LO STRANO CASO DEL NUMERO 3.5   (I)

UN ALTRO SASSOLINO NELLA SCARPA

(per matematici che imparano a camminare)

(Ricordo che intendo per sassolino nella scarpa un risultato matematico evidentemente semplice da ottenersi, sovente dato per scontato in un testo di matematica, che però a me non è mai stato chiaro del tutto, anche perché impararlo a memoria può esser costato meno fatica che comprenderlo).

Qui trattiamo **lo strano caso del numero 3.5, che ci porterà a parlare del "secondo più importante teorema del calcolo delle probabilità", il teorema centrale del limite (TCL).**

Supponiamo che vi si proponga di scommettere sul lancio, non di un dado, ma di due dadi, e di scommettere sulla somma dei valori delle due facce superiori. In molti giochi basati su quello che era un tempo chiamato "l'utile e dilettevole gioco dell'oca", da Monopoli in avanti, si lanciano appunto due dadi, ma, se ci si fa attenzione, non si scommette di norma sul risultato. Perché? Perché chiunque abbia fatto l'esperimento sa che se si lancia un solo dado, i valori da 1 a 6 possono uscire con uguale probabilità, ma se si lanciano due dadi c'è un valore più probabile per la somma dei due valori, che è ***sette***. Controintuitivamente, quanto maggiore è il numero n di dadi lanciati insieme, tanto più facile è prevedere il risultato della somma dei valori dati sulle n facce.

Se lanciamo n dadi, di quelli soliti di sei facce, ci possiamo chiedere
1) quale sia il risultato più probabile, M, per la somma S dei valori dati sulle n facce superiori;
2) quale il valor più probabile medio, m, risultante per ogni dado;
3) quale sia la probabilità che su N lanci otteniamo il valore più probabile.

Se è noto il primo risultato, naturalmente, il secondo segue subito dividendo per n il valore più probabile della somma: m= M/n. Ed è qui che abbiamo il risultato curioso: *il valor medio risultante per ogni dado è 3.5, praticamente indipendente da* ***n.*** Ricordando a memoria questo numero magico, abbiamo, se giochiamo ai dadi, la seguente tabella:

numero dadi, n	Valore medio, m, *più probabile per dado*	Valore più probabile, M, della somma dei valori di ogni dado
1		(equiprobabile da 1 a 6)
2	3.5	7
3	3.5	10.5 (cioè tra 10 e 11)
4	3.5	14
5	3.5	17.5 (cioè tra 17 e 18)

Questo numero magico è utile a sapersi, perché non molti sanno calcolarsi il valore più probabile della somma dei valori risultanti dal lancio di più di due dadi (colonna 3). Quelli che hanno qualche intuizione matematica direbbero che il valore più probabile della somma è circa metà del massimo valore possibile, che è dato dal numero di dadi moltiplicato 6. Nel caso di tre dadi, il valore sarebbe quindi 9. Come vediamo dalla tabella, ciò è quasi vero.

I nostri risultati sono il frutto di un'applicazione del cosiddetto **Teorema Centrale del Limite (TCL)**, forse il secondo più importante teorema della teoria della probabilità, **dopo la legge dei grandi numeri** ([1]) . Qui abbiamo il solito dubbio: non è più semplice ricordar a memoria il numero 3.5 , che non comprendere l'enunciato del teorema centrale del limite (non parliamo neppure della dimostrazione, che darò nella seconda parte per "matematici ciclisti")? Il fatto è che il TCL serve in una quantità di applicazioni, mentre il nostro numero 3.5 è utile soltanto a calcolare medie e probabilità riguardanti dadi di 6 facce.

Che dice in questo caso il teorema centrale del limite?

Nel caso dei dadi ci dice: supponiamo di lanciare n dadi di 6 facce. Ogni dado avrà eguale probabilità (1/6) di produrre un risultato da 1 a 6 sulla faccia superiore. L'istogramma delle probabilità sarà quindi dato dal primo istogramma in figura 1. Uniforme.
I lanci sono indipendenti, nel senso che il risultato del dado 1 non influenza i risultati degli altri dadi. Le distribuzioni di probabilità, uniformi, sono le stesse per tutti i dadi. La somma dei valori delle facce superiori dei dati, S, può variare tra *n* (il numero dei dadi, che è anche il valore minimo di S, perché ogni dado contribuisce almeno 1 alla somma S) e *6n* (il valore massimo). Ci sarà un solo modo di ottenere *n* e un solo modo di ottenere *6n*. Per calcolare la probabilità di ottenere un valore intermedio dovremo calcolare ***in quanti modi*** può essere ottenuto ogni valore tra *n* e *6n*. **Se disegneremo l'istogramma risultante, vedremo che al crescere di *n* esso assomiglierà sempre più ad una gaussiana, cioè alla famosa curva di Gauss, o "curva a campana", simmetrica e centrata su una certa media M, e con una certa "grassezza**" (Fig.1). La somiglianza sarà ancora migliore se divideremo ogni ordinata dell'istogramma per P, il numero totale di modi per ottenere tutti i valori tra n e 6n. Avremo così la probabilità p(S) di ottenere un dato valore della somma S. Come si è detto, M sarà il valore più probabile della somma S dei valori delle facce superiori dei dadi e p(M) sarà la probabilità di ottenerlo.

Nel caso di due dadi il numero di modi per ottenere un valore S viene calcolato assegnando in successione ad un dado i valori d che esso può assumere, e all'altro la differenza S-d, tenendo naturalmente presente che per ottenere, ad esempio, 8, la somma 1+7 non è valida, in quanto i valori delle facce di un dado vanno soltanto da 1 a 6. Quindi, otterremo 8 soltanto con 2+6, 3+5, 4+4, 5+3, 6+2, cinque modi in tutto.

Quindi, per il lancio di due dadi si può far vedere facilmente , calcolando tutti i modi in cui possono essere dati i valori della somma da 2 a 12, che m vale effettivamente 3.5. come dato nella nostra tabella: 3.5 =7/2, dove 7 è il valore più probabile, e 2 è il numero di dadi. I valori possibili della somma sono 12

Valore della somma, S	2	3	4	5	6	7	8	9	10	11	12
S/n	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
Modi di ottenere il valore S	1	2	3	4	5	6	5	4	3	2	1

Valore più probabile, M, di S : 7
Numero di modi per ottenerlo : 6
Numero totale di modi per ottenere tutti i valori possibili di S: 36
Probabilità p(M), normalizzata, di ottenere il valore massimo 7: 6/36 = 1/6
Valore più probabile (7) diviso il numero di dadi (2), cioè valore medio per dado: m = 3.5
Fig.1 riporta (in giallo) le frequenze osservate nel caso di 10000 lanci.

Per ottenere il massimo della nostra distribuzione, ovvero il numero M più probabile, tutto sta nel piazzare il valore centrale della quasi-Gaussiana, che corrisponde al massimo della medesima. Qui il TCL viene usato in modo surrettizio, sfruttando il fatto che la gaussiana è una curva simmetrica. Il nostro istogramma, però, non parte da 0, ma da *n*, perché ogni dado contribuisce almeno un valore 1 alla somma. Quindi, la simmetria della gaussiana ci dice che la somma più probabile tenderà ad essere piazzata a metà fra il valore minimo, n (numero dei dadi per il valore di faccia minimo), ed il valore massimo, 6n (numero di dadi per il valore di faccia massimo). Quindi il valore che cerchiamo sarà

$$M = n + \frac{6n - n}{2} = \frac{7n}{2} = 3.5\, n$$

Tutto qui. Dividendo per *n*, si trova il valor medio *m* più probabile *per dado*, che è 3.5.

Lanciando 1000 dadi, quindi, la media sarà 3500.

Abbiamo calcolato M=3.5 con successo per n =2, anche se le regole non scritte dei probabilisti ci dicono che per andare sul sicuro n dovrebbe essere vicino a 30. Il punto è che i 6 valori della faccia superiore dei 6 dadi sono equiprobabili, e quindi la distribuzione è regolare. Vediamo però qui una delle bellezze del teorema centrale del limite: indipendentemente dalle distribuzioni dei valori dei dadi nelle singole somme, otteniamo sempre una distribuzione quasi normale quando calcoliamo la distribuzione del valori delle somme.

Resta da calcolare la forma della gaussiana. Abbiamo dunque la media, che è sempre data dalla somma dei valori delle facce (qui da 1 ad 6) diviso 2. Cioè 3.5.

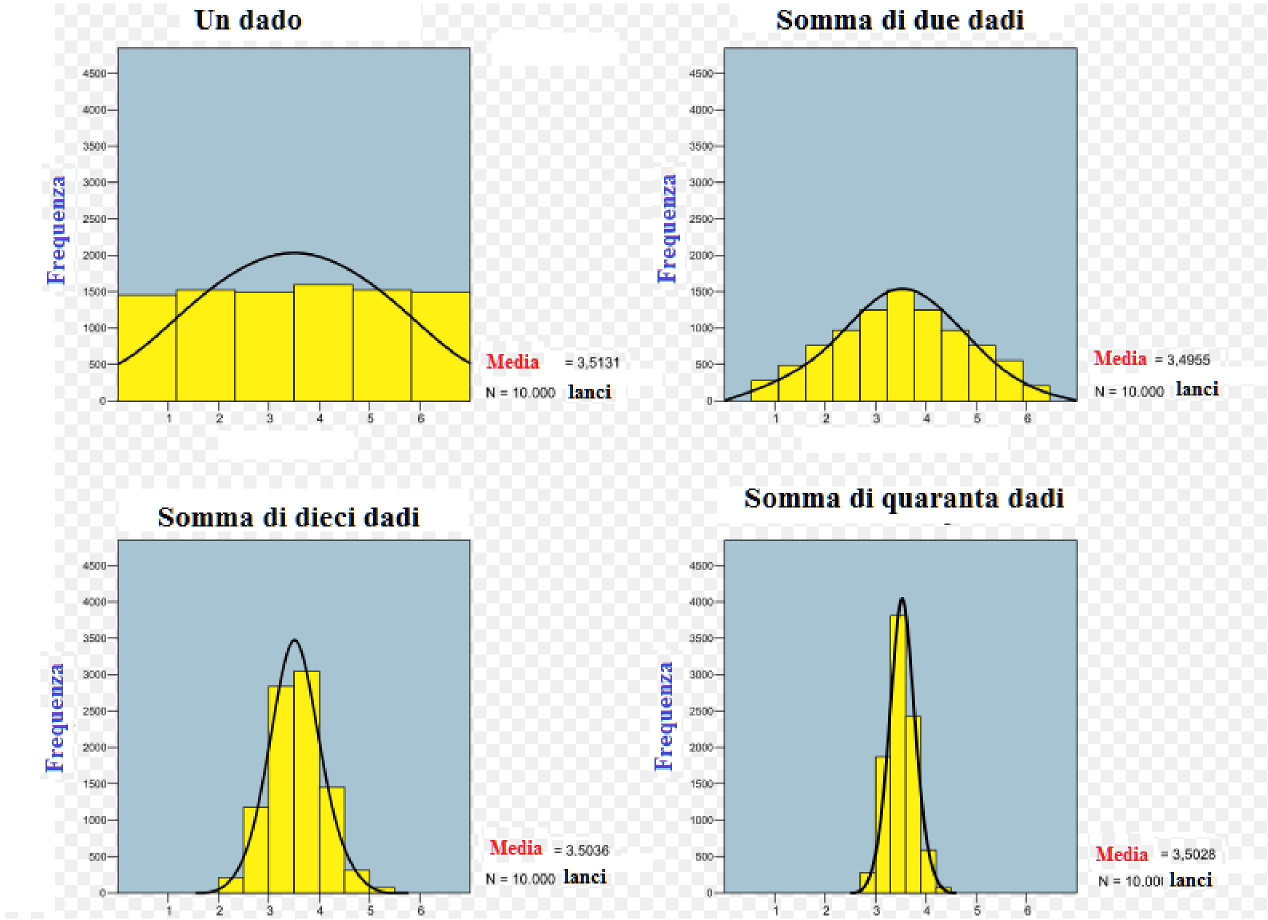

Fig.1 (da Wikipedia)

https://upload.wikimedia.org/wikipedia/commons/7/7a/Teorema_central_do_limite.svg
By Fernando Lang da Silveira (Own work) [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons

Per calcolare la forma esatta della Gaussiana ci occorrerà conoscere la “deviazione standard”, cioè la “grassezza della curva”, generalmente misurata prendendo la larghezza della curva a mezza alteszza del massimo. Per fare questo, però, occorre derivare il teorema centrale del limite, cosa che farò alla prossima puntata, derivazione riservata ai matematici ciclisti (II-III corso universitario di analisi). Ad ogni modo, dalla figura si vede che la gaussiana non solo viene sempre meglio approssimata, ma diventa sempre più affilata al crescere del numero di dadi.

Però dalla figura si vede bene che partendo da una distribuzione di probabilità piatta o uniforme per i valori da 1 a 6 dei risultati di un singolo dado, la distribuzione dei valori della somma dei risultati si avvicinerà sempre più ad una gaussiana al crescere di n.

In realtà il teorema centrale del limite ci assicura che ***qualunque sia la distribuzione di probabilità iniziale di un grande numero di variabili aleatorie, purché risponda a certi criteri relativamente poco restrittivi, la distribuzione delle medie (che non sono altro che le somme dei valori delle***

variabili diviso il numero n di variabili) si avvicina sempre di più ad una gaussiana al crescere di n.

Questo spiegherebbe perché la curva a campana è così importante in statistica: in molti fenomeni in cui si desidera conoscere la distribuzione di probabilità di una certa variabile, questa è a nostra insaputa o in modo occulto, la media di altri valori che possono avere le più strane distribuzioni di probabilità. Il TCL ci assicura che nondimeno ne risulterà una curva a campana.

Un caso caratteristico è facile da spiegare. Supponiamo di avere una classe di allievi a cui è assegnato una verifica o compito d'esame in diverse domande, per esempio dieci. Vengono dati tre punteggi per domanda, 2 se la risposta è corretta, 1 se la risposta ha qualche errore, ma non troppi, 0 se la risposta è sbagliata. Poi si fa la somma. Se l'insegnante ha fatto bene il suo lavoro di insegnante, in generale le frequenze dei punteggi dei vari studenti si distribuiscono in una curva a campana, che assomiglierà abbastanza ad una gaussiana. La curva a campana, con un poco di ragionamento, permetterà all'insegnante di capire se l'esame era troppo facile o troppo difficile. Lui/lei potrà comunque dare dei voti basandosi non su una scala assoluta (10 a chi fa tutto giusto), ma in base alla media apparente dal diagramma.
Ad ogni modo, in questo caso, il motivo per cui si ha una gaussiana è evidente: si sono fatte somme di punti per ogni allievo, e il TCL ci assicura che queste somme tendono a disporsi in una curva a campana.
Anche le spese fatte dai clienti di un supermercato in un dato giorno tendono a disporsi su una simile curva per lo stesso motivo, che la spesa totale fatta da un cliente è la somma di diverse spese parziali.
In altri casi, non è facile vedere dove si nasconda questa somma. Anche l'altezza o il peso degli italiani di una data età e sesso, per esempio, si distribuiscono secondo una curva a campana. Ma anche se non c'è evidenza di un'applicazione del TCL, le distribuzioni a campana sono talmente frequenti che questa curva è detta anche "curva normale". Oggi bisogna andarci piano con l'uso di questo nome, che sembra indicare che gli individui più alti di un certo valore ai limiti superiori della curva a campana siano "anormali". Comunque, la normalità in matematica è un nome come un altro e può essere definita in modo chiaro, non offensivo.

Vorrei ora dare una **traccia della dimostrazione del TCL** (la più semplice che ho trovato, che riprodurrò per esteso in seguito). Un matematico "ciclista" dovrebbe poter ricostruire il teorema secondo queste linee.

1) Una distribuzione di probabilità p(x) della variabile x possiede una "funzione caratteristica" f(k), che è la sua "trasformata di Fourier" e dà molte informazioni utili sulla p(x). Questa "funzione caratteristica" può essere sviluppata in serie ed i vari termini danno i cosiddetti "momenti" della distribuzione (termine derivato per analogia colla meccanica classica), i quali descrivono vari aspetti della p(x).

3) Noi vogliamo calcolare la funzione caratteristica F(k) della distribuzione di probabilità P(X) **della media X di n valori di variabili indipendenti x**. E' quasi immediato provare che essa è data dal prodotto delle n funzioni caratteristiche f(k) delle distribuzioni di probabilità delle variabili p(x). Nel caso del lancio di n dadi si tratta di n funzioni identiche . Essa è dunque eguale alla **potenza ennesima** dello sviluppo in serie della funzione caratteristica f(k) della distribuzione di probabilità

di una delle variabili. Questa potenza, avendo l'accortezza di portare il massimo della gaussiana sull'origine (cioè ponendo M = 0), ciò che si può sempre fare, può essere scritta come:

$$\left(1-\frac{Ak^2}{n^2}+\ altri\ \ termini\right)^n$$

Qui **gli "altri termini" presentano al denominatore potenze di n di grado maggiore di 2, e quindi tendono a zero più rapidamente del secondo termine, per n tendente ad infinito.**

Ignorando gli "altri termini" e utilizzando il noto limite per n tendente ad infinito di

$$(1+\frac{1}{m})^m$$

risulta quindi che la funzione caratteristica cercata F(k) è del tipo:

$$\exp\left(-\frac{Ak^2}{n}\right)$$

4) La **funzione caratteristica** F(k) della distribuzione di probabilità cercata P(X) avrà dunque la forma di una gaussiana. Per ottenere la distribuzione di probabilità P(X) occorrerà applicare l'inversione della F(k) la quale, come sappiamo, la trasformata di Fourier di P(X). Ma l'inversione della trasformata di Fourier è un'altra trasformata di Fourier, che risulterà a sua volta essere una gaussiana, in quanto la trasformata di Fourier di una gaussiana è ancora una gaussiana opportuna.

Il che è ciò che volevamo dimostrare.

Più chiaro di così'...

Arrivederci.

NOTA:

([1]) Osservo che in inglese il nome è ambiguo e potrebbe essere "teorema del limite centrale" o "teorema centrale del limite". Lo stesso dubbio esiste in francese. In tedesco ("zentraler Grenzwertsatz") e altre lingue, incluso l'italiano, invece è certamente "teorema centrale del limite". Penso che questa sia la forma corretta, introdotta intorno al 1920 dal matematico Polya, che scriveva in tedesco e si riferiva all'importanza o centralità del teorema nella teoria dalla probabilità..