un esempio concreto della formula di Gandhi: per $n=3$ si trova $P(3)=2\cdot 3\cdot 5=30$. I divisori di 30 sono 1, 2, 3, 5, 6, 10, 15 e 30 e di conseguenza, tenendo diligentemente conto dei valori della funzione μ, la somma $S(30)$ vale

$$\begin{aligned} S(30)&=\sum_{d|30}\frac{\mu(d)}{2^d-1}=1-\frac{1}{3}-\frac{1}{7}-\frac{1}{31}+\frac{1}{63}+\frac{1}{1023}\\ &\quad +\frac{1}{32\,767}-\frac{1}{1\,073\,741\,823}=\frac{545\,925\,250}{1\,073\,741\,823}\\ &=\frac{2^1+2^7+2^{11}+2^{13}+2^{17}+2^{19}+2^{23}+2^{29}}{2^{30}-1}. \end{aligned} \tag{1}$$

Dunque $1-\log_2\big(S(30)-\frac{1}{2}\big)\approx 7.889822\ldots$. Tutto ciò per determinare che $p_4=7$: un bello spreco di risorse .... Una forma meno precisa ma più esplicita della formula di Gandhi è

$$S(P(n))-\frac{1}{2}\approx 2^{-p_{n+1}}. \tag{2}$$

Daremo la dimostrazione generale in Appendice, nel §C. Per il momento vediamo una stima approssimativa della grandezza delle quantità in gioco e cioè quanti addendi ha la somma $S(P(n))$ e qual è la dimensione massima dei relativi denominatori, ma già il caso particolare dato dalla formula (1) illustra bene la situazione che troveremo in generale: moltissimi addendi, denominatori enormi per ottenere un risultato finale che ha un ordine di grandezza ben più piccolo.

5.1 – *Esame critico della formula di Gandhi*

Se $n=10$ abbiamo $P(10)=2\cdot 3\cdot 5\cdot 7\cdot 11\cdot 13\cdot 17\cdot 19\cdot 23\cdot 29=6\,469\,693\,230$. Il denominatore di $S(P(10))$ vale $2^{P(10)}-1$ che è un numero di circa 1 947 571 725 cifre decimali. La somma $S(P(10))$ contiene $2^{10}=1024$ addendi. In generale, per sapere, anche solo approssimativamente, la dimensione del denominatore nella formula di Gandhi è necessario conoscere il Teorema dei Numeri Primi 8.1: una delle forme equivalenti afferma che $\log P(n)\sim p_n\sim n\log n$ quando $n\to+\infty$. In definitiva, il denominatore di $S(P(n))$ vale $2^{P(n)}-1$ che è *doppiamente esponenziale* in n, mentre il numero di addendi presenti nella somma vale 2^n. Dunque la formula di Gandhi richiede almeno 2^n iterazioni perché $S(P(n))$ ha altrettanti addendi. Infine, ciascun addendo deve essere calcolato con una precisione di p_{n+1} "bit," come mostriamo nella Figura 3. Tutto questo deve essere confrontato con il fatto che $p_{n+1}\sim n\log n$. Dunque, la complessità computazionale della formula di Gandhi e la intrinseca necessità di memorizzare termini con una grande precisione la rendono del tutto inutilizzabile dal punto di vista pratico: come abbiamo visto nel §2, il Crivello di Eratostene è molto più efficiente e semplice da realizzare come programma per computer.

Sia la macchina di Conway che la Formula di Gandhi producono i numeri primi in ordine crescente: sono dunque inadatte entrambe alla ricerca di primi "grandi" come quelli che si usano in crittografia.

6. – Polinomi e numeri primi

La formula di Gandhi descritta nel paragrafo precedente non ha certo il pregio della semplicità: fornisce tutti i numeri primi in ordine crescente, questo è vero, ma ha tutti i difetti che abbiamo

elencato dettagliatamente qui sopra. Potremmo decidere di rinunciare ad ottenere *tutti* i numeri primi con una sola formula, accontentandoci di ottenerne uno diverso per ogni valore della variabile indipendente. Naturalmente, la nuova formula dovrà essere molto semplice, più semplice di quella di Gandhi o simili. La prima idea che può venire in mente è quella di scegliere un polinomio.

In una variabile non c'è verso: i valori di un polinomio, presi modulo n per ogni intero n fissato, sono periodici con periodo n o un suo divisore, e questo impedisce che un polinomio non costante in una variabile assuma solo valori primi. La dimostrazione dipende dal fatto che i monomi hanno questa proprietà, e non è difficile vedere che è sufficiente dimostrare che tutti i coefficienti di $(x+n)^k - x^k$ sono divisibili per n per ogni $k \in \mathbb{N}$: questo segue immediatamente dallo sviluppo del binomio.

Un esempio famoso è il polinomio di Eulero $q(n) = n^2 - n + 41$, che dà numeri primi per $n = 0, \ldots, 40$. Ma $q(41n) = 41(41n^2 - n + 1)$ è divisibile per 41 qualunque sia n intero, e quindi può essere primo solo quando $41n^2 - n + 1$ vale esattamente ± 1: per un polinomio di secondo grado, questa cosa può accadere al massimo per 4 valori di n. *Mutatis mutandis*, questa idea mostra che un polinomio non costante in una variabile assume infiniti valori composti. In generale, poniamo $q(0) = p$: se $p = 0$ allora il polinomio $q(n)$ è divisibile algebricamente per n ed i suoi valori sono tutti composti da un certo punto in poi. Se $p \neq 0$, abbiamo visto che $q(kp) \equiv 0 \bmod p$ per ogni $k \in \mathbb{Z}$ e quindi q può assumere solo valori primi solo se p è un numero primo e inoltre $q(kp) = \pm p$ *qualunque* sia l'intero k. Se q non è costante ed ha grado $d \geq 1$, ciascuna delle due equazioni $q(n) = \pm p$ ha al massimo d soluzioni, e questo contraddice l'ipotesi. La stessa dimostrazione vale, più in generale, per polinomi in più variabili. Ma per questi polinomi c'è anche il Teorema di Matijasevič: esiste un polinomio in più variabili i cui valori *positivi* sono tutti e soli i numeri primi. Anche in questo caso, smontato il procedimento, si scopre che si tratta di un risultato di grande interesse teorico soprattutto in Logica Matematica, più che uno strumento d'indagine adatto allo studio dei numeri primi. La dimostrazione, che è molto tecnica, può essere basata sul Teorema di Wilson 3.1. Notiamo che questa dimostrazione è costruttiva, non semplicemente di pura esistenza, come succede tante volte in matematica: è possibile scrivere esplicitamente un polinomio con le caratteristiche richieste.

Potremmo dunque pensare di aver risolto ogni problema: scegliamo a caso le variabili del polinomio e lo valutiamo; se il risultato è un numero positivo abbiamo la garanzia che si tratta di un numero primo, in caso contrario scegliamo a caso di nuovo, e ricominciamo. In realtà, le cose non stanno così, almeno dal punto di vista delle applicazioni pratiche. Infatti, il polinomio dato nell'articolo di Jones et al. [8] ha la forma

$$p(\boldsymbol{x}) = (x_1 + 2)\big(1 - p_1(\boldsymbol{x})^2 - p_2(\boldsymbol{x})^2 - \ldots - p_{14}(\boldsymbol{x})^2\big)$$

dove $\boldsymbol{x} = (x_1, x_2, \ldots, x_{26})$ e $p_1, \ldots, p_{14}$ indicano opportuni polinomi a coefficienti interi (dati esplicitamente) in 26 variabili, di grado che non supera 12, per cui il grado di p risulta essere 25. Basta una rapida occhiata a p per rendersi conto del fatto che p assume un valore positivo *se e solo se*

$$p_1(\boldsymbol{x}) = p_2(\boldsymbol{x}) = \cdots = p_{14}(\boldsymbol{x}) = 0. \tag{3}$$

In definitiva, il Teorema di Matijasevič equivale all'affermazione che il sistema (3) in 26 variabili intere e 14 equazioni ha una soluzione in interi positivi $x_1, x_2, \ldots, x_{26}$ se e solo se $x_1 + 2$ è un numero primo. Una cosa ben diversa da quella che poteva sembrare all'inizio: si tratta, ripetiamo, di un interessante teorema di logica, non di uno strumento pratico per generare numeri primi. In termini di logica matematica, l'insieme dei numeri primi è diofanteo. Notiamo per concludere che è possibile esibire esplicitamente altri polinomi con le stesse proprietà, con molte variabili o con grado molto alto, e che questo risultato è un sottoprodotto della soluzione del decimo problema di Hilbert.

7. – Il Crivello di Eratostene e la Formula di Legendre

Resta il problema di dare una forma quantitativa al Crivello di Eratostene descritto nel §2: è il contenuto della Formula di Legendre, che ci permette di determinare il numero dei numeri primi che non superano un certo numero reale x, e che denotiamo con $\pi(x)$, conoscendo individualmente *tutti* i numeri primi che non superano $x^{1/2}$.

TEOREMA 7.1 [Formula di Legendre]. – *Per $x \geq 1$ sia p_n il massimo numero primo che non supera $x^{1/2}$. Allora si ha*

$$\pi(x) - \pi\big(x^{1/2}\big) + 1 = \pi(x) - n + 1 = \sum_{d \mid P(p_n)} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor. \tag{4}$$

Notiamo che, per definizione, p_n è il massimo numero primo adoperato nel crivello. Ricordiamo inoltre che $P(p_n)$ indica $p_1 \cdots p_n$, e cioè il prodotto dei primi n numeri primi.

La dimostrazione della Formula di Legendre (4) è di natura combinatoria: si tratta di un esempio concreto di quello che si chiama Principio di Inclusione–Esclusione. Una dimostrazione alternativa di cui parliamo più avanti dipende, in ultima analisi, dal Lemma C.1. Vediamo la dimostrazione combinatoria: ci sono esattamente $\lfloor x \rfloor$ interi $\leq x$, che corrispondono al termine $d = 1$. Ogni primo $p \leq x^{1/2}$ divide $\lfloor x/p \rfloor$ di questi interi: dunque dobbiamo sottrarre dal termine già trovato un addendo del tipo $\lfloor x/p_j \rfloor$, per $j \in \{1, \ldots, n\}$. Ma così facendo abbiamo indebitamente sottratto due volte tutti gli interi che sono divisibili per due o più numeri primi distinti. Dobbiamo dunque sommare il contributo dei numeri primi $p_1, \ldots, p_n$ presi a due a due, poi sottrarre il contributo degli stessi primi, presi tre a tre, e così via. In definitiva, sommiamo su tutti i divisori d di $P(p_n)$ un addendo $\lfloor x/d \rfloor$, cioè il numero degli interi $\leq x$ divisibili per d, con un "segno" $\mu(d)$.

Per esempio, se prendiamo $x = 100$ allora $n = \pi(100^{1/2}) = 4$ e $P(p_4) = 2 \cdot 3 \cdot 5 \cdot 7 = 210$. La Formula di Legendre ci dice quanto vale $\pi(100) - \pi(10) + 1 = \pi(100) - 3$:

$$\begin{aligned}
\sum_{d \mid 210} \mu(d) \left\lfloor \frac{100}{d} \right\rfloor &= \left\lfloor \frac{100}{1} \right\rfloor - \left\lfloor \frac{100}{2} \right\rfloor - \left\lfloor \frac{100}{3} \right\rfloor - \left\lfloor \frac{100}{5} \right\rfloor \\
&\quad + \left\lfloor \frac{100}{6} \right\rfloor - \left\lfloor \frac{100}{7} \right\rfloor + \left\lfloor \frac{100}{10} \right\rfloor + \left\lfloor \frac{100}{14} \right\rfloor \\
&\quad + \left\lfloor \frac{100}{15} \right\rfloor + \left\lfloor \frac{100}{21} \right\rfloor - \left\lfloor \frac{100}{30} \right\rfloor + \left\lfloor \frac{100}{35} \right\rfloor \\
&\quad - \left\lfloor \frac{100}{42} \right\rfloor - \left\lfloor \frac{100}{70} \right\rfloor - \left\lfloor \frac{100}{105} \right\rfloor + \left\lfloor \frac{100}{210} \right\rfloor \\
&= 100 - 50 - 33 - 20 + 16 - 14 \\
&\quad + 10 + 7 + 6 + 4 - 3 \\
&\quad + 2 - 2 - 1 - 0 + 0 \\
&= 22,
\end{aligned} \tag{5}$$

ed infatti $\pi(100) - \pi(10) + 1 = 25 - 4 + 1 = 22$.

Per rendere più chiaro perché funziona la formula di Legendre, proviamo ad "inseguire" alcuni numeri, come per esempio 12, 13 e 22, cioè vogliamo determinare il loro contributo totale alla somma all'estrema sinistra della (5). Il numero $m = 12$ contribuisce un'unità al termine $\lfloor 100/1 \rfloor$ in quanto multiplo di 1, un'unità al termine $\lfloor 100/2 \rfloor$ in quanto multiplo di 2, un'unità al termine $\lfloor 100/3 \rfloor$ in quanto multiplo di 3, un'unità al termine $\lfloor 100/4 \rfloor$ in quanto multiplo di 4, un'unità al termine $\lfloor 100/6 \rfloor$ in quanto multiplo di 6, un'unità al termine $\lfloor 100/12 \rfloor$ in quanto multiplo di 12. Tenendo conto del fatto che i termini citati sono preceduti dal segno $\mu(d)$, il contributo totale di 12 è

$$\mu(1) + \mu(2) + \mu(3) + \mu(4) + \mu(6) + \mu(12) = 0$$

per il Lemma C.1. Se ripetiamo lo stesso ragionamento per $m = 13$, vediamo che questo contribuisce una sola unità al termine $\lfloor 100/1 \rfloor$ in quanto multiplo di 1, e dunque contribuisce in totale 1, mentre $m = 22$ contribuisce un'unità al termine $\lfloor 100/1 \rfloor$ in quanto multiplo di 1, un'unità al termine $\lfloor 100/2 \rfloor$ in quanto multiplo di 2, ed il suo contributo totale vale $\mu(1) + \mu(2) = 0$ ancora per il Lemma C.1. Si noti che $(12, 210) = 6$, $(13, 210) = 1$ e $(22, 210) = 2$ rispettivamente. Stiamo dunque usando la formula (9). Infatti, il contributo vale 1 solo per quegli interi nell'intervallo $[1, x]$ che non hanno fattori in comune con $P(p_n)$: questo accade solo per $m = 1$ e per i numeri primi nell'intervallo $(x^{1/2}, x]$.

7.1 – *Riordinamento della formula di Legendre: il crivello combinatorio*

Il calcolo della formula di Legendre in cui si ordina d in modo crescente come nella (5) non dà risultati particolarmente illuminanti: se invece si ordina d mettendo insieme prima tutti i valori che hanno esattamente 0 fattori primi, poi quelli che ne hanno esattamente 1, e così via, si ottiene una cosa molto interessante:

$$
\begin{aligned}
\sum_{d\mid 210}\mu(d)\Big\lfloor\frac{100}{d}\Big\rfloor &= \Big\lfloor\frac{100}{1}\Big\rfloor \\
&\quad - \Big(\Big\lfloor\frac{100}{2}\Big\rfloor+\Big\lfloor\frac{100}{3}\Big\rfloor+\Big\lfloor\frac{100}{5}\Big\rfloor+\Big\lfloor\frac{100}{7}\Big\rfloor\Big) \\
&\quad + \Big(\Big\lfloor\frac{100}{6}\Big\rfloor+\Big\lfloor\frac{100}{10}\Big\rfloor+\Big\lfloor\frac{100}{14}\Big\rfloor \\
&\quad + \Big\lfloor\frac{100}{15}\Big\rfloor+\Big\lfloor\frac{100}{21}\Big\rfloor+\Big\lfloor\frac{100}{35}\Big\rfloor\Big) \\
&\quad - \Big(\Big\lfloor\frac{100}{30}\Big\rfloor+\Big\lfloor\frac{100}{42}\Big\rfloor+\Big\lfloor\frac{100}{70}\Big\rfloor+\Big\lfloor\frac{100}{105}\Big\rfloor\Big) \\
&\quad + \Big\lfloor\frac{100}{210}\Big\rfloor \\
&= 100-(50+33+20+14) \\
&\quad +(16+10+7+6+4+2) \\
&\quad -(3+2+1+0)+0 \\
&= 100-117+45-6+0=22.
\end{aligned}
\tag{6}
$$

Naturalmente il risultato non è cambiato rispetto a prima, dato che ci siamo limitati a permutare l'ordine degli addendi della formula (5), facendo attenzione ai segni. L'osservazione importante è che le somme parziali, ordinate opportunamente (sul numero dei fattori primi di d piuttosto che sul valore di d) forniscono approssimazioni dall'alto e dal basso per il valore di $\pi(100)$. Infatti, se prendiamo il solo primo termine all'estrema destra della relazione (6) troviamo un'approssimazione dall'alto, molto grossolana, per il numero che vogliamo determinare, e cioè 100. Se prendiamo i primi due termini, troviamo invece un'approssimazione dal basso, ancor più grossolana dato che il risultato è negativo! Le cose cominciano a mostrare il loro interesse quando prendiamo i primi 3 termini, che ci danno 28 come approssimazione dall'alto per il risultato giusto che è, lo ripetiamo, 22. Al passaggio successivo troviamo l'approssimazione dal basso 22, in definitiva corretta, dato che l'ultimo addendo vale 0 e non può modificare il valore così trovato.

Per renderci conto della complessità, anche combinatoria, della formula di Legendre, prendiamo ora $x=1000$, in modo che $x^{1/2}\approx 31.6$ ed $n=\pi(31.6)=11$. La formula che corrisponde alla (6) ha un totale di $2048=2^{11}$ addendi, che devono essere disposti in 12 gruppi, partendo dall'unico addendo in cui d ha 0 divisori primi (cioè $d=1$). Le somme parziali, ordinate come nella (6), sono 1000, -560, 414, 135, 158 (gli addendi successivi sono tutti nulli) ed in effetti $\pi(1000)=168$, coerentemente con questo risultato ($158=168-11+1$).

In astratto, si ottengono informazioni non banali su $\pi(x)$ considerando un numero relativamente piccolo di addendi nella formula di Legendre, ma scelti con oculatezza: in particolare, segue che $\pi(x)=O(x/\log(\log x))$, che è naturalmente un risultato molto più debole del Teorema dei Numeri Primi 8.1 ma allo stesso tempo non ovvio.

Varianti di questa formula permettono oggi di calcolare il valore esatto di $\pi\big(10^{25}\big)$ senza conoscere individualmente tutti i numeri primi conteggiati. Per farsi un'idea del numero di termini non nulli nella Formula di Eratostene-Legendre, si noti che dal Teorema dei Numeri Primi 8.1 segue, essenzialmente, che $d\mid P(p_n)$ è $>x$ non appena il numero di fattori primi distinti di d supera $(1+o(1))(\log x)/(\log(\log x))$.

Queste considerazioni suggeriscono che si possono ottenere ragionevoli stime dall'alto per $\pi(1000)$, e più in generale per $\pi(x)$, se invece di considerare *tutti* i numeri primi che non superano 31, in generale $x^{1/2}$, utilizziamo un sottoinsieme di questi. Nella tabella che segue abbiamo riportato la limitazione $f(n)$ per $\pi(1000)-\pi(31)+1$ che si ottiene utilizzando solo i primi n numeri primi, quando n assume i valori 1, 2, ... , 11. Naturalmente solo l'ultimo valore è esatto, ma già quando $n=6$ o 7 si trova una limitazione ragionevolmente vicina al vero.

n	1	2	3	4	5	6	7	8	9	10	11
$f(n)$	500	333	266	228	207	190	179	170	163	160	158

Se $n=1$ consideriamo solo $p_1=2$ e quindi la somma ha due soli addendi, e cioè diventa $1000-500$; se $n=2$ allora $P=6$ e la somma ha quattro addendi corrispondenti ai suoi divisori 1, 2, 3 e 6 e vale $1000-500-333+166=333$. Se invece $n=3$ allora $P=30$ e la somma ha otto addendi: i precedenti quattro e quelli che corrispondono ai divisori 5, 10, 15 e 30, e quindi bisogna aggiungere la quantità $-200+100+66-33$, ottenendo complessivamente 266. In parole povere, stiamo dicendo che nell'intervallo $[1,1000]$ vi sono

266 interi che *non hanno* fattori primi in comune con $P = 30$. Uno di questi 266 interi è il numero 1, che non consideriamo primo; dobbiamo poi ricordare che ci sono i numeri primi 2, 3 e 5 che non sono stati conteggiati nei 266; in definitiva ricaviamo la maggiorazione $\pi(1000) \leq 266 + 3 - 1$. Prendendo n più grande la formula diventa più complicata perché il numero di addendi cresce, e allo stesso tempo la limitazione ottenuta diventa più precisa.

In generale, possiamo introdurre il *crivello combinatorio*, che si basa sull'identità

$$\sum_{i=0}^{m} (-1)^i \binom{k}{i} = (-1)^m \binom{k-1}{m}$$

estendendo il dominio del coefficiente binomiale, in modo che valga per ogni $m \geq 0$. Questa formula è una generalizzazione della (9) qui sotto.

Il crivello combinatorio è sufficiente a dimostrare una versione "forte" di quanto detto sopra a proposito dei valori primi assunti da un polinomio non costante in una variabile: per la precisione, "quasi tutti" i valori assunti sono composti. Nel caso particolare $q(n) = n$ si trova una versione debole del Teorema dei Numeri Primi 8.1.

Notiamo che i numeri primi non vengono determinati individualmente e poi contati, ma si rende quantitativo il Crivello di Eratostene. Risultati quantitativi deboli, ma non banali, si ottengono scegliendo $z = \log(x)$ nel crivello combinatorio, facendo in pratica un crivello con i numeri primi fino a z, come abbiamo fatto sopra con la variabile intera n. Questo permette di avere un numero di addendi relativamente piccolo, evitando i problemi ai quali è dedicato il prossimo paragrafo. L'idea di ottenere approssimazioni invece del numero esatto vede la luce con Gauss.

7.2 – *Critica della formula di Legendre*

Il difetto principale della formula di Legendre è che contiene troppi termini per essere utilizzabile come strumento pratico per il calcolo. La formula di Gandhi ha un numero enorme di termini quasi tutti piccolissimi, mentre la formula di Eratostene-Legendre ha lo stesso numero di termini, quasi tutti nulli! Possiamo dare una motivazione quantitativa alle considerazioni qualitative dei paragrafi precedenti: si potrebbe pensare di ottenere una buona approssimazione per il secondo membro della (4) eliminando le parti intere, cioè calcolando

$$\sum_{d|P(p_n)} \mu(d)\frac{x}{d} = x \prod_{p \leq x^{1/2}} \left(1 - \frac{1}{p}\right) \sim \frac{2e^{-\gamma}x}{\log x},$$

per una delle formule di Mertens citate sopra e per una proprietà della funzione di Möbius. Quella che si ottiene è una quantità dell'ordine di grandezza giusto, ma, alla luce del Teorema dei Numeri Primi 8.1 con la costante sbagliata, e cioè $2e^{-\gamma}$ al posto di 1.

8. – Per chi vuole approfondire: materiale e spunti per letture ulteriori

8.1 – *Applicazioni del Crivello di Eratostene*

Come detto nel §2, si può utilizzare un procedimento di crivello, in combinazione con altri algoritmi, per determinare numeri primi molto grandi. Inoltre, una variante del crivello è il cuore dell'algoritmo di fattorizzazione detto "crivello quadratico", proprio in virtú della sua efficienza computazionale e della possibilità di distribuire il calcolo su diversi processori (parallelismo). Si trova una descrizione del crivello quadratico nel §5.4 di [9], insieme a una panoramica sugli aspetti computazionali che riguardano i numeri primi.

8.2 – *Aspetti quantitativi della distribuzione dei numeri primi*

Il risultato più importante è il Teorema dei Numeri Primi, dimostrato nel 1896 indipendentemente da Hadamard e de la Vallée Poussin; qui ne diamo una forma semplificata.

TEOREMA 8.1. – *Si ha*

$$\lim_{x \to +\infty} \frac{\pi(x)}{x/\log x} = 1.$$

Nell'analisi della Formula di Gandhi usiamo la forma equivalente

$$\lim_{x \to +\infty} \frac{1}{x} \sum_{p \leq x} \log p = 1.$$

La dimostrazione di entrambe le versioni si trova in Hardy & Wright [7], Capitolo 22. Nello stesso capitolo si possono trovare anche le Formule di Mertens: sono i Teoremi 429 e 427. Per approfondire la questione da vari punti di vista, si possono consultare [14] e Zagier [15].

8.3 – *Condizioni necessarie e sufficienti per la primalità*

La più conosciuta condizione necessaria per la primalità deriva dal cosiddetto "Piccolo" Teorema di Fermat: se p è un numero primo ed a è un intero non divisibile per p, allora $a^{p-1} \equiv 1 \bmod p$. Si tratta, in effetti, del Teorema di Lagrange per il gruppo finito $(\mathbb{Z}/p\mathbb{Z})^*$. Purtroppo, è ben noto che questa condizione non è sufficiente: poiché il calcolo di potenze modulo n, anche con esponenti grandi, può essere realizzato in modo relativamente efficiente, sono state a lungo ricercate condizioni supplementari che permettano di rendere quella enunciata qui sopra anche sufficiente per la primalità. Uno dei criteri più usati è quello di Miller & Rabin: se ne trova una descrizione dettagliata nel contesto dei problemi discussi in questo paragrafo nel Capitolo 11 di [9]. Il criterio di Agrawal, Kayal & Saxena [1], che ha la sua origine nel Teorema di Fermat, fornisce la condizione sufficiente richiesta ed è computazionalmente efficiente: si tratta di prendere congruenze modulo polinomi di grado relativamente basso. L'enunciato e una dimostrazione semplificata ed accessibile di questo criterio si possono trovare in Granville [5].

8.4 – *La macchina di Conway*

Una descrizione dettagliata della macchina di Conway si trova nel bellissimo articolo di Guy [6], dove ogni singolo pezzo è letteralmente "smontato" e spiegato nei suoi minimi particolari. Per un'operazione analoga di dissezione dei meccanismi, si possono utilmente consultare la pagina Fractran di wikipedia all'indirizzo http://en.wikipedia.org/wiki/Fractran, e quella di MathWorld http://mathworld.wolfram.com/FRACTRAN.html. Ulteriori informazioni si trovano nella Online Encyclopedia Integer Sequences, in particolare cercando la successione A007542 http://oeis.org/A007542. Una divertente animazione della macchine di Conway si trova sulla pagina web di Andrew Granville all'indirizzo http://www.dms.umontreal.ca/~andrew/conwaymachine.html.

8.5 – *Formula di Gandhi*

La dimostrazione che presentiamo in Appendice è quella di Vanden Eynden [11], mentre l'interpretazione come crivello è dovuta a Golomb [4].

8.6 – *Altre formule per i numeri primi*

Il lettore interessato alle formule per i numeri primi ne può trovare in quantità nei libri di Dickson [2] (Capitolo XVIII, in particolare fra le pagine 429 e 435) e Ribenboim [10] (Capitolo 3). Ripetiamo l'invito alla cautela fatto alla fine del nostro §3 prima di dare un altro semplice esempio. Si tratta della formula scoperta indipendentemente da S. Wigert (1895) e G. Andreoli (1912): il luogo di zeri della funzione $f(x) = \sin^2(\pi(\Gamma(x)+1)/x) + \sin^2(\pi x)$ definita per $x > 0$ è l'insieme $\{1\} \cup \{p : p \text{ è primo}\}$. Si veda Dickson, Ch. XVIII, p. 432 e 434. Anche questa è una conseguenza del Teorema di Wilson 3.1: infatti, $f(x) \geq 0$ per ogni $x > 0$, ed inoltre $f(x) = 0$ se e solo se $\sin(\pi(\Gamma(x)+1)/x) = \sin(\pi x) = 0$. L'ultima condizione è soddisfatta se e solo se x è un intero positivo; in questo caso, $\Gamma(x) = (x-1)!$ e dunque, per il Teorema di Wilson, anche $(\Gamma(x)+1)/x$ è un intero positivo se e solo se x vale 1 oppure x è un numero primo. In linea di principio, usando il Teorema dei Residui, si può dare una formula per $\pi(x)$ integrando $f'(x)/f(x)$ lungo un opportuno cammino nel piano complesso: anche questa non risulta particolarmente maneggevole e non è per questa strada che è stato dimostrato il Teorema 8.1.

8.7 – *Il Teorema di Matijasevič*

Abbiamo visto che i primi sono il luogo di zeri simultaneo di opportuni polinomi in più variabili a coefficienti interi. Si veda Jones et al [8] per un esempio concreto e per la relativa discussione. Una trattazione accessibile si trova nel §3.III di Ribenboim [10].

8.8 – *Miscellanea*

Per le applicazioni dei numeri primi alla crittografia si vedano per esempio [12] e [9]. La giustificazione della formula per i "decimali" periodici usata nella Figura 3 è in [13], appendice.

9. – Conclusioni

I numeri primi sono così complessi, nonostante la definizione apparentemente molto semplice, che non possono essere "catturati" efficientemente in modo meccanico (in senso stretto, o nel senso di una formula).

Mettere a confronto le strutture additiva e moltiplicativa di $\mathbb{N}$ non è un mero artificio retorico. Il problema di determinare quanti interi distinti fra 1 ed N^2 compaiono nella tavola pitagorica con N righe ed N colonne è stato proposto e parzialmente risolto da Erdős nel 1955, ma l'ordine di grandezza esatto è stato determinato solo recentemente da Ford [3] (si badi bene: solo l'ordine di grandezza, *non* la formula asintotica; vedi il Corollario 3 dell'articolo appena citato). Il corrispondente problema per la tavola pitagorica additiva è noioso: compaiono tutti gli interi fra 2 e $2N$, e l'intero $n \in [2, 2N]$ compare esattamente $N - |N + 1 - n|$ volte.([1]) Dunque, quando N è sufficientemente grande (cioè nella tavola pitagorica additiva infinita) l'intero $n \geq 2$ compare $n - 1$ volte. Il numero di "rappresentazioni additive" dell'intero n è una funzione crescente di n; il numero di "rappresentazioni moltiplicative" di n è la funzione che conta i divisori di n che è molto irregolare, e proprio per questo motivo molto interessante. Si tratta di una delle funzioni più studiate della Teoria dei Numeri, e ciononostante continua ad avere alcuni aspetti relativamente misteriosi: non a caso, era una delle funzioni preferite di Pál Erdős.

([1]) La formula è piú semplice di quello che sembri a prima vista!

Appendice A
Qualche dettaglio sulla macchina di Conway

Prima di dare la descrizione di una semplice macchina, notiamo che in definitiva Conway ha inventato un rudimentale linguaggio di programmazione, noto come Fractran. In generale, una macchina Fractran è data da una stringa di frazioni insieme ad un intero n_0 che rappresenta l'*input*, sul quale è talvolta indispensabile fare qualche ipotesi. È dunque possibile costruire macchine di Conway per un gran numero di problemi diversi.

La macchina di Conway per i numeri primi, introdotta nel §4, può essere descritta da 10 *registri* $a_1, \ldots, a_{10}$ che contengono interi non negativi. Lo *stato* della macchina è dato da $2^{a_1} \cdot 3^{a_2} \cdot 5^{a_3} \cdot 7^{a_4} \cdot 11^{a_5} \cdot 13^{a_6} \cdot 17^{a_7} \cdot 19^{a_8} \cdot 23^{a_9} \cdot 29^{a_{10}}$. Il *valore iniziale* $n_0 = 2$ corrisponde ad avere $a_1 = 1$ ed $a_2 = a_3 = \cdots = a_{10} = 0$: rappresenteremo questo fatto mediante il vettore $s_0 = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$. A ciascuna delle frazioni $A, \ldots, N$ possiamo associare un vettore incremento $v_1, \ldots, v_{14}$, secondo lo schema seguente:

$$
\begin{aligned}
v_1 &= (0, 0, 0, -1, 0, -1, 1, 0, 0, 0)\\
v_2 &= (1, 1, -1, 0, 0, 1, -1, 0, 0, 0)\\
v_3 &= (0, -1, 0, 0, 0, 0, -1, 1, 0, 0)\\
v_4 &= (-1, 0, 0, 0, 0, 0, 0, -1, 1, 0)\\
v_5 &= (0, -1, 0, 0, -1, 0, 0, 0, 0, 1)\\
v_6 &= (0, 0, 0, 1, 1, 0, 0, 0, 0, -1)\\
v_7 &= (0, 0, 1, 0, 0, 0, 0, 1, -1, 0)\\
v_8 &= (0, 0, 0, 1, 1, 0, 0, -1, 0, 0)\\
v_9 &= (0, 0, 0, 0, 0, 0, -1, 0, 0, 0)\\
v_{10} &= (0, 0, 0, 0, 1, -1, 0, 0, 0, 0)\\
v_{11} &= (0, 0, 0, 0, -1, 1, 0, 0, 0, 0)\\
v_{12} &= (-1, 1, 1, -1, 0, 0, 0, 0, 0, 0)\\
v_{13} &= (-1, 1, 1, 0, 0, 0, 0, 0, 0, 0)\\
v_{14} &= (0, 0, 1, 0, 1, 0, 0, 0, 0, 0).
\end{aligned}
$$

Useremo l'abbreviazione $\mathbf{0}$ per indicare il vettore $(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ e scriveremo $\boldsymbol{v} \geq \mathbf{0}$ per dire che ogni componente del vettore $\boldsymbol{v}$ è non negativa. Il corpo del programma consiste nel verificare in sequenza una serie di condizionali:

$$s_0 + v_1 \geq \mathbf{0} \qquad s_0 + v_2 \geq \mathbf{0} \ldots$$

fino a che uno non sia vero: male che vada, l'ultimo lo è certamente poiché $v_{14} \geq \mathbf{0}$. Partendo da 2, cioè dal vettore s_0 definito sopra, il primo condizionale vero è il penultimo

$$s_0 + v_{13} = (0, 1, 1, 0, 0, 0, 0, 0, 0, 0) \geq \mathbf{0},$$

perché $2M = 15$. A questo punto aggiorniamo ponendo $n_1 = 15$ e ricominciamo. Si noti che la presenza della frazione N in coda (un numero intero) garantisce che la macchina di Conway non termina mai!

In definitiva, si capisce meglio il funzionamento della macchina di Conway se classifichiamo i suoi 10 registri come segue:

- Registri contabili (i primi 4)
- Variabili (*flag*) di stato (i successivi 4)
- Variabili locali (gli ultimi 2)

In fondo, la macchina sfrutta in modo essenziale l'unicità della fattorizzazione per tenere separati (segregati) in modo efficiente i valori dei registri. Chi ha imparato a programmare i microprocessori noterà le evidenti analogie.

Appendice B

Una semplice macchina di Conway

Ci limitiamo a descrivere in dettaglio una macchina costituita da una sola frazione, la quale somma il contenuto del registro b al registro a prima di fermarsi:

$$\frac{2}{3}: \quad \begin{cases} \text{input}: \ n_0 \ = \ 2^a \, 3^b \, m \text{ con } (m, \, 2 \cdot 3) \ = \ 1 \\ \text{output}: \ n_b \ = \ 2^{a \, + \, b} \, m. \end{cases}$$

Infatti, se n_0 ha la forma data e $b > 0$, dobbiamo porre $n_1 = \frac{2}{3} n_0 = 2^{a+1} 3^{b-1} \cdot m$. Se $b - 1 > 0$ a sua volta, calcoliamo $n_2 = \frac{1}{2} n_1 = 2^{a+2} 3^{b-2} m$. Induttivamente, troviamo che se $b \geq k$ allora dopo k passi abbiamo $n_k = 2^{a+k} 3^{b-k} m$. Al passo b-esimo la macchina si ferma restituendo il valore $n_b = 2^{a+b} m$.

È possibile costruire "macchine" che eseguono altre operazioni elementari: per esempio, la moltiplicazione si realizza mediante addizioni ripetute ed utilizzando come "subroutine" la macchina descritta sopra. La divisione con resto si ottiene mediante sottrazioni ripetute, e così via.

La cosa più difficile, che rende quasi miracolosa l'esistenza della macchina di Conway, è garantire che mettendo insieme alcune semplici macchine per costruirne di più complicate, non vi siano effetti collaterali indesiderati, quali l'azzeramento di qualche registro. Un'altra difficoltà brillantemente superata da Conway è quella di simulare cicli annidati: per ottenere questo scopo ha utilizzato i numeri primi "grandi," i cui rispettivi registri assumono solo il valore 0 ed 1 nell'esecuzione del programma, come accennato alla fine del paragrafo precedente. Si tratta di una conseguenza non ovvia dell'ipotesi sull'input.

Appendice C

Dimostrazione della formula di Gandhi

L'ultima uguaglianza nella formula (1) è particolarmente degna di nota. Per prima cosa rende evidente, o almeno plausibile, il risultato finale: il numeratore della frazione differisce poco dalla somma dei suoi ultimi 2 addendi, il denominatore differisce pochissimo da 2^{30}. In definitiva, $S(30) \approx (2^{23} + 2^{29})/2^{30} = 2^{-7} + 1/2$; in altre parole, $S(30) - 1/2 \approx 2^{-7}$.

Inoltre, è importante osservare che il numeratore contiene la somma delle potenze di 2 con esponente r nell'intervallo $[1, 30]$ per cui $(r, 30) = 1$. Questo naturalmente è un fatto generale e non contingente: tra le tante dimostrazioni possibili, ne scegliamo una che dipende da una semplice identità algebrica e da uno dei "trucchi" più usati in Teoria dei Numeri, e cioè lo scambio fra due somme annidate. Useremo l'identità

$$(7) \qquad \frac{x^a - 1}{x - 1} = 1 + x + x^2 + \cdots + x^{a-1},$$

che è valida per ogni x reale diverso da 1 e per ogni intero positivo a. Nella dimostrazione scriviamo S in luogo di $S(P(n))$ per brevità. Dunque abbiamo

$$S = \sum_{d|P(n)} \frac{\mu(d)}{2^d - 1} = (2^{P(n)} - 1)^{-1} \sum_{d|P(n)} \mu(d) \, \frac{2^{P(n)} - 1}{2^d - 1}.$$

Le quantità all'estrema destra sono numeri interi, come si vede applicando l'identità (7) qui sopra con $x = 2^d$ ed $a = P(n)/d$. Dunque

$$\begin{aligned} (2^{P(n)} - 1) S &= \sum_{d|P(n)} \mu(d)(1 + 2^d + 2^{2d} + \cdots + 2^{P(n)-d}) \\ &= \sum_{d|P(n)} \mu(d) \sum_{r=0}^{P(n)/d-1} 2^{rd}. \end{aligned}$$

Chi abbia la pazienza di scrivere esplicitamente i primi addendi di questa somma nel caso $n = 3$ discusso sopra, noterà che alcune potenze di 2 compaiono diverse volte, precedute da un "segno" $\mu(d)$. Raggruppiamo dunque le potenze di 2 con lo stesso esponente, chiamando $m = rd \in \{0, 1, \dots, P(n) - 1\}$ il valore comune, ottenendo

$$(2^{P(n)} - 1)S = \sum_{m=0}^{P(n)-1} 2^m \sum_{d|P(n),\, d|m} \mu(d). \tag{8}$$

La condizione all'estrema destra implica che d divide il massimo comun divisore fra m e $P(n)$. È in questo punto che avviene lo scambio delle somme. La somma risultante ha un ruolo di primo piano in molte applicazioni, e la sua valutazione ha la dignità di Lemma.

LEMMA C.1. – *Posto*

$$\Sigma(n) = \sum_{d|n} \mu(d),$$

si ha $\Sigma(1) = 1$ *e* $\Sigma(n) = 0$ *per* $n > 1$.

Daremo una dimostrazione di natura combinatoria tralasciando qualche dettaglio che i lettori più pignoli potranno divertirsi ad integrare. Possiamo evidentemente supporre che $n > 1$ abbia la fattorizzazione canonica $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ con $p_1 < \ldots < p_k$ dove gli α_i sono interi positivi e $k \geq 1$. Posto $n_0 = p_1 \cdots p_k$, abbiamo per prima cosa che $\Sigma(n) = \Sigma(n_0)$: infatti, i termini diversi da zero nella somma $\Sigma(n)$ sono tutti e soli quelli per cui d divide n_0. Ricordiamo che $d \mid n$ se e solo se $d = p_1^{\beta_1} \cdots p_k^{\beta_k}$, dove $\beta_i \in \{0, \dots, \alpha_i\}$ per $i \in \{1, \ldots, k\}$. Ma, per definizione, se $\beta_i \geq 2$ per qualche i allora $\mu(d) = 0$. Dunque, un divisore d di n dà un contributo non nullo alla somma $\Sigma(n)$ solo se $d \mid n_0$, come affermato sopra. In altre parole, se $d \mid n$ e $\mu(d) \neq 0$ allora esiste $I \subseteq \{1, \ldots, k\}$ tale che $d = \prod_{i \in I} p_i$ e quindi $\mu(d) = (-1)^{|I|}$. In definitiva

$$\begin{aligned} \Sigma(n_0) = \sum_{d|p_1 \cdots p_k} \mu(d) &= \sum_{I \subseteq \{1,\ldots,k\}} (-1)^{|I|} \\ &= 1 + (-1)\binom{k}{1} + (-1)^2\binom{k}{2} \\ &\quad + \cdots + (-1)^k\binom{k}{k} \\ &= (1-1)^k = 0, \end{aligned} \tag{9}$$

perché $k \geq 1$. Notiamo che in effetti questo equivale a

$$\Sigma(n) = \sum_{d|n} \mu(d) = (1 + \mu(p_1)) \cdots (1 + \mu(p_k)),$$

perché si può "espandere" il prodotto qui sopra a destra e notare che si ottengono tutti gli addendi non nulli che compaiono in $\Sigma(n)$, e cioè tutti gli addendi che corrispondono ai divisori di n_0. Naturalmente, ogni fattore vale 0 se $n > 1$.

Riprendiamo dunque il calcolo dalla (8): la quantità all'estrema destra vale 1 se $\big(d, P(n)\big) = 1$, e 0 altrimenti. In definitiva

$$\begin{aligned} (2^{P(n)} - 1)S &= \sum_{m=0}^{P(n)-1} 2^m \sum_{d|P(n),\, d|m} \mu(d) \\ &= \sum_{m=0}^{P(n)-1} 2^m \begin{cases} 1 & \text{se } (m,\ P(n)) = 1, \\ 0 & \text{se } (m,\ P(n)) > 1, \end{cases} \end{aligned} \tag{10}$$

che è quanto volevamo dimostrare. Si confronti con la quantità all'estrema destra di (1).

Non ci resta che la deduzione finale: l'addendo più grande nell'ultima somma vale $2^{P(n)-1}$. Qual è il precedente? È quello per cui $m = P(n) - p_{n+1}$, poiché tutti gli interi fra $P(n) - p_{n+1} + 1$ e $P(n) - 2$ hanno un fattore primo in comune con $P(n)$. Da questo segue immediatamente che

$$\begin{aligned} S &\geq \frac{2^{P(n)-1} + 2^{P(n)-p_{n+1}}}{2^{P(n)} - 1} \\ &> \frac{2^{P(n)-1} + 2^{P(n)-p_{n+1}}}{2^{P(n)}} = \frac{1}{2} + \frac{1}{2^{p_{n+1}}}. \end{aligned} \tag{11}$$

Procedendo analogamente ed usando la (7) con $x = 2$ ed $a = P(n) - p_{n+1} + 1$, vediamo che

$$\begin{aligned} (2^{P(n)} - 1)S &\leq 2^{P(n)-1} + 2^{P(n)-p_{n+1}} \\ &\quad + 2^{P(n)-p_{n+1}-1} + \ldots + 2^1 + 2^0 \\ &= 2^{P(n)-1} + 2^{P(n)-p_{n+1}+1} - 1. \end{aligned}$$

Dividendo troviamo in conclusione

$$S \leq \frac{2^{P(n)-1} + 2^{P(n)-p_{n+1}+1} - 1}{2^{P(n)} - 1} < \frac{1}{2} + \frac{2}{2^{p_{n+1}}}, \tag{12}$$

poiché $p_{n+1} \geq 3$. Dalle disuguaglianze (11) e (12)

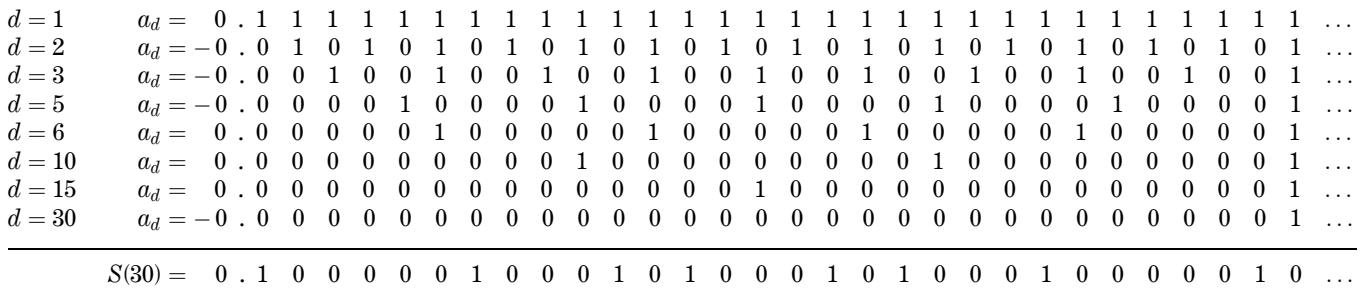

FIGURA 3. – Nella somma $S(P(n))$ in basso, il primo *bit* a destra della virgola vale sempre 1, mentre il successivo *bit* non nullo indica la posizione del numero primo cercato, p_{n+1}. Qui vediamo in azione lo scambio di due somme descritto nel testo: prima scriviamo ciascun a_d come numero "binario" infinito in orizzontale, poi sommiamo in verticale i *bit* nella stessa posizione. Si osservi che i 30 *bit* di ciascun a_d e di $S(30)$ mostrati in figura si ripetono periodicamente con periodo 30 proprio perché il vero denominatore di $S(30)$ è $2^{30}-1$ e non 2^{30}, come nella (1). Notiamo esplicitamente che per calcolare il valore delle somme in verticale si usa il Lemma C.1.

segue che

$$1-\log_2\Big(S(P(n))-\frac{1}{2}\Big)\in\big(p_{n+1},p_{n+1}+1\big)$$

che implica la tesi.

I lettori più attenti avranno notato che in uno degli ultimi passaggi abbiamo usato implicitamente il fatto che $P(n)\geq p_{n+1}$. Per la precisione, questo è vero solo per $n\geq 2$, ma la verifica che $S(1)=1$ ed $S(2)=\frac{2}{3}$ è immediata e mostra che la Formula di Gandhi è vera in generale per ogni $n\geq 0$.

"Smontando" questa dimostrazione e rimontandone diversamente i pezzi si può vedere che, in generale, per $m\geq 1$ si ha

$$S(m)=\sum_{k\geq 1}\frac{1}{2^k}\Sigma\big((k,m)\big),$$

dove (k,m) indica il massimo comun divisore fra k ed m. Volendo, questa identità può essere dedotta dalla (10) con l'ausilio della (7). Usando il Lemma C.1, per $m\geq 2$ diventa evidente che

$$S(m)\in\Big(\frac{1}{2}+2^{-f(m)},\frac{1}{2}+2^{1-f(m)}\Big)$$

dove $f(m)$ indica il minimo intero $k\geq 2$ tale che $(m,k)=1$, che è una versione precisa della (2). In definitiva questa relazione dice che

$$f(m)=\Big\lfloor 1-\log_2\Big(S(m)-\frac{1}{2}\Big)\Big\rfloor,$$

e la Formula di Gandhi è il caso particolare $m=P(n)$, poiché $f(P(n))=p_{n+1}$. Qui abbiamo preferito evitare l'uso di somme infinite, al costo di qualche complicazione di dettaglio.

Concludiamo con un'interpretazione interessante, ripresa nella Figura 3. La formula di Gandhi "corrisponde" a fare un crivello con i fattori primi di $P(n)$, scrivendo in base 2 le quantità $a_d=\mu(d)/(2^d-1)$ e sommando in colonna, *bit* per *bit*. A parte il segno, a_d ha la rappresentazione come numero binario $0.\overline{0\ldots01}$ (dove sono presenti esattamente $d-1$ zeri nel periodo), perché, essenzialmente, stiamo usando la formula (7) con $x=2^{-d}$ ed a arbitrariamente grande.

RIFERIMENTI BIBLIOGRAFICI

[1] AGRAWAL M., KAYAL N. & SAXENA N., *PRIMES is in P*, Ann. Math. **160** (2004), 781–793.

[2] DICKSON L. E., *History of the Theory of Numbers (3 volumes)*, Carnegie, 1919–1923, Reprint Chelsea–AMS, 1999.

[3] FORD K., *The distribution of integers with a divisor in a given interval*, Ann. Math. **168** (2008), 367–433.

[4] GOLOMB S. W., *A direct interpretation of Gandhi's formula*, Amer. Math. Monthly **81** (1974), 752–754.

[5] GRANVILLE A., *It is easy to determine whether a given integer is prime*, Bull. Amer. Math. Soc. **42** (2005), 3–38, http://www.ams.org/bull/2005-42-01/S0273-0979-04-01037-7/home.html.

[6] GUY, R. K., *Conway's prime-producing machine*, Math. Mag. **56** (1983), 26–33.

[7] HARDY G. H. & WRIGHT E. M., *An Introduction to the Theory of Numbers*, fifth ed., Oxford Science Publications, Oxford, 1979.

[8] JONES J. P., SATO D., WADA H. & WIENS D., *Diophantine representation of the set of prime numbers*, Amer. Math. Monthly **83** (1976), no. 6, 449–464.

[9] LANGUASCO A. & ZACCAGNINI A., *Manuale di crittografia*, Ulrico Hoepli Editore, Milano, 2015.

[10] RIBENBOIM P., *The New Book of Prime Numbers Records*, Springer, New York, 1996.
[11] VANDEN EYNDEN C., *A proof of Gandhi's formula for the n-th prime*, Amer. Math. Monthly **79** (1982), 625.
[12] ZACCAGNINI A., *L'importanza di essere primo*, Ricordando Franco Conti (a cura di A. Abbondandolo, M. Giaquinta & F. Ricci), Scuola Normale Superiore, Pisa, 2004, http://people.math.unipr.it/alessandro.zaccagnini/psfiles/papers/importanza.pdf, pp. 343–354.
[13] ZACCAGNINI A., *La calcolatrice e le sue limitazioni*, L'Educazione Matematica, Anno XXVII, Serie VII **2** (2007), 35–45.
[14] ZACCAGNINI A., *Breve storia dei numeri primi*, Ithaca: Viaggio nella Scienza **III** (2014), 67–83, http://ithaca.unisalento.it/nr-03_04_14/index.html.
[15] ZAGIER D., *The first 50 million prime numbers*, The Mathematical Intelligencer **0** (1977), 7–19.

Alessandro Zaccagnini

Professore associato di Analisi Matematica presso l'Università di Parma. Si occupa di Teoria Analitica dei Numeri. Ha scritto più di 30 lavori scientifici, oltre a una dozzina di articoli divulgativi e due libri di testo di crittografia.

TROVATO IL NUMERO PIÙ GRANDE

Così titolava un suo trafiletto di molti anni fa un autorevole quotidiano italiano. Il testo dell'articolo provvedeva a chiarire come il record si riferisse non alla totalità dei numeri naturali, ma "solo" a quelli primi. In realtà neppure in questo caso poteva trattarsi di un massimo assoluto, visto che già Euclide aveva mostrato l'infinità dei primi, semmai di un numero primo maggiore di quelli fino ad allora conosciuti. D'altra parte trovare primi sempre nuovi e sempre più estesi consente sorprendenti applicazioni pratiche, per esempio in crittografia. Uno dei metodi per riuscirci, tra i più famosi se non tra i più fruttuosi, consiste nel ricorrere alla successione dei numeri di Mersenne $M(n) = 2^n - 1$ quando n varia ancora tra primi. Non che ogni numero $M(n)$ così ottenuto sia per ciò stesso primo, anzi i primi di Mersenne che per ora si conoscono sono appena 49. L'ultimo però è stato scoperto a inizio anno, il 7 gennaio 2016. Si tratta di $M(274207281)$, dunque di $2^{274207281} - 1$. È lui per adesso il "numero più grande", perché non si conosce primo che lo superi nemmeno all'esterno della successione di Mersenne. Consiste di 22338618 cifre in base 10, cinque milioni in più del record precedente. Ad annunciarne la scoperta è il sito di GIMPS, Great Internet Mersenne Prime Search, il software open source che permette a chiunque di cimentarsi in questo genere di sfida. Tutti infatti, e non solo i matematici, sono invitati ad utilizzarlo per cercare di battere il primato di M(274207281). In effetti si può dubitare che queste manifestazioni di potenza informatica siano davvero "matematica". È comunque un dato di fatto che il 42-mo primo di Mersenne fu trovato grazie a GIMPS nel 2005 da un chirurgo tedesco degli occhi appassionato alla questione, Martin Nowak. Il record attuale invece è stato invece stabilito da Curtis Cooper, professore di Computer Science alla University of Central Missouri.

La Redazione