Giornale di Tecniche Nefrologiche & Dialitiche 2014; 26 (3): 290-293
DOI: 10.5301/GTND.2014.12676

La distribuzione normale - Parte III: una semplice introduzione al Teorema del Limite Centrale

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

a nome del Comitato Scientifico SIN-RIDT

The normal distribution - Part III: a simple introduction to Central Limit Theorem

Abstract. The Central Limit Theorem is presented showing how a uniform distribution, generated with a dice, tends to converge to a normal distribution when the number of thrown dices becomes large enough.

Key words: Uniform distribution, Dices, Normal distribution, Central limit theorem

Conflict of interest: None.

Financial support: None.

Accettato: 8 Settembre 2014

Lancio di un dado: la distribuzione uniforme

Se lanciamo un dado da gioco non truccato, sappiamo già in anticipo che i risultati possibili (che da adesso in poi chiameremo anche *eventi* o *realizzazioni*) possono variare tra 1 e 6, e che tali risultati hanno la stessa probabilità di verificarsi. La probabilità identica di ogni evento è infatti pari a 1/6, cioè al 16.67% circa, quindi lanciando il dado sappiamo che è identicamente probabile realizzare uno qualsiasi degli eventi da 1 a 6, e useremo la notazione $\{n\}$ per individuare la realizzazione del generico evento n.
Se produciamo su un istogramma la probabilità della realizzazione di ciascun evento possibile, otteniamo un grafico dal profilo piatto e uniforme e, appunto, uniforme è detta una distribuzione di probabilità come quella che descrive gli eventi realizzati dal lancio di un dado e in tutti i casi in cui i vari eventi attesi sono equiprobabili. Un secondo esempio è il gioco della *roulette*, in cui sono tutte equiprobabili le differenti 37 realizzazioni che si possono ottenere, cioè i numeri che vanno da 0 a 36.
Osservando l'istogramma notiamo immediatamente che il valore mediano delle possibili realizzazioni è 3.5, che è anche il valore medio, essendo la somma delle possibili realizzazioni, divisa per il possibile numero delle stesse realizzazioni, pari a $(1 + 2 + 3 + 4 + 5 + 6) / =3.5$. Il punteggio che si ottiene con il lancio di un dado è ovviamente pari alla realizzazione.

Lancio di due dadi

Le cose diventano molto più interessanti quando si lanciano due dadi simultaneamente. Ogni realizzazione sarà un coppia di numeri compresi tra 1 e 6, e la somma della coppia sarà equivalente al punteggio ottenuto con ogni lancio. Le realizzazioni possibili diventano allora $6 \times 6 = 36$ e possono sommare da 2 a 12, dato che il punteggio minimo è 2, ottenuto con l'evento $\{1,1\}$, e quello massimo è 12, che si ottiene con l'evento $\{6,6\}$.
Per essere più chiari, elenchiamo i punteggi possibili in funzione degli eventi che li generano:

Punteggio	*Eventi*	*Numero di eventi*
2	{1,1}	1
3	{1,2}, {2,1}	2
4	{1,3}, {2,2}, {3,1}	3
5	{1,4}, {2,3}, {3,2}, {4,1}	4
6	{1,5}, {2,4}, {3,3}, {4,2}, {5,1}	5
7	{1,6}, {2,5}, {3,4}, {4,3}, {5,2}, {6,1}	6
8	{2,6}, {3,5}, {4,4}, {5,3}, {6,2}	5
9	{3,6}, {4,5}, {5,4}, {6,3}	4
10	{4,6}, {5,5}, {6,4}	3
11	{5,6}, {6,5}	2
12	{6,6}	1

punteggio realizzato con il secondo dado						
6	7	8	9	10	11	12
5	6	7	8	9	10	11
4	5	6	7	8	9	10
3	4	5	6	7	8	9
2	3	4	5	6	7	8
1	2	3	4	5	6	7
	1	2	3	4	5	6

punteggio realizzato con il primo dado

Fig. 1 - Tavola sinottica con i punteggi generati dal lancio dei due dadi.

La situazione è riassunta in modo grafico dalla Figura 1.
Si osserva che il punteggio mediano e medio coincidono e sono pari a

$$\frac{2+3+4+5+6+7+8+9+10+11+12}{11} = \frac{77}{11} = 7$$

per cui sono possibili 11 punteggi differenti. Osserviamo anche che mentre gli eventi sono ancora tutti equiprobabili, per cui, per esempio, l'evento {1,1} ha la stessa probabilità di verificarsi dell'evento {3,4}, i punteggi (che erano equiprobabili con il lancio di un solo dado) ora non lo sono più: infatti, lanciando due dadi, il punteggio con la massima probabilità è 7, che viene generato da 6 differenti eventi ({1,6}, {2,5}, {3,4}, {4,3}, {5,2} e {6,1}), mentre i punteggi con la minima probabilità sono 2 e 12, ciascuno generato da un unico evento ({1,1} e {6,6}, rispettivamente). Questo è il motivo per cui il profilo delle probabilità, che era uniforme, ovvero "piatto", per il lancio di un dado, mostra ora una netta crescita fino al suo massimo, che si osserva a 7, ovvero al punteggio che mostra il massimo numero di eventi che lo generano (la moda), per poi tornare a calare, con un profilo nettamente triangolare, come osservabile nella Figura 2.

Lancio di tre dadi

Se ora si lanciano simultaneamente tre dadi, la situazione si fa ancora più interessante: gli eventi sono ora $6 \times 6 \times 6 = 216$, e i punteggi possono variare dal valore minimo 3, generato dall'unico evento {1,1,1}, al valore massimo 18, generato dall'unico evento {6,6,6}. In pratica, quindi, 216 eventi generano un numero limitato di punteggi (16, per la precisione), ma, a differenza di quanto osservato con due soli dadi, il numero degli eventi che generano un dato punteggio non cresce (o decresce) di un'unità alla volta (per esempio, con due dadi ci sono 4 eventi che producono il punteggio 5, e 5 eventi che producono il punteggio 6), ma la crescita è molto diversa, e per certi versi notevolmente più complicata. Infatti, a solo titolo di esempio, vediamo che gli eventi che generano i punteggi 3, 4, 5 e 6 sono rispettivamente 1 ({1,1,1}), 3 ({1,1,2}, {1,2,1} e {2,1,1}), 6 ({1,1,3}, {1,2,2}, {1,3,1}, {2,1,2}, {3,1,1} e {2,2,1}) e 10 ({1,1,4}, {1,2,3}, {1,3,2}, {1,4,1}, {2,1,3}, {2,2,2}, {2,3,1}, {3,1,2}, {3,2,1} e {4,1,1}), per cui al crescere unitario del punteggio gli eventi che lo generano crescono in modo non omogeneo. Non è ancora il momento di dire a quale legge obbedisce la variazione del numero di eventi: ne parleremo in un prossimo numero, quando ci saremo appropriati degli strumenti del calcolo combinatorio, tuttavia possiamo tabulare queste informazioni per vedere come la crescita del numero di eventi tenda ad aumentare, poi a rallentare e raggiunto il massimo inizi quindi a diminuire, prima lentamente e poi velocemente.

Punteggio con tre dadi	*Numero di eventi*
3	1
4	3
5	6
6	10
7	15
8	21
9	25
10	27
11	27
12	25
13	21
14	15
15	10
16	6
17	3
18	1

Il risultato di questa situazione è evidente sul profilo della probabilità, che assume una forma a campana abbastanza evidente (Fig. 2).

Lancio di quattro dadi

Lanciando simultaneamente quattro dadi, il profilo a campana della probabilità si fa più evidente. In questo caso, gli eventi

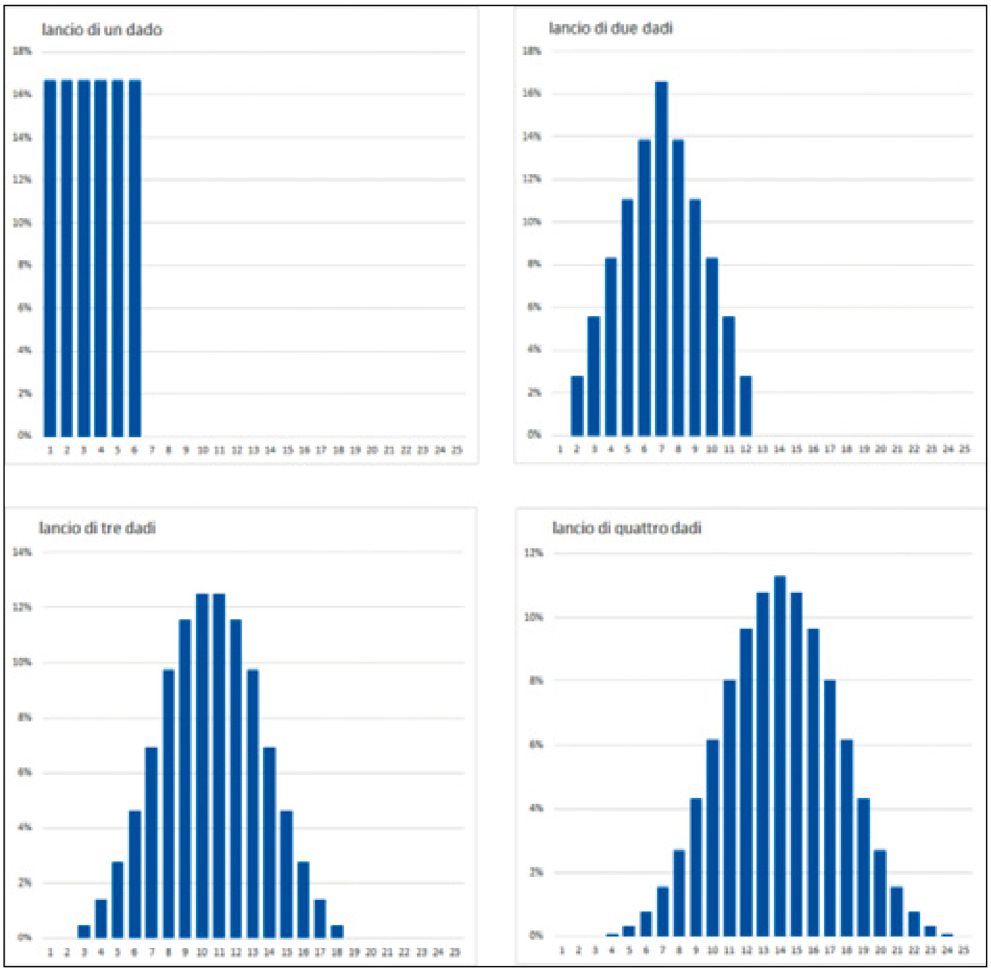

Fig. 2 - Variazione delle curve di probabilità per i punteggi, che si osservano al lancio di 1, 2, 3 e 4 dadi, rispettivamente. Si notano la sempre maggiore somiglianza con la curva normale, la progressiva traslazione della curva verso valori di ascissa crescenti e l'allargamento della sua base.

sono $6 \times 6 \times 6 \times 6 = 6^4 = 1296$, e i punteggi possono variare dal valore minimo 4, generato dall'unico evento {1,1,1,1}, al valore massimo 24, generato dall'unico evento {6,6,6,6}. Con quattro dati, quindi, 1296 eventi possibili generano 21 punteggi differenti, e questo complica di molto le cose rispetto al lancio di tre dadi.

Per comprendere quello che succede, conviene ricorrere alla tabulazione dei punteggi e del numero di eventi che li realizzano, come fatto per i tre dadi: in questo caso, tuttavia, visto che ben 7 punteggi sui 21 possibili sono realizzati da più di cento eventi, è di gran lunga preferibile evitare l'elenco pedissequo di ogni singolo evento. Come semplice esercizio, i lettori volonterosi potrebbero provare a ricercare tutti gli eventi che generano, per esempio i punteggi 8 e 22.

Quello che sembra abbastanza evidente è l'aumento della somiglianza della curva con una curva normale (Fig. 2).

Punteggio con quattro dadi	*Numero di eventi*
4	1
5	4
6	10
7	20
8	35
9	56
10	80
11	104
12	125
13	140
14	146
15	140
16	125
17	104
18	80
19	56
20	35
21	20
22	10
23	4
24	1

Tutte le strade portano a Roma: il teorema del limite centrale

Generalizzando i risultati visti finora possiamo verificare che all'aumentare del numero di dadi lanciati, la distribuzione si fa via via sempre più somigliante a una distribuzione normale. Lanciando n dadi, otterremo sempre un punteggio minimo pari a n, ed un punteggio massimo pari a $6n$, con un numero complessivo di eventi dato da 6^n. Il numero dei differenti punteggi ottenibili è pari a $6n - (n - 1) = 5n + 1$; per esempio, con 4 dadi si ottengono $5 \times 4 + 1 = 21$ punteggi differenti (tra 4 e 24, estremi inclusi), e la curva che descrive la probabilità di ottenere un dato punteggio tende sempre più a somigliare a una curva normale. Questo è un fenomeno del tutto generale, che vale per qualsiasi distribuzione (per adesso ne conosciamo molto poche, ma in futuro ne conosceremo altre, quindi, almeno per il momento, il lettore si fidi).

Questo tendere di ogni distribuzione a diventare sempre più simile a una distribuzione normale, entro un numero più o meno grande di repliche (per ora chiamiamole così), ha un nome: *Teorema del Limite Centrale* (in inglese *Central Limit Theorem*, spesso abbreviato in CLT). Su questo argomento torneremo diffusamente nei prossimi numeri; al momento ci basti sapere che il teorema del limite centrale afferma che ogni distribuzione tenderà ad assumere una forma sempre più simile alla normale, se applicata per un numero sufficiente di volte.

Si tratta di un teorema dimostrabile solo utilizzando un formalismo matematico molto complesso, decisamente improponibile per un medico, e quindi eviteremo di farlo (eventualmente, a richiesta dei più curiosi, potremo postare la dimostrazione sul sito della rivista). Comunque, il lettore che voglia accedere al sito http://www.math.csusb.edu/faculty/stanton/m262/central_limit_theorem/clt.html potrà verificarlo da sé, in modo pratico (nessuna formula) e divertente.

Riassunto

Viene presentato il teorema del limite centrale mostrando come la distribuzione uniforme generata dal lancio di un dado tenda a convergere a una distribuzione normale, al crescere del numero dei dadi lanciati.

Parole chiave: Distribuzione uniforme, Dadi, Distribuzione normale, Teorema del limite centrale

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

Contributi economici agli Autori: Gli Autori dichiarano di non avere 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

