Alcune proprietà dei numeri primi, I

Alessandro Languasco & Alessandro Zaccagnini

In questa serie di lavori sull'aritmetica presenteremo alcune idee elementari riguardanti i numeri primi ed alcune loro applicazioni. Con il termine "elementari" intendiamo specificare che le tecniche che utilizzeremo non fanno uso dell'analisi matematica o dell'algebra lineare e non che i risultati che presenteremo siano semplici o banali. In particolare cercheremo di coniugare il rigore con la semplicità dell'esposizione poiché siamo convinti che gli argomenti trattati siano comprensibili a chiunque abbia voglia di dedicarci un po' di tempo ed anche che, d'altra parte, per capire davvero le cose sia necessario guardarle dall'alto, cioè con maggiore generalità. Per questo motivo, abbiamo incluso alcuni approfondimenti, nella speranza che i lettori non si fermino al primo livello, ma cerchino di andare oltre.

Le argomentazioni che presenteremo sono classiche, e sono diventate rilevanti anche dal punto di vista pratico dopo il 1975, ossia dall'avvento della crittografia a chiave pubblica. Cercheremo quindi in questa sede di introdurre le proprietà dei numeri primi che hanno rilevanza crittografica, senza però dimenticare di inquadrarle in una teoria più generale e completa, che ha un interesse intrinseco anche se, al momento, non trova applicazioni pratiche.

Uno degli aspetti più affascinanti, a nostro avviso, della teoria dei numeri primi sta nel fatto che ha la sua origine in un contesto discreto (i numeri interi positivi) ma, per averne una comprensione non superficiale, è necessario introdurre concetti dell'analisi matematica, che tratta prevalentemente grandezze continue. Per poi averne una comprensione profonda, si deve fare ricorso all'analisi complessa, apparentemente molto remota dal problema originale. In queste note tratteremo parte delle interrelazioni dell'aritmetica con l'analisi reale.

Per motivi di spazio, non parliamo di congruenze, pur utilizzando la nozione e le proprietà: rimandiamo alla trattazione che si trova nel Capitolo 2 di Conway e Guy [3], oppure a quella in [11].

Per quanto possibile, daremo riferimenti bibliografici in lingua italiana, ma desideriamo ricordare ai Lettori che non ce ne sono moltissimi. Quando non è possibile fare altrimenti, e limitatamente agli approfondimenti suggeriti, daremo qualche riferimento anche in inglese.

1 Notazioni

Per prima cosa fissiamo alcune notazioni che useremo in seguito.

Notazione 1.1 (Divisibilità) *Diremo che l'intero a divide l'intero b se esiste un intero c tale che $a \cdot c = b$, e in questo caso scriveremo $a \mid b$.*

Osserviamo che non chiediamo che a o b sia positivo, né diverso da zero. Consideriamo note le proprietà elementari della divisibilità. Per completezza, ricordiamo anche la definizione di congruenza, osservando al tempo stesso che è una relazione di equivalenza.

Notazione 1.2 (Congruenza) *Dato un intero positivo n, diremo che l'intero a è congruo all'intero b modulo n, e scriveremo $a \equiv b \bmod n$, se $n \mid a - b$.*

2 Cosa sono i numeri primi

Cominciamo la nostra discussione sui numeri primi, che sono i "mattoni" dell'aritmetica dal punto di vista della moltiplicazione. La nostra definizione è la seguente.

Definizione 2.1 (Numero primo) *Un intero positivo n si dice* primo *se ha esattamente due divisori positivi.*

Questa definizione di numero primo è diversa da quella che la maggior parte delle persone ricorda dalle Scuole Medie: "un intero positivo n si dice primo se è divisibile solo per 1 e per sé stesso". Il motivo principale per cui diamo questa definizione è che vogliamo escludere il numero 1 dall'insieme dei numeri primi: daremo qualche giustificazione per la nostra scelta nel §2.1. In effetti, su alcuni testi si trova la definizione di numero primo nella forma (equivalente alla nostra): "un intero $n \geq 2$ si dice primo se è divisibile solo per 1 e per sé stesso". Sono dunque primi i numeri 2, 3, 5, 7, 11, 13, ..., mentre non sono primi i numeri 4, 6, 8, 9, 10, 12, 14, 15, 16, ... Il numero 1 non è classificato.

I numeri primi sono importanti perché sono alla base della struttura moltiplicativa dei numeri naturali: il Teorema Fondamentale dell'Aritmetica 3.1 assicura che ogni numero naturale si può ottenere moltiplicando fra loro opportuni numeri primi in *uno ed un solo modo*, a parte l'ordine in cui i fattori sono presi. Per questo motivo, gli interi $n \geq 2$ che non sono numeri primi si dicono *composti*.

Il nostro scopo qui è anche quello di indicare altri motivi di interesse (teorico e pratico) dei numeri primi, non immediatamente evidenti.

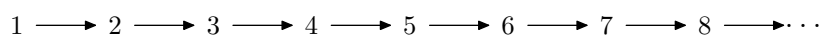

Figura 1: L'ordinamento dei numeri naturali positivi indotto dalla funzione successore: per motivi evidenti questo ordinamento si chiama *lineare* o *totale*.

La Figura 1 mostra l'ordinamento standard dell'insieme dei numeri naturali positivi $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$: a partire da 1, ogni intero si ottiene dal precedente aggiungendo 1, ed in questo modo si ottengono *tutti* gli interi positivi. In questa figura come nella successiva, indichiamo solo il successore o i successori immediati di ogni intero, e quindi la relazione di ordine si considera prolungata per transitività: se $a < b$ e $b < c$ allora $a < c$. In altre parole, poniamo *per definizione* che $n < n+1$ (esattamente come suggerito dalle frecce) e poi estendiamo questa definizione dicendo che, dalle due disuguaglianze $n < n+1$ ed $n+1 < (n+1)+1 = n+2$, segue $n < n+2$, e cosí via.

La Figura 2 mostra invece un possibile ordinamento *non standard* dei numeri interi positivi: dati due numeri naturali distinti a e b con $0 < a < b$, diciamo che $a \prec b$ (*a divide b*) se $a \mid b$, e in questo caso li colleghiamo mediante una linea orientata da a verso b. È immediatamente evidente che questo secondo ordinamento è assai più complesso del primo, e di conseguenza assai più interessante. In particolare osserviamo che gli interi che seguono immediatamente 1 (senza altri interi intermedi) sono i numeri primi: in altre parole, i numeri primi sono i *primi numeri* che seguono 1 in questo ordinamento. Notiamo inoltre che con questo ordinamento non è sempre possibile *confrontare* due interi positivi qualsiasi: a questo proposito, si può comunque notare che ogni coppia di interi ha un *massimo predecessore comune* ed un *minimo successore comune*, che sono, naturalmente, il *massimo comun divisore* ed il *minimo comune multiplo* rispettivamente.

È molto importante osservare che, in questo caso, massimo e minimo si riferiscono entrambi alla relazione $\prec$ e *non* alla relazione $<$. Facciamo un esempio concreto: scelti $a = 30$ e $b = 36$, gli insiemi dei predecessori di a e di b rispettivamente sono dati da

$$\{d \in \mathbb{N}^* : d \preceq a\} = \{1, 2, 3, 5, 6, 10, 15, 30\},$$
$$\{d \in \mathbb{N}^* : d \preceq b\} = \{1, 2, 3, 4, 6, 9, 12, 18, 36\}.$$

Dunque $\{d \in \mathbb{N}^* : d \preceq a \wedge d \preceq b\} = \{1, 2, 3, 6\}$, e 6 è il massimo dell'ultimo insieme rispetto alla relazione $\prec$: è essenziale notare che in quest'ultimo insieme *ogni* elemento è un predecessore di 6, o, in altre parole, che ogni elemento è un divisore di 6. Indicheremo con (n,m) il massimo comun divisore fra n ed m.

Notiamo anche che $\mathbb{N}^*$ è generato additivamente da un solo elemento, e cioè 1, mentre (si veda il Teorema Fondamentale dell'Aritmetica 3.1) è generato moltiplicativamente dall'insieme infinito dei numeri primi: questo è un altro modo per mettere in evidenza la straordinaria ricchezza della struttura moltiplicativa degli

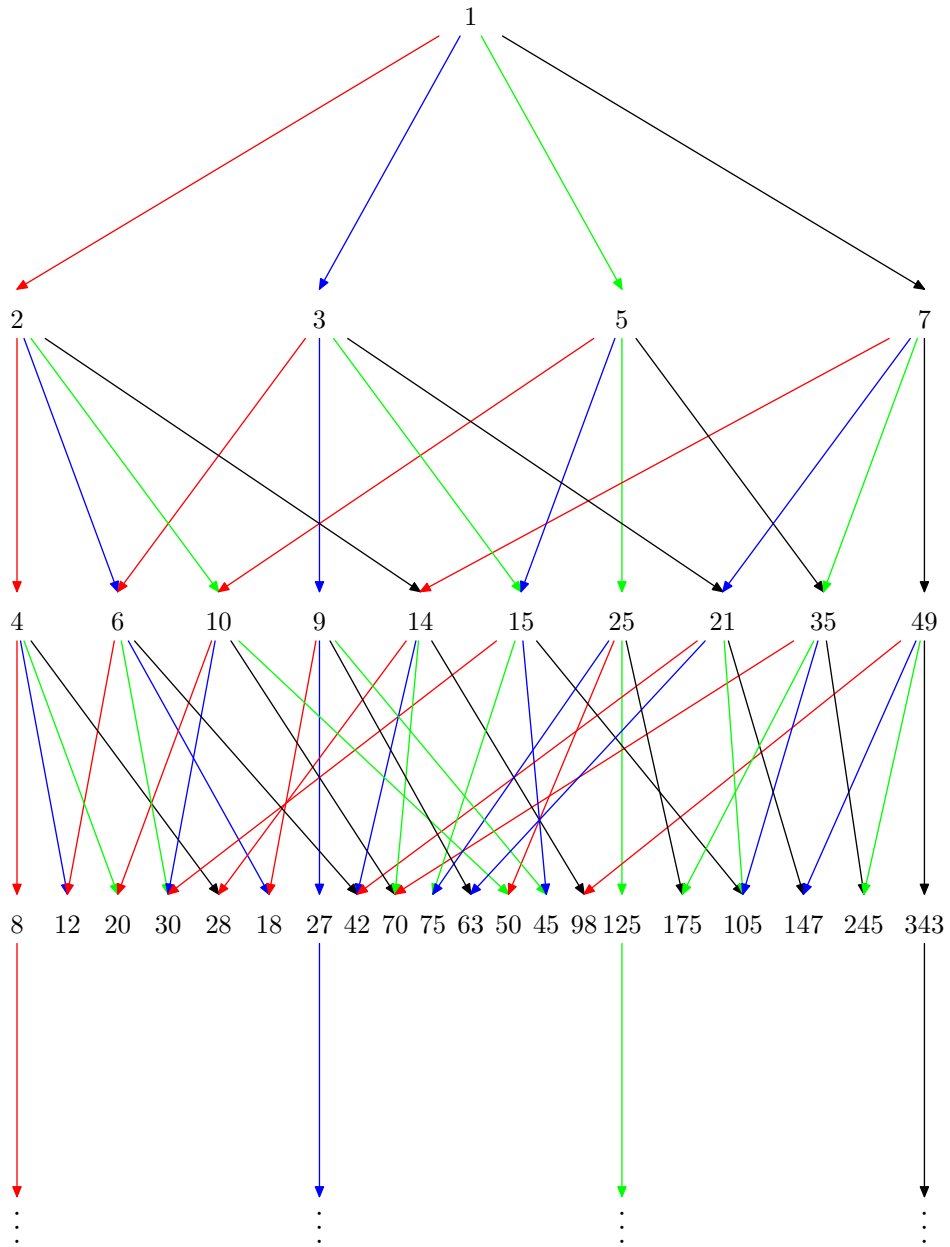

Figura 2: L'ordinamento dei numeri naturali indotto dalla relazione di divisibilità. La figura dovrebbe contenere infinite colonne, una per ogni numero primo, ed essere illimitata sia verso destra che verso il basso. Abbiamo messo nella stessa riga gli interi che hanno lo stesso numero di fattori primi, non necessariamente distinti. Si noti che non è detto che due interi qualsiasi siano confrontabili, anche se hanno sempre almeno un predecessore comune, ed infiniti successori comuni. L'evidente complessità di questo ordinamento è alla base dell'interesse dei numeri primi. La freccia che parte da n e raggiunge m è rossa se m/n è 2, blu se questo rapporto vale 3, verde se vale 5 e nera se vale 7.

interi, paragonata alla relativa povertà di quella additiva. Naturalmente, questa ricchezza si riflette sia nell'interesse astratto per i numeri primi, sia sulle loro applicazioni concrete.

2.1 Perché 1 non è un numero primo

Vogliamo qui spiegare perché si suole escludere 1 dall'insieme dei numeri primi: prima di passare al dettaglio, è bene osservare che in Matematica si cerca di dare definizioni utili e generali, anche al costo di darle in modo apparentemente "non naturale." Non è certamente pensabile che i Matematici siano costretti a conservare la definizione vista nelle Scuole Medie, quando questa confligge con il principio generale appena esposto: speriamo di convincere anche i nostri Lettori con gli esempi qui sotto.

Veniamo dunque al nostro problema; il numero 1 non viene considerato primo per vari motivi, fra i quali citiamo quelli che riteniamo più importanti.

1. Il numero 1 ha un solo divisore, mentre tutti i numeri primi ne hanno due. Questo è solo un esempio di un fenomeno generale: per molte funzioni aritmetiche assolutamente naturali, come la funzione φ di Eulero definita dalla cardinalità dell'insieme degli interi $0 \leq a < n$ tali che $(a,n) = 1$, sarebbe necessario avere due formule distinte, una valida per 1 e l'altra per i numeri primi $p \geq 2$. In questo caso, infatti, dovremmo dire che $\varphi(p) = p-1$ per *tutti* i $p \geq 2$, ma $\varphi(1) = 1$.

2. Molti teoremi, per esempio il Teorema Fondamentale dell'Aritmetica 3.1, dovrebbero essere enunciati in un modo molto più complicato per tener conto delle proprietà speciali di 1.

3. Nel crivello di Eratostene, descritto nel §5, se il numero 1 fosse considerato primo si cancellerebbero *tutti* i numeri tranne lo stesso 1 al primo passo.

4. Un'altra funzione importante è la funzione Ω, che conta il numero *totale* dei fattori primi di un intero positivo n: in altre parole, se n ha la fattorizzazione della (1), la funzione $\Omega(n)$ vale $\alpha_1 + \alpha_2 + \cdots + \alpha_k$. Se 1 fosse primo, potremmo includere nel membro destro della (1) la potenza 1^α, con $\alpha \in \mathbb{N}$ arbitrario, e quindi $\Omega(n)$ sarebbe indeterminato.

5. Nell'Algebra, gli elementi invertibili degli anelli (cioè quegli elementi a per i quali si può risolvere l'equazione $ax = 1$) hanno uno *status* speciale. In $\mathbb{N}$ l'unico elemento invertibile è proprio 1; esso va quindi trattato a parte. Per non distogliere l'attenzione dal nostro obiettivo principale, ne parliamo di nuovo in maggiore dettaglio nell'Appendice B.

In definitiva, possiamo riassumere la nostra argomentazione cosí: se decidessimo di considerare primo anche 1, dovremmo rassegnarci a fare continue eccezioni perfino nelle definizioni o nei teoremi più semplici. Per economia, dunque, preferiamo dare una definizione che a prima vista può sembrare meno naturale, ma con la quale non c'è questa necessità. In effetti si tratta di un principio generale della Matematica: l'utilità e la versatilità delle definizioni sono decidibili solo "a posteriori", cioè solo dopo averle viste all'opera e confrontate con possibili definizioni alternative.

3 Il Teorema Fondamentale dell'Aritmetica

Il risultato che vedremo qui di seguito è cosí importante da contenere l'aggettivo "fondamentale" nel nome: fra i vari enunciati possibili, scegliamo quello in cui la rappresentazione in forma di prodotto è effettivamente unica. Una dimostrazione si trova nel §3.1 di [11]. Il nostro enunciato dipende in modo essenziale dal fatto che in $\mathbb{N}$ è possibile ordinare gli elementi: negli insiemi in cui questo non è possibile, ma vale un enunciato analogo, è necessario dare una formulazione diversa. Riprenderemo l'argomento nell'Appendice B.

Teorema 3.1 *Ogni numero naturale $n \geq 2$ può essere espresso nella forma*

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_k^{\alpha_k} = \prod_{i=1}^{k} p_i^{\alpha_i} \tag{1}$$

dove i p_i sono numeri primi con $p_1 < p_2 < \cdots < p_k$, ed $\alpha_i \in \mathbb{N}^*$ per $i = 1, \ldots, k$, in un unico modo.

Osserviamo che esiste un enunciato alternativo di questo Teorema in cui non si chiede che i numeri primi nella (1) siano elencati in ordine crescente, ma siano semplicemente distinti fra loro: in questo caso, a causa della proprietà commutativa della moltiplicazione, è necessario dire che la rappresentazione è unica, a meno dell'ordine con cui sono presi i fattori. Per esempio, $8911 = 7 \cdot 19 \cdot 67$, e questa decomposizione è essenzialmente unica. Come conseguenza del Teorema 3.1, non è necessario eseguire esplicitamente la moltiplicazione $59 \cdot 151$ per sapere che il risultato *non è* uguale a 8911: la situazione è completamente diversa se i fattori nel prodotto non sono primi, come mostra l'esempio $27 \cdot 330 = 10 \cdot 891$.

Il Teorema Fondamentale dell'Aritmetica può sembrare una banalità perché fino dalle Scuole Medie siamo abituati ad effettuare la scomposizione di un intero nel prodotto delle sue potenze prime. Si potrebbe dunque pensare che debba essere vero per tutti gli insiemi numerici. Ma invece esistono insiemi numerici semplici quanto gli interi in cui esso non è valido: ad esempio l'insieme $\mathcal{H}$ degli

interi della forma $4k+1$, con $k \in \mathbb{N}$. L'insieme $\mathcal{H}$ è un sistema chiuso rispetto alla moltiplicazione (cioè se $a, b \in \mathcal{H}$ allora $ab \in \mathcal{H}$) ma in esso non vale il Teorema Fondamentale dell'Aritmetica. Infatti si verifica facilmente che

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

e queste due diverse scomposizioni di 693 sono entrambe formate da "primi" in $\mathcal{H}$. Infatti $9,\ 77,\ 21$ e 33 non ammettono una fattorizzazione non banale in $\mathcal{H}$ (ossia tali numeri hanno esattamente due divisori aventi la forma $4k+1$).

Peraltro risultati analoghi al Teorema Fondamentale dell'Aritmetica valgono in strutture più sofisticate degli interi; esempi semplici sono i polinomi a coefficienti reali $\mathbb{R}[x]$ o a coefficienti complessi $\mathbb{C}[x]$. In generale le strutture algebriche per cui vale una proprietà di questo tipo vengono dette *a fattorizzazione unica*.

4 Quanti sono i numeri primi?

In questo paragrafo ci chiediamo quanti sono i numeri primi: a parte il Teorema di Euclide 4.1 e la sua dimostrazione, talmente belli che non riusciamo a resistere alla tentazione di includerli, vogliamo anche fare un discorso di "densità" dei primi nella successione dei numeri naturali. In altre parole, vi sono molte successioni interessanti di interi, come i quadrati perfetti, le potenze di 2, i numeri di Fibonacci, e molte altre, e tutte queste, per la loro stessa definizione, hanno infiniti termini, ma se andiamo a contare il numero dei loro termini nell'intervallo $[1, N]$, con N grande, troviamo risultati piuttosto diversi. Alla luce del Teorema di Euclide, ha senso porsi questa domanda anche per la successione dei numeri primi.

Teorema 4.1 (Euclide) *Esistono infiniti numeri primi.*

Come accennavamo precedentemente, la dimostrazione del Teorema di Euclide è considerata una delle "gemme" della matematica antica; per tale ragione la riportiamo in questa sede.

Supponiamo dunque per assurdo che esista solamente un numero finito di primi e che questi siano esattamente $p_1 < p_2 < \cdots < p_k$. Allora il numero

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

non può essere anch'esso primo (perché non è presente nell'elenco precedente); d'altronde, siccome N non è divisibile per alcun p_i, $i = 1, \ldots, k$, deve anch'esso essere primo. Il che è chiaramente una contraddizione. Dunque l'insieme dei numeri primi non può essere finito e quindi il Teorema di Euclide è dimostrato.

Dopo aver dimostrato che esistono infiniti numeri primi è importante capire quale sia la loro densità negli interi. Il saper rispondere a questo problema non solo è significativo da un punto di vista teorico, ma ha anche riflessi applicativi poiché alcuni tra i più usati metodi crittografici moderni si basano sulla "facilità" di costruire numeri primi e sulla "difficoltà" di determinare la fattorizzazione di interi. Se i primi fossero "pochi" sarebbero difficili da costruire e di conseguenza il determinare la fattorizzazione di un intero dato risulterebbe facile.

Il primo passo nel capire quale fosse la densità dei primi fu fatto alla fine del XVIII secolo da Gauss il quale congetturò che il numero dei primi fino ad N, N molto "grande", fosse

$$\pi(N) \sim \frac{N}{\log N} \qquad \text{per } N \to +\infty,$$

dove

$$\pi(N) = \text{numero dei primi fino a } N$$

e la notazione $F(N) \sim G(N)$ indica che il rapporto $F(N)/G(N)$ ha limite 1 quando N tende a $+\infty$. Il fatto sorprendente è che Gauss elaborò la propria congettura basandosi solamente sulle (scarne) informazioni fornite dalle tavole dei numeri primi disponibili all'epoca.

La congettura di Gauss fu dimostrata indipendentemente nel 1896 da Hadamard e de la Vallée Poussin (e dopo allora assunse il nome di Teorema dei Numeri Primi) usando l'idea fondamentale introdotta da Riemann nel 1858: studiare la distribuzione dei primi mediante l'analisi complessa.

Non ci soffermiamo sulle idee di Riemann e sul ruolo fondamentale che la sua funzione ζ ha avuto nello sviluppo della Teoria dei Numeri perché questo discorso ci porterebbe troppo lontano. Notiamo solamente che, da un punto di vista euristico, il Teorema dei Numeri Primi consente di aspettarci che, per N abbastanza grande, esista un primo ogni $\log N$ interi circa, ossia che esistano "tanti" numeri primi.

Facciamo anche notare che la densità dei primi fino ad N è nettamente maggiore rispetto a quelle delle successioni menzionate nella prima parte di questo paragrafo; infatti nell'intervallo $[1,N]$ ci sono approssimativamente $\sqrt{N}$ quadrati perfetti, circa $\log_2 N$ potenze di 2 ed approssimativamente $\log_\Phi N$ numeri di Fibonacci, dove $\Phi = \frac{1+\sqrt{5}}{2}$ è il numero aureo.

Le definizioni e proprietà discusse qui sopra danno origine ad una certa quantità di interessanti problemi teorici e pratici, che discuteremo nel resto di questo articolo e in quelli che lo seguiranno: ne diamo una breve descrizione complessiva, prima di passare alla trattazione dettagliata.

Come generare numeri primi. In molte applicazioni, soprattutto di tipo crittografico, è necessario determinare numeri primi "grandi," con un numero di

cifre decimali dell'ordine di 100 o più: vedremo che un metodo di calcolo che proviene dall'antichità classica, il Crivello di Eratostene, è ancora essenzialmente il modo migliore per determinare tutti i numeri primi in un intervallo di interi consecutivi, ed è di valido ausilio nella ricerca di numeri primi grandi privi di forma particolare.

Come riconoscere i numeri primi. Dato un intero "grande" ci si può chiedere se questo sia primo o meno; non è necessario verificare esplicitamente la definizione, come vedremo, ma è possibile rispondere a questa domanda in modo piuttosto efficiente usando importanti risultati teorici che permettono di "riconoscere" i numeri primi fra tutti gli interi.

Come scomporre in fattori primi. In alcune situazioni non ci si può accontentare della semplice conoscenza del fatto che un certo intero non è primo, ma è necessario conoscerne esplicitamente tutti i fattori primi. Viceversa, dato che alcuni sistemi crittografici importanti (per esempio, RSA) si basano sulla presunta difficoltà di determinare i fattori primi di interi composti grandi scelti in modo opportuno, è interessante conoscere i limiti degli attuali algoritmi di fattorizzazione, per poter scegliere i parametri dei crittosistemi avendo un buon margine di sicurezza.

Come convincere della primalità di un intero. Si tratta di un curioso problema pratico: supponiamo di aver dimostrato in qualche modo che un certo intero è effettivamente un numero primo. È possibile convincere un'altra persona della correttezza della dimostrazione senza doverla replicare interamente? In altre parole, esiste un "certificato" che garantisca la primalità? Non è un problema ozioso, in quanto gli utenti dei crittosistemi moderni hanno bisogno di uno o più numeri primi, e spesso non hanno i mezzi teorici e pratici per generarli (cioè per generarne di grandezza adeguata all'applicazione crittografica che hanno in mente), e quindi sono costretti ad "acquistarli" da qualcuno. Per questo è stato introdotto il concetto di "certificato succinto" che garantisca l'acquirente, senza che questi abbia la necessità di ripetere integralmente il calcolo eseguito dal venditore.

5 Come generare numeri primi

In questo paragrafo cominciamo a rispondere alle domande poste alla fine del paragrafo precedente, e precisamente cominciamo dal problema di determinare tutti i numeri primi nell'intervallo $[1, N]$ dove N è un intero "grande". Un metodo possibile è quello di scorrere la lista di tutti questi interi, e verificarne singolarmente

l'eventuale primalità: Eratostene scoprí che questa procedura è estremamente dispersiva, nel senso che gran parte del calcolo è inutilmente ripetuta più volte, e quindi suggerí una strategia alternativa che risulta di gran lunga più efficiente. Discuteremo questo aspetto quantitativo nell'Appendice A.2.

La procedura di Eratostene è nota con il nome di "crivello" (che vuol dire "setaccio"): si passano i numeri interi positivi attraverso un opportuno setaccio, e quelli che restano sono solo i numeri primi. La Figura 3 illustra il Crivello di Eratostene applicato all'intervallo $[1,225]$. È relativamente semplice spiegare il procedimento in forma algoritmica: supponiamo di voler eseguire il crivello sugli interi nell'intervallo $[1,N]$, dove N è un parametro a nostra scelta.

1. Si scrivono tutti gli interi da 1 ad N.
2. Si parte da $p=2$ (il più piccolo numero primo).
3. Si cancellano tutti i multipli di p partendo da p^2 fino ad N.
4. Si cerca il più piccolo intero $q>p$ non ancora cancellato. Se $q^2>N$ la procedura termina.
5. Si pone $p=q$ e si torna al passo 3.

Si noti che l'operazione di cancellazione di cui al punto 3 può essere effettuata in modo estremamente efficiente (ed è esattamente qui l'essenza dell'algoritmo) partendo da p^2 ed aggiungendo sempre p all'ultimo valore trovato, fino a superare N. Infatti i multipli di p fra $2p$ e p^2-p, estremi inclusi, sono stati cancellati tutti nei passi precedenti, poiché hanno almeno un fattore primo $<p$. Si veda anche la Figura 4.

A questo punto il nostro obiettivo è dimostrare che i numeri "sopravvissuti" a questa operazione sono il numero 1 e tutti i numeri primi fino ad N. Il numero 1, evidentemente, non è stato mai cancellato (il più piccolo numero cancellato è 4, alla prima iterazione). I numeri primi nell'intervallo $[1,N^{1/2}]$ non sono stati cancellati, perché il più piccolo multiplo del primo p che è effettivamente cancellato è $2p$, e i numeri primi nell'intervallo $[N^{1/2},N]$ non sono stati cancellati perché non hanno divisori fra i numeri con i quali abbiamo effettuato le cancellazioni.

Infine, tutti i numeri composti nell'intervallo $[1,N]$ sono stati cancellati: infatti, sia $n\le N$ un numero composto. Dunque esistono interi a, b con $1<a\le b<n$ e tali che $n=ab$. Se a e b fossero entrambi $>N^{1/2}$, il loro prodotto n dovrebbe essere a sua volta $>N$. Dunque n è divisibile per almeno un numero primo $\le N^{1/2}$ (ogni fattore primo di a va bene), e di conseguenza è stato eliminato.

Invitiamo i Lettori a convincersi della correttezza dell'algoritmo eseguendo materialmente queste operazioni sui numeri da 1 a 100: conviene prima disporre i

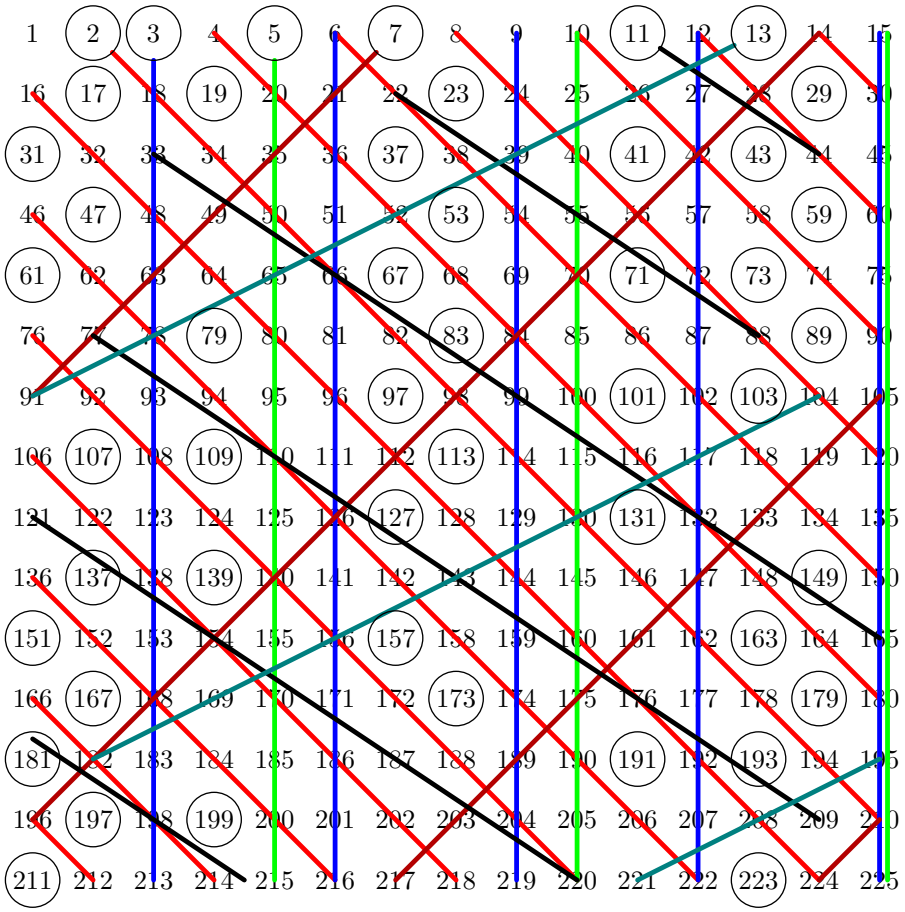

Figura 3: Il crivello di Eratostene. Il meccanismo di funzionamento è tutto sommato piuttosto semplice: saltato il numero 1, che è speciale come abbiamo spiegato nel §2.1, si prende il primo numero che soddisfa la Definizione 2.1, e cioè 2, lo si evidenzia mettendolo dentro un cerchio e se ne eliminano tutti i multipli fino al limite desiderato N, in questo caso 225. Si cerca quindi l'intero successivo non ancora cancellato, che è 3, lo si evidenzia e si eliminano tutti i suoi multipli. Si ripete la stessa procedura per tutti i numeri primi fino a $N^{1/2}$: ciò che resta sono il numero 1 e tutti i numeri primi nell'intervallo $[2, N]$. Le linee colorate "cancellano" i multipli dei numeri primi 2, 3, 5, 7, 11 e 13.

IL CRIVELLO DI ERATOSTENE, VERSIONE DI BASE

```
 1 function crivello(N)
 2 boolean c[N] = {vero * N}
 3 int p ← 2
 4 while (p^2 ≤ N)
 5       int n ← p^2
 6       while (n ≤ N)
 7             c[n] ← falso
 8             n ← p + n
 9       endwhile
10       repeat
11             p ← p + 1
12       until (c[p] = vero)
13 endwhile
```

Figura 4: Lo pseudo-codice per il Crivello di Eratostene.

numeri in uno schema quadrato come nella Figura 3. Si noterà come i multipli di 2, 3, 5, 7 compaiono lungo opportuni segmenti.

Questo è lo schema di base: naturalmente, quando si esegue questa procedura al computer, non è necessario "scrivere" esplicitamente gli interi fra 1 ed N, e sono anche possibili alcuni miglioramenti che ora andiamo a discutere. Questo schema può essere scritto in "pseudo-codice" (cioè in una forma intermedia fra il linguaggio naturale ed un linguaggio di programmazione vero e proprio) e proponiamo la nostra versione nella Figura 4. Le istruzioni nelle righe 6–9 realizzano la fase di "cancellazione" mentre le righe 10–12 eseguono la ricerca del primo intero non cancellato successivo a p.

Se intendiamo scrivere una procedura per computer che esegua il Crivello di Eratostene, dobbiamo cominciare con l'assegnare un certo spazio di memoria per rappresentare i numeri dell'intervallo $[1, N]$: normalmente questo viene realizzato creando un `array` (cioè una matrice ad una dimensione, altrimenti detto vettore) con N posizioni. Dato che quello che ci interessa è sapere se il numero intero corrispondente alla k-esima posizione di questo array è primo oppure no, questo array conterrà inizialmente il valore `vero` in tutte le sue N posizioni (si veda la riga 2), e l'operazione di *cancellazione* descritta nel passo 3 corrisponderà a trasformare questo valore in `falso` (riga 7). In linguaggio più tecnico, abbiamo un array di variabili *logiche* (dette anche `booleani`), e possiamo pensare che il valore `vero` corrisponda, alla fine del calcolo, ai numeri primi ed al numero 1.

La prima cosa che si nota è che c'è un certo "spreco": infatti, tutte le posizioni

dell'array corrispondenti ad interi pari ≥ 4 contengono numeri che certamente non sono primi, e quindi stiamo sprecando metà circa dell'array per contenere informazioni inutili, e per sovrammercato stiamo anche sprecando del tempo per ottenere queste stesse informazioni!

Dunque, cerchiamo un modo per sfruttare questo fatto, ed utilizzare la metà circa delle posizioni necessarie seguendo alla lettera la "ricetta" data qui sopra. Tutto si basa sul Lemma seguente, che è la generalizzazione dell'osservazione, in sé piuttosto banale, che tutti i numeri primi, a parte il numero 2, sono dispari.

Lemma 5.1 *Sia $M \geq 1$ un numero intero. Se p è un numero primo, allora $p \mid M$ oppure $(p, M) = 1$.*

La prima modifica che possiamo apportare allo schema dato sopra è questa: se abbiamo N posizioni dell'array, invece di far corrispondere la k-esima posizione al numero k (dove k assume i valori 1, 2, 3, $\ldots N$), la facciamo corrispondere al numero $2k-1$. Questo significa che utilizzando N posizioni, siamo in grado di eseguire il Crivello sugli interi da 1 a $2N-1$, oppure, che possiamo eseguire il crivello sugli interi da 1 ad N utilizzando circa $N/2$ posizioni. In entrambi i casi, abbiamo un risparmio del 50% circa.

Dobbiamo subito notare che il passo 2 deve essere sostituito da "Si parte da $p = 3$" (in un certo senso, il passo con $p = 2$ è implicito nella nostra costruzione), che l'operazione di cancellazione dei multipli di p deve tener conto del fatto che p^2 (il primo intero da cancellare) corrisponde alla posizione $\frac{1}{2}(p^2+1)$, e che non vi sono multipli pari da cancellare, ma solo quelli dispari. Questo comporta una certa complicazione nei dettagli della realizzazione pratica, che qui omettiamo.

Il discorso appena fatto si riferisce al caso $M = 2$ del Lemma 5.1, ma vi sono valori più grandi di M per cui il rapporto tra risparmio e complicazione nei dettagli rimane vantaggioso. Ne descriviamo uno particolarmente interessante.

Se scegliamo $M = 30$ nel Lemma 5.1, tutti i numeri primi, a parte 2, 3, 5, non hanno fattori comuni con 30, e quindi giacciono in una delle progressioni aritmetiche $1+30k$, $7+30k$, $11+30k$, $13+30k$, $17+30k$, $19+30k$, $23+30k$, $29+30k$. La cosa interessante dal punto di vista "informatico" è che il numero di queste progressioni è 8, e quindi possiamo pensare di associare ogni intervallo di 30 interi consecutivi ad un byte, (diciamo che associamo l'intervallo di estremi $30k$ e $30(k+1)$ al k-esimo byte per $k \in \mathbb{N}$), e associamo il j-esimo bit del byte in questione all'intero $a_j + 30k$, dove gli a_j, in ordine, valgono 1, 7, 11, 13, 17, 19, 23, 29, quando j va da 0 a 7. In questo modo non c'è "spreco" e si usano i singoli bit, con la convenzione che `vero` corrisponde al valore 1 e `falso` al valore 0.

Ovviamente il numero 30 non è speciale, ed è possibile usare, per esempio, $N = 210$, nel quale caso vi sono 48 progressioni da esaminare e la faccenda diventa più complicata.

Come osservazione finale, notiamo che sostanzialmente la stessa procedura funziona se si vogliono determinare i numeri primi nell'intevallo $[M, M+N]$, dove M è un intero grande. C'è una certa complicazione nel dettaglio, dovuta al fatto che il primo intero da eliminare non è necessariamente p^2 come nel nostro schema, ma l'idea di base rimane la stessa.

Questa osservazione ha rilevanza pratica: se si vuole determinare un numero primo "grande" p, diciamo dell'ordine di grandezza di M, si sceglie N dell'ordine di grandezza di $2\log M$ (in modo che vi sia una ragionevole speranza che l'intervallo $[M, M+N]$ contenga almeno un numero primo, come spiega l'argomentazione nel §4), e si opera un crivello con tutti i numeri primi relativamente piccoli. Evidentemente, questo non garantisce che i numeri sopravvissuti al crivello siano effettivamente primi, ma in questo modo si eliminano dall'intervallo in questione, in modo molto efficiente, tutti i numeri che non hanno alcuna speranza di essere primi. A questo punto, si sottopongono i numeri rimasti, che sono relativamente pochi, a dei test di primalità (che descriveremo nella seconda parte), che sono più onerosi dal punto di vista computazionale.

In definitiva, il vantaggio risiede nel fatto che la stragrande maggioranza degli interi vengono eliminati con un basso "costo unitario" e i pochi interi residui possono essere attaccati singolarmente, ad un costo individuale maggiore.

A Risultati quantitativi

A.1 Euristica basata sulla Formula di Stirling

In questo paragrafo daremo una dimostrazione "quantitativa" dell'esistenza di infiniti numeri primi, che parte da una delle più importanti, e a nostro giudizio anche una delle più belle formule della Matematica, la Formula di Stirling:

$$N! \sim \sqrt{2\pi N}\left(\frac{N}{e}\right)^N.$$

Ricordiamo che la notazione $F(N) \sim G(N)$ indica che il rapporto $F(N)/G(N)$ ha limite 1 quando N tende a $+\infty$. In realtà, per i nostri scopi è sufficiente una formula più debole, che dimostreremo per induzione nel Lemma seguente.

Lemma A.1 *Per ogni intero $N \geq 1$ si ha*

$$\log N! \geq N\log N - N. \tag{2}$$

Prima della dimostrazione, osserviamo che questa formula è equivalente a $N! \geq (N/e)^N$, e che è piuttosto ragionevole dato che ci si può aspettare che

$$\log N! = \sum_{n=1}^{N} \log n \sim \int_1^N \log t\,\mathrm{d}t = N\log N - N + 1.$$

Dim. La tesi è evidentemente vera per $N = 1$. Supponiamo dunque che la (2) valga per un certo intero N e dimostriamo che vale per $N+1$. Per ipotesi induttiva

$$\log(N+1)! = \log(N+1) + \log N! \geq \log(N+1) + N\log N - N. \tag{3}$$

Posto $f(x) = x - \log(1+x)$ per $x > -1$, osserviamo che $f'(x) = x/(1+x) \geq 0$ per $x \geq 0$, e quindi $f(x) \geq f(0) = 0$ per ogni $x \geq 0$. Prendendo $x = 1/N$, da questa ultima disuguaglianza deduciamo

$$N\log\left(1+\frac{1}{N}\right) \leq 1 \qquad \Longrightarrow \qquad N\log N \geq N\log(N+1) - 1.$$

Sostituendo nella (3) ricaviamo

$$\log(N+1)! \geq N\log(N+1) - 1 - N + \log(N+1) = (N+1)\log(N+1) - (N+1),$$

che è la tesi. □

Per inciso, notiamo che la disuguaglianza $N\log\big(1+1/N\big) \leq 1$ è una conseguenza diretta del fatto che la successione $a_N = \big(1+1/N\big)^N$, il cui limite per definizione è il numero di Nepero e, è monotona crescente.

Il nostro obiettivo ora è quello di scomporre in fattori primi $N!$: è evidente che nella scomposizione interverranno tutti e soli i numeri primi fra 2 ed N, ma quello che ci interessa è determinare l'esponente esatto con cui ciascun numero primo compare.

Per esempio, qual è la potenza di 2 che divide 100!? Si può ragionare cosí: ogni numero pari fra 1 e 100 contribuisce un fattore 2 al prodotto 100!, e questo significa che l'esponente di 2 deve essere almeno 50, ma non dobbiamo dimenticarci che i multipli di 4 contribuiscono un ulteriore fattore 2, i multipli di 8 un altro ancora, e cosí via. In altre parole, l'esponente di 2 è dato da

$$\begin{aligned}\alpha(2) &= \left\lfloor\frac{100}{2}\right\rfloor + \left\lfloor\frac{100}{4}\right\rfloor + \left\lfloor\frac{100}{8}\right\rfloor + \left\lfloor\frac{100}{16}\right\rfloor + \left\lfloor\frac{100}{32}\right\rfloor + \left\lfloor\frac{100}{64}\right\rfloor \\ &= 50+25+12+6+3+1 = 97.\end{aligned}$$

A questo punto ci possiamo fermare, perché non vi sono multipli di $128 = 2^7$ che siano minori o uguali a 100. Notiamo che vale una semplice disuguaglianza per $\alpha(2)$, che è una conseguenza della formula che dà la somma dei termini della

progressione geometrica di ragione $\frac{1}{2}$:

$$\begin{aligned}\alpha(2) &\le \frac{100}{2}+\frac{100}{4}+\frac{100}{8}+\frac{100}{16}+\frac{100}{32}+\frac{100}{64}\\ &< \frac{100}{2}+\frac{100}{4}+\frac{100}{8}+\frac{100}{16}+\frac{100}{32}+\frac{100}{64}+\frac{100}{128}+\frac{100}{256}+\frac{100}{512}+\cdots\\ &= 100\cdot\frac{1}{2}\cdot\left(1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\cdots\right)\\ &= 100\cdot\frac{1}{2}\cdot\frac{1}{1-\frac{1}{2}} = 100.\end{aligned}$$

Un discorso del tutto analogo vale per qualunque numero primo p e intero positivo N: per completezza riportiamo le due formule. Cominciamo con $\alpha(p)$:

$$\alpha(p) = \left\lfloor \frac{N}{p} \right\rfloor + \left\lfloor \frac{N}{p^2} \right\rfloor + \left\lfloor \frac{N}{p^3} \right\rfloor + \cdots + \left\lfloor \frac{N}{p^{m(p)}} \right\rfloor, \tag{4}$$

dove $m(p)$ indica il massimo intero m per cui $p^m \le N$: in altre parole, $m(p) = \lfloor (\log N)/\log p \rfloor$. Inoltre,

$$\begin{aligned}\alpha(p) &\le \frac{N}{p}+\frac{N}{p^2}+\frac{N}{p^3}+\cdots+\frac{N}{p^{m(p)}}\\ &< \frac{N}{p}\cdot\left(1+\frac{1}{p}+\frac{1}{p^2}+\frac{1}{p^3}+\cdots\right)\\ &= \frac{N}{p}\cdot\frac{1}{1-1/p} = \frac{N}{p-1}.\end{aligned} \tag{5}$$

In definitiva, la scomposizione in fattori primi di $N!$ è

$$N! = \prod_{p\le N} p^{\alpha(p)}$$

dove $\alpha(p)$ è dato dalla (4), e quindi

$$\log N! = \sum_{p\le N} \alpha(p)\log p \le N \sum_{p\le N} \frac{\log p}{p-1} \tag{6}$$

per la disuguaglianza (5).

Dopo questo *tour de force*, siamo pronti a dimostrare che i numeri primi sono "tanti": infatti, mettendo insieme la (2) e la (6), troviamo

$$\sum_{p\le N} \frac{\log p}{p-1} \ge \log N - 1. \tag{7}$$

Osserviamo che il secondo membro di questa relazione tende all'infinito per N che tende a $+\infty$, e questo implica che i numeri primi non finiscono mai! In realtà, essa fornisce qualche ulteriore informazione: infatti, l'analoga della relazione (7) con i quadrati perfetti al posto dei primi è

$$\sum_{1<m^2\leq N} \frac{\log(m^2)}{m^2-1} \leq \sum_{m=2}^{+\infty} \frac{\log(m^2)}{m^2-1}$$

per ogni $N \geq 1$, dove la serie a destra è convergente (si usi il fatto che per m grande l'addendo m-esimo è minore di $m^{-3/2}$) e questo significa che i quadrati perfetti sono molto meno "densi" dei numeri primi nella successione dei numeri naturali.

Concludiamo questo faticoso paragrafo osservando che è possibile modificare questa argomentazione in modo da ottenere una relazione più precisa della (7): per la precisione, si può ottenere che la differenza fra primo e secondo membro è una funzione *limitata* dall'alto e dal basso. Ci asteniamo dallo sviluppare qui la matematica necessaria, avvertendo i Lettori che si basa su una versione *ad hoc* della formula di integrazione per parti. Purtroppo non esistono veri e propri testi didattici in italiano, per cui il riferimento standard è al Capitolo 22 del libro di Hardy e Wright [7], che peraltro ci sentiamo di consigliare in ogni caso. Su rete è possibile trovare dispense di corsi, che possono fare da introduzione ad alcuni di questi problemi: si veda per esempio [16].

A.2 Complessità computazionale del Crivello di Eratostene

Esaminando attentamente la procedura del Crivello di Eratostene, vediamo che quando eliminiamo i multipli del numero primo p dobbiamo accedere $\lfloor N/p \rfloor$ volte alla memoria. Dato che dobbiamo ripetere la stessa operazione per tutti i numeri primi nell'intervallo $[1, N^{1/2}]$, il numero di accessi alla memoria è

$$\sum_{p\leq N^{1/2}} \left\lfloor \frac{N}{p} \right\rfloor. \tag{8}$$

Ripetendo l'argomentazione delineata nel paragrafo precedente, ed utilizzando qualche trucco di analisi matematica, scopriamo che la quantità in (8) può essere stimata in modo piuttosto accurato, ed ha ordine di grandezza $N \log\log N$. In media, dunque, il *costo unitario* del Crivello è dell'ordine di $\log\log N$: in altre parole, è una procedura estremamente efficiente.

A.3 Un'argomentazione euristica basata sul Crivello

Un aspetto interessante del Crivello di Eratostene sta nel fatto che, oltre a fornire una procedura per determinare in modo efficiente i numeri primi, suggerisce

un'argomentazione in sostegno della congettura che $\pi(N)$ abbia l'ordine di grandezza corretto, e cioè $N/\log N$. Fissiamo dunque N grande, e consideriamo le operazioni da eseguire nel Crivello: per prima cosa cancelliamo i multipli di 2, tranne il numero 2 stesso, e quindi cancelliamo

$$\left\lfloor \frac{N}{2} \right\rfloor - 1 \approx \frac{N}{2}$$

interi. Poi cancelliamo i multipli di 3: dato che le congruenze modulo primi distinti sono indipendenti fra loro (per il Teorema Cinese del Resto, vedi il Teorema 2.1.3 di [11]), è piuttosto ragionevole aspettarsi che circa $1/3$ degli interi non ancora cancellati (e cioè degli interi dispari) venga eliminato a questo passo, lasciando

$$\approx \left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)N$$

interi. Essenzialmente la stessa argomentazione suggerisce che per ogni primo p da considerare si debba moltiplicare quest'ultimo prodotto per $1-1/p$, ottenendo

$$\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\cdots\left(1-\frac{1}{p_k}\right)N \tag{9}$$

dove p_k è il massimo numero primo che non supera $N^{1/2}$. Il prodotto sui numeri primi può essere valutato in modo preciso e vale approssimativamente $c/\log N$ dove $c>1$ è un'opportuna costante. Saremmo quindi indotti a fare la congettura (errata)

$$\pi(N) \sim \frac{cN}{\log N} \qquad \text{per } N \to +\infty.$$

Il motivo per cui questa argomentazione fallisce è duplice: da una parte, vi sono delle piccole approssimazioni (non cancelliamo *esattamente* metà degli interi al primo passo, né *esattamente* un terzo al secondo e cosí via). Il numero di queste approssimazioni è uguale al numero di numeri primi $\le N^{1/2}$, e questo numero è piuttosto grande e può generare un effetto cumulativo distorcente. D'altra parte, e forse questo è un motivo ancora più importante, non è del tutto corretto dire che le congruenze modulo primi distinti sono indipendenti: se questo è indubbiamente vero in $\mathbb{Z}$, può non esserlo in un intervallo limitato come $[1,N]$. Per esempio, consideriamo tre primi distinti q_1, q_2 e q_3 appartenenti all'intervallo $[N^{1/3},N^{1/2}]$: evidentemente non esiste *nessun* intero nell'intervallo $[1,N]$ che sia simultaneamente divisibile per tutti questi primi, mentre la nostra ipotesi di indipendenza richiede che ce ne siano circa $N/(q_1q_2q_3)$. [1] Considerando il fatto che esistono

[1] È vero che questo numero è <1, ma è *molto vicino* ad 1 per moltissime scelte dei numeri primi q.

moltissime scelte di primi siffatti, anche piccoli errori di questo tipo danno un effetto cumulativo che distorce il risultato.

Una discussione elementare più generale di questi fenomeni si trova in [14]. Concludiamo osservando che l'affermazione sul vero ordine di grandezza della funzione a sinistra nella (7), della funzione (8) e del prodotto (9) sono essenzialmente "equivalenti" fra loro: si veda per esempio [16].

B Anelli

In questa Appendice forniamo qualche definizione formale di strutture che generalizzano $\mathbb{Z}$ ossia di insiemi con due operazioni che interagiscano così come accade in $\mathbb{Z}$ per la somma ed il prodotto usuale. Per prima cosa dobbiamo definire il concetto di *gruppo* ossia di insieme dotato di una operazione che verifica certe proprietà.

Definizione B.1 (Gruppo) *Un insieme G dotato dell'operazione $\circ$ si dice* gruppo *se*

- *per ogni $g, h \in G$ si ha $g \circ h \in G$;*
- *esiste $e \in G$ tale che $e \circ g = g \circ e = g$ per ogni $g \in G$; e si dice* elemento neutro *o* identità*;*
- *per ogni $g \in G$ esiste $h \in G$ tale che $g \circ h = h \circ g = e$; h si dice* inverso *di g e si indica con g^{-1};*
- *per ogni $g, h, j \in G$ si ha $g \circ (h \circ j) = (g \circ h) \circ j$; questa viene detta* proprietà associativa*.*

Se inoltre $g \circ h = h \circ g$ per ogni $g, h \in G$, allora G si dice gruppo commutativo *o* abeliano*.*

Esempio B.2 *I gruppi più semplici sono $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$ con l'operazione di addizione, $\mathbb{Q}^*$, $\mathbb{R}^*$ o $\mathbb{C}^*$ con la moltiplicazione (ed anche $\mathbb{Q}^+$ o $\mathbb{R}^+$). Altri esempi sono:*

$$\mathbb{Z}_m = \{a \in \mathbb{Z} : 0 \leq a \leq m-1\}$$

con l'addizione modulo m, ed anche

$$\mathbb{Z}_m^* = \{a \in \mathbb{Z}_m : (a,m) = 1\}.$$

Tutti questi sono gruppi abeliani.

Sfruttiamo ora il concetto di gruppo per definire gli *Anelli*.

Definizione B.3 (Anello commutativo con identità) *Un insieme R dotato delle due operazioni $+$ e $\cdot$ si dice* anello commutativo con identità *se R con l'operazione $+$ è un gruppo abeliano con elemento neutro 0, l'operazione $\cdot$ ha elemento neutro $1 \neq 0$, è associativa e commutativa ed inoltre vale la proprietà distributiva: per ogni $x, y, z \in R$ si ha $(x+y)\cdot z = x\cdot z + y\cdot z$.*

Sono anelli gli insiemi $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ e $\mathbb{C}$ ed anche $\mathbb{Z}_m$, qualunque sia l'intero positivo m. Si noti che in alcuni anelli del tipo $\mathbb{Z}_m$ non vale la legge di annullamento del prodotto; la prossima definizione servirà a distinguere gli anelli con questa proprietà dagli altri.

Definizione B.4 (Divisore di zero) *Dato un anello R, un suo elemento $x \neq 0$ si dice* divisore di zero *se esiste $y \in R$ con $y \neq 0$ tale che $x\cdot y = 0$. Un anello privo di divisori di zero si dice* integro. *Indicheremo con R^* l'insieme degli elementi di R diversi da 0 e dai divisori di zero.*

Gli anelli integri sono precisamente quelli in cui vale la *legge di annullamento del prodotto* (se $x\cdot y = 0$ allora $x = 0$ o $y = 0$). Fra quelli che abbiamo considerato finora, sono integri gli anelli $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ e $\mathbb{C}$, oltre a tutti gli $\mathbb{Z}_p$ quando p è un numero primo. In $\mathbb{Z}_m$, m composto, $m = ab$, si ha che a e b sono divisori dello zero perché

$$ab = m \equiv 0 \bmod m$$

e quindi $\mathbb{Z}_m$, m composto, non è un anello integro.

Un altro concetto importante in un anello è quello di elemento invertibile.

Definizione B.5 (Unità) *Un elemento $u \in R$ si dice* unità *se è invertibile, cioè se esiste $v \in R$ tale che $u\cdot v = 1$.*

In $\mathbb{Z}_m$ gli elementi invertibili sono gli $a \in \mathbb{Z}_m$ tali che $(a,m) = 1$; in $\mathbb{Z}$ l'unico elemento invertibile è 1 mentre in $\mathbb{Q}$, $\mathbb{R}$ e $\mathbb{C}$ tutti gli elementi non nulli sono invertibili.

Classificheremo gli elementi di un anello secondo le loro proprietà rispetto alla moltiplicazione: per primi consideriamo 0 ed i suoi divisori, poi le eventuali *unità*. È fra tutti gli altri elementi che cercheremo i numeri primi: questo è il motivo astratto per cui il numero 1 non può essere considerato primo a cui facevamo riferimento nel §2.1.

Definizione B.6 (Associato di un elemento; divisore) *Due elementi x ed $y \in R$ si dicono* associati *se esiste un'unità $u \in R$ tale che $x = u\cdot y$; diremo che x è un* divisore *di y, e scriveremo $x \mid y$, se esiste un elemento $z \in R$ tale che $y = z\cdot x$.*

Definizione B.7 (Elemento irriducibile; elemento primo) *Diremo che x è* irriducibile *se x non è un'unità di R, ed i suoi divisori sono solo i suoi associati e le unità di R; diremo che p è* primo *se non è un'unità, e se $p \mid x \cdot y$ implica $p \mid x$ oppure $p \mid y$.*

Come si vede, è necessario distinguere fra *irriducibilità* e *primalità*, perché esistono anelli in cui i due concetti sono distinti. In $\mathbb{Z}$ i due concetti coincidono.

D'ora in poi scriveremo per definizione $a^1 = a$, $a^{n+1} = a \cdot a^n$ per ogni $n \in \mathbb{N}$. Scriveremo anche $a^0 = 1$ per ogni $a \neq 0$.

Vediamo un paio di esempi che illustrano come sia possibile costruire altri anelli a partire da quelli che abbiamo visto sopra.

Esempio B.8 *Prendiamo $\mathbb{Z}$, e consideriamo il* più piccolo *anello che contiene $\mathbb{Z}$ ed anche il numero reale $\sqrt{2}$. Questo anello si indica con $\mathbb{Z}[\sqrt{2}]$, e non è troppo difficile convincersi del fatto che $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : \ a, b \in \mathbb{Z}\}$. Non è neppure difficile vedere che questo insieme è effettivamente un anello: più complicato (e molto più interessante che in $\mathbb{Z}$) è il problema di determinarne le unità. Si può dimostrare che esistono infinite unità, come per esempio $1 + \sqrt{2}$, $3 + 2\sqrt{2}$, $7 + 5\sqrt{2}$, ..., e, più in generale, $(1 + \sqrt{2})^n$ per ogni $n \in \mathbb{Z}$, dato che $(1 + \sqrt{2})^{-1} = \sqrt{2} - 1 \in \mathbb{Z}[\sqrt{2}]$.*

Esempio B.9 *Un altro esempio classico di anello interessante è quello detto degli* interi di Gauss*: aggiungiamo a $\mathbb{Z}$ l'unità immaginaria* i*, e quindi abbiamo $\mathbb{Z}[\mathrm{i}] = \{a + b\mathrm{i} : \ a, b \in \mathbb{Z}\}$.*

A questo punto sorge dunque il problema di capire quali fra questi anelli hanno in comune con $\mathbb{Z}$ le proprietà più familiari (l'unicità della fattorizzazione, per fare un esempio). Per questo motivo introduciamo la definizione che segue.

Definizione B.10 (Anello euclideo) *Un anello integro R si dice* euclideo *se esiste un'applicazione $\delta : R^* \to \mathbb{N}$ detta* grado *tale che*

- *per ogni $a, b \in R^*$ si ha $\delta(a) \leq \delta(ab)$;*
- *per ogni $a \in R$ e per ogni $b \in R^*$ esistono $q, r \in R$ tali che*

  1. $a = q \cdot b + r$;
  2. $r = 0$ *oppure* $\delta(r) < \delta(b)$.

In sostanza, gli anelli euclidei sono quelli dove è possibile fare la *divisione euclidea* o *divisione con resto*: il più semplice esempio di anello euclideo è infatti $\mathbb{Z}$, con $\delta(n) = |n|$. In effetti, gli anelli euclidei sono quelli più simili a $\mathbb{Z}$, anche nel senso del prossimo teorema (che è l'analogo del Teorema Fondamentale dell'Aritmetica 3.1).

Teorema B.11 (Fattorizzazione unica negli anelli euclidei) *Sia $x \in R^*$ dove R è un anello euclideo. Se x non è un'unità, esistono $k \in \mathbb{N}$ e k elementi* primi *di R, p_1, …, p_k tali che $x = p_1 \cdots p_k$. Questa decomposizione è unica a meno dell'ordine dei fattori, e del cambiamento di qualcuno dei p_j in uno dei suoi associati.*

Esempio B.12 *Il Teorema di fattorizzazione unica non vale nell'anello $\mathbb{Z}[\mathrm{i}\sqrt{5}]$, come mostra l'esempio $6 = 2 \cdot 3 = (1+\mathrm{i}\sqrt{5}) \cdot (1-\mathrm{i}\sqrt{5})$. Dato che 2 è irriducibile e non divide nessuno dei fattori a destra, in questo anello 2 non è primo.*

Definizione B.13 (Anello dei polinomi) *Dato un anello R ed una* indeterminata *$x \notin R$, indicheremo con $R[x]$ l'insieme dei* polinomi *a coefficienti in R, e cioè il più piccolo anello che contenga R ed x.*

Dunque, sono polinomi (cioè elementi di $R[x]$) tutti le espressioni del tipo

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d = a_0 + \sum_{k=1}^{d} a_k x^k, \tag{10}$$

dove $a_k \in R$ per $k = 0, \ldots, d$. Infatti, dalla definizione di anello segue che se $x \in R[x]$, allora anche $x^2 \in R[x]$, e quindi, per esempio, anche $x^2 + x \in R[x]$, e così via. È necessario specificare "il più piccolo" nella definizione, perché, per esempio, $R[x]$ è contenuto in $R[x,y]$, l'insieme dei polinomi a coefficienti in R in *due* indeterminate.

Osserviamo anche che le notazioni $\mathbb{Z}[\sqrt{2}]$ e $\mathbb{Z}[x]$ sono coerenti: infatti, si tratta in ogni caso di anelli di polinomi, la cui differenza sta nel fatto che nel primo dei due abbiamo che $(\sqrt{2})^2 = 2 \in \mathbb{Z}$, e quindi possiamo semplificare tutte le espressioni che coinvolgono potenze di $\sqrt{2}$ di esponente ≥ 2.

Definizione B.14 (Grado e primo coefficiente di un polinomio) *Si dice* grado *di un polinomio $P \in R[x]$, e si indica con $\partial(P)$, il massimo intero k tale che nella rappresentazione (10) si ha $a_k \neq 0$. In questo caso, a_k si chiama* primo coefficiente *o* coefficiente direttivo *di P. Se $a_k = 0$ per ogni $k \in \mathbb{N}$, il polinomio P è il polinomio nullo che si indica con 0 ed al quale non si assegna né grado né primo coefficiente.*

Osserviamo che in $\mathbb{Z}_{12}$ (che non è integro visto che, ad esempio, 3 e 4 sono divisori dello zero) il polinomio $x^2 - 1$ ha più di una fattorizzazione in elementi irriducibili:

$$x^2 - 1 \equiv (x-1) \cdot (x+1) \equiv (x-5) \cdot (x+5) \bmod 12.$$

Una condizione sufficiente per avere fattorizzazione unica in anello di polinomi $R[x]$ è, come abbiamo visto avere che tale anello di polinomi sia *euclideo* tramite la funzione grado di un polinomio. Si può dimostrare che ciò certamente accade se R è una struttura che ammette ancora più proprietà di un anello.

Definizione B.15 (Campo) *Un anello commutativo con identità R si dice* campo *se $R \setminus \{0\}$ è un gruppo rispetto alla moltiplicazione.*

In tal caso si ha il seguente

Teorema B.16 *Se K è un campo, scelti $f \in K[x]$ e $g \in K[x] \setminus \{0\}$, esistono unici $q, r \in K[x]$ tali che $f(x) = q(x)g(x) + r(x)$, ed inoltre $r = 0$ oppure $\partial(r) < \partial(g)$. In altre parole, $K[x]$ è un anello euclideo.*

Allora, per il Teorema B.11, si ha che $K[x]$ è a fattorizzazione unica. Ad esempio $\mathbb{Q}[x]$, $\mathbb{R}[x]$, $\mathbb{C}[x]$ e $\mathbb{Z}_p[x]$, p numero primo, sono tutti anelli a fattorizzazione unica.

Tornando ora all'esempio dell'insieme $\mathcal{H}$ considerato nel §3, anche se può sembrare a prima vista piuttosto artificioso, è invece molto interessante dato che è precisamente l'insieme degli interi positivi dispari che possono essere decomposti come somma di due quadrati perfetti. Questa proprietà è una conseguenza di un altro Teorema di Fermat (dimostrato in modo tutto sommato elementare nel Capitolo 8 del libro di Conway e Guy [3]) secondo il quale ogni numero primo p della forma $4k+1$ è a sua volta decomponibile come somma di due quadrati, e dell'identità algebrica

$$(a^2+b^2)(c^2+d^2) = (ad \pm bc)^2 + (ac \mp bd)^2,$$

dalla quale si deduce che anche l'insieme dei numeri rappresentabili come somma di due quadrati è moltiplicativamente chiuso.

L'esempio dato nel testo è solamente uno di un'infinità di esempi possibili: presi p e q primi distinti della forma $4k+3$, il numero p^2q^2 può essere decomposto in $\mathcal{H}$ come $(pq)^2$ e come $p^2 \cdot q^2$, ed i numeri pq, p^2, q^2 sono primi di $\mathcal{H}$.

Insiemi aventi solamente la proprietà di chiusura rispetto alla moltiplicazione si chiamano *semigruppi* (che è quindi una struttura ancora più semplice di quella di gruppo), e la Figura 2 rappresenta una parte del semigruppo generato da 2, 3, 5, 7, cioè dell'insieme moltiplicativamente chiuso degli interi positivi i cui fattori primi appartengono tutti all'insieme $\{2, 3, 5, 7\}$.

C Letture ulteriori

Proprietà dei numeri primi Una trattazione introduttiva molto interessante e non tecnica delle proprietà dei numeri primi si trova nel Capitolo 5 del libro di Conway e Guy [3]. Per la chiarezza dell'esposizione e la presenza di numerosi esempi riteniamo che questo testo costituisca una buona risorsa per l'appassionato. Inoltre consigliamo di leggere l'articolo di Pomerance [12] che, sebbene

datato per quanto riguarda le ultime scoperte, introduce con chiarezza le tipologie di problemi e le strategie di attacco da utilizzare per ricercare i numeri primi. Per chi vuole conoscere gli ultimi risultati (e non ha problemi con la lingua inglese) consigliamo l'articolo di Granville [6] in cui viene presentato anche il recente risultato di Agrawal-Kayal-Saxena [1] sull'esistenza di un algoritmo di primalità avente complessità polinomiale nel numero delle cifre dell'intero sotto esame. Una presentazione in italiano di tale algoritmo si può anche trovare in [11].

Per chi vuole andare oltre un livello puramente amatoriale e si vuole avvicinare più professionalmente al mondo della Teoria dei Numeri, consigliamo il libro di Davenport [5] per un'introduzione generale ed il libro (in lingua inglese) di Crandall e Pomerance [4] per gli aspetti computazionali sui numeri primi. Parte di tali aspetti sono anche presentati in [11].

Applicazioni alla crittografia Negli ultimi anni sono stati pubblicati diversi articoli in lingua italiana che presentano a livello introduttivo vari aspetti delle applicazioni della Teoria dei Numeri alla Crittografia. In realtà questi lavori hanno tra loro delle intersezioni non banali, ma siccome il punto di vista adottato è usualmente diverso da caso a caso pensiamo che sia utile citarli tutti. In ordine di tempo ricordiamo l'articolo di Schoof [13] in cui viene presentato il collegamento tra metodo RSA e la fattorizzazione di interi; Languasco e Perelli [9]-[10] in cui si fornisce una succinta storia di alcuni risultati sulla distribuzione dei numeri primi e sulla crittografia a chiave pubblica nonché sulla firma digitale; Zaccagnini [15] in cui si analizza il ruolo della primalità nei metodi crittografici basati sul logaritmo discreto ed Alberti [2] in cui viene proposta una presentazione adatta agli studenti delle Scuole Superiori dei concetti base dell'Aritmetica finita e della crittografia a chiave pubblica.

Una trattazione ben più estesa delle problematiche relative alle applicazioni teorico-numeriche si trova nella recente monografia [11] in cui, dopo aver fornito le nozioni matematiche di base necessarie, sono presentati vari algoritmi crittografici basati sia sulla fattorizzazione che sul problema del logaritmo discreto nonché diversi algoritmi di primalità e fattorizzazione. Inoltre sono descritti alcuni protocolli crittografici (senza dettagli implementativi) e uno strumento di calcolo (PARI/Gp) che permette di svolgere calcoli didatticamente significativi senza dover ricorrere a veri e propri linguaggi di programmazione.

Per chi non ha problemi con la lingua inglese consigliamo anche il libro di Koblitz [8] che rappresenta un'ottima introduzione al mondo delle applicazioni della Teoria dei Numeri alla Crittografia.

Siti web Anche in questo caso notiamo che la maggior parte delle informazioni sul web non sono in lingua italiana. A costo di doverci auto-citare di nuovo,

facciamo presente che sono a disposizione dei nostri Lettori alcune pagine web, accessibili dai link
www.math.unipd.it/~languasc/crittografia/Crittografia.html
www.math.unipr.it/~zaccagni/crittografia/Crittografia.html
nelle quali abbiamo raccolto link a materiale interessante (principalmente dal punto di vista crittografico, ma anche per le questioni più strettamente legate ai problemi generali qui trattati), e del software commentato che permette di vedere delle realizzazioni pratiche degli algoritmi discussi in questa serie di articoli.

Riferimenti bibliografici

[1] M. Agrawal, N. Kayal, N. Saxena, "PRIMES is in P", in corso di pubblicazione, 2002, disponibile sul sito http://www.cse.iitk.ac.in/news/primality_v3.pdf.

[2] G. Alberti, "Aritmetica finita e crittografia a chiave pubblica - Un percorso didattico per gli studenti delle Scuole Medie Superiori", in A. Abbondandolo, M. Giaquinta, F. Ricci (a cura di), *Ricordando Franco Conti*, pagine 1-29, Scuola Normale Superiore, Pisa, 2004.

[3] J.H. Conway, R.K. Guy, *Il libro dei numeri*, Hoepli, Milano, 1999.

[4] R. Crandall, C. Pomerance, *Prime numbers. A computational perspective*, Springer-Verlag, Berlin, Heidelberg, New York, 2001.

[5] H. Davenport, *Aritmetica Superiore*, Zanichelli, Bologna, 1994.

[6] A. Granville, "It is easy to determine whether a given integer is prime", *Bulletin A.M.S.*, 42, pagine 3-38, 2005, Disponibile all'indirizzo http://www.ams.org/bull/2005-42-01/S0273-0979-04-01037-7/home.html.

[7] G.H. Hardy, E.M. Wright, *An Introduction to the Theory of Numbers*, Oxford Science Publications, Oxford, quinta edizione, 1979.

[8] N. Koblitz, *A Course in Number Theory and Cryptography*, Springer-Verlag, Berlin, Heidelberg, New York, seconda edizione, 1994.

[9] A. Languasco, A. Perelli, "Numeri Primi e Crittografia", in M. Emmer (a cura di), *Matematica e Cultura 2000*, pagine 227-233, Venezia, 2000, Springer-Verlag, Milano, trad. inglese in *Mathematics and Culture I*, Springer-Verlag, Berlin, Heidelberg, New York, 2003.

[10] A. Languasco, A. Perelli, "Crittografia e firma digitale", in M. Emmer, M. Manaresi (a cura di), *Matematica, Arte, Tecnologia, Cinema*, pagine 99-106, Bologna, 2002, Springer-Verlag, Milano, trad. inglese in *Mathematics, Art, Technology, and Cinema*, Springer-Verlag, Berlin, Heidelberg, New York, 2003.

[11] A. Languasco, A. Zaccagnini, *Introduzione alla Crittografia*, Ulrico Hoepli Editore, Milano, 2004.

[12] C. Pomerance, "Alla ricerca dei numeri primi", *Le Scienze*, 174, pagine 86-94, febbraio 1983.

[13] R. Schoof, "Fattorizzazione e criptosistemi a chiave pubblica", *Didattica delle Scienze*, 137, pagine 48-54, 1988, Disponibile all'indirizzo `http://www.mat.uniroma2.it/~eal/eal2005.html`.

[14] A. Zaccagnini, Variazioni Goldbach: problemi con numeri primi, *L'Educazione Matematica, Anno XXI, Serie VI*, 2, pagine 47–57, 2000, `http://www.math.unipr.it/~zaccagni/psfiles/papers/Goldbach_I.pdf`.

[15] A. Zaccagnini, "L'importanza di essere primo", in A. Abbondandolo, M. Giaquinta, F. Ricci (a cura di), *Ricordando Franco Conti*, pagine 343-354, Scuola Normale Superiore, Pisa, 2004.

[16] A. Zaccagnini, Lezioni di Teoria dei Numeri, 2005, Dispense del Corso di *Teoria dei Numeri*, A. A. 2004-2005. Disponibili all'indirizzo `http:/www.math.unipr.it/~zaccagni/psfiles/lezioni/tdn2005.pdf`.

Alessandro Languasco
Dipartimento di Matematica Pura e Applicata,
via Belzoni 7, 35131 Padova
e-mail: languasco@math.unipd.it
pagina web: http://www.math.unipd.it/~languasc

Alessandro Zaccagnini
Dipartimento di Matematica,
via Massimo d'Azeglio, 85/a, 43100 Parma
e-mail: alessandro.zaccagnini@unipr.it
pagina web: http://www.math.unipr.it/~zaccagni/home.html