



Queste osservazioni, come indicato da una nota nel libro, completano alcuni accenni contenuti nel testo originale – e anche nei commenti all’edizione italiana – di *How to Lie with Statistics* di Darrell Huff.

La “curva logica” o “campana di Gauss”

Giancarlo Livraghi – maggio 2007

Nel capitolo 2 e nel capitolo 9 del libro di Darrell Huff si parla di quella “curva logica”, o “distribuzione normale”, o “campana di Gauss”, che può comparire in vari punti interessanti di un’analisi statistica. Prende il nome da Carl Friedrich Gauss, studioso tedesco del diciannovesimo secolo che è considerato “uno dei più grandi matematici di tutti i tempi”.

Fra i suoi molti studi c’è la definizione della “distribuzione gaussiana degli errori” (o “variabile casuale normale”) che si traduce graficamente in una “curva a campana”.



Non è questa la sede per entrare nelle complessità matematiche dell’argomento o definire le formule algebriche da cui deriva una curva di questo genere.¹ Ma anche una semplice occhiata alla sua struttura può permetterci di capire come si collocano valori o fattori diversi in rapporto alla loro probabilità.

Non si tratta solo del calcolo di probabilità dell’errore, cosa che interessa ai matematici e agli statistici di professione. Ciò che interessa a tutti, cioè a ogni persona che cerchi di capire il significato di una statistica, è che la “curva normale a campana”, come la chiama Darrell Huff, ci aiuta a sapere come valori di ogni genere (in particolare in biologia, antropologia, sociologia) tendono a distribuirsi in modo non omogeneo.²

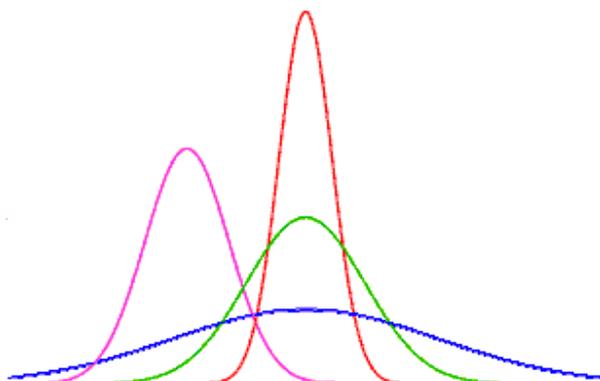
Per fare un esempio semplice, la statura delle persone è distribuita in modo che molti sono vicini alla media, pochi ai livelli più alti e più bassi. In un’infinità di situazioni, anche più complesse e meno ovvie, è utile sapere che questo genere di distribuzione esiste – e il concetto è facilmente “memorizzabile” attraverso la rappresentazione visiva della “campana”.

Fra le più interessanti analisi di Darrell Huff (in particolare nel capitolo 5) c’è la descrizione dei molti modi in cui un grafico può essere deformato (intenzionalmente o per errore) dando una percezione sbagliata del significato dei dati. Quelle anomalie sono meno facili nel caso di una curva tracciata nella sua semplicità, senza enfasi od orpelli. Ma qualsiasi cosa può essere manipolata – per esempio falsando la dimensione delle coordinate e così dando alla curva un’inclinazione più o meno accentuata.

1 Il concetto è spiegato in molti testi di matematica e statistica. Una sintesi online si trova, per esempio, in http://it.wikipedia.org/wiki/variabile_casuale_normale

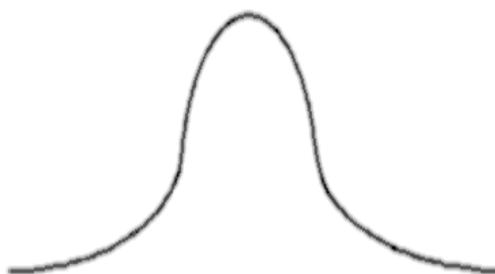
2 Fra le varie equazioni che determinano curve di questo genere, una sola produce quella che è definibile, in senso stretto, “campana di Gauss”. Ma è diffusa l’abitudine di chiamare “gaussiane” anche altre “curve logiche” con un andamento “a campana”, benché derivate da equazioni diverse.

Naturalmente, anche quando la curva è tracciata in modo corretto, il suo andamento cambia secondo il caso che stiamo osservando. Può essere più “montagnoso” o più “pianeggiante”, come vediamo in questa immagine.



Cioè in pratica è utile sapere che la curva c'è, ma i dati (e perciò il tracciato del grafico) possono essere molto diversi secondo la situazione che si sta esaminando e i criteri con cui viene analizzata.

Inoltre, mentre la “campana di Gauss” (nel senso stretto della definizione) è necessariamente simmetrica, altre curve utili in diversi tipi di analisi possono non esserlo. Ci possono essere situazioni in cui i due lati della “campana” hanno un andamento diverso. Come in questo esempio.



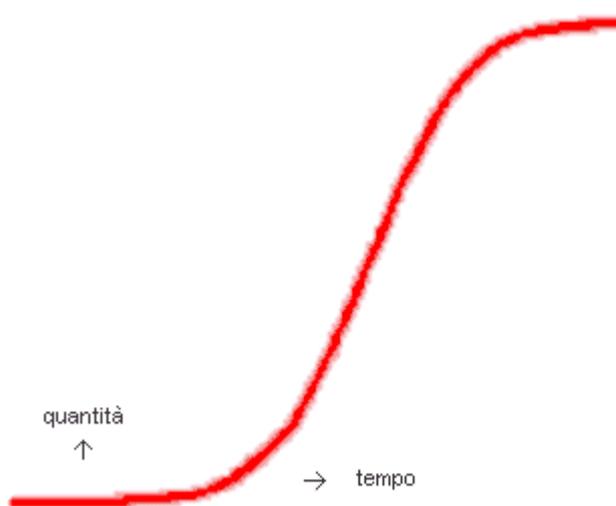
In quest'altro esempio (è il tracciato di un'analisi statistica su dati reali) vediamo una curva ancora più spostata rispetto al “profilo ideale” simmetrico, ma la struttura è analoga a quella della “campana”.



L'immagine simmetrica (pagina 2) è quella che più facilmente ricordiamo per capire il significato della “curva a campana” – anche se, in pratica, non è sempre così. Per chi non si occupa di studiare scientificamente il calcolo delle probabilità, basta sapere che, nelle specifiche applicazioni, il concetto è lo stesso, ma la curva può avere un'altra forma.

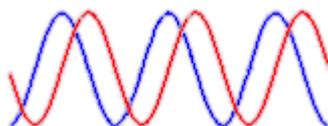
C'è una curva analoga che si può applicare in vari modi, ma è spesso un'efficace rappresentazione dello sviluppo di un fenomeno nel tempo.

Si tratta di un quadrante di “coordinate cartesiane” dove nelle ascisse (asse orizzontale) si indica il tempo e nelle ordinate (asse verticale) la quantità.



Dalla sua forma è evidente il motivo per cui, oltre che “curva logica” o “normale” o “curva a S”, può essere chiamata “semicampana di Gauss” o più genericamente “gaussiana” (anche se, da un punto di vista algebrico, quella definizione può non essere esatta).

Talvolta questo genere di curva è chiamato “sinusoidale” perché la “campana” somiglia a una porzione di una “sinusoide” logaritmica – o di una “curva d'onda”.

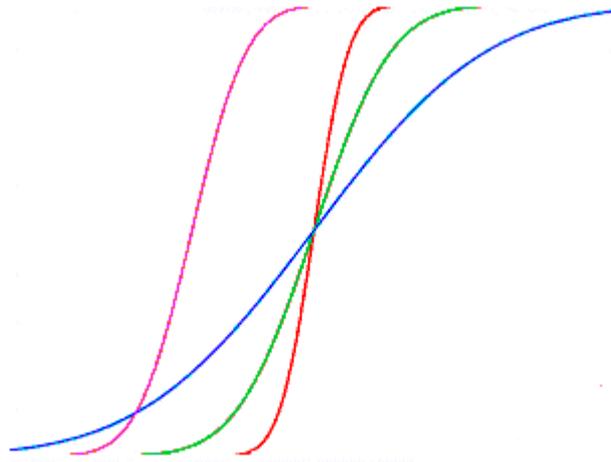


Sempre senza entrare nelle complessità scientifiche della matematica e della fisica, non è una coincidenza che diverse curve “naturali” abbiano andamenti analoghi.

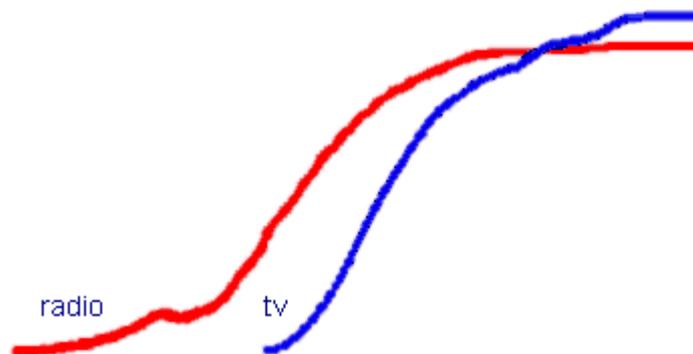
Molti dati in biologia, sociologia, economia, eccetera, come nelle attività o situazioni di una singola persona o impresa, tendono a collocarsi secondo una curva di questo genere. Cioè lo sviluppo di qualcosa è inizialmente lento, poi ha una progressiva accelerazione, poi di nuovo rallenta quando si avvicina a un “punto di saturazione”.

(Sembra che alcuni abbiano capito male. Quando parlano di “ciclo di vita” di un comportamento, di una tendenza, di un prodotto, eccetera, talvolta sembrano avere in mente la “campana” nella sua interezza – cioè ragionano come se i criteri di distribuzione probabilistica fossero “analogamente” applicabili all’evoluzione nel tempo – e perciò tutto ciò che sale fosse automaticamente condannato a scendere nello stesso modo e in un uguale periodo di tempo. È ovviamente possibile che a una fase di crescita ne segua una di declino, ma in molti casi non è così – e comunque l’andamento in discesa è raramente descrivibile o “proiettabile” in base alle “curve normali” di cui stiamo parlando).

Se è spesso prevedibile che un andamento di questo genere ci sia, è difficile “proiettare” in modo attendibile il ciclo di evoluzione. Le quantità e le velocità possono essere molto diverse, come è simbolicamente rappresentato in questa immagine.



Un esempio fra tanti possibili... questo grafico (approssimativo e molto semplificato) mostra la diffusione della radio e della televisione in Italia dal 1926 al 2003. Vediamo due curve con la stessa struttura, ma con diverse evoluzioni nel tempo. ³



³ Per dati e analisi su queste tendenze vedi *Storia dei sistemi di informazione e comunicazione in Italia* <http://gandalf.it/storia/> e per analoghi studi su un periodo più recente *Evoluzione complessa fra cambiamenti e continuità* <http://gandalf.it/cens2006/>

Anche in molte altre situazioni si rileva come gli andamenti reali possano riflettere una curva di questo genere, ma con variazioni dovute a una varietà di circostanze e con evoluzioni spesso diverse da ciò che precedentemente sembrava “prevedibile”. In generale – il concetto è sostanzialmente valido. È spesso concretamente verificato nei fatti. L’idea è semplice e chiara. Le applicazioni a singoli casi, invece, sono meno ovvie e possono essere molto più complesse.

I motivi per cui, dopo una fase iniziale di stasi o di crescita lenta, c’è un’accelerazione, possono essere vari. Ma spesso si tratta del modo in cui si sviluppano le relazioni. Qualcosa (un’idea, un’informazione, un comportamento, un prodotto) è conosciuto da poche persone. Quando si arriva al punto in cui molti conoscono qualcuno che sa o ha quella cosa, si supera la “soglia iniziale” e si innesca un’accelerazione. La crescita sarà veloce fino al raggiungimento della “soglia di saturazione”, quando tutte le persone che possono essere interessate fanno o hanno la “cosa” di cui stiamo valutando la diffusione. Da lì in avanti lo sviluppo è molto più lento e può essere “tendente a zero”.

(Lo stesso processo si verifica per ogni sorta di fenomeni, compresi quelli preoccupanti, come la diffusione non solo di malattie contagiose, ma anche di comportamenti socialmente dannosi. Quando manca un’adeguata prevenzione tendono a crescere fino al loro “limite di saturazione”, che può essere pericolosamente alto).

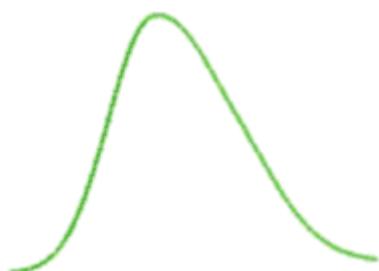
Il tracciato successivo della curva è difficilmente proiettabile. Raggiunto il “livello di saturazione”, che cosa potrà succedere dopo? Non c’è una “curva normale” applicabile in tutti i casi, perché le evoluzioni possono essere molto diverse.

In alcuni casi la situazione si stabilizza: arrivata a una certa dimensione, che è il limite strutturale, la quantità rimane statica o ha una leggera crescita dovuta a fattori generali (per esempio l’aumento della popolazione). In altri casi ci può essere un crollo precipitoso: perché tramonta qualche labile moda o perché qualcos’altro arriva a sostituire ciò che stiamo misurando. La “curva” risultante può avere un aspetto come questo.



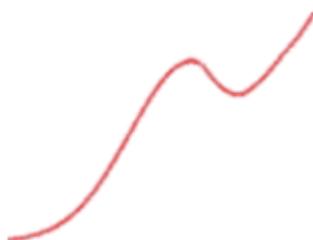
Questo esempio è del tutto immaginario, ma somiglia ad alcuni percorsi che si rilevano in situazioni reali.

Oppure ci può essere una fase lenta di “esaurimento” in cui la spinta propulsiva è venuta a mancare, ma fattori di inerzia rallentano il declino. In situazioni di quel genere ci si può trovare davanti a una curva “assimmetrica” come questa.



Anche questo esempio è immaginario,
ma ci sono casi reali in cui la “curva” ha questo aspetto.

Ci sono anche situazioni in cui, raggiunta la “saturazione” secondo certi parametri, si aprono nuovi orizzonti per cui una certa cosa trova usi diversi, o un certo comportamento si diffonde ad altre persone in modo diversamente percepito, eccetera – e così nasce una nuova fase di sviluppo che può avere un nuovo andamento analogo (ma probabilmente non identico) a quello della “curva” iniziale. Cioè si verifica una situazione come questa.



Anche questa “curva” è immaginaria,
ma ci sono casi reali di evoluzioni strutturalmente analoghe.

Insomma... nessun criterio di analisi può essere applicato meccanicamente in modo generico. La diffusione e l’evoluzione vanno sempre studiate nella specificità di ogni singolo fenomeno. Ma conoscere e ricordare le “curve naturali” può aiutarci a capire e interpretare ogni sorta di dati, situazioni, avvenimenti e comportamenti.