

Frequenze anomale e falsi in Bilancio

LE SCIENZE – MARZO 2005 – PIERGIORGIO ODIFREDDI

Le lettere dell'alfabeto occorrono, nelle varie lingue, con una frequenza caratteristica della lingua stessa. Per esempio, nell'italiano scritto le lettere più comuni sono «a», «e» e «i», con frequenze rispettive pari a 0,114 (11,4 per cento), 0,111 e 0,104: ovvero, una lettera su tre, all'incirca, è una delle tre vocali. Le lettere meno comuni, invece, sono «z» e «q», con frequenze pari a 0,008 e 0,006: ovvero, solo una lettera su cento, all'incirca, è una delle due consonanti.

Conoscendo l'intera lista delle occorrenze, si può dunque facilmente decifrare un testo sufficientemente lungo in cui le lettere siano indicate con segni convenzionali, invece che con i soliti simboli alfabetici. E conoscendo le liste di occorrenze di varie lingue, si può anche indovinare in quale lingua è scritto il testo. Questo genere di procedimento sta alla base della **crittanalisi statistica**, e rende praticamente inutile la codifica di un testo con semplici metodi di occultamento dell'alfabeto.

È meno noto, invece, che anche le cifre delle liste di **numeri non casuali** occorrono con una precisa frequenza, che va sotto il nome di «**legge di Benford**». Più precisamente, **n** appare come prima cifra con una frequenza pari al logaritmo in base 10 di $1 + 1/n$: ovvero, all'incirca, 1 appare con frequenza del 30 per cento, 2 col 18, eccetera, via via fino all'8 e al 9, che hanno solo una frequenza del 5 per cento. In particolare, circa il 60 per cento dei numeri di una sequenza iniziano con 1, 2 o 3.

Il nome della legge deriva dal fisico Frank Benford, che la scoprì nel 1934, analizzando dati relativi a vari fenomeni naturali, dai bacini dei fiumi alle statistiche del baseball, e la chiamò «**legge dei numeri anomali**». Ma essa era già stata scoperta nel 1881 dall'astronomo Simon Newcomb, che aveva notato come le pagine del suo manuale dei logaritmi fossero stranamente più sgualcite all'inizio che alla fine, in maniera uniformemente decrescente.

Si tratta ovviamente di una **legge empirica**, che però si può derivare in alcuni casi in maniera teorica. Infatti, forse perché la risposta dei nostri sensi è logaritmica, tendiamo a scegliere unità di misura in modo che i dati siano distribuiti in maniera uniforme rispetto all'ordine di grandezza, e non al valore assoluto. Due tipici esempi sono le note della scala musicale e i gradi della scala dei terremoti.

Se i dati sono distribuiti in maniera logaritmica la quantità di numeri che iniziano con la cifra **n** sarà proporzionale alla differenza tra i logaritmi di $n+1$ e di **n**, e dunque al logaritmo del rapporto $(n+1)/n = 1+1/n$: cioè, appunto, alla legge di Benford.

Una giustificazione in un certo senso opposta è stata data nel 1961 dal matematico Roger Pinkham, che ha mostrato come la legge di Benford sia la naturale conseguenza dell'invarianza di scala dei dati: se si vuole ottenere la stessa frequenza anche dopo che tutti i numeri della lista sono stati moltiplicati per una costante arbitraria, per esempio quando gli stessi dati monetari sono espressi in valute diverse, allora la frequenza deve essere quella della legge di Benford.

La legge, come abbiamo detto, non è universale: non vale per liste di numeri casuali, che producono solo frequenze uniformi, e neppure per liste di numeri generate da distribuzioni specifiche, che possono invece produrre qualunque tipo di frequenza. Vale però, sostanzialmente, per **ogni lista naturale** intermedia, che contenga una quantità sufficiente di numeri abbastanza grandi, e che non abbia limiti superiori o inferiori prefissati. E queste condizioni si applicano in molte situazioni: dai numeri primi a quelli di Fibonacci, dalla borsa alle statistiche demografiche, dai dati scientifici ai **bilanci**.

Il che significa che anche la Guardia di Finanza può ormai usare una specie di crittanalisi statistica, che permette di scoprire se i dati sono veritieri o inventati: nel secondo caso, la frequenza delle prime cifre sarà diversa da quella stabilita dalla legge di Benford, e il falso potrà essere scoperto. La California usa ormai da anni programmi informatici di controllo delle dichiarazioni dei redditi basati su questo metodo.