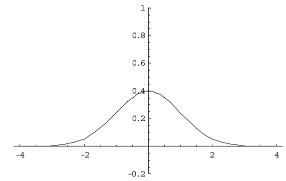
MatematicaMente

Pubblicazione mensile della sezione veronese della MATHESIS – Società Italiana di Scienze Matematiche e Fisiche – Fondata nel 1895 – Autorizzazione del Tribunale di Verona n. 1360 del 15 – 03 – 1999 – I diritti d'autore sono riservati. Direttore: Luciano Corso - Redazione: Luciano Corso, Elisabetta Capotosto, Carlo Marchiori – Via IV Novembre, 11/b – 37126 Verona – tel e fax (045) 8344785 – 338 6416432
e-mail: lcorso@iol.it – Stampa in proprio - Numero 132 – Uscito il 22 – 12 – 2008



Saggio sulla distribuzione Normale
Parte III

di Luciano Corso

[*Segue dal numero* 131] **La dimostrazione di Gauss** [**]

Iohann Carl Friedrich Gauss (1777 – 1855) nel 1809 arrivò alla Normale, in modo indipendente e per altra via. Egli usò la distribuzione Normale per lo studio e l'analisi dei dati provenienti dalle osservazioni astronomiche. Sia Gauss, sia Laplace considerarono la Normale come curva degli errori.

Gauss, arrivò alla Normale staccandosi dallo schema delle prove ripetute di De Moivre-Laplace, impostando il ragionamento in modo originale e molto generale attraverso la formula di Bayes (Thomas Bayes, britannico, 1702 - 1761) della probabilità delle cause a posteriori [B.9, pagg. 205, 212]. La dimostrazione di Gauss che porta da una serie di osservazioni sperimentali alla Normale è sorprendente [si trova in *Theoria motus corporum coelestium* del 1809]. Vale la pena presentarla in chiave moderna per comprendere la sua genialità (si rinvia a [B.9] per una verifica completa di quanto qui riportato).

Data una v. a. X che in un esperimento assume una serie di numeri aleatori $x_1, x_2, \dots, x_n$ inerenti a n misurazioni accurate di una certa grandezza fisica di «valore vero» x (ignoto, che si presume esista), si consideri la sequenza di errori (scarti)

$$z_1 = x - x_1,\ z_2 = x - x_2,\ \dots,\ z_n = x - x_n. \qquad (15)$$

Posta la condizione che la probabilità di commettere un generico errore z sia funzione solo dell'entità dell'errore stesso in ogni prova effettuata (e non del valore vero x), allora:

$$P(z \le Z \le z+dz) = \varphi(z)\cdot d(z). \qquad (16)$$

Per il teorema delle probabilità composte per eventi indipendenti, la probabilità congiunta relativa all'evento che ciascuno degli n errori Z_k cada, rispettivamente, nell'intervallo differenziale $((z_k\,,\, z_k + dz_k]$, sapendo che la misura della grandezza X è x, è data da:

$$P\left(\bigcap_{k=1}^{n}\left(z_k < Z_k \le z_k + dz_k \middle| X = x\right)\right) = \prod_{k=1}^{n}\varphi(z_k)\cdot d(z_k). \qquad (17)$$

Nella moderna terminologia statistica questa espressione rappresenta quanto è verosimile (*likelihood*) che gli errori siano compresi nei vari intervalli differenziali $(z_k, z_k+dz_k]$ se la vera misura della grandezza X è x. A questo punto della dimostrazione Gauss utilizza un'impostazione tipicamente bayesiana per determinare la probabilità a posteriori $\Pi(x)\cdot dx$ che l'ignoto valore x della grandezza X sia compreso nell'intervallo $(x, x+dx]$, sapendo che il vettore aleatorio $\mathbf{X}$ delle n misure $(X_1, X_2, \dots, X_n)$ ha dato luogo alla n-pla di valori $(x_1, x_2, \dots, x_n)$. Tenendo presente la particolare versione del teorema di Bayes nel caso di un numero aleatorio X e di un vettore di numeri aleatori $\mathbf{X}$, entrambi continui, in virtù della quale

$$f_X(x|\mathbf{X}=\mathbf{x}) = \frac{f_X(x)\cdot\prod_{k=1}^{n} f_{X_k}(x_k|\,X=x)}{\int_{-\infty}^{+\infty} f_X(x)\cdot\prod_{k=1}^{n} f_{X_k}(x_k|\,X=x)\cdot dx}$$

Risulta:

$$\Pi(x)\cdot dx = P(x \le X \le x+dx | X_1 = x_1, \dots, X_n = x_n) =$$

$$\frac{P(x\le X\le x+dx)\cdot P(X_1 = x_1, \dots, X_n = x_n | x \le X \le x+dx)}{P(X_1 = x_1, \dots, X_n = x_n)} =$$

$$\frac{P(x\le X\le x+dx)\cdot P(z_1 \le Z_1 \le z_1 + dz_1, \dots, z_n \le Z_n \le z_n + dz_n)}{P(X_1 = x_1, \dots, X_n = x_n)}.$$

Posto $P(x \le X \le x+dx) = \omega(x)\cdot dx$ il teorema porta a:

$$\Pi(x)\cdot dx = \frac{\omega(x)\cdot dx\cdot[\varphi(z_1)\cdot dz_1\cdot\varphi(z_2)\cdot dz_2\cdot\ldots\cdot\varphi(z_n)\cdot dz_n]}{\int_{-\infty}^{+\infty}\omega(x)\cdot\varphi(z_1)\cdot dz_1\cdot\varphi(z_2)\cdot dz_2\cdot\ldots\cdot\varphi(z_n)\cdot dz_n\cdot dx}$$

$$\Pi(x)\cdot dx = \frac{\omega(x)\cdot dx\cdot[\varphi(z_1)\cdot dz_1\cdot\varphi(z_2)\cdot dz_2\cdot\ldots\cdot\varphi(z_n)\cdot dz_n]}{dz_1\cdot dz_2\cdot\ldots dz_n\cdot\int_{-\infty}^{+\infty}\omega(x)\cdot\varphi(z_1)\cdot\varphi(z_2)\cdot\ldots\cdot\varphi(z_n)\cdot dx}$$

Semplificando i differenziali si ottiene:

$$\Pi(x) = \frac{\omega(x)\cdot[\varphi(z_1)\cdot\varphi(z_2)\cdot\ldots\cdot\varphi(z_n)]}{\int_{-\infty}^{+\infty}\omega(x)\cdot\varphi(z_1)\cdot\varphi(z_2)\cdot\ldots\cdot\varphi(z_n)\cdot dx} \qquad (19)$$

In cui $\Pi(x)\cdot dx$ dipende da x. Come è noto il denominatore è una costante. Poniamo

$$C_0 = \left(\int_{-\infty}^{+\infty}\omega(x)\cdot\varphi(z_1)\cdot\varphi(z_2)\cdot\ldots\cdot\varphi(z_n)\cdot dx\right)^{-1} \qquad (20)$$

e perciò la (19) diventa:

$$\Pi(x) = C_0\cdot\omega(x)\cdot\varphi(z_1)\ldots\ \varphi(z_n). \qquad (21)$$

$\omega(x)$ è distribuzione di probabilità a priori di x. Gauss ipotizza (2° ipotesi) che tale funzione sia una costante, il che vale a dire che a priori tutti i possibili valori della v.a. x abbiano uguale densità di probabilità. Da ciò, conglobando $\omega(x)$ in C_0, la (21) diventa:

$$\Pi(x) = C\cdot\varphi(z_1)\ \ldots\ \varphi(z_n). \qquad (22)$$

Su x, di cui non si sa quasi niente, si può dire solo che è più probabile che abbia un valore compreso tra il minimo e il massimo dei valori sperimentali ottenuti. A questo punto Gauss fa una 3° ipotesi: postula che $\Pi(x)$ abbia il massimo in corrispondenza della media aritmetica degli n valori sperimentali $x_1, x_2, \dots, x_n$:

$$Max(\Pi(x)) \Rightarrow \left(\bar{x} = \frac{1}{n}\cdot\sum_{i=1}^{n} x_i\right). \qquad (23)$$

Pertanto in corrispondenza di $\bar{x}$ si deve annullare la derivata prima di $\Pi(x)$. Gauss fa prima una trasformazione logaritmica di $\Pi(x)$ e poi calcola la derivata prima di tale trasformazione ponendola uguale a zero. Inoltre, ammette che le proprietà della funzione φ incognita siano comuni a tutte le n-ple di misure $x_1, x_2, \dots, x_n$.

Con $X = \bar{x}$ gli errori

$$z_k = \bar{x} - x_k$$

soddisfano il sistema:

$$\begin{cases} \dfrac{d}{dx} Ln\,\Pi(x) = \dfrac{\varphi'(z_1)}{\varphi(z_1)} + \dfrac{\varphi'(z_2)}{\varphi(z2)} + \ldots + \dfrac{\varphi'(z_n)}{\varphi(z_n)} = 0 \\ z_1 + z_2 + \ldots + z_n = 0\,, \end{cases} \qquad (24)$$

essendo

$$Ln[\Pi(x)] = Ln(C) + Ln[\varphi(z_1)] + \ldots + Ln[\varphi(z_n)].$$

Allora, posto

$$\psi(z) = \frac{\varphi'(z)}{\varphi(z)}\ , \qquad (25)$$

otteniamo:

$$\begin{cases} \psi(z_1)+\psi(z_2)+...+\psi(z_n)=0 \\ z_1+z_2+...+z_n=0 \end{cases} . \qquad (26)$$

Questa equazione ha un'unica soluzione data da:

$$\psi(z)=a\cdot z\ , \qquad (27)$$

ove a è una costante arbitraria. Vediamo come si arriva a questo risultato. Consideriamo il fatto che

$\psi(z)+\psi(-z)=0$

Per ogni scarto z appare quindi una simmetria polare di ψ rispetto a z.

Poiché il sistema (26) deve essere soddisfatto per ogni insieme di valori $z_1, z_2, \dots, z_n$ la cui somma è zero, allora sarà soddisfatto, in particolare, per la seguente serie di $\nu+\mu$ misure:

$z_1 = z_2 = \dots = z_\nu = s_0/\nu$ e $z_{\nu+1} = z_{\nu+2} = \dots = z_{\nu+\mu} = -s_0/\mu$,

dove s_0 è un numero reale qualsiasi.

È possibile quindi scrivere la (24) nel modo seguente:

$$\nu\cdot\psi(s_0/\nu)+\mu\cdot\psi(-s_0/\mu)=0\ , \qquad (28)$$

da cui

$$\nu\cdot\psi(s_0/\nu)-\mu\cdot\psi(s_0/\mu)=0\ ,$$

$$\psi(s_0/\nu)=\frac{\mu}{\nu}\cdot\psi(s_0/\mu)\ .$$

Infine otteniamo:

$$\psi(z)=\frac{\mu}{\nu}\cdot\psi(s_0/\mu)=a\cdot z\ ,$$

in quanto $\psi(s_0/\mu) = a$ è una costante, essendolo s_0/μ, e $\mu/\nu = z$ razionale.

Il rapporto μ/ν di z, vista la continuità di ψ, può essere esteso al campo dei numeri reali. Ritornando alla (25) si ha:

$$\psi(z)=\frac{\varphi'(z)}{\varphi(z)}=a\cdot z\ .$$

Integrando si ottiene perciò:

$$Ln(\varphi(z))=\frac{a}{2}\cdot z^2+c \qquad (29)$$

e quindi

$$\varphi(z)=\exp\left(\frac{a}{2}\cdot z^2+c\right)=k\cdot e^{\left(\frac{a}{2}\cdot z^2\right)}=k\cdot e^{-h^2\cdot z^2}\ . \qquad (30)$$

Le due costanti k e $h^2 = (a/2)^2$ sono positive in quanto φ è positiva per ogni z e ha un massimo in $z = 0$, posto, come fa Gauss, l'ipotesi che $n = 1$. La $\varphi(z)$ è ammessa su tutto l'asse reale, anche quindi per valori molto grandi di z. Si integra perciò su tutto l'asse reale e tale integrale deve essere uguale a 1:

$$\int_{-\infty}^{+\infty}\varphi(z)\cdot dz=k\cdot\int_{-\infty}^{+\infty}e^{-h^2\cdot z^2}\cdot dz=1\ . \qquad (31)$$

L'integrale (31), che possiamo risolvere in coordinate polari, è uguale a:

$$\int_{-\infty}^{+\infty}e^{-h^2\cdot z^2}\cdot dz=\frac{\sqrt{\pi}}{h} \qquad (32)$$

da cui $k = h/\sqrt{\pi}$. Ottieniamo così l'importante risultato che la probabilità di commettere un errore compreso tra z e $z+dz$ nella misura di una grandezza è espressa dalla relazione:

$$\frac{h}{\sqrt{\pi}}\cdot e^{-h^2\cdot z^2}\cdot dz \qquad (33)$$

Dove h è una opportuna costante che dipende dall'accuratezza delle misure fatte.

Per dare corpo al risultato finale di Gauss, si dimostra ora che h è uguale a

$$h=\frac{1}{\sigma\cdot\sqrt{2}}\ , \qquad (34)$$

ove σ è lo scarto quadratico medio [si veda [B.9] pagg. 211 e segg.]. Il parametro h viene detto grado di precisione delle misurazioni. Maggiore è h e maggiore è a probabilità di commettere piccoli errori di misura e, corrispondentemente, minore è la probabilità di fare errori grandi. Se consideriamo lo scarto quadratico medio σ, per valutare la dispersione media di queste misure rispetto al valore vero, otteniamo:

$$\sigma^2=M(z^2)-M^2(z)=\left(\frac{h}{\sqrt{\pi}}\cdot\int_{-\infty}^{+\infty}z^2\cdot e^{-h^2\cdot z^2}dz\right)-0=\frac{1}{2\cdot h^2}$$

e quindi:

$$\sigma=\frac{1}{h\cdot\sqrt{2}}\ .$$

Si noti che lo scarto (o errore) quadratico medio è inversamente proporzionale alla precisione h. Sostituendo (34) in (33) otteniamo, infine, il risultato cui è giunto Gauss:

$$\varphi(x_k)=\frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot e^{-\frac{(x_k-x)^2}{2\cdot\sigma^2}} \qquad (35)$$

con dominio $x_k\in\mathbf{R}$ (si confronti con la (11)). Questo risultato è la prima espressione completa e moderna della funzione di densità di probabilità Normale.

Osservo che Gauss, rispetto alla nostra sensibilità dimostrativa fa dei salti. Egli, infatti, dà per scontati alcuni passaggi che invece presentano delle difficoltà concettuali (peraltro viste alla luce del XXI secolo). Nonostante ciò, la dimostrazione di Gauss costituisce il fondamento della (11), non essendoci nessun altra dimostrazione precedente così organica e controllabile nel suo sviluppo costruttivo.

[*Segue al numero* 135]

[**] La presente dimostrazione è una mia personale interpretazione della vera dimostrazione di Gauss. Rispetto a quella di Gauss, e anche alla interpretazione riportata da Guido Castelnuovo in [B.9] ho migliorato, in senso moderno, alcuni passaggi riguardanti la simbologia usata, l'interpretazione del processo costruttivo che ha portato alla dimostrazione e l'organicità – anche in senso didattico – dell'applicazione del teorema di Bayes.

Ringrazio il collega Alberto Guerrini per la revisione di questa mia interpretazione della dimostrazione di Gauss e per il contributo datomi a migliorare l'esposizione di alcuni passaggi delicati.

Una sintesi di questo lavoro è stata presentata, il 7 dicembre, al congresso nazionale della Mathesis tenutosi a Lecce, dal 5 all'8 dicembre 2008, presso la Facoltà di Lingue. Pubblichiamo per concessione della Mathesis Nazionale.

L'induzione … degli economisti

Le storie – che parlino di corvi neri, di mucche o di quant'altro – sono sempre le stesse: invenzioni fantasiose! Haddon (Mark Haddon, *Lo strano caso del cane ucciso a mezzanotte*, Einaudi, Torino, 2003) per dimostrare quanto siano fallaci le induzioni racconta la sua storia: un economista, un logico e un matematico viaggiano su un treno, in Scozia, e vedono dal finestrino una mucca bruna. L'economista afferma, di conseguenza, che le mucche sono brune in Scozia; il logico dichiara che almeno una mucca è bruna in Scozia; il matematico, infine, dice che in Scozia c'è almeno una mucca con almeno una parte (quella che si vede) bruna. L'induzione porta a tre conclusioni distinte: ovviamente quella del matematico è sicura, quella del logico è precisa ma non certa, quella dell'economista appare la più ingenua. L'esempio di Haddon ha almeno un errore: pensare che gli economisti ragionino ponendo al primo posto la logica. Il problema è che un economista non ragiona solo secondo categorie di verità e soprattutto non pone al primo posto la logica nella gerarchia delle scelte del metodo di valutazione degli enunciati. La logica, infatti, è importante, ma lo è di più – per lui – il calcolo di convenienza economica. Un economista, di fronte alla necessità di dichiarare una regola generale da un solo caso particolare, l'osservazione in treno che una mucca in Scozia appare bruna, si pone *questo semplice, ma assoluto, problema*: la mia dichiarazione che conseguenze avrà sotto l'aspetto economico? Egli, di conseguenza, farà quella dichiarazione che garantisce il massimo vantaggio perché *il guadagno domina la verità nella specie Homo Sapiens sapiens*. Potrà, poi, giustificare bene la sua scelta perché, come è noto: *data una decisione, esiste sempre una teoria coerente in grado di giustificarla, a partire da opportuni principi*. Lo sforzo, in questo caso, è di tirar fuori i giusti principi che rendono valida qualunque "fesseria". I politici conoscono bene questo enunciato. Insomma un economista non ha il mito dei valori di verità da assegnare alle proposizioni logiche. Ha invece il difetto, chiamiamolo così, di porre nella gerarchia dei criteri decisionali al primo posto il valore economico delle conseguenze delle decisioni prese. È tutta un'altra storia. (*Elisabetta Capotosto*)