Problemi classici e moderni in teoria dei numeri

Roberto Dvornicich

Dipartimento di Matematica
Università di Pisa

17 novembre 2009

Outline

Outline

Outline

Outline

Introduzione

Scopo di questa conversazione è di presentare la situazione della ricerca relativamente ad alcuni problemi classici della teoria dei numeri.

Ma che cos'è, esattamente, la *Teoria dei numeri*?

Data l'allargamento delle ricerche in questo settore, una definizione esaustiva è assai difficile. È molto meglio evidenziare dei filoni di ricerca, che hanno spesso interazioni tra loro:

1. lo studio dei *numeri primi*, delle loro proprietà e della loro *distribuzione*;
2. lo studio delle *equazioni diofantee* (cioè le equazioni per cui non si ricercano tutte le soluzioni reali, ma **solo** quelle con numeri interi).
3. lo studio dell'*approssimazione diofantea* (cioè della possibilità di approssimare un numero non razionale, per esempio π, mediante frazioni);
4. lo studio delle *proprietà aritmetiche di insiemi di numeri più complessi dei numeri interi*, ma che hanno caratteristiche simili, sia finiti che infiniti;
5. le *applicazioni* dell'aritmetica ai *sistemi di trasmissione digitale di dati* (codici e crittografia).

1. lo studio dei *numeri primi*, delle loro proprietà e della loro *distribuzione*;
2. lo studio delle *equazioni diofantee* (cioè le equazioni per cui non si ricercano tutte le soluzioni reali, ma **solo** quelle con numeri interi).
3. lo studio dell'*approssimazione diofantea* (cioè della possibilità di approssimare un numero non razionale, per esempio π, mediante frazioni);
4. lo studio delle *proprietà aritmetiche di insiemi di numeri più complessi dei numeri interi*, ma che hanno caratteristiche simili, sia finiti che infiniti;
5. le *applicazioni* dell'aritmetica ai *sistemi di trasmissione digitale di dati* (codici e crittografia).

1. lo studio dei *numeri primi*, delle loro proprietà e della loro *distribuzione*;
2. lo studio delle *equazioni diofantee* (cioè le equazioni per cui non si ricercano tutte le soluzioni reali, ma **solo** quelle con numeri interi).
3. lo studio dell'*approssimazione diofantea* (cioè della possibilità di approssimare un numero non razionale, per esempio π, mediante frazioni);
4. lo studio delle *proprietà aritmetiche di insiemi di numeri più complessi dei numeri interi*, ma che hanno caratteristiche simili, sia finiti che infiniti;
5. le *applicazioni* dell'aritmetica ai *sistemi di trasmissione digitale di dati* (codici e crittografia).

1. lo studio dei *numeri primi*, delle loro proprietà e della loro *distribuzione*;
2. lo studio delle *equazioni diofantee* (cioè le equazioni per cui non si ricercano tutte le soluzioni reali, ma **solo** quelle con numeri interi).
3. lo studio dell'*approssimazione diofantea* (cioè della possibilità di approssimare un numero non razionale, per esempio π, mediante frazioni);
4. lo studio delle *proprietà aritmetiche di insiemi di numeri più complessi dei numeri interi*, ma che hanno caratteristiche simili, sia finiti che infiniti;
5. le *applicazioni* dell'aritmetica ai *sistemi di trasmissione digitale di dati* (codici e crittografia).

1. lo studio dei *numeri primi*, delle loro proprietà e della loro *distribuzione*;
2. lo studio delle *equazioni diofantee* (cioè le equazioni per cui non si ricercano tutte le soluzioni reali, ma **solo** quelle con numeri interi).
3. lo studio dell'*approssimazione diofantea* (cioè della possibilità di approssimare un numero non razionale, per esempio π, mediante frazioni);
4. lo studio delle *proprietà aritmetiche di insiemi di numeri più complessi dei numeri interi*, ma che hanno caratteristiche simili, sia finiti che infiniti;
5. le *applicazioni* dell'aritmetica ai *sistemi di trasmissione digitale di dati* (codici e crittografia).

SOMMARIO
degli argomenti trattati

- l'ipotesi di Riemann sulla distribuzione dei numeri primi
- le congetture "famose" (primi gemelli e congettura di Goldbach)
- i numeri "famosi" (e e π)
- equazioni diofantee: l'ultimo teorema di Fermat ed altro
- test di primalità e algoritmi di fattorizzazione
- applicazioni

L'ipotesi di Riemann

L'ipotesi di Riemann (ma sarebbe meglio dire *congettura* di Riemann) è probabilmente il più grande problema aperto della teoria dei numeri. Questa congettura, dovuta appunto a Riemann, risale al 1859.
Per capire di cosa si tratta, facciamo un passo indietro.

Quanti sono i numeri primi? Tutti sanno che sono infiniti. Ma la domanda ha ancora un senso se viene posta in maniera più precisa. Per ogni numero reale positivo x, definiamo $\pi(x)$ il numero dei primi che sono compresi fra 1 e x.
Come varia $\pi(x)$ in funzione di x?

Per esempio, abbiamo la seguente tabella:

x	$\pi(x)$	$x/\pi(x)$
1000	168	6.0
1000000	78498	12.7
1000000000	50847534	19.7

Si può notare che:

1. la *percentuale* dei numeri primi fra 1 e x descresce con x;
2. approssimativamnete, il rapporto $x/\pi(x)$ è una funzione lineare del numero di cifre di x.

Per esempio, abbiamo la seguente tabella:

x	$\pi(x)$	$x/\pi(x)$
1000	168	6.0
1000000	78498	12.7
1000000000	50847534	19.7

Si può notare che:

1. la *percentuale* dei numeri primi fra 1 e x descresce con x;
2. approssimativamnete, il rapporto $x/\pi(x)$ è una funzione lineare del numero di cifre di x.

In effetti, esiste un certo argomento euristico, basato su un modello naturale di probabilità (essenzialmente, si assume che, per un generico intero k, gli eventi *k è divisibile per m* e *k è divisibile per n*, dove m ed n sono primi fra loro, siano eventi indipendenti), che induce alla seguente

Congettura

La probabilità che un numero n sia primo è $\frac{1}{\log n}$, dove il logaritmo è il logaritmo naturale, ossia fatto rispetto alla base e (costante di Nepero).

Questo fatto è stato dimostrato, ed è il succo del cosiddetto **Teorema dei numeri primi**.

Teorema

$$\pi(x) \sim \int_2^x \frac{1}{\log t} dt \sim \frac{x}{\log x}$$

Il simbolo $\sim$ sta a denotare un'approssimazione, ma un'approssimazione molto precisa.

Il teorema dei numeri primi si può enunciare in una forma del tutto equivalente "pesando" ogni numero primo p con un peso uguale a $\log p$:

Teorema

$\theta(x) := \sum_{p \leq x} \log p \sim x.$

Questo fatto è stato dimostrato, ed è il succo del cosiddetto **Teorema dei numeri primi**.

Teorema

$$\pi(x) \sim \int_2^x \frac{1}{\log t}\,dt \sim \frac{x}{\log x}$$

Il simbolo $\sim$ sta a denotare un'approssimazione, ma un'approssimazione molto precisa.

Il teorema dei numeri primi si può enunciare in una forma del tutto equivalente "pesando" ogni numero primo p con un peso uguale a $\log p$:

Teorema

$\theta(x) := \sum_{p \le x} \log p \sim x.$

L'ipotesi di Riemann riguarda la bontà di questa approssimazione.

Essa dice: se il modello probabilistico che abbiamo inventato funziona, esso dovrebbe seguire le leggi della probabilità. La probabilità (legge dei grandi numeri) dice per esempio che se facciamo una serie successiva di n lanci di monete (testa o croce) non solo ci aspettiamo che circa la metà dei lanci diano testa e metà croce, ma anche che la scostamento rispetto a questo valore atteso sia piccolo. Lo scostamento previsto è circa $\sqrt{n}$.

Sarà vero anche nel nostro caso? È vero, cioè, che nella nostra formula approssimata l'errore che facciamo è al massimo quello che dovrebbe essere, ossia all'incirca $\sqrt{\frac{x}{\log x}}$? O, nella formulazione equivalente, $\sqrt{n}$?

La validità di questa tesi sembra molto plausibile, ed avrebbe conseguenze rilevanti per la conoscenza di un gran numero di problemi collegati ai numeri primi.

Purtroppo, a tutt'oggi, non conosciamo la verità. Non sappiamo se la congettura di Riemann sia vera o falsa, anche se ci sono vari indizi a favore. Il primo indizio è di carattere "filosofico": un modello probabilistico che si discosta da quello "naturale" dovrebbe essere motivato da fenomeni distorsivi speciali, che nessuno ha mai riscontrato.
Ma ci sono indizi più convincenti dal punto di vista numerico.
Ci sono vari modi equivalenti di formulare la congettura di Riemann: il più famoso di essi e che un insieme infinito di punti del piano (gli zeri non banali della funzione zeta di Riemann) sia in realtà costituito da punti che giacciono tutti *su una medesima retta verticale*.

Allora uno può verificare se almeno qualcuno di questi punti giace effettivamente su questa retta. Si sono fatti dei calcoli:

I primi 24.000.000.000.000 punti controllati giacciono effettivamente sulla retta!

Per molti osservatori esterni questa evidenza numerica costituisce una prova oggettiva, ma purtroppo non è così per i matematici. Un'eccezione è sempre possibile, e non sarebbe il primo caso nella storia quello in cui si è trovata un'eccezione non prevista anche quando tutti credevano che ormai non fosse più ragionevolmente possibile pensare che ci fosse.

I numerosissimi tentativi di dimostrazione della congettura hanno dato solo risultati parziali, come per esempio quello di garantire che almeno il 40% dei punti giace sulla retta (bisogna però specificare cosa vuol dire il 40% di *infiniti* punti).

VOLETE DIMOSTRARLA VOI?

A parte lo scherzo implicito in questa domanda, ecco una formulazione alternativa semplice ma assolutamente esatta dell'ipotesi di Riemann, che tutti possono capire. Consideriamo il minimo comune multiplo di tutti di numeri fra 1 ed n e chiamiamolo $M(n)$. Si sa, dal teorema dei numeri primi, che

$$M(n) \sim e^n$$

ossia che

$$\log M(n) \sim n.$$

È vero che l'approssimazione scritta sopra rispetta le leggi normali della probabilità, ossia che la differenza fra $\log M(n)$ *ed* n *è al massimo (all'incirca)* $\sqrt{n}$*?*
BUONA FORTUNA!

VOLETE DIMOSTRARLA VOI?

A parte lo scherzo implicito in questa domanda, ecco una formulazione alternativa semplice ma assolutamente esatta dell'ipotesi di Riemann, che tutti possono capire. Consideriamo il minimo comune multiplo di tutti di numeri fra 1 ed n e chiamiamolo $M(n)$. Si sa, dal teorema dei numeri primi, che

$$M(n) \sim e^n$$

ossia che

$$\log M(n) \sim n.$$

È vero che l'approssimazione scritta sopra rispetta le leggi normali della probabilità, ossia che la differenza fra $\log M(n)$ *ed* n *è al massimo (all'incirca)* $\sqrt{n}$?
BUONA FORTUNA!

I primi gemelli

Il problema dei primi gemelli è quello di stabilire se esistano infinite coppie di numeri primi della forma $(n, n+2)$. Per esempio, le coppie

$$(3,5),(5,7),(11,13),(17,19),(29,31),(41,43)$$

sono coppie di primi gemelli.
Purtoppo, anche questo è un problema aperto. È interessante solo notare che, se il modello probabilistico dei numeri naturali delineato prima fosse adeguato, allora si potrebbe non solo dimostrare che esistono infinite coppie di primi gemelli, ma anche dire “quante” sono.

Denotiamo con $\pi_2(x)$ il numero di coppie di primi gemelli $(n, n+2)$ fatte con muneri minori o uguali a x. Allora si ha la formula euristica (congetturale)

$$\pi_2(x) = 1,320326... \times \frac{x}{\log^2 x}.$$

Anche qui si sono fatti alcuni calcoli; l'ultimo risultato disponibile riguarda un'analisi di tutti i numeri che hanno fino a 15 cifre decimali. L'errore percentuale che dà la formula congetturale è inferiore a un milionesimo!!!

La congettura di Goldbach

La congettura di Goldbach dice che

Congettura

Ogni numero pari maggiore o uguale a 4 si può scrivere come la somma di due numeri primi.

Esiste una congettura analoga per i numeri dispari:

Congettura

Ogni numero dispari maggiore o uguale a 7 si può scrivere come somma di tre numeri primi.

L'attenzione attuale, specialmente da parte dei dilettanti della matematica, è rivolta verso il problema che riguarda i numeri pari. Infatti, in un senso che andiamo a precisare, il problema relativo ai numeri dispari è "sostanzialmente" risolto.

IL CASO DISPARI: $2n + 1 = p_1 + p_2 + p_3$?

Nel 1937 il matematico russo Vinogradov ha dimostrato il seguente

Teorema

Tutti i numeri dispari "abbastanza grandi" si possono scrivere come somma di numeri primi.

L'attenzione attuale, specialmente da parte dei dilettanti della matematica, è rivolta verso il problema che riguarda i numeri pari. Infatti, in un senso che andiamo a precisare, il problema relativo ai numeri dispari è "sostanzialmente" risolto.

$$\text{IL CASO DISPARI: } 2n+1 = p_1 + p_2 + p_3 ?$$

Nel 1937 il matematico russo Vinogradov ha dimostrato il seguente

Teorema

Tutti i numeri dispari "abbastanza grandi" si possono scrivere come somma di numeri primi.

Cosa vuol dire “abbastanza grandi”? Analizzando la dimostrazione di Vinogradov, si vede che essa funziona per tutti i numeri maggiori o uguali di una costante incredibilmente grande, $3^{3^{15}}$ (un numero con quasi 7 milioni di cifre decimali).
Recentemente, raffinando i metodi di Vinogradov, si è trovata una dimostrazione che funziona per tutti i numeri maggiori di 2×10^{1346} (quindi con almeno 1347 cifre decimali).
Quindi, basta usare il computer per trattare tutti i numeri più piccoli ....

Sembra facile, ma non lo è. Vediamo perché.
Suppomiamo di voler verificare tramite un computer che tutti i numeri dispari che la dimostrazione lascia in sospeso si possono esprimere come somma di tre numeri primi. Se $x = p_1 + p_2 + p_3$ allora è chiaro che almeno uno degli addendi deve essere maggiore o uguale di $x/3$. Questo significa che dobbiamo avere a disposizione una tavola di numeri primi che arrivi almeno fino a $2/3 \times 10^{1346}$. Ammesso che abbiamo a disposizione i mezzi teorici per farlo, c'è un problema:

GLI ATOMI DELL'UNIVERSO SONO "SOLO" 10^{90} !

Però si ha questo teorema (1997):

Teorema

se l'ipotesi di Riemann **generalizzata** *è vera, allora TUTTI i numeri dispari sono sicuramente somma di tre numeri primi.*

Sembra facile, ma non lo è. Vediamo perché.
Suppomiamo di voler verificare tramite un computer che tutti i numeri dispari che la dimostrazione lascia in sospeso si possono esprimere come somma di tre numeri primi. Se $x = p_1 + p_2 + p_3$ allora è chiaro che almeno uno degli addendi deve essere maggiore o uguale di $x/3$. Questo significa che dobbiamo avere a disposizione una tavola di numeri primi che arrivi almeno fino a $2/3 \times 10^{1346}$. Ammesso che abbiamo a disposizione i mezzi teorici per farlo, c'è un problema:

GLI ATOMI DELL'UNIVERSO SONO "SOLO" 10^{90} !

Però si ha questo teorema (1997):

Teorema

se l'ipotesi di Riemann **generalizzata** *è vera, allora TUTTI i numeri dispari sono sicuramente somma di tre numeri primi.*

Sembra facile, ma non lo è. Vediamo perché.
Suppomiamo di voler verificare tramite un computer che tutti i numeri dispari che la dimostrazione lascia in sospeso si possono esprimere come somma di tre numeri primi. Se $x = p_1 + p_2 + p_3$ allora è chiaro che almeno uno degli addendi deve essere maggiore o uguale di $x/3$. Questo significa che dobbiamo avere a disposizione una tavola di numeri primi che arrivi almeno fino a $2/3 \times 10^{1346}$. Ammesso che abbiamo a disposizione i mezzi teorici per farlo, c'è un problema:

GLI ATOMI DELL'UNIVERSO SONO "SOLO" 10^{90} !

Però si ha questo teorema (1997):

Teorema

se l'ipotesi di Riemann **generalizzata** *è vera, allora TUTTI i numeri dispari sono sicuramente somma di tre numeri primi.*

IL CASO PARI: $2n = p_1 + p_2$?

Considerato, a ragione o a torto, risolto il caso dispari, è su questo caso che si concentra l'attenzione di molti appassionati. Infatti per questo caso non esiste un teorema analogo a quello del caso dispari. I fatti definitivamente dimostrati hanno una validità minore. Ecco due esempi. Il primo riguarda una variazione del problema:

Teorema

Ogni numero pari "abbastanza grande" si può scrivere come somma di due numeri dei quali uno è sicuramente primo e l'altro o è primo oppure è il prodotto di due numeri primi.

Anche qui "abbastanza grande" è *troppo* grande per poter verificare tutti i casi esclusi dal teorema.

Il secondo esempio riguarda il numero di possibili eccezioni alla validità della congettura. Definiamo

$$E(x) := \#\{n \leq x \,|\, 2n \neq p_1 + p_2\}.$$

Teorema

Per ogni $\varepsilon > 0$ esiste una costante $C = C(\varepsilon)$ tale che

$$E(x) \leq Cx^{\frac{1}{2}+\varepsilon}.$$

Da questo teorema di deduce che:
"quasi tutti" i numeri pari si possono scrivere come somma di due numeri primi.

Nota per i prossimi "solutori" della congettura di Goldbach.

È abbastanza naturale cercare di risolvere il problema prima per i numeri piccoli, tentando poi di scoprire delle regole, magari delle formule generali, che valgano per tutti i numeri. Finora nessuno è riuscito in questo intento, ed è assai probabile che formule generali non esistano affatto.
Per questo motivo i tentativi di soluzione non partono dai numeri piccoli, ma dai numeri sufficientemente grandi. I metodi usati, però sono *analitici* e non *aritmetici*. Si tratta di esprimere, tramite un integrale, il *numero di modi* in cui un numero, per esempio dispari, si può scrivere come somma di tre numeri primi; e di dimostrare poi, tramite disuguaglianze (che però sono vere solo per numeri grandi) che questo integrale dà come risultato un numero positivo.

I numeri famosi

I numeri famosi *e* e π (ma il discorso vale per molti altri numeri di uso quotidiano in matematica) non si possono scrivere con esattezza usando il sistema decimale, perché si avrebbe bisogno di infinite cifre. Non si può nemmeno specificare una regola che permetta di calcolare tutte le cifre, come per esempio per

$$\frac{1}{11} = 0,090909090909...$$

perché tale regola non esiste.
Il motivo risiede nel fatto che essi sono definiti con processi di limite e non semplicemente tramite le usuali quattro operazioni.
La domanda è: si possono definire esattamente questi numeri usando strumenti puramente algebrici (come le quattro operazioni, i radicali, eccetera) ma senza strumenti analitici?

Per specificare il problema abbiamo bisogno della seguente definizione:

Definizione

Un numero (reale o complesso) si dice algebrico *se è radice di un polinomio non nullo a coefficienti interi.*

Per esempio, $\sqrt[n]{3}$ è radice del polinomio $x^n - 3$ e un numero α che soddisfi la relazione $\alpha^5 - \alpha - 1 = 0$ è algebrico (anche se non si riesce a scrivere tramite radicali).

Il problema quindi diventa:

Problema

I numeri e e π sono algebrici?

La risposta è NO per entrambi i casi.

Il risultato non è inatteso, nel senso che si può verificare che, preso un numero “a caso” è quasi certo (la probabilità è uguale a 1) che la risposta sio no per questo numero.
Tuttavia, come è facilmente intuibile, è assai complicato escludere che un certo numero possa essere radice di uno qaualsiasi fra gli infiniti polinomi a coefficienti interi.

La soluzione del problema relativo ad e è datata 1873, quella relativa a π è datata 1882. Quest’ultima ha conseguenze su un problema posto già dagli antichi greci, il problema della *quadratura del cerchio*.

La quadratura del cerchio

Dato un cerchio di raggio 1, è possibile costruire con riga e compasso un quadrato la cui area sia uguale a quella del cerchio dato (e cioè π)?
Si può dimostrare abbastanza facilmente che, in un sistema di riferimento cartesiano, le coordinate di tutti i punti che si riescono a costruire con riga e compasso SONO soluzioni di un'equazione $f(x) = 0$, dove $f(x)$ è un polinomio a coefficienti interi.
Se si potesse quadrare il cerchio, si potrebbe costruire un quadrato di lato $\sqrt{\pi}$. Ma né π né la sua radice quadrata (questa è una conseguenza relativamente semplice) sono soluzioni di alcuna equazione di questo tipo.

Ma i problemi aperti sono sempre più di quelli risolti ...

Come detto, il problema relativo ad e e π è stato risolto, ma, in pratica, solo per loro e per pochissimi altri numeri. Per esempio, non si sa risolvere il problema nemmeno per le combinazioni più semplici che si possono fare con questi due numeri, quali $e + \pi$, $e \cdot \pi$, etc.

Ma i problemi aperti sono sempre più di quelli risolti ...

Come detto, il problema relativo ad e e π è stato risolto, ma, in pratica, solo per loro e per pochissimi altri numeri. Per esempio, non si sa risolvere il problema nemmeno per le combinazioni più semplici che si possono fare con questi due numeri, quali $e + \pi$, $e \cdot \pi$, etc.

L'equazione di Fermat

L'equazione diofantea più conosciuta è quella di Fermat:

$$x^n + y^n = z^n.$$

Fermat affermava che, se $n \geq 3$, questa equazione non ha alcuna soluzione con numeri interi, se si eccettuano quelle "banali", ossia quelle in cui una delle variabili è uguale a zero (per esempio, $0^5 + 3^5 = 3^5$). La storia di questa equazione è molto lunga, e molto si è speculato sul fatto che Fermat avesse in mente una soluzione del problema da lui stesso posto.
È probabile (ma certamente non è sicuro) che Fermat NON avesse una soluzione. Sta di fatto che il problema è stato risolto solo 350 anni dopo la sua proposizione (Wiles, 1995).

Innanzitutto: qual è l'interesse di sapere se un'equazione come quella di Fermat ha soluzioni, ed eventualmente di conoscere quali?
A questa domanda si potrebbe tranquillamente rispondere: NESSUNO. La stessa cosa si può dire di moltissimi, per non dire quasi tutti, i problemi di matematica. Storicamente, i problemi di matematica sono stati studiati in quanto interessanti *per se stessi*, indipendentemente dalle loro applicazioni pratiche. È altresì vero che molti risultati della matematica *hanno poi avuto* applicazioni pratiche, ma molto spesso applicazioni che non rientravano nell'obiettivo di coloro che vi hanno contribuito, e che non erano nemmeno nella loro immaginazione.

Nel caso del cosiddetto Ultimo Teorema di Fermat (la ricerca delle soluzioni dell'equazione diofantea $x^n + y^n = z^n$) l'interesse puramente speculativo del problema è quello che ha mosso migliaia e migliaia di matematici, professionisti o dilettanti, a dedicarcisi. Col senno di poi si può dire che questo ha contribuito a enormi sviluppi del pensiero matematico, alcuni dei quali hanno avuto *anche* ricadute dal punto di vista delle applicazioni.
Come certamente saprete, la dimostrazione di Wiles dell'Ultimo Teorema di Fermat è estremamente lunga e tecnica, e non si può raccontare se non ad un pubblico molto esperto.
Perciò sono costretto a limitarmi ad alcuni cenni.

Innanzitutto, si tratta di una dimostrazione *per assurdo*.

In secondo luogo, essa usa dei risultati profondi di *geometria*. Che cosa ha a che fare la geometria con un problema puramente aritmetico come questo?

Già negli anni ’50 il matematico Frei aveva avuto l’idea di legare l’equazione di Fermat all’equazione di una *curva*. Una curva, nel piano, si può descrivere tramite un’equazione in due variabili: per esempio, l’equazione $x^2 + y^2 = 1$ descrive i punti di una *circonferenza* (di centro l’origine e di raggio 1).

Frei argomentava così: supponiamo, per assurdo, che l'equazione di Fermat abbia una soluzione (non banale), e che a, b, c siano tre numeri positivi tali che $a^p + b^p = c^p$ (qui l'esponente p è un numero prino diverso da 2, ma si può facilmente vedere che questo è il caso cruciale).

Consideriamo l'equazione

$$y^2 = (x - a^p)(x - b^p)(x + c^p).$$

Le soluzioni di questa equazione formano appunto una curva *algebrica*, di un genere speciale: una *curva ellittica*.
I geometri classificano le curve algebriche secondo il loro grado di complessità, un invariante chiamato genere. Le coniche, che sono le curve più semplici, hanno genere 0; le curve ellittiche, che rappresentano il livello successivo di difficoltà, hanno genere 1.

Sulle curve ellittiche si sa moltissimo: in particolare, si sa quando due diverse equazioni definiscono curve ellittiche dello stesso *tipo* (=isomorfe), e si sanno classificare tutti i tipi possibili di curve ellittiche.
Sapendo i coefficienti dell'equazione che descrive la curva ellittica, se ne può dedurre il "tipo".
La dimostrazione consiste, essenzialmente, nel far vedere che, se effettivamente si potessero trovare a, b, c come sopra e quindi si costruisse la curva ellittica relativa, questa curva *non sarebbe di nessun tipo possibile*.

L'interazione fra geometria ed aritmetica, sviluppata enormemente a partire dalla seconda metà del secolo scorso, è uno dei grossi risultati che si sono avuti anche e per merito dello studio dell'Ultimo Teorema di Fermat. In particolare, oggi le equazioni diofantee non si studiano più *una alla volta*, ma si raggruppano in famiglie che descrivono insiemi geometrici dello stesso tipo. È stato per esempio dimostrato il seguente teorema:

Teorema

Sia $f(x, y) = 0$ *l'equazione di una curva di genere* > 1. *Allora esistono solo un numero finito di soluzioni dell'equazione* $f(x, y) = 0$ *con* x, y *numeri razionali.*

L'equazione di Catalan

Un altro spettacolare risultato recente consiste nella soluzione dell'equazione di Catalan. Catalan (1844) considerava i quadrati

$$1, 4, 9, 16, 25, 36, 49, \ldots,$$

i cubi

$$1, 8, 27, 64, 125, \ldots,$$

le quarte potenze

$$1, 16, 81, 256, \ldots$$

e così via, per riunirli in un'unica successione:

$$1, 4, 8, 9, 16, 25, 27, 36, 49, 64, \ldots$$

Catalan notava che in questa successione ci sono due numeri consecutivi, e cioè 8 e 9.

Ci sono altre coppie di numeri consecutivi in questa successione? Catalan pensava di no.

Ebbene, Mihajlescu nel 2001 ha dimostrato che Catalan aveva ragione:

Teorema

se consideriamo tutti i numeri della forma a^b, dove b è un esponente maggiore o uguale a 2, l'unica coppia di numeri consecutivi è costituita da 8 e 9.

Ci sono altre coppie di numeri consecutivi in questa successione? Catalan pensava di no.

Ebbene, Mihajlescu nel 2001 ha dimostrato che Catalan aveva ragione:

Teorema

se consideriamo tutti i numeri della forma a^b, dove b è un esponente maggiore o uguale a 2, l'unica coppia di numeri consecutivi è costituita da 8 e 9.

La spettacolarità della dimostrazione consiste nel fatto che invece, questa volta, si tratta di una dimostrazione *puramente aritmetica*, ed in fondo basata su idee dovute a Kummer intorno alla metà del secolo diciannovesimo (Kummer è lo stessa persona che aveva fatto i primi importanti progressi nello studio dell'ultimo teorema di Fermat).
Tuttavia, l'aritmetica dei numeri interi *non basta*: bisogna costruire un'aritmetica su strutture più complesse (i cosiddetti *campi ciclotomici*) ed è lì che si può risolvere il problema.

Risolveremo tutte le equazioni?

Le soluzioni di problemi così antichi in tempi recenti possono far pensare che siamo vicini a risolvere il problema di tutte le equazioni diofantee.
Non è così. Un problema che Hilbert, nel congresso mondiale dei matematici del 1900, aveva posto in una lista di problemi per il ventesimo secolo era il seguente:
È possibile trovare un *algoritmo* per risolvere tutte le eqauzioni diofantee?
Matijaševic, nel 1970, ha risposto di NO. Non esiste, né potrà mai esistere, un modo per risolvere TUTTE le equazioni diofantee.

La dimostrazione di Matijaševic si inquadra nell'ambito della *logica matematica*.
Nel 1936 K. Gödel aveva dimostrato che, nell'usuale sistema di assiomi della matematica, ma anche con qualsiasi altro sistema che si potesse inventare, la matematica ha dei limiti: ci sono degli enunciati di cui non potremo mai dimostrare né che sono veri, né che sono falsi. Si tratta degli enunciati che Gödel ha chiamato *indecidibili*.
Matijaševic ha fatto vedere che esistono dei particolari tipi di equazioni diofantee per cui la questione se abbiano o meno soluzioni è indecidibile.

Primalità e fattorizzazione

Uno dei problemi basilari della teoria dei numeri consiste nel determinare se un certo numero n è un numero primo; nel caso in cui non lo sia, di determinare la sua scomposizione in fattori primi. Come vedremo, questo problema assolutamente teorico e astratto ha incredibili conseguenze pratiche.
È chiaro che, se il numero dato n è relativamente piccolo, chiunque, o a mano o con l'aiuto di un calcolatore, può rispondere alla domanda. Il problema quindi si pone in termini di complessità: dato un numero n di k cifre, quante sono le operazioni necessarie per dare una risposta?
In pratica, *quanto tempo* ci vuole?

Gli algoritmi che riguardano i numeri interi vengono classificati in classi che corrispondono a diversi gradi di complessità (tempo necessario per la loro esecuzione).
La classe **P** è la classe dei problemi per i quali il numero di passi necessario per eseguire l'algoritmo è *polinomiale* rispetto al numero di cifre dei numeri interi che si esaminano.
Nel nostro caso, considerando un numero n con k cifre, un algoritmo che decida se n è primo oppure no è polinomiale se si può effettuare con un numero di passi non superiore a una *potenza* di k, per esempio k^2 oppure anche k^{100}.

Esaminiamo l'algoritmo più naturale per decidere se un numero n è primo oppure no:
dividiamo n per 2, per 3, per 4, per 5, e così via: se ad un certo punto troveremo una divisione esatta (con resto zero), allora il numero non sarà primo (ed avremo trovato un fattore di n); se invece tutte le divisioni per numeri minori di n (ma in realtà basta fermarsi alla radice quadrata di n) danno resto diverso da zero, allora il numero è primo.
Quindi l'esecuzione dell'algoritmo, almeno nel caso in cui n sia un numero primo, richiede di fare circa $\sqrt{n}$ divisioni. Se n ha k cifre, diciamo che n è dell'ordine di grandezza di 10^k, ci vorranno quindi circa $10^{k/2}$ divisioni. Per k grande, questo numero è molto superiore a una potenza (qualsiasi) di k.
Se ne deduce che l'algoritmo naturale non è polinomiale, ma *esponenziale*.

Sono stati studiati vari altri algoritmi che “accorciano” il tempo di esecuzione: alcuni *deterministici*, ossia che danno la risposta con assoluta certezza, altri *probabilistici*, ossia che hanno una altissima probabilità di dare la risposta esatta, Questi ultimi algoritmi sono ovviamente più veloci dei primi, ma bisogna accontentarsi di un grado, se pur minimo, di incertezza.
Tutti i tipi di algoritmi deterministici conosciuti fino a pochissimo tempo fa erano di tipo esponenziale; una combinazione ingegnosa dei tipi deterministico e probabilistico porta a degli algoritmi che, nella grande maggioranza dei casi, si possono eseguire in tempo polinomiale, ma che lasciano un numero di eccezioni in cui è necessario un tempo esponenziale.

Tra la sorpresa generale dei matematici, tre indiani, M. Agraval, N. Kayal e N. Saxena (in seguito AKS), hanno trovato nel 2002 un algoritmo deterministico per stabilire se un numero è primo oppure no che funziona in tempo polinomiale.
Di questi matematici solo il primo aveva una certa notorietà internazionale, ma forse più per i suoi studi informatici che per quelli matematici; gli altri due sono suoi giovanissimi allievi.
Ma il vero motivo di sorpresa è un altro: l'idea che sta alla base della formulazione dell'algoritmo è così semplice che sarebbe potuta venire in mente a un qualsiasi studente del primo biennio di matematica.
Invece, nel corso di secoli, non era venuta in mente a nessuno!

L'IDEA DELL'ALGORITMO AKS

Si parte da due fatti legati fra loro e noti da secoli. Primo fatto:

Teorema

Se p è un numero primo, allora vale la congruenza

$$(x+y)^p \equiv x^p + y^p \pmod{p}.$$

Inoltre, la questa congruenza non vale se al posto di p si prende un numero non primo.

Ricordiamo che due numeri si dicono *congrui* modulo p se divisi per p danno lo stesso resto. La congruenza enunciata sopra dice che, se p è un numero primo, il polinomio $(x+y)^p$ ha un termine uguale a x^p, un termine uguale a y^p e tutti gli altri suoi termini hanno coefficienti divisibili per p.

Secondo fatto (piccolo teorema di Fermat):

Teorema

Se p è un numero primo, allora per ogni intero m vale la congruenza

$$m^p \equiv m \pmod{p}.$$

Da questi due fatti elementari AKS deducono il loro teorema, che è un semplice esercizio per un normale studente:

Teorema

(AKS) *Siano n ed a due numeri interi senza fattori comuni. Allora*

$$(x + a)^n \equiv x^n + a \pmod{n}$$

SE E SOLO SE n è un numero primo.

La difficoltà di ottenere questo teorema, come detto, non consiste affatto nella sua dimostrazione, ma nella sua *invenzione*: bisogna IMMAGINARE l'enunicato e le sue possibili applicazioni.

Dal teorema AKS è abbastanza chiaro quello che bisogna fare: dato n, provare a vedere se la cosa è vera, per esempio, per $a = 1$. Detto così, questo richiede ancora un tempo troppo elevato, perché bisognerebbe calcolare tutti i coefficienti del polinomio $(x + a)^n$.
Ma, contando su idee presenti in algoritmi precedentemente sviluppati, si vede che in realtà non occorre considerare i coefficienti uno per uno, ma solo un numero assai più limitato di combinazioni fra di loro, e provare a vedere che cosa succede di queste combinazioni se le si divide per numeri piccoli.
Questo porta ad un algoritmo polinomiale.

Nonostante il risultato teorica sia straordinario, l'algoritmo AKS non viene ancora usato nella pratica. Come mai?

Il fatto è che, per testare se un numero con k cifre è primo oppure no, ci vogliono un numero di passi che è circa $C \cdot k^{7.5}$, dove C è una costante molto grande. Anche se il numero di passi necessario per gli altri algoritmi è dato da una formula che è sicuramente peggiore per k molto grande, questa formula risulta dà un risultato migliore, per via della costante C, quando k è relativamente piccolo (un punto di riferimento attuale è $k = 200$).

Nonostante il risultato teorica sia straordinario, l'algoritmo AKS non viene ancora usato nella pratica. Come mai?

Il fatto è che, per testare se un numero con k cifre è primo oppure no, ci vogliono un numero di passi che è circa $C \cdot k^{7.5}$, dove C è una costante molto grande. Anche se il numero di passi necessario per gli altri algoritmi è dato da una formula che è sicuramente peggiore per k molto grande, questa formula risulta dà un risultato migliore, per via della costante C, quando k è relativamente piccolo (un punto di riferimento attuale è $k = 200$).

Gli algoritmi di fattorizzazione

Quando si usa un test di primalità del tipo di AKS, e si ottiene la risposta "n non è un numero primo", non si individua necessariamente la fattorizzazione di n. Si sa solo che n non soddisfa le proprietà che sono proprie dei numeri primi.
Per avere un algoritmo di fattorizzazione bisogna fare un passo in più. Quello che in realtà serve, se si scopre che un numero n non è primo, è di individuare un suo divisore proprio (cioè diverso da 1 e da n). Infatti, se a è un divisore proprio di n, e dunque $n = ab$ per qualche intero b, si può ripetere l'algoritmo per i numeri a e b al fine di scoprire se essi sono primi o hanno dei divisori propri. Ripetendo questo ragionamento, con numeri via via più piccoli, si riesce a determinare la scomposizione di n in fattori primi.

Gli algoritmi di fattorizzazione oggi disponibili sono sicuramente molto più efficienti dell'algoritmo "naturale" descritto precedentemente. È forse interessante notare che, tra gli algoritmi più efficienti conosciuti, uno fa un uso sistematico delle curve ellittiche, che abbiamo già incontrato nella discussione a proposito dell'ultimo teorema di Fermat. Tuttavia, se si eccettuano gli algoritmi *ad hoc* che funzionano solo per numeri di una forma molto speciale, la loro complessità di tutti gli algoritmi noti è sempre *esponenziale*. Anche se non si può avere una prova sicura che qualcuno non scopra, prima o poi, un algoritmo di fattorizzazione di complessità polinomiale, al giorno d'oggi la fattorizzazione di un numero rimane uno dei problemi più complessi (nel senso di "*time-consuming*"), ed è su questa convinzione che si basa una delle applicazioni della teoria dei numeri più diffusa, i cui andiamo a discutere.

La crittografia

La crittografia si occupa di trovare dei metodi efficienti per trasmettere dei messaggi, o comunque delle informazioni, in modo codificato, in modo tale che una persona che sia in possesso di uno strumento (chiave di lettura) per decodificare le informazioni le possa decifrare, ma una persona che non conosca la chiave di lettura no.

L'uso della crittografia è storicamente provato fin dal tempo degli antichi romani, per scopi militari.

Oggi se ne fanno diversi usi: insieme a quello militare e di spionaggio, quelli preponderanti sono per le transazioni di carattere economico, per garantire la privacy, per un controllo di sicurezza dell'identità degli individui ammessi a certi servizi.

Una carta bancomat, un acquisto on-line con una carta di credito, l'uso della password nell'aprire un computer, per scaricare files, per leggere la posta elettronica o per entrare in alcuni siti internet sono esempi quotidiani dell'uso della crittografia.

Oggi se ne fanno diversi usi: insieme a quello militare e di spionaggio, quelli preponderanti sono per le transazioni di carattere economico, per garantire la privacy, per un controllo di sicurezza dell'identità degli individui ammessi a certi servizi.

Una carta bancomat, un acquisto on-line con una carta di credito, l'uso della password nell'aprire un computer, per scaricare files, per leggere la posta elettronica o per entrare in alcuni siti internet sono esempi quotidiani dell'uso della crittografia.

Un sistema crittografico *efficiente* deve rispondere ai seguenti requisiti:

1. rendere *facile* l'uso del sistema da parte dei suoi utenti autorizzati; in particolare, per il mittente di un messaggio deve essere facile codificarlo, per il ricevente deve essere facile decodificarlo;
2. rendere *estremamente difficile*, per non dire impossibile, decodificare dei messaggi se non si conosce la chiave di interpretazione.

Un sistema crittografico *efficiente* deve rispondere ai seguenti requisiti:

1. rendere *facile* l'uso del sistema da parte dei suoi utenti autorizzati; in particolare, per il mittente di un messaggio deve essere facile codificarlo, per il ricevente deve essere facile decodificarlo;
2. rendere *estremamente difficile*, per non dire impossibile, decodificare dei messaggi se non si conosce la chiave di interpretazione.

Relativamente al primo obiettivo, la matematica risponde che un modo facile di codificare e decodificare i messaggi consiste:

- innanzitutto nello scrivere i messaggi in forma *digitale*, ossia trasformare le lettere o le parole che si usano in numeri; la legge di trasformazione deve essere poi il più possibile facile ed immediata (per esempio, $A = 1$, $B = 2$, $C = 3$, ...);
- in secondo luogo, nell'usare, sia per la codificazione che per la decodificazione, operazioni numeriche semplici, quali per esempio quelle fornite dalle quattro operazioni.

Il secondo obiettivo si può ottenere in due modi, “filosoficamente” diversi tra loro.
Il primo modo (detto a chiave segreta) è quello di usare un sistema di codificazione che, pur eventualmente conosciuto da altri, dipenda da uno o più parametri, dando luogo ad un numero altissimo di combinazioni possibili. In questo caso sarà molto difficile, per una persona estranea al sistema, scoprire quale delle combinazioni possibili è stata usata.
Il secondo modo (detto a chiave pubblica) è quello di usare un sistema di codificazione chiaro, anche accessibile a tutti, ma che deve seguire una regola: che anche conoscendo il sistema di codificazione, *non sia ragionevolmente possibile* scoprire, in tempi adeguati, come si fa a decodificare.

Il secondo modo è quello più moderno ed è attualmente usato in quasi tutte le applicazioni.
La strategia per inventare un sistema crittografico di questo tipo è quella di usare:

- per la codificazione (operazione diretta), un'operazione aritmetica semplice;
- per la decodificazione (operazione inversa), un'operazione aritmetica estremamente complicata.

Operazioni di questo genere, che fatte in un verso risultano semplici e nell'altro complicate, esistono: quella su cui ci si basa maggiormente è quella del prodotto di numeri primi, in un verso, e della scomposizione di un numero in fattori primi, nell'altro verso.

Immaginate per esempio che vi si chieda di moltiplicare fra loro i numeri primi 317 e 593. O a mano, com un po' di pazienza, o con l'uso di un calcolatore, in una frazione di secondo piccolissima, otterrete il risultato: 187981.
Immaginate invece che vi dica: il mumero 412861 è il prodotto di due numeri primi. Quali? I tentativi che dovete fare sono già un numero molto elevato; anche con l'uso di un calcolatore, vi ci vorrà un po' per ottenere una risposta, e magari dovrete impostare un programma. Come abbiamo visto, gli algoritmi di fattorizzazione conosciuti hanno una complessità esponenziale.

PS: la risposta alla domanda è $412861 = 467 \times 883$.

Immaginate per esempio che vi si chieda di moltiplicare fra loro i numeri primi 317 e 593. O a mano, com un po' di pazienza, o con l'uso di un calcolatore, in una frazione di secondo piccolissima, otterrete il risultato: 187981.
Immaginate invece che vi dica: il mumero 412861 è il prodotto di due numeri primi. Quali? I tentativi che dovete fare sono già un numero molto elevato; anche con l'uso di un calcolatore, vi ci vorrà un po' per ottenere una risposta, e magari dovrete impostare un programma. Come abbiamo visto, gli algoritmi di fattorizzazione conosciuti hanno una complessità esponenziale.

PS: la risposta alla domanda è $412861 = 467 \times 883$.

Il metodo crittografico RSA

Il sistema *RSA* (dai nomi dei suoi inventori, Rivest, Shamir e Adleman, nel 1978) è un sistema di codificazione di messaggi a chiave pubblica.
Dopo la digitalizzazione delle parole, è necessario, per evitare di ottenere numeri eccessivamente grandi con le operazioni che si usano, usare l'*aritmetica modulare*. Esattamente come in un orologio, in cui quando si raggiungono le 12 o le 24 ore si ricomincia daccapo, nell'aritmetica modulare, con un modulo N, quando si raggiunge il numero N si ricomincia daccapo, e si utilizzano solo numeri più piccoli di N. In pratica, se con le operazioni usate si trova un numero più grande di N, lo si sostituisce con il suo *resto nella divisione per* N, un numero più piccolo di N e che è congruo al numero precedente modulo N.

Il funzionamento del sistema RSA si basa su osservazioni aritmetiche relativamente semplici:

Proposizione

Se p è un numero primo, vi sono infiniti interi positivi k per i quali $x^k \equiv x \pmod{p} \quad \forall x \in \mathbb{Z}$.

Un caso banale si ha scegliendo $k = 1$; un secondo caso si ha per $k = p$ (piccolo teorema di Fermat). In generale osserviamo che

$$x^{n+p-1} \equiv x^{n-1+p} \equiv x^{n-1}x^p$$
$$x^{n-1}x \equiv x^n \pmod{p}$$

e quindi abbiamo $x^k \equiv x \; \forall x \in \mathbb{Z}$ per tutti i k della forma

$$k = 1 + (p-1)t \quad \text{con } t \in \mathbb{Z}.$$

Proposizione

Più in generale, se $N = p_1 p_2 \dots p_s$ è un numero "libero da quadrati", ossia è un prodotto di primi distinti $p_1, p_2, \dots, p_s$, allora esistono infiniti interi positivi k per cui
$x^k \equiv x \pmod{N} \quad \forall x \in \mathbb{Z}$.

Infatti, si ha che $x^k \equiv x \pmod{N}$ se e solo se

$$x^k \equiv x \pmod{p_i} \quad \forall i = 1, \dots, s\,.$$

Se $k - 1$ è un un multiplo comune di tutti i numeri $p_1 - 1, \dots, p_s - 1$, allora tutte le equazioni del sistema sono soddisfatte, per cui il sistema è soddisfatto. Concludendo, una successione di interi che risponde al problema è

$$k = 1 + Mt \quad (m \in \mathbb{Z})$$

dove $M = (p_1 - 1) \times \cdots \times (p_s - 1)$.

Passiamo ora ad illustrare il funzionamento del sistema.

- Il responsabile dell'uso del sistema crittografico sceglie un numero $N = pq$ che sia uguale al prodotto di due numeri primi distinti (p e q avranno, nei sistemi più sofisticati, circa 50 cifre decimali ciascuno). *Il numero N viene reso pubblico, ma i fattori p e q vengono tenuti segreti*.
- Se gli utenti del sistema sono le persone $P_1, \ldots, P_n$, il responsabile assegna a ciascuno degli utenti un "indirizzo", che è costituito da un numero. Gli "indirizzi" $A_1, \ldots, A_n$ devono soddisfare la seguente proprietà aritmetica: il massimo comune divisore fra A_i e N è uguale a 1. *I numeri $A_1, \ldots, A_n$ vengono resi pubblici*.

- Per ogni persona P_i e per ogni indirizzo A_i, il responsabile calcola un numero B_i tale che

$$A_i B_i \equiv 1 \pmod{M},$$

ossia tale che $A_i B_i$ sia della forma $1 + Mt$. (L'equazione $A_i x \equiv 1 \pmod{M}$ hanno soluzione perché il massimo comune divisore fra A_i ed M è uguale a 1). *I numeri B_i sono le "passwords" degli utenti, e pertanto vengono comunicate solo a loro*.

Si noti che la scelta di A_i e B_i è fatta in modo che si abbia sempre $x^{A_i B_i} = x$ nell'aritmetica modulare con modulo N.

LA CODIFICAZIONE DI UN MESSAGGIO.
Supponiamo che P_1 voglia mandare un messaggio codificato a P_1. Egli userà il "modulo" N e l'"indirizzo" A_2 di P_2 per codificare il suo messaggio x nel modo seguente:

$$x \to y = x^{A_2} \quad \text{modulo } N.$$

Il ricevente userà la sua "password" per decodificare il messaggio y ricevuto. Egli farà l'operazione

$$y \to y^{B_2} = (x^{A_2})^{B_2} = x^{A_2 B_2} = x.$$

Il sistema è sicuro? Ossia, un estraneo che sia in possesso dei dati *pubblici* N e degli indirizzi *pubblici* A_i, può riuscire a determinare che "passwords" B_i degli utenti?
Per determinare una password, il responsabile del sistema doveva risolvere un'equazione del tipo $A_i x \equiv 1 \pmod{M}$, quindi doveva conoscere M, dove

$$M = (p-1)(q-1) = N - (p+q) + 1.$$

Ne segue che conoscere M, una volta noto N, equivale a conoscere la somma $p+q$; d'altra parte, se si conoscessero $p+q$ e pq si ricaverebbero facilmente p e q, cioè i fattori di N, risolvendo un'equazione di secondo grado. *Quindi il problema di determinare M è equivalente al problema di fattorizzare N.*

AUTENTICAZIONE

Se usato in maniera ingenua, il metodo RSA può dare luogo a problemi di identificazione dell'autore di un messaggio. Infatti, a priori, chiunque potrebbe firmare il proprio messaggio con il mome di un altro. Per far fronte a questo si può usare un sistema per **autenticare** un messaggio.
Supponiamo di nuovo P_1 voglia spedire un messaggio a P_2 *facendosi riconoscere come autore del messaggio*. Allora egli inserirà, in una sezione separata del messaggio, la codificazione della propria "firma" f secondo lo schema

$$f \to g = f^{A_2 B_1}.$$

A questo punto P_2, per verificare che l'autore del messaggio è proprio P_1 decodificherà la sua firma secondo lo schema

$$g \to g^{A_1 B_2}.$$

I codici a correzione di errore

Al contrario dei metodi crittografici, i codici a correzione di errore servono a trasmettere informazioni e messaggi mella maniera piú chiara possibile e in modo aperto a tutti.
Poiché, a causa dell'imperferzione dei materiali o di disturbi nel canale di comunicazione, non si può essere del tutto certi che tutti i bit di informazione vengano trasmessi in maniera corretta, si inserisce un'informazione suppletiva, tramite la quale si è in grado di scoprire se ci sono stati errori di informazione, e, possibilmente, di **correggerli automaticamente**.

Una precauzione elementare per controllare la correttezza dell'infromazione è quella di inserire un **codice di controllo**, come per esempio quello fornito dalla prova del 9 per le operazioni aritmetiche o quello utilizzato nel codice fiscale. Questo metodo tuttavia non funziona sempre e, anche quando funziona, non permette di risalire all'errore, ma soltanto di stabilire che c'è un errore; in pratica, anche quando si stabilisce che certamente c'è stato un errore, il metodo richiede una seconda prova di trasmissione.

I codici a correzione di errore sono uno strumento molto più potente, anche se nemmeno essi in grado di far fronte a *qualsiasi eventualità*. Ci mettono però in grado di correggere automaticamente un numero di errori di trasmissione "ragionevole" e sono basati su metodi probabilistici. Ossia, se si ritiene che la probabilità di trasmissione errata di un bit sia p (per esempio l'1%), si costruisce un codice che sia in grado di correggere un messaggio con probabilità q (per esempio il 99,9999%); la sicurezza assoluta non esiste, ma una sicurezza elevatissima, anzi elevata quanto si desidera, sì.

Nota: C'è un unico caso in cui non è possibile costruire alcun codice che corregga gli errori, e si ha quando la probabilità p è ESATTAMENTE uguale al 50%. In questo caso, la trasmissione dei bit segue regole del tipo di quella del lancio di monete (testa o croce) e dunque una sequenza di bit trasmessi apparirà come completamente casuale, e quindi sarà impossibile ricostruire l'informazione originaria.

Il principio dei codici a correzione di errore è, come si è detto, di fornire un'informazione supplementare. Nei codici a correzione di errore si utilizzano dunque informazioni più lunghe dello stretto necessario, ma con l'idea che tra tutte le stringhe possibili solo alcune vengano considerate corrette, mentre altre sono certamente sbagliate.
Le stringhe corrette devono essere nettamente distinguibili l'una dall'altra, in modo da non confondersi tra loro tramite un errore di trasmissione. Esse devono, cioè, essere "distanti" l'una dall'altra, dove la distanza tra due stringhe è data dal numero di bit in cui le due stringhe sono diverse.

Se, date due qualsiasi stringhe corrette distinte, la loro distanza è di almeno $2e + 1$ bit, allora una stringa qualsiasi non può essere vicina (al più e bit di differenza) a due o più stringhe corrette, ma AL MASSIMO AD UNA. Questa sarà considerata la stringa corretta più probabile nelle intenzioni del trasmettitore e questa sarà l'informazione ricevuta.
In pratica, supponiamo di voler spedire un'informazione di n bit e di sapere che, anche se aumentiamo n di un pochino, disciamo fino a $n + k$, la probabilità di trasmettere più di e errori in un messaggio di lunghezza $n + k$ sia da considerarsi irrilevante.

Allora costruiamo un “codice” nel modo seguente: aggiungiamo a tutti i messaggi di lunghezza n un numero di informazioni supplementari uguale a k, in modo tale che le stringhe di lunghezza $n + k$ risultanti siano distanti fra loro almeno $2e + 1$. Allora ogni stringa di lunghezza $n + k$ ricevuta sarà riconducibile ad un’unica stringa corretta.

EFFICACIA ED AFFIDABILITA' DI UN CODICE

Perché un codice sia *efficace*, cioè di facile uso, il numero k di informazioni supplementari richieste non deve essere troppo grande rispetto al numero n di informazioni necessarie. Il rapporto $R = \frac{n}{n+k}$ misura l'efficacia di un codice.
Un codice è affidabile se la probabilità di ricostruire gli errori di trasmissione è vicina a 1 quanto si desideri.

Teorema

Se $R < 1 + p \log_2 p + (1-p)\log_2(1-p)$, *si possono costruire codici affidabili.*

Notiamo che la probabilità si avvicina tanto più a 1 quanto più il codice è lungo, ossia quanti più bit utilizza; come esempio numerico, nel caso in cui $p = 1\%$, si può scegliere R uguale a circa il 92%.

I CODICI DI HAMMING

I codici di Hamming sono solo un esempio per messaggi relativamente piccoli.
Supponiamo di voler mandare un messaggio di 4 bit essendo sicuri che anche se un bit viene trasmesso in modo sbagliato, il ricevente correggerà l'errore.
Scegliamo di mandare un'informazione supplementare nel modo seguente. Formiamo una tabella come segue:

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

La regola è quella di scrivere nelle colonne tutti i numeri da 1 a 7 scritti in forma binaria.

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Utilizziamo un'informazione supplementare di 3 bit, quindi mandando stringhe di 7=4+3 bit. La regola di composizione delle stringhe di 7 bit è quella della risoluzione di un sistema di equazioni lineari, dove le equazioni sono intese nell'aritmetica modulare con modulo 2.

$$\begin{cases} x_4 + x_5 + x_6 + x_7 = 0 \\ x_2 + x_3 + x_6 + x_7 = 0 \\ x_1 + x_3 + x_5 + x_7 = 0 \end{cases}$$

In pratica, i tre bit supplementari x_5, x_6, x_7 si ricavano dai 4 bit di informazione vera risolvendo il sistema.

Osserviamo che la distanza tra due qualsiasi "parole" del codice è di almeno 3 bit.
Infatti, se la distanza fra due parole distinte $(x_1, \ldots, x_7)$ e $(y_1, \ldots, y_7)$ fosse di soli due bit (o un solo bit) si avrebbe, ponendo $z_i = x_i - y_i$ che $(z_1, \ldots, z_7)$ sarebbe una soluzione del sistema con al massimo due bit diversi da zero. Ma questo è impossibile ... (se per esempio di due bit diversi da zero fossero z_1 e z_2 si avrebbe $z_1 = z_2 = 0$, assurdo; il segreto sta nell'aver costruito una tabella con le colonne a due a due distinte).

Pertanto il codice è in grado di correggere 1 errore.

Come correggerlo?

Se si riceve una parola del codice corretta, non c'e' niente da correggere; se no, sia $(t_1, \ldots, t_7)$ la parola ricevuta, e si calcolino i valori ottenuti inserendo $(t_1, \ldots, t_7)$ nel sistema. Si avrà:

$$\begin{cases} t_4 + t_5 + t_6 + t_7 = a \\ t_2 + t_3 + t_6 + t_7 = b \\ t_1 + t_3 + t_5 + t_7 = c \end{cases}$$

dove

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

è una ed una sola delle colonne della tabella. Se colonna è la i-esima, l'errore sta nella posizione i, e il gioco è fatto!

INTERLEAVING

Se l'informazione è registrata su supporti materiali, come compact disks (ma anche hard disks), si rende necessaria un'ulteriore precauzione. Infatti i supporti materiali possono avere difetti di costruzione e quindi rendere invisibile o sbagliata una lunga serie consecutiva di bit, tale da mettere a rischio l'efficacia del codice costruito.

Con la tecnica di interleaving i bit vengono prima messi in una matrice riempiendo le righe ad una ad una, poi vengono trasmessi colonna per colonna; a questo punto gli eventuali errori consecutivi vengono distribuiti lontano fra loro, e possono essere corretti; infine si ricostruisce la matrice riga per riga.

 L. CHILDS - *Algebra, un'introduzione concreta*, ETS Editrice.

 H. DAVENPORT - *Aritmetica superiore: un'introduzione alla teoria dei numeri*, Zanichelli.

 M. CERASOLI, F. EUGENI, M. PROTASI - *Elementi di matematica discreta*, Zanichelli.

 R.L. GRAHAM, D.E. KNUTH, O. PATASHNIK - *Matematica discreta (Principi matematici per l'informatica)*, Hoepli.

 G.H. HARDY AND E.M. WRIGHT, - *An introduction to the theory of numbers*, Oxford University Press.

 N. KOBLITZ, *A course in number theory and cryptography*, Springer Verlag.

 P. RIBEMBOIM, - *The new book of prime number records*, Springer Verlag.