

# La distribuzione normale - Parte I: Introduzione

Michele Nichelatti, Maurizio Nordio, Umberto Maggiore, Maurizio Postorino, Aurelio Limido, Anteo Di Napoli

a nome del Comitato Scientifico SIN-RIDT

---

## THE NORMAL DISTRIBUTION – PART I: INTRODUCTION

**Abstract.** The Gaussian function is presented as one of the first possible options to describe the distribution of data that have a high frequency in their middle values and a low frequency in their extreme values. The Euler's number  $e$  and its derivation are also here presented.

**Key words:** Normal distribution, Gaussian function, Euler's number

**Conflict of interest:** None.

**Financial support:** None.

---

Accettato: 18 Febbraio 2014

## In medio stat virtus

Fin dai tempi della scuola dell'obbligo, ci siamo abituati ad ascoltare e ad accettare il motto latino "*in medio stat virtus*" (che, in realtà, è figlio del "*μέσων τε καὶ ἄριστων*" di Aristotele), una specie di quintessenza della saggezza e sobrietà degli antichi nostri padri. Poi, magari, abbiamo provato a pensarci sopra e abbiamo capito che non è vero niente e che quel motto latino è solo una gossissima sciocchezza.

Noi siamo degli ex cacciatori carnivori, che l'evoluzione ha in parte convertito alla pastorizia (che, se vogliamo, è un *proxy* della caccia) e all'agricoltura, ma, nell'animo, siamo rimasti, appunto, dei cacciatori: l'istinto dell'uomo, come quello dei cacciatori, è essenzialmente la curiosità per le cose strane e per le cose nuove. Fare della ricerca è come andare a caccia e un bravo ricercatore deve avere lo stesso atteggiamento di un provetto cacciatore come è il gatto: forte curiosità, istinto per la sperimentazione (il gatto annusa e poi tocca con la zampa gli oggetti che attirano la sua attenzione: è il suo metodo sperimentale), molta prudenza, e un po' di saggio e disincantato scetticismo. Difficile che un gatto venga incuriosito da un oggetto qualsiasi, che vede tutti i giorni.

Ecco, quindi, che in *in medio stat virtus* non è il motto ispiratore né per chi fa ricerca né per l'uomo in generale: siamo attirati dalle cose strane, guardiamo con curiosità un uomo o una donna bellissimi, ma anche un uomo e una donna bruttissimi. Un *business-man* come Phineas Taylor Barnum nel 1872 è diventato ricco mettendo in piedi un circo di fenomeni da baraccone ("The Greatest Show on Earth"). Per la precisione, il Circo Barnum schierava: 1) l'uomo più piccolo del mondo, 2) la donna barbata, 3) due nani gemelli, 4) l'uomo con la faccia di cane, 5) la donna con la criniera sulla schiena, 6) due gemelli siamesi e 7) il gigante del Congo.

Fenomeni estremi, se così vogliamo chiamarli, ma che attira-

vano la gente. Certo, qui siamo più vicini a un morboso voyeurismo che alla curiosità scientifica, ma il Nostro sapeva benissimo come far soldi; non per niente, oltre che essere uno scaltro uomo d'affari, Barnum era anche un matematico: uno scienziato, insomma.

Infine, pensiamo ai giornali di gossip (vendutissimi), che altro non sono che una trasposizione moderna e lievemente edulcorata del circo Barnum: nessuno comprenderebbe un giornale che parla della vita banale del condomino del piano di sotto o della suocera del geometra che lavora nell'amministrazione della nostra azienda. Siamo, invece, incuriositi dalle gesta estreme di principesse decotte che nominano erede universale il proprio cagnolino o di attori in disarmo che fanno *outing* o di ex pornstar che si convertono e diventano monache di clausura.

## Un piccolo esercizio mentale

Pensiamo a un nostro modo molto comune di ragionare o di esporre delle idee: esprimere concetti di tutti i giorni, come *Tizio è alto* oppure *Caio è grasso* o, ancora, *Sempronia è molto bella*, implica che, nella nostra testa, abbiamo almeno delineato il concetto di altezza media, di peso medio e di bellezza media di una popolazione di riferimento e che le caratteristiche somatiche di Tizio, Caio e Sempronia sono sufficientemente diverse dalle medie di riferimento che abbiamo in testa, tanto da colpire la nostra attenzione e da farci classificare tali caratteristiche, se non come fenomeni estremi, almeno come fenomeni degni di nota (tanto da ritenerli qualificanti), utilizzando, nella descrizione, l'aspetto che li differenzia dalla media.

Prendiamo l'altezza delle persone, e, mentalmente, facciamo un esercizio semplicissimo: rispondiamo, cioè, alle seguenti domande e poi ragioniamo (sempre mentalmente) sulle risposte:

1. Quando definiamo una persona *alta*, *molto alta*, *bassa* o *molto bassa*?
2. Qual è l'intervallo delle altezze all'interno del quale non utilizziamo alcuna classificazione, proprio perché l'altezza la consideriamo *normale* e quindi non degna di nota?
3. L'intervallo di cui al punto precedente lo consideriamo identico per i due sessi oppure variabile a seconda del sesso?

### La distribuzione degli eventi e la curva di Gauss

Qualunque sia la nostra altezza, citiamo quella di un'altra persona solo se la sua ha qualcosa di particolare, che la rende degna di nota: se la persona ha, invece, un'altezza che riteniamo "normale" o, comunque, "non degna di particolare attenzione", viene naturale non citarla, perché dire che il geometra Bianchi ha un'altezza normale è una cosa perfettamente superflua; dire che, invece, è molto alto o molto basso aggiunge una certa qualità all'informazione.

Da tempo siamo, quindi, abituati a prestare attenzione agli eventi insoliti, rari o particolari. Consideriamo "normale" che una persona abbia una statura "media" e consideriamo più rari gli eventi estremi associati alla statura elevata e alla statura bassa. Proprio da qui parte il concetto di "distribuzione normale": è normale che gli eventi estremi (in una direzione o nell'altra) siano più rari degli altri. Sin dall'origine della statistica, anche sulla scorta delle considerazioni fatte precedentemente, si è, quindi, cercata una rappresentazione matematica in grado di esprimere la frequenza degli eventi routinari e la rarità degli eventi straordinari: per questo motivo, una curva a forma di campana era considerata il modo migliore di raffigurare la distribuzione degli eventi. La prima curva a campana è stata studiata da Gauss<sup>1</sup>, che l'ha definita tramite la formula

$$y = e^{-x^2}, \quad (1)$$

dove  $e = 2.71828\dots$  è il numero di Euler<sup>2</sup>, ovvero la base dei logaritmi naturali, che si può esprimere come

$$e = \lim_{n \rightarrow +\infty} \left(1 + \frac{1}{n}\right)^n, \quad (2)$$

oppure come somma della serie infinita

$$e = \sum_{n=0}^{+\infty} \frac{1}{n!} = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \dots$$

La curva di Gauss è molto interessante, per tanti motivi, in particolare perché l'area che sottende (i matematici la chiamerebbero l'*integrale della curva*) è pari a  $\sqrt{\pi}$ , e questo è uno dei tanti fatti curiosi<sup>3</sup> che collegano i numeri  $e$  e  $\pi$ , ma, per gli statistici, è importante (per motivi che saranno resi più chiari

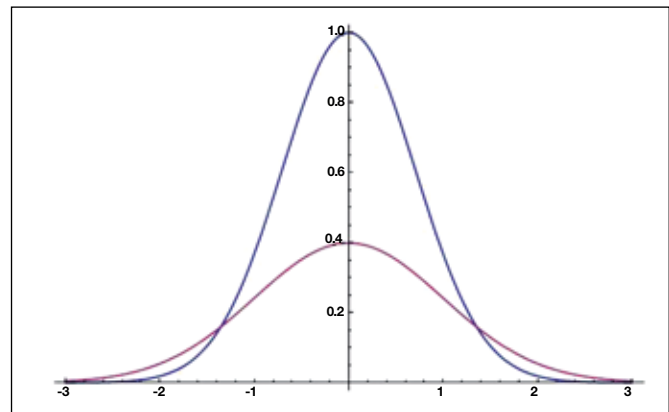


Fig. 1 - In blu, è presentato il grafico della funzione di Gauss, data dall'equazione (1), mentre in rosso si rappresenta la funzione "modificata" data dall'equazione (3). L'area sottesa dalla curva blu (cioè il suo integrale) è pari a  $\sqrt{\pi}$ , mentre quella sottesa dalla curva rossa è uguale a 1.

nei prossimi numeri) che la curva sottenda un'area pari a 1 e, quindi, la curva di Gauss va sottoposta a una lieve modifica (ce ne saranno altre), trasformandola in

$$y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}; \quad (3)$$

in questo modo, l'area sottesa dalla curva viene "forzata" a essere pari a 1 e la curva non ha più il punto di massimo quando  $y = 1$ , bensì quando  $y = 1/\sqrt{2} \approx 0.39894$  (vedi Fig. 1).

Il grafico della curva (3) è costruito per assomigliare al profilo di un istogramma che descrive dei dati "normali" che raccogliamo tutti i giorni: il peso dei ragazzi della classe di un Liceo, l'altezza dei colleghi che lavorano in ospedale e la pressione sistolica dei pazienti che abbiamo visitato la scorsa settimana. Ma su questo torneremo nei prossimi numeri; ora conviene capire cosa sia quel numero  $e$ , il numero di Euler che abbiamo appena incontrato.

### Che cos'è e da dove deriva il numero $e$

Il numero di Euler viene da lontano e, soprattutto, da un posto davvero inatteso: una banca. Nel '500 si teorizzava quale fosse il massimo guadagno che si potesse realizzare, per la banca e il correntista, con gli interessi bancari, a parità di deposito. Questo problema era stato già studiato da altri matematici (tra cui Napier), ma, solo nella seconda metà del '600, Bernoulli<sup>4</sup> è riuscito a trovare la definizione esatta di  $e$ , ovvero quella dell'equazione (2).

In pratica, il problema si riconduce a questo: depositando in banca 1 euro, a quanto ammonterà il conto corrente, dopo un anno, se il tasso corrisposto è del 100%? La risposta, fino a qui, è banale, perché nel conto, a fine anno, vi saranno 2 euro. Tuttavia, le cose cambiano sensibilmente se ragioniamo in termini di interesse composto. Se, anziché corrispondere il 100% a 12 mesi, la banca corrispondesse il 50% a 6 mesi e

<sup>1</sup> Carl Friedrich Gauss, matematico tedesco (1777-1855)

<sup>2</sup> Leonhard Euler, matematico svizzero (1707-1783)

<sup>3</sup> In realtà non si tratta per niente di fatti curiosi, bensì di motivi matematici evidenti e chiarissimi

<sup>4</sup> Jakob Bernoulli, matematico svizzero (1654-1705)

il rimanente 50% a 12 mesi, come cambierebbero le cose? In questo caso, al capitale di 1 euro si aggiungerebbero 50 centesimi di interesse a 6 mesi, e, quindi, a 12 mesi, l'interesse corrisposto sarebbe il 50% di 1.50 euro, cioè 0.75 euro, per cui, a fine anno, il capitale sarebbe complessivamente di 2.25 euro. Continuando il ragionamento, se la banca pagasse gli interessi del 25% per 4 volte all'anno, cioè ogni 3 mesi, il capitale a 3 mesi sarebbe di 1.25 euro, a 6 mesi sarebbe di 1.5625, a 9 mesi sarebbe di 1.95313 e allo scadere dell'anno sarebbe di 2.44141 euro. E, se gli interessi venissero corrisposti al tasso dell'8.33% una volta ogni mese (circa), a fine anno ci ritroveremmo con un capitale di 2.61304 euro. Continuando in questo modo, con interessi pagati una volta alla settimana, una volta al giorno e una volta all'ora, il capitale a un anno sarebbe rispettivamente pari a 2.69260, 2.71457, 2.71813 e 2.71828 euro. Non importa aumentare la finezza della discriminazione temporale: abbiamo già capito che il guadagno non è illimitato e che il capitale raggiunge un valore massimo che non può essere in alcun modo superato, per quanto frequentemente siano pagati gli interessi. Il valore massimo del capitale a fine anno, dopo il deposito di 1 euro, sarà di 2.71828... euro, cioè di  $e$  euro.

Per le sue proprietà matematiche,  $e$  è diventato la base dei logaritmi naturali, e la funzione esponenziale  $e^x$  è caratterizzata dal fatto di avere una velocità di crescita pari a se stessa, unica tra tutte le funzioni conosciute dell'analisi matematica. Soprattutto, più in generale, il numero  $e$ , e le funzioni esponenziali li ritroveremo tutte le volte che avremo a che fare con fenomeni di crescita o decrescita. In pratica, le dinamiche di una popolazione umana o animale, del numero di ceppi virali in coltura, delle cellule cancerose che invadono un organismo e di qualsiasi ensemble che possa cambiare il numero dei suoi componenti sono determinate e descritte da funzioni che contengono  $e$ .

## Riassunto

Si presenta la funzione di Gauss come primo provvisorio modello usato dagli statistici per descrivere la distribuzione di dati in cui sono più frequenti i valori medi di una variabile e meno frequenti quelli estremi. Viene presentato anche il numero di Euler e e la sua derivazione.

**Parole chiave:** Distribuzione normale, Funzione di Gauss, Numero di Euler

**Dichiarazione di conflitto di interessi:** Gli Autori dichiarano di non avere conflitto di interesse.

**Contributi economici agli Autori:** Gli Autori dichiarano di non aver ricevuto sponsorizzazioni economiche per la preparazione dell'articolo.

### Indirizzo degli Autori:

Dr. Michele Nichelatti  
 Servizio di Biostatistica  
 Dipartimento di Ematologia e Oncologia  
 Ospedale Niguarda Ca' Granda  
 Piazza Ospedale Maggiore 3  
 20162 Milano  
[michele.nichelatti@ospedaleniguarda.it](mailto:michele.nichelatti@ospedaleniguarda.it)