Il teorema dei numeri primi

a cura di Flavio Cimolin

(ultimo aggiornamento: 20/02/2008)
Pubblicato su Matematicamente.it Magazine n.5, Gennaio 2008

Da millenni i numeri primi e le loro bizzarre proprietà stuzzicano la curiosità di matematici e appassionati. I problemi in cui essi vengono tirati in ballo appaiono infatti talmente semplici nel modo in cui possono essere espressi da risultare un continuo sberleffo nei confronti degli eminenti matematici di tutto il mondo che tentano invano di risolverli. Basti pensare alle famosissime congetture tuttora irrisolte di Goldbach (ogni numero pari maggiore di due si può esprimere come somma di due primi) e dei primi gemelli (esistono infinite coppie di primi a distanza due). Solo recentemente e con grande clamore è stata annunciata la dimostrazione del cosiddetto Ultimo Teorema di Fermat (non esistono "terne pitagoriche" con esponenti maggiori di due), la quale tuttavia si basa su artifici matematici talmente complessi da richiedere mesi e mesi di studio ai più esperti matematici solo per intuire il percorso della dimostrazione!

I numeri primi non sono altro che una "base moltiplicativa" dell'insieme dei numeri naturali, alternativa a quella additiva sulla quale esso è definito (la definizione formale dei numeri naturali segue infatti il principio di induzione: ad ogni numero N ne segue un altro $N+1$, dato dalla somma del precedente più uno. Così nascono i numeri 1, 2, 3, 4, 5, …, N, $N+1$, …). Nonostante i primi compaiano in maniera assolutamente lineare nonappena si introduca nell'insieme dei naturali l'operazione di moltiplicazione, per qualche oscura ragione essi sono come avvolti da un alone di mistero che impedisce di comprendere a fondo i loro speciali intrecci. Fra i numerosissimi problemi che si possono formulare riguardo ai numeri primi, ci occupiamo qui di presentarne uno che ha a che fare con il loro "conteggio": quanti sono i numeri primi? E, se sono infiniti, con quale tipo di legge crescono? La risposta a questa domanda è stata trovata e va sotto il nome di Teorema dei Numeri Primi. L'avventura che ha condotto eminenti matematici dei tre secoli scorsi, come Gauss, Riemann, Hadamard e De la Vallèe Poussin, verso la dimostrazione del tutt'altro che semplice risultato è stata talmente trasversale da chiamare in causa addirittura i numeri complessi e le più avanzate tecniche analitiche che operano su di essi, tanto da condurre a formulare un altro celebre problema, passato alla storia col nome di Ipotesi di Riemann.

Anzitutto è ragionevole domandarsi quanti siano i numeri primi, nel caso essi fossero solo un numero finito. Su questo punto già Euclide nel IV secolo a.C. sollevò il primo velo, mostrando che esistono infiniti numeri primi. La sua dimostrazione classica (a cui ne sono seguite decine di altre nel corso dei secoli) è talmente semplice da poter essere esposta per intero. Supponiamo per assurdo che esista solo un numero finito di numeri primi. Allora li possiamo elencare tutti dal più piccolo al più grande: chiamiamoli $p_1, p_2, p_3, \ldots, p_n$. Adesso andiamo a calcolare il numero

$$N = p_1 \cdot p_2 \cdot p_3 \cdot \ldots \cdot p_n + 1$$

N è divisibile per p_1? No, perché la parte ove c'è il prodotto lo è, ma ad essa è aggiunto un 1, quindi la divisione di N per p_1 dà resto 1. E' divisibile per p_2? No, per lo stesso motivo di prima, in quanto il resto della divisione di N per p_2 è 1. Questo vale evidentemente per tutti i p_i, quindi N non può essere divisibile per nessuno dei numeri primi che abbiamo elencato. Allora deve essere un numero primo esso stesso (oppure il prodotto di altri numeri primi diversi da quelli elencati), ma ciò va contro la nostra ipotesi iniziale di aver potuto elencare tutti i numeri primi. Conclusione: esistono infiniti numeri primi! Vale la pena osservare che la tecnica di Euclide serve a dimostrare che, dato un insieme di numeri primi, esiste sempre un numero più grande che sia coprimo con essi (ovvero che non abbia fattori comuni con alcuno di essi). Questo non significa che il numero N così costruito sia esso stesso primo, come mostra chiaramente il seguente controesempio:

$$2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 + 1 = 30031 = 59 \cdot 509$$

Bene, ora che abbiamo dimostrato che i numeri primi sono infiniti, ha senso chiederci quanti siano rispetto ai numeri naturali, nel senso che andremo subito a precisare. Ci sono vari modi di confrontare tra loro insiemi infiniti, ad esempio seguendo la teoria di Cantor si potrebbe dimostrare che i numeri primi hanno la stessa cardinalità dei numeri naturali; questo tuttavia ci sarebbe di poco aiuto, in quanto per conoscere effettivamente qualcosa in più sulla distribuzione dei primi abbiamo bisogno di un risultato quantitativo. Occorre cioè trovare un modo formale per indicare che, ad esempio, i numeri pari sono "la metà" dei naturali, i multipli di tre sono "un terzo", e così via, in modo da poter dire anche "quanti sono" i quadrati perfetti, i cubi o i numeri primi appunto.

Il modo corretto di affrontare la questione dal punto di vista matematico è quello di definire un "contatore" della tipologia di numeri che si vuole considerare, ovvero una funzione che indichi quante occorrenze di numeri di quella specie ci sono fino ad un valore dato. Chiamiamo ad esempio $S_{pari}(n)$ la funzione che indica quanti sono i numeri pari da 1 fino ad n. Se n è pari, $S_{pari}(n) = n/2$, se n è dispari $S_{pari}(n) = (n-1)/2$. Supponendo ora che n diventi sempre più grande si può approssimare considerando solo l'ordine di grandezza principale e dire che $S_{pari}(n) \approx n/2$. Una procedura di questo genere è detta in gergo una "stima asintotica", in quanto non fornisce il valore esatto, ma fornisce un valore che è tanto più prossimo a quello giusto quanto più i numeri coinvolti crescono. Questo modo di procedere costituisce esattamente la formalizzazione matematica di quello che abbiamo in mente quando pensiamo che "i numeri pari sono la metà dei naturali". E' semplice osservare che per i numeri dispari varrà analogamente $S_{dispari}(n) \approx n/2$, per i multipli di tre varrà $S_{tre}(n) \approx n/3$, e così via. Seguendo le regole delle stime asintotiche, non è difficile mostrare che per i quadrati perfetti vale $S_{quadrati}(n) \approx sqrt(n)$ e per le potenze di due $S_{potenze2}(n) \approx log_2(n)$.

Un modo perfettamente speculare di vedere la stessa cosa è quello di chiedersi quanto valga l'n-esimo numero di un certo tipo. Chiamiamo $N_{pari}(n)$ l'n-esimo numero pari. Invertendo la relazione per $S_{pari}(n)$ avremo che $N_{pari}(n) \approx 2n$. Volutamente non abbiamo inserito il simbolo di uguaglianza, ma quello di approssimazione, in quanto, nell'ottica di sostituire i pari con i numeri primi, non pretendiamo di conoscere esattamente il valore dell'n-esimo numero, ma ci accontentiamo di averne una sua stima. Ecco perché, facendo lo stesso con i dispari, avremmo $N_{dispari}(n) \approx 2n$; e poi $N_{tre}(n) \approx 3n$, $N_{quadrati}(n) \approx n^2$, $N_{potenze2}(n) \approx 2^n$. Le funzioni N_{xxx} sono esattamente le inverse delle funzioni S_{xxx}, capito come funziona il gioco?

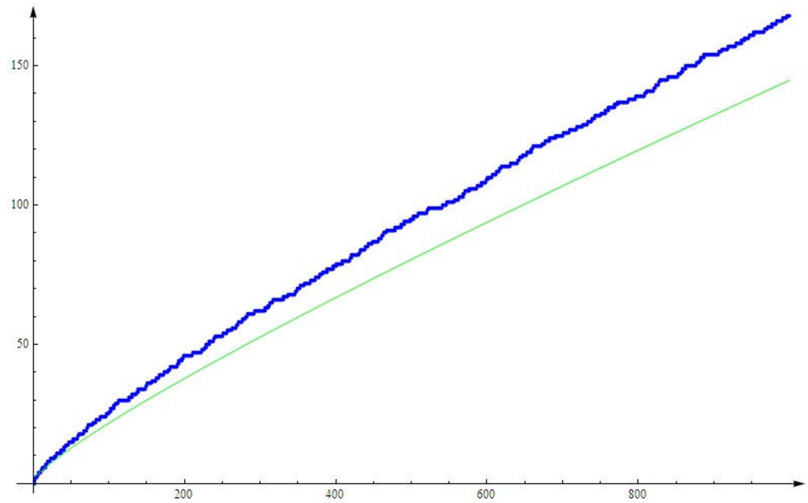
Allora è finalmente venuto il momento di chiederci: "Quanto valgono $S_{primi}(n)$ e $N_{primi}(n)$"? La tutt'altro che banale risposta a questa domanda va sotto il nome di Teorema dei Numeri Primi, che afferma che:

$$\boxed{S_{primi}(n) \approx \frac{n}{\log(n)} \qquad N_{primi}(n) \approx n \cdot \log(n)}$$

dove $log(n)$ indica il logaritmo naturale di n.

Questo significa che il numero di primi che ci aspettiamo di trovare da *1* fino al numero *n* è circa *n/log(n)*, o viceversa che l'*n*-esimo numero primo è approssimativamente *n·log(n)*. Scegliete voi quale delle due formule (equivalenti) vi piace di più per indicare "quanti sono i numeri primi" rispetto ai naturali. Anche se le due formule appaiono in perfetta simmetria, l'una con una divisione per il logaritmo, l'altra con una moltiplicazione per lo stesso, non è del tutto ovvio il fatto che siano l'una l'inversa asintotica dell'altra, cosa che invece era direttamente verificabile negli esempi precedenti.

Per convincersi di cosa significhi esattamente una stima asintotica per dei numeri, come i primi, che notoriamente incontrano difficoltà a inquadrarsi in regole precise, proviamo a confrontare l'andamento della funzione *x/log(x)* (in verde) con i valori effettivi di $S_{primi}(n)$ (in blu) per i primi *1000* valori di *n*.



Come si vede la corrispondenza è buona, in quanto la crescita della funzione approssimante segue bene l'andamento della funzione che conta i primi, ma ben lungi dall'essere perfetta: come abbiamo già detto, i numeri primi non ne vogliono sapere di sottostare a regole troppo precise! Andando a guardare valori sempre più grandi di *n* si scopre come le due curve si avvicinino (relativamente al valore di *n*) sempre di più. Vedremo più avanti che in realtà esistono diverse altre formulazioni più potenti del Teorema dei Numeri Primi, grazie alle quali si raggiungono livelli di precisione incredibilmente accurati, talmente in accordo che risulterebbe difficile visualizzare in un grafico come il precedente la differenza fra la curva esatta e quella approssimata, per numeri grandi.

La storia del Teorema dei Numeri Primi è decisamente affascinante, e comincia inesorabilmente con i fallimenti di due dei più grandi matematici di tutti i tempi, che costruirono le basi della teoria dei numeri: Eulero e Gauss. Era il 1751 quando il matematico svizzero Leonhard Euler affermò che

Scoprire qualche ordine nella progressione dei numeri primi è un mistero che lo spirito umano non sarà mai in grado di penetrare

A posteriori possiamo dire che qualcosa siamo stati effettivamente in grado di scoprirlo, ma tutti i matematici concordano sul fatto che Eulero avesse pienamente ragione riguardo all'impossibilità di svelare completamente il mistero dei numeri primi. Chi per primo riuscì ad intuire l'andamento del tipo *x/log(x)* fu il grande matematico e fisico tedesco Carl Friedrich Gauss (1777-1855), che in realtà non fece altro che calcolare un bel po' di valori della funzione $S_{primi}(n)$ e poi, usando la sua brillante mente matematica, individuare quale tipo di curva si avvicinasse di più ad essa. Senza dubbio il fatto che un genio come Gauss non sia riuscito a dimostrare il teorema dei numeri primi la dice lunga su quanto la cosa sia più complicata!

Si è dovuto aspettare la fine del XIX secolo per vedere finalmente risolta la questione, per giunta in un modo assolutamente inatteso che passa addirittura attraverso l'analisi complessa. Il passo decisivo grazie al quale è stato possibile "attaccare" il problema è stato compiuto dal matematico tedesco Bernhard Riemann (1826-1866), allievo di Gauss, morto ahimè prematuramente per via della tubercolosi, ma dallo spirito talmente geniale da risultare il vero e proprio iniziatore non solo della rivoluzione che andremo a descrivere riguardo al legame fra i numeri primi e l'analisi complessa, ma anche di tutti quegli strumenti di geometria differenziale senza i quali Albert Einstein mezzo secolo dopo non sarebbe mai riuscito a dare una formulazione matematica alla sua teoria della relatività generale.

La stragrande maggioranza dei contribuiti di Riemann alla matematica si colloca nei campi della geometria differenziale e nel modo di approcciarsi ad essa, così come nell'analisi matematica (basti pensare alla formalizzazione del cosiddetto "Integrale di Riemann"). Egli dedicò alla teoria dei numeri solo un unico saggio di una decina di pagine, pubblicato nel 1859 sulle note dell'Accademia di Berlino, nel quale espresse in tutta la sua eleganza il rapporto fra la teoria dei numeri e l'analisi complessa e nel quale formulò uno dei problemi irrisolti più famosi di tutti i tempi, che va sotto il nome di *Ipotesi di Riemann*.

Il legame fra l'analisi complessa e la teoria dei numeri non è affatto ovvio e la dimostrazione completa del Teorema dei Numeri Primi è lunga e complicata. Proviamo comunque a dare una idea di come mai i numeri primi siano legati all'analisi complessa, partendo da quella che Eulero aveva chiamato funzione Zeta:

$$\zeta(x) = \sum_{k=1}^{+\infty} \frac{1}{k^x} = \frac{1}{1^x} + \frac{1}{2^x} + \frac{1}{3^x} + \frac{1}{4^x} + \frac{1}{5^x} + \ldots$$

Si tratta di una funzione (di una variabile reale x) data della somma infinita dei reciproci di tutti i numeri naturali elevati alla potenza x. La funzione è ben definita solo per gli $x > 1$, infatti è ben noto che la somma dei reciproci dei numeri naturali cresce indefinitamente, seppure in maniera estremamente lenta. Per qualsiasi potenza anche solo leggermente maggiore di 1 invece si ha convergenza della serie numerica. In alcuni casi speciali è addirittura possibile calcolare il valore esatto della somma infinita, come ad esempio accade per la bellissima formula che lega la somma dei reciproci dei quadrati al numero pi greco:

$$\zeta(2) = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \ldots = \frac{\pi^2}{6}$$

Vediamo ora quale sia il legame fra la funzione Zeta e i numeri primi. Lo possiamo vedere subito trasformando la somma dei reciproci in un prodotto che coinvolga somme relative ai soli numeri primi. Per semplicità mostriamo la relazione algebrica valore unitario dell'esponente, anche se in corrispondenza di tale valore la serie non converge: è banale verificare che la relazione vale in realtà per qualsiasi esponente (ovvero elevando i denominatori di tutte le frazioni all'esponente x).

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \ldots = \left(1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \ldots\right) \cdot \left(1 + \frac{1}{3} + \frac{1}{3^2} + \frac{1}{3^3} + \ldots\right) \cdot \left(1 + \frac{1}{5} + \frac{1}{5^2} + \frac{1}{5^3} + \ldots\right) \cdot \ldots$$

La somma dei reciproci dei naturali può cioè essere scomposta come il prodotto delle somme dei reciproci delle potenze dei numeri primi! Se vi sembra solo un modo insulso di complicare le cose, provate a svolgere alcuni prodotti e capirete immediatamente come funziona il meccanismo, che è direttamente legato al teorema di fattorizzazione unica dei numeri naturali: per ogni frazione *1/n* sulla sinistra, c'è una e una sola combinazione di frazioni sulla destra che la generano, le quali si deducono dalla scomposizione in fattori primi di *n*. Sfruttando la formula che consente di trovare la somma di serie geometriche, la relazione precedente si può scrivere in un modo ancora più semplice, che illustra precisamente come la *somma* dei reciproci dei naturali si possa esprimere come un *prodotto* che coinvolge solo i numeri primi:

$$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\ldots=\frac{1}{1-\frac{1}{2}}\cdot\frac{1}{1-\frac{1}{3}}\cdot\frac{1}{1-\frac{1}{5}}\cdot\ldots=\prod_p\frac{1}{1-\frac{1}{p}}$$

Quanto abbiamo visto non è nulla di più che un'intuizione puramente qualitativa del possibile legame fra la funzione Zeta e i numeri primi. Di qui alla dimostrazione del Teorema dei Numeri Primi la strada sarebbe ancora molto molto lunga. Il grande passo avanti fu compiuto nel momento in cui Riemann mostrò che in realtà l'esponente che compare come argomento della funzione Zeta di Eulero può essere anche inteso come un numero complesso, ovvero caratterizzato da una parte reale e una parte immaginaria! L'analisi complessa è una delle discipline più affascinanti della matematica moderna e si occupa precisamente di studiare le funzioni, con le usuali regole di derivazione e integrazione, nel campo dei numeri complessi anziché nel campo dei numeri reali. Il bello è che le funzioni di variabile complessa possiedono proprietà molto più curiose e interessanti di quelle di variabile reale, ad esempio in alcuni casi è possibile conoscere tutti i valori di una funzione in un'intera zona conoscendo solamente i valori della funzione stessa su di un cerchio!

La scoperta fondamentale che compì Riemann nello studio della funzione Zeta (che da allora in poi assunse il nome di funzione Zeta di Riemann), fu che essa poteva essere estesa ad un dominio più grande di quello che appariva originariamente nello studio di Eulero. Se ricordate abbiamo detto che l'esponente x doveva essere un numero maggiore di *1*; ebbene Riemann mostrò invece come questo esponente poteva essere un qualsiasi numero complesso diverso da *1*! Questa scoperta ha dell'incredibile, infatti non è apparentemente razionale pensare che abbia senso parlare di numeri finiti del tipo:

$$\zeta(-1)=\frac{1}{1^{-1}}+\frac{1}{2^{-1}}+\frac{1}{3^{-1}}+\ldots=1+2+3+\ldots \qquad \zeta(-2)=\frac{1}{1^{-2}}+\frac{1}{2^{-2}}+\frac{1}{3^{-2}}+\ldots=1^2+2^2+3^2+\ldots$$

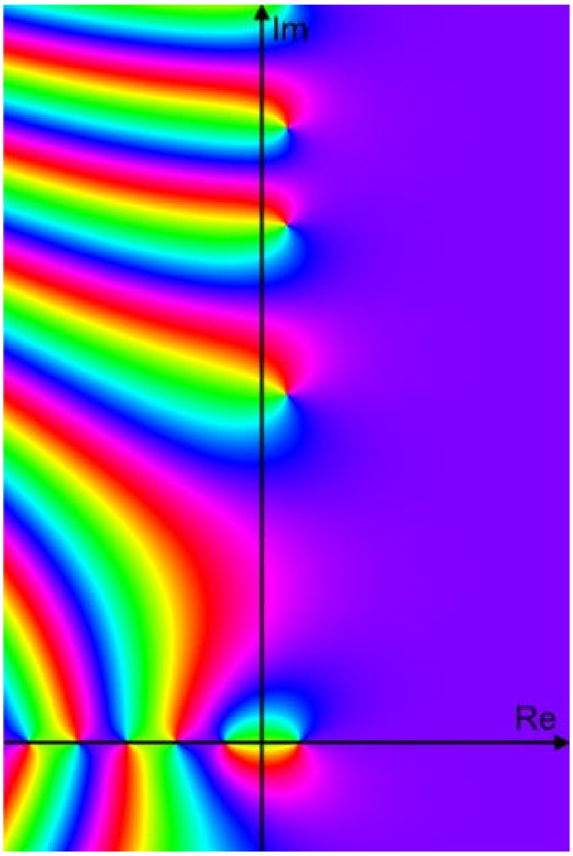
Il modo corretto di approcciare questa estensione è differente: Riemann dimostrò che la funzione Zeta di Eulero è esattamente coincidente con un'altra funzione (detta Zeta di Riemann appunto) che risulta ben definita per qualsiasi numero complesso diverso da *1*. Dato esiste un teorema di analisi complessa che afferma che se due funzioni complesse coincidono su di una porzione di dominio allora devono coincidere su tutto il dominio, allora ecco che possiamo pensare alla scrittura come sommatoria (Zeta di Eulero classica) come ad un possibile modo di scrivere la funzione più generale (Zeta di Riemann) quando l'esponente ha parte reale maggiore di *1*.

Purtroppo è impossibile fornire una descrizione vera e propria dei passaggi che, dalla cosiddetta estensione analitica della Zeta di Riemann, conducono alla dimostrazione del Teorema dei Numeri Primi, perché questo richiederebbe, oltre ad una perfetta dimestichezza con le tecniche dell'analisi complessa, anche un discreto numero di pagine. E' però interessante raccontare ancora qualche risvolto "coreografico" della vicenda, che infatti è tutt'altro che chiusa.

Va detto infatti che Riemann nel suo articoletto usò un'ipotesi (che in seguitò passò alla storia con il nome Ipotesi di Riemann) che, per quanto potesse sembrare ragionevole, non fu in grado di dimostrare. La "scontatezza" della verità di questa ipotesi ingannò la comunità scientifica per oltre mezzo secolo, dato che tutti i più eminenti teorici dei numeri erano assolutamente convinti che si trattasse di poca cosa e che la dimostrazione del teorema fosse effettivamente stata ottenuta. In realtà l'ipotesi di Riemann regge ancora oggi ed è uno dei più grandi problemi insoluti della matematica moderna. Centinaia di articoli di ricercatori attivi nel campo della teoria dei numeri da un secolo a questa parte iniziano con la dicitura "Se l'ipotesi di Riemann fosse vera, allora…".

L'ipotesi di Riemann afferma che gli zeri non banali della funzione Zeta di Riemann (che come abbiamo detto è definita sui numeri complessi) hanno tutti parte reale esattamente uguale a *1/2*. Tutti gli zeri della funzione Zeta aventi parte reale negativa sono semplici da calcolare e sono per questo detti "banali", ma finora nessuno è stato in grado di valutare esattamente quelli che si collocano nella sottile striscia compresa fra parte reale *0* e parte reale *1*. Se si va a guardare come la Zeta interviene nella dimostrazione del Teorema dei Numeri Primi si scopre che è esattamente questa la parte fondamentale che consente di ottenere il risultato, che risulta tanto migliore quanto più gli zeri sono vicini all'asse *1/2*. E' curioso notare che il motivo principale per cui Riemann formulò la sua famosa "ipotesi" era essenzialmente di tipo estetico: tutta la teoria risulterebbe particolarmente elegante se gli zeri della funzione Zeta avessero tutti parte reale *1/2* e si otterrebbe immediatamente il migliore risultato possibile; quindi per Riemann l'ipotesi non poteva che essere vera!

E' possibile calcolare numericamente i punti in cui la Zeta si annulla, cosa che è stata effettivamente fatta per migliaia di valori. Ma come ben sanno i matematici, una congettura non è dimostrata fino a quando non si dimostra che essa è valida per *tutti* i valori! E' curioso "guardare" la funzione Zeta disegnata nel piano complesso, rappresentata nella figura a lato. Ovviamente per darne un grafico occorrerebbe usare uno spazio a 4 dimensioni, che non possediamo (le funzioni a variabile complessa sono caratterizzate da due numeri, la parte reale e la parte immaginaria, e il dominio di definizione è composto esso stesso da altri due numeri). Con alcuni trucchi si può però dare ugualmente una rappresentazione, ad esempio mostrando solo la parte reale o solo quella immaginaria. Nella figura qui a fianco invece si usa una rappresentazione per mezzo di colori dell'argomento del numero complesso: i punti in cui convergono tutti i colori sono zeri oppure poli (si distinguono a seconda del verso di rotazione dei colori). E si vede chiaramente che, salvo il polo e gli zeri sull'asse reale, gli unici altri zeri appaiano ben allineati sulla retta *Re=1/2*!



Ma torniamo al Teorema dei Numeri Primi vero e proprio. Nonostante nessuno sia riuscito a dimostrare l'ipotesi di Riemann, nel frattempo i teorici dei numeri si sono dati un gran daffare per trovare strade alternative che consentissero di dimostrare il teorema *senza* quella ipotesi. Nel 1896, finalmente, il matematico francese Jacques Hadamard contemporaneamente al belga Charles Jean de la Vallée Poussin, indipendentemente, riuscirono a completare la dimostrazione, sempre utilizzando metodi di analisi complessa simili a quelli introdotti da Riemann. La portata della dimostrazione, che come abbiamo detto si basa su strumenti estremamente evoluti di analisi complessa per stabilire una relazione riferita ai soli numeri naturali, è brillantemente esplicitata nello storico aforisma di Hadamard:

La via più breve fra due verità sulla retta dei numeri reali passa attraverso il piano complesso

Con il passare degli anni i matematici sono riusciti a semplificare le dimostrazioni proposte alla fine del XVIII secolo, ma solo verso la metà del XX secolo con i matematici Paul Erdos e Atle Selberg si è riusciti finalmente ad uscire dal campo dei numeri complessi, fornendo una dimostrazione elementare del Teorema dei Numeri Primi. Nel contesto della teoria dei numeri l'aggettivo "elementare" non è sinonimo di "semplice", bensì indica che la dimostrazione non necessita di concetti di analisi complessa. E infatti tali dimostrazioni sono terribilmente lunghe e artificiose, tanto da non fornire alcun aiuto ulteriore alla comprensione del problema (come spesso accade purtroppo quando si attacca in modo "elementare" un problema che può essere risolto più agevolmente in campo complesso).

In precedenza avevamo accennato al fatto che il Teorema dei Numeri Primi si può esprimere in una forma ancora più precisa rispetto all'andamento di tipo $x/\log(x)$ per la funzione $S_{primi}(n)$. Molti matematici del secolo scorso hanno speso buona parte del loro tempo per trovare piccoli miglioramenti della formula asintotica e del suo "resto", ovvero della stima dell'errore che si commette con tale approssimazione. Sappiamo certamente che tale resto non è nullo, ma è molto interessante capire quanto sia "buona" la stima che siamo in grado di fornire rispetto alla curva esatta. La migliore stima possibile per la funzione $S_{primi}(n)$ è definita tramite quello che viene detto Logaritmo Integrale e indicato con la funzione $li(n)$:

$$S_{primi}(n) \approx li(n) = \int_2^n \frac{1}{\log(x)}\,dx$$

Si può dimostrare che il logaritmo integrale è effettivamente un "miglioramento" della formula $x/\log(x)$, infatti quest'ultima compare come termine principale del suo sviluppo in serie asintotico. Il risvolto più interessante di questa nuova formula consiste però in una curiosa interpretazione di tipo statistico che essa consente di dare al problema del conteggio dei numeri primi.

La cosiddetta "Interpretazione Statistica" del Teorema dei Numeri Primi parte dall'assunzione (totalmente falsa!) che un numero sia primo oppure no con una certa probabilità. Sulla base di ricerche prevalentemente empiriche, già Gauss aveva notato che era ragionevole supporre che, dato un numero X, la probabilità che X fosse primo si aggira intorno al valore $1/\log(X)$. Dando per valida questa ipotesi, che difficilmente si potrebbe giustificare in maniera formale (per questo si parla di "interpretazione", e non di "teorema"), ecco che il Teorema dei Numeri Primi nella sua formulazione con il Logaritmo Integrale diventa di immediata comprensione. Per "contare" i numeri primi minori o uguali a un certo numero x non dobbiamo fare altro che sommare tutte le probabilità associate ai numeri da 1 a x. La somma su un numero grande si può per definizione approssimare con un integrale, e quindi ecco che il Teorema dei Numeri Primi si può leggere come una "media statistica" dei numeri primi minori di un certo numero:

$$S_{primi}(n) \approx \sum_{i=2}^{n} p(i) = \sum_{i=2}^{n} \frac{1}{\log(x)} \approx \int_2^n \frac{1}{\log(x)}\,dx$$

E' decisamente inusuale in una disciplina matematica così rigorosa come la teoria dei numeri utilizzare ragionamenti poco formali come questo, tuttavia mai come in questo caso l'interpretazione statistica riesce a rendere intuitiva una formula che altrimenti non lo sarebbe stata. Da notare che essa non ha nulla a che vedere con i test di primalità di tipo statistico, che si basano invece su teoremi rigorosi e che consentono di valutare in maniera il più rapida possibile, con un grado di certezza che può essere reso piccolo a piacere, se un numero è primo oppure no. Più i numeri in gioco diventano grandi e più diventa infatti difficile stabilire con certezza se un numero è effettivamente primo oppure no. Nell'interpretazione della nostra formula, invece, la "probabilità di essere primo" è un concetto vago che si applica, per intenderci, anche ai numeri pari, che primi non lo sono di sicuro.

Si potrebbe ancora continuare oltre con altri risultati di tipo statistico, che diventano sempre più interessanti man mano che l'intervallo di numeri che si prende in considerazione diventa più piccolo. Non si può non citare un'altra splendida "perla" secondo cui il numero di fattori primi di n è "in media" $log(log(n))$. Più precisamente, i matematici Hardy e Ramanujan hanno dimostrato che il numero di fattori primi di un numero n "scelto a caso" è distributo secondo una variabile aleatoria gaussiana avente media $log(log(n))$ e varianza ancora $log(log(n))$!

Per concludere, è interessante provare a vedere cosa succede se anziché studiare i numeri primi, ne analizziamo i loro reciproci. Si può dimostrare che, così come la somma dei reciproci dei numeri naturali (detta serie armonica), anche la somma dei reciproci dei numeri primi diverge:

$$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\ldots=+\infty \qquad 1+\frac{1}{2}+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\ldots=+\infty$$

Questa è, se vogliamo, un'altra dimostrazione del fatto che i numeri primi sono infiniti, e afferma, in un senso diverso da tutti gli altri visti finora, che i numeri primi sono ancora "tanti" rispetto ai numeri naturali. Se la somma dei loro reciproci fosse stata un numero finito, allora avremmo avuto un'indicazione ben diversa sulla loro "quantità". Per esempio la somma dei reciproci dei quadrati dei naturali (lo abbiamo già visto quando abbiamo scritto $\zeta(2)$), o delle potenze di 2, sono dei numeri finiti, pur essendo essi stessi in numero infinito. Per avere una misura della differenza fra il comportamento delle somme dei reciproci dei naturali e dei numeri primi, bisogna ancora una volta ricorrere ad un'analisi asintotica, in base alla quale si ottiene la bella quanto inaspettata coppia di formule:

$$\lim_{n\to\infty}\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\ldots+\frac{1}{n}-\log(n)\right)=\gamma=0.577216\ldots$$

$$\lim_{p\to\infty}\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\ldots+\frac{1}{p}-\log(\log(p))\right)=M=0.261497\ldots$$

dove γ e M prendono rispettivamente il nome di costante di Eulero-Mascheroni e di Meissel-Mertens. Le formule affermano quindi che i reciproci dei naturali crescono con un andamento di tipo $log(x)$, mentre le somme dei reciproci dei primi crescono con un andamento di tipo $log(log(n))$. Certo la velocità con cui cresce la funzione $log(log(n))$ è davvero lenta, tanto che il teorico dei numeri americano John Selfridge soleva scherzare affermando:

Si sa che log(log(n)) cresce all'infinito, ma nessuno l'ha mai visto arrivarci.