

6 Il Teorema del Limite Centrale

6.1 Cenni storici

La storia del teorema del limite centrale è forse una tra le più interessanti in ambito matematico, o almeno in ambito probabilistico. Come noto, il calcolo delle probabilità nasce essenzialmente nel 1654, con il carteggio tra Pascal e Fermat in merito ad alcuni semplici problemi di gioco d'azzardo e calcolo combinatorio. Dal loro lavoro e da quelli successivi, in particolare ad opera dei Bernoulli, si arriva alla formulazione della legge binomiale: ripetendo in modo indipendente n esperimenti in ciascuno dei quali si campioni una Bernoulliana con probabilità di successo pari a p , la probabilità di avere k successi ($0 \leq k \leq n$) è data da $Pr\{X_n = k\} = \binom{n}{k} p^k (1-p)^{n-k}$. Mentre questo risolve completamente il problema dal punto di vista teorico, dal punto di vista pratico la formula ottenuta è del tutto inapplicabile quando il valore di n diventa nell'ordine di 100 o superiore, anche utilizzando gli attuali supercomputer nel 1700 il calcolo avrebbe dovuto essere effettuato a mano e sarebbe stato verosimilmente infattibile anche per valori di n molto bassi.

Il primo a cercare di ottenere una approssimazione numerica per queste probabilità fu Abraham De Moivre, che tra il 1733 ed il 1738 si rese conto che i valori della legge binomiale si potevano approssimare con la funzione

$$Pr\{X_n = k\} \approx const \times \exp\left(-\frac{1}{2(np(1-p))}(k - np)^2\right);$$

il calcolo delle costante di normalizzazione fu ottenuto circa 50 anni da Laplace, che fu quindi il primo a scrivere la densità nella funzione Gaussiana nella forma che oggi conosciamo,

$$\phi_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad \mu = np, \sigma^2 = np(1-p).$$

L'accuratezza di questa approssimazione può essere verificata con alcuni semplici giochi da tavoli (la cosiddetta Galton Board).

Spiegazione Semplice

Cosa hanno scoperto De Moivre e Laplace?

- **Il Problema:** La formula **Binomiale** (es. la probabilità di ottenere esattamente 53 teste su 100 lanci) è un incubo da calcolare a mano. Richiede fattoriali enormi (es. 100!).
- **La Soluzione (De Moivre):** De Moivre si accorse che se si disegnava l'istogramma della distribuzione Binomiale (es. per $n = 100$), questo assumeva una forma a "campana" molto regolare.
- **La Formula (Laplace):** Laplace completò il lavoro trovando la formula esatta di questa curva a campana: la **distribuzione Gaussiana** (o "Normale").

In pratica, hanno scoperto che per n grande, invece di usare la complicata formula Binomiale, si può usare la molto più semplice curva Gaussiana per *approssimare* la probabilità. La media (μ) della campana è np e la sua larghezza (σ^2) è $np(1-p)$. Questo è il primo esempio storico del Teorema del Limite Centrale.

A questo punto la densità Gaussiana sembrava poco più di un utile strumento di approssimazione numerica; iniziò a occuparsene però Gauss, nell'ambito del suo lavoro come astronomo reale all'Osservatorio di Berlino intorno al 1810. In particolare, Gauss si trovò di fronte al problema di dover confrontare le traiettorie dei pianeti previste dalla meccanica Newtoniana con le osservazioni da lui raccolte; esistevano naturalmente alcune discrepanze dovute principalmente ad errori di misura, e Gauss pensò di diminuirne l'effetto prendendo semplicemente la media aritmetica di un maggior numero di osservazioni successive. Si pose quindi una domanda quale legge dovrebbero seguire gli errori di misurazione affinché prendere la media aritmetica sia la scelta ottimale? Per rispondere a questa domanda, bisogna prima specificare cosa si intenda per "ottimale" Gauss scelse un criterio che fu ufficialmente codificato più di 100 anni dopo di lui da Ronald Fisher, la cosiddetta Massima Verosimiglianza, tuttora alla base di quasi tutta la teoria degli stimatori. L'idea della massima verosimiglianza è semplice: si prende come valore "stimato" del parametro quello che rende massimilmente probabile osserva quello che ho effettivamente osservato. Ad

esempio, supponiamo di avere di fronte una scatola che contiene palline bianche e nere; non sappiamo in quale proporzione però sappiamo che se la scatola è di tipo "1" ci saranno 90 palline bianche ed 10 nere, se la scatola è di tipo "2" ci saranno 10 bianche e 90 nere. Effettuiamo una estrazione, che produce una pallina bianca come stima del parametro della scatola ovviamente prendiamo il valore 1, perché rende molto più probabile osservare quello che abbiamo effettivamente osservato.

Spiegazione Semplice

Cos'è la "Massima Verosimiglianza" (Maximum Likelihood)?

È un'idea fondamentale in statistica per "indovinare" un parametro ignoto.

L'idea: Se osservo un certo dato, qual è il modello (quale parametro) che rende quel dato il *più probabile* possibile?

Esempio del testo:

- **Dato osservato:** Estraggo una pallina **bianca**.
- **Ipotesi 1 (Scatola "1"):** La probabilità di estrarre una bianca è $90/100 = 90\%$.
- **Ipotesi 2 (Scatola "2"):** La probabilità di estrarre una bianca è $10/100 = 10\%$.

Conclusione (di Massima Verosimiglianza): Scelgo l'ipotesi 1 ("è la scatola 1") perché rende il dato che ho osservato (la pallina bianca) massimamente probabile (90% è molto più di 10%). Sto "massimizzando la verosimiglianza" del dato.

Proseguendo con questa ragionamento, Gauss suppone che gli errori seguano una certa legge ignota (oggi diremmo densità di probabilità) che possiamo chiamare $f(x)$ avendo effettuato n misurazioni indipendenti, conclude che quella che oggi chiameremmo la loro legge congiunta deve avere forma

$$f(X_1)f(X_2)\dots f(X_n).$$

E' immediato verificare che la media aritmetica minimizza in c la funzione $-\sum_{i=1}^n (X_i - c)^2$; quindi la media aritmetica costituisce uno stimatore ottimale quando si ha

$$f(X_1)f(X_2)\dots f(X_n) = \text{const} \times \exp\{-\text{const} \times \sum_{i=1}^n (X_i - c)^2\},$$

cioè esattamente la densità congiunta di n Gaussiane indipendenti, a meno di costanti. Non sappiamo se Gauss abbia anche verificato empiricamente il fatto che gli errori seguissero questa legge, nel qual caso sarà rimasto senz'altro sbalordito: come se la Natura stesse cercando segretamente di aiutarlo, gli errori di osservazione seguono proprio quella legge che rende ottimale prendere come stimatore la media aritmetica, nonostante avesse scelto questa procedura senza sapere a priori che fosse ottimale. In ogni caso, il fatto che la stessa legge introdotta da De Moivre esclusivamente per motivazioni numeriche avesse anche questa proprietà importantissima di ottimalità per la media numerica, oltre ad essere quella effettivamente seguita empiricamente dagli errori di osservazione, deve aver avuto sicuramente un effetto notevole su Gauss ed i suoi successori (Galton diceva che se l'avessero scoperto i Greci avrebbero aggiunto la legge Gaussiana all'Olimpo dei loro dei).

Spiegazione Semplice

La scoperta "magica" di Gauss

Gauss si è chiesto: "Io uso la media aritmetica per stimare il valore vero. Quale forma deve avere la distribuzione degli errori $f(x)$ affinché la media aritmetica sia la stima 'migliore' (secondo la Massima Verosimiglianza)?"

1. Ha impostato il problema matematicamente.
2. Ha scoperto che l'unica distribuzione $f(x)$ che rende la media aritmetica la stima "ottimale" è proprio la curva a campana, la **distribuzione Gaussiana**.
3. La cosa sbalorditiva è che, in pratica, gli errori di misurazione *seguono* davvero quella distribuzione!

Quindi, la Gaussiana non è solo un trucco matematico per approssimare la Binomiale (come pensava De Moivre), ma sembra essere una legge fondamentale della natura per gli errori e, come vedremo, per quasi tutto ciò che è una somma di piccoli effetti casuali.

Il passo successivo si ha con Maxwell nel 1860, quando si scopre che la Gaussiana determina la legge delle velocità delle molecole in gas perfetti. Da quel momento la Gaussiana comincia ad apparire in una infinità di campi diversi, dalla biologia alla termodinamica, dalla finanza alla medicina. Ad esempio, Einstein nel 1905 lega l'equazione termodinamica della diffusione del calore ad un modello probabilistico di diffusione di particelle; nello stesso anno Bachelier crea di fatto la finanza matematica modellando con una Gaussiana le fluttuazioni dei rendimenti dei titoli di borsa. Le applicazioni della Gaussiana nel corso del ventesimo e ventunesimo secolo sono troppe per poter essere elencate: sul sito arxiv.org risultano circa 450 articoli nel 2022 che hanno il termine "Central Limit theorem" nell'abstract e più di 6500 che hanno il termine "Gaussian". Queste applicazioni sono centrali a tutte le aree della matematica; qui sotto alcuni esempi un po' inaspettati.

Example 61 (Teoria dei Numeri). *E' ben noto che ogni numero intero $n \in \mathbb{N}$ si fattorizza in modo univoco nella somma dei suoi fattori primi, $n = p_1 \times p_2 \times \dots \times p_k$. Quanti sono questi fattori per un numero scelto "a caso", nel limite in cui $n \rightarrow \infty$? si può dimostrare che vale un Teorema del Limite Centrale, ed in particolare scrivendo $\omega(J_n)$ per il numero di fattori primi di J_n , con J_n uniformemente distribuita tra 1 e n , si ha*

$$\frac{\omega(J_n) - \log \log n}{\sqrt{\log \log n}} \xrightarrow{d} N(0, 1),$$

si veda ad esempio P. Erdos and M. Kac. *The Gaussian law of errors in the theory of additive number theoretic functions.*, Amer. J. Math., 62:738-742, 1940, Chen, Louis H. Y.; Jaramillo, Arturo Yang, Xiaochuan A generalized Kubilius-Barban-Vinogradov bound for prime multiplicities. ALEA Lat. Am. J. Probab. Math. Stat. 20 (2023), no. 1, 713-730.

Example 62 (Geometria in alta dimensione). *Scegiamo un punto a caso (x_1, \dots, x_d) su una sfera unitaria di dimensione d , S^d qui per "a caso" intendiamo in maniera uniforme, cioè tutte le regioni con la stessa area hanno la stessa probabilità di essere estratte. Fissiamo un intero k , e consideriamo le primi k coordinate di questo punto; allora per una opportuna sequenza di normalizzazione $\sigma(d)$ si ha*

$$\frac{1}{\sigma(d)}(x_1, \dots, x_k) \xrightarrow{d} (\bar{Z}_1, \dots, Z_k),$$

dove le Z_i sono variabili aleatorie Gaussiane standard indipendenti tra loro. In particolare, qualsiasi coordinate tende (dopo una rinormalizzazione per l'opportuno fattore di scala) ad una Gaussiana standard. (Freedman e Diaconis, Poincaré...)

Example 63 (Polinomi trigonometrici). *Fissiamo un insieme di pesi a_k, b_k $k = 1, 2, \dots, n$; per fissare le idee potremmo anche prendere tutti questi pesi uguali a 1. Consideriamo ora il polinomio trigonometrico (deterministico!)*

$$p_n(x) = \sum_{k=1}^n a_k \cos(kx) + b_k \sin(kx), \quad x \in [-\pi, \pi].$$

Dopo una opportuna normalizzazione $\sigma(n)$, per quasi tutte (nel senso di con probabilità 1, se si vedono questi pesi come variabili aleatorie) le scelte dei pesi (a_k, b_k) si ha

$$\frac{1}{2\pi} \text{Misura}\{x : c_1 \leq \frac{p_n(x)}{\sigma(n)} \leq c_2\} \xrightarrow{n \rightarrow \infty} \int_{c_1}^{c_2} \frac{\exp(-x^2/2)}{\sqrt{2\pi}} dx = \Pr\{Z \in [c_1, c_2]\},$$

cioè in altre parole i valori (deterministici) presi dal polinomio trigonometrico si distribuiscono come una Gaussiana standard: facendo un plot di questi valori si otterrebbe la classica curva a campana. (Per alcune referenze relative a questo risultato, si veda ad esempio il classico articolo R. Salem and A. Zygmund. "Some properties of trigonometric series whose terms have random signs". In: Acta Math. 91 (1954), pp. 245-301, oppure i più recenti Jürgen Angst and Guillaume Poly "Variations on Salem Zygmund results for random trigonometric polynomials: application to almost sure nodal asymptotics". In: Electronic Journal of Probability 26 (2021), pp. 1-36, e Louis Gass, Almost-sure asymptotics for Riemannian random waves. In Bernoulli 29, (2023), no. 1, 625-651.) Un risultato analogo, ma forse più

semplice da enunciare, è il seguente: sia n_k una sequenza di numeri interi tale per cui $1 < n_{k+1} - n_k = O(1)$; definiamo il polinomio trigonometrico deterministico

$$q_n(x) = \frac{\sqrt{2}}{\sqrt{n}} \sum_{k=1}^n \cos(n_k x).$$

Allora quando $n \rightarrow \infty$ si ha

$$\frac{1}{2\pi} \text{Misura}\{x : c_1 \leq q_n(x) \leq c_2\} \xrightarrow{n \rightarrow \infty} \int_{c_1}^{c_2} \frac{\exp(-x^2/2)}{\sqrt{2\pi}} dx = \Pr\{Z \in [c_1, c_2]\}$$

Quindi i valori deterministici del polinomio $q_n(x)$ si distribuiscono seguendo una curva Gaussiana standard. Per questo risultato si veda ad esempio Katusi Fukuyama, A central limit theorem to trigonometric series with bounded gaps, *Probab. Theory Related Fields*, 149, no. 1-2, 139-148, (2011).

Spiegazione Semplice

Il punto chiave di questi esempi

Non è necessario capire la matematica complessa di questi tre esempi (Teoria dei Numeri, Geometria, Polinomi).

Il punto è mostrare l'incredibile **universalità** del Teorema del Limite Centrale. La stessa curva a campana (Gaussiana) spunta fuori:

- nel numero di fattori primi di un intero;
- nelle coordinate di un punto casuale su una sfera in alta dimensione;
- nei valori assunti da una funzione trigonometrica (che è puramente deterministica!).

Il motivo è sempre lo stesso: tutti questi fenomeni, in fondo, sono il risultato della *somma* di tanti piccoli pezzi indipendenti (o quasi), e il Teorema del Limite Centrale dice che la somma di tanti pezzi casuali, indipendentemente da come sono fatti, finisce per assomigliare a una Gaussiana.