Teorema del limite centrale

Teorema del limite centrale

Il teorema del limite centrale (TLC) afferma che la somma (o la media) di un grande numero di variabili aleatorie indipendenti e dotate della stessa distribuzione è approssimativamente normale, indipendentemente dalla distribuzione soggiacente.

Il TLC costituisce il fondamento di un grande numero di procedure statistiche.

Teorema del limite centrale

Il TLC implica che, se la grandezza di un campione è "grande", allora la distribuzione della somma S_n di n variabili aleatorie indiependenti (o, in maniera equivalente, della media M_n) sarà approssimativamente normale.

Teorema del limite centrale

Questo significa che la distribuzione di alcune statistiche (per esempio, la media del campione) diventa nota, **anche se non sappiamo nulla a proposito della forma della distribuzione della popolazione** da cui i campioni sono stati tratti.

Teorema del limite centrale

Naturalmente, il termine "grande" è relativo.

Tanto più la distribuzione della popolazione è diversa dalla normale, tanto maggiore deve essere la dimensione del campione (n) affinché sia sensato applicare il TLC.

Teorema del limite centrale

La regola euristica è che un campione con $n \geq 30$

sia sufficientemente grande da giustificare l'applicazione del TLC, anche se per molte distribuzioni non normali un campione più piccolo si dimostra sufficiente.

Teorema del limite centrale

Un problema nasce quando la distribuzione della popolazione è discreta. In questo caso, l'applicazione del TLC ci porta ad approssimare una distribuzione discreta con una continua.

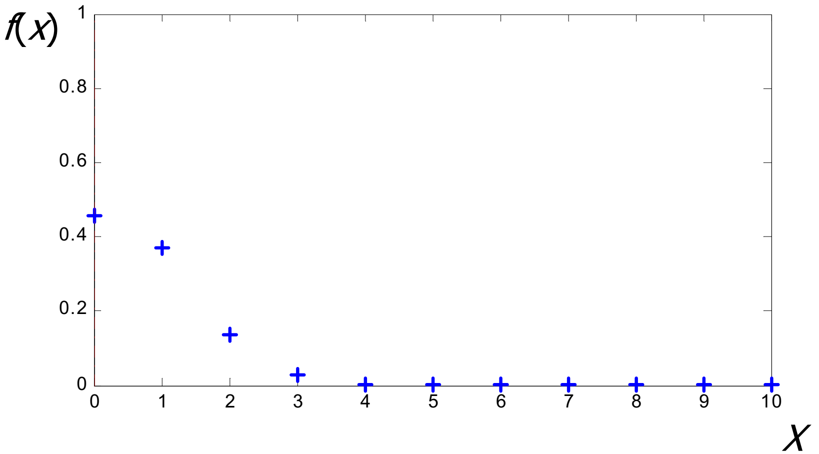
Questo problema si risolve introducendo quella che viene chiamata la "correzione per la continuità".

Teorema del limite centrale: esempio numerico

Supponiamo di estrarre dei campioni da una popolazione che segue la **distribuzione binomiale** con i seguenti parametri ***n* = 10**, ***p* = 0,075**.

La popolazione avrà dunque al seguente distribuzione.

Teorema del limite centrale: esempio numerico



Teorema del limite centrale: esempio numerico

Si noti che la variabile *X* assume i valori 0, 1, 2, …, 10. Questi valori però non sono presenti nella popolazione nella stessa misura.

Dato che p = 0,075, i valori più bassi sono rappresentati nella popolazione in maniera molto superiore ai più valori alti.

Teorema del limite centrale: esempio numerico

In base alle proprietà della distribuzione binomiale, questa popolazione avrà una media uguale a np, ovvero:

$E(X) = np = 10 \times 0{,}075 = 0{,}75$.

Teorema del limite centrale: esempio numerico

Supponiamo di estrarre da questa popolazione dei **campioni casuali di grandezza *n* = 5** e di calcolare la <u>media</u> di ciascun campione.

Teorema del limite centrale: esempio numerico

Un particolare campione sarà, per esempio:

$C_1 = \{0, 3, 0, 1, 0\}$

Ciascuno di questi numeri rappresenta un valore estratto a caso dalla distribuzione binomiale con $n = 10$ e $p = 0{,}075$.

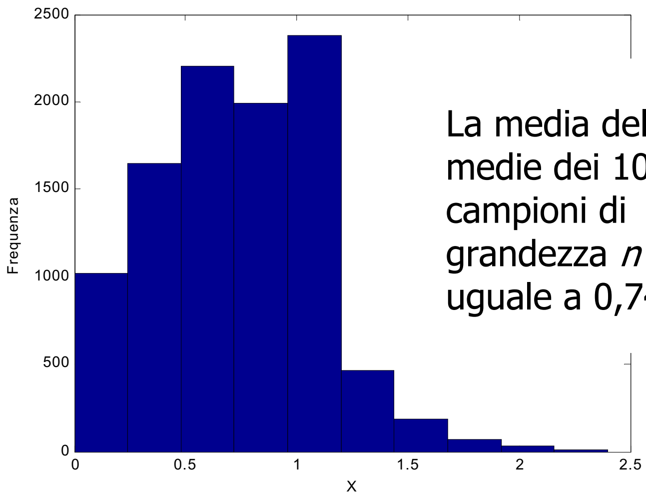
Teorema del limite centrale: esempio numerico

$C_1 = \{0, 3, 0, 1, 0\}$

Per ciascun campione calcoliamo la media. Per il campione C_1 avremo $\overline{X}_1 = 0{,}8$.

Ripetiamo questo processo 10.000 volte e costruiamo un istogramma che rappresenta la distribuzione delle medie dei 10.000 campioni.

Teorema del limite centrale: esempio numerico



La media delle medie dei 10.000 campioni di grandezza $n = 5$ è uguale a 0,7466.

Teorema del limite centrale: esempio numerico

Si noti che la media delle medie dei campioni (0.7466) è molto simile alla media della popolazione (0,75) da cui i campioni sono tratti.

Se il numero dei campioni estratti dalla popolazione fosse più grande (nel caso presente è 10.000), l'approssimazione sarebbe migliore.

Teorema del limite centrale: esempio numerico

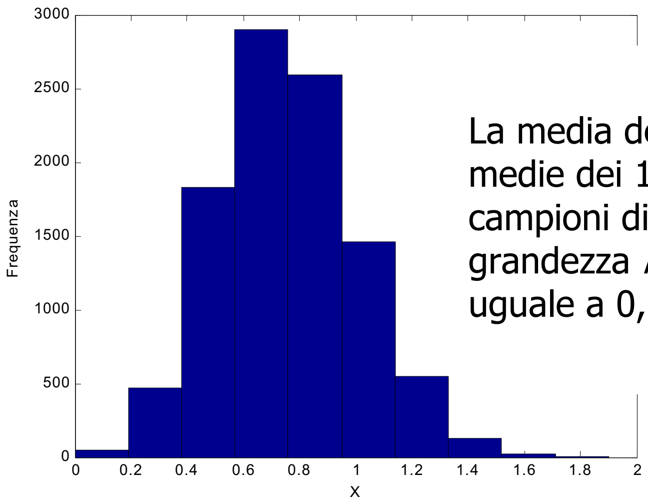
Si noti però che **la forma della distribuzione** non è affatto normale. E' "meno asimmetrica" della distribuzione della popolazione di partenza, ma sicuramente non è per nulla normale.

Teorema del limite centrale: esempio numerico

Il TLC afferma che **la forma della distribuzione delle medie dei campioni si approssima alla normale con l'aumentare della grandezza dei campioni**.

Ripetiamo dunque tutto il processo usando però campioni di grandezza $n = 10$.

Teorema del limite centrale: esempio numerico



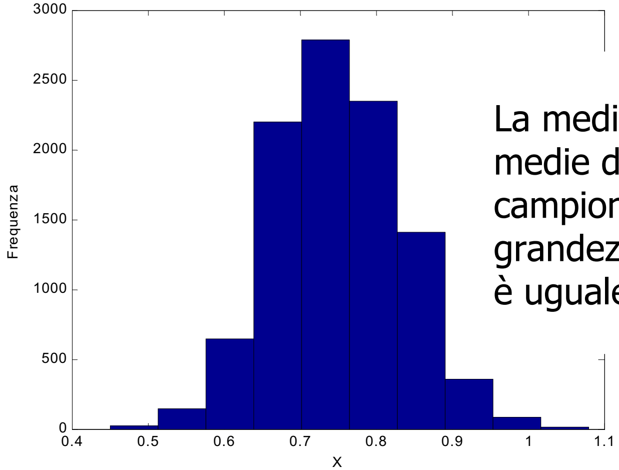
La media delle medie dei 10.000 campioni di grandezza $n = 10$ è uguale a 0,7477.

Teorema del limite centrale: esempio numerico

In questo caso la forma della distribuzione si approssima molto meglio alla forma della distribuzione normale.

Aumentiamo ancora la grandezza del campione: $n = 100$.

Teorema del limite centrale: esempio numerico



La media delle medie dei 10.000 campioni di grandezza n = 100 è uguale a 0,7502.

Teorema del limite centrale: esempio numerico

Con n = 100 l'approssimazione alla normale è molto buona, anche se la distribuzione di partenza era molto asimmetrica.

Approssimazione normale alla binomiale

Questo esempio numerico suggerisce dunque che la distribuzione normale può essere concepita come il limite a cui tende la distribuzione binomiale per un valore fisso di p e $n \rightarrow \infty$.

Sommario

La cosa importante da notare è che, per campioni sufficientemente grandi, la distribuzione della media dei campioni tende alla distribuzione normale indipendentemente dalla forma della distribuzione della popolazione da cui i campioni sono tratti.

Approssimazione normale alla binomiale

Consideriamo nuovamente la distribuzione binomiale alla luce del TLC. La distribuzione normale, infatti, può essere concepita come il limite a cui tende la distribuzione binomiale per un valore fisso di p e $n \to \infty$.

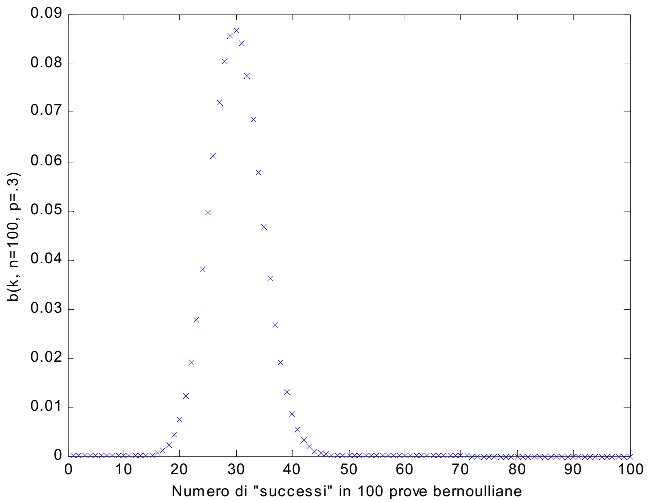
Approssimazione normale alla binomiale

(continua)

Esaminiamo in dettaglio la distribuzione binomiale con parametri $p = 0,3$ e $n = 100$.

Approssimazione normale alla binomiale

(continua)



Approssimazione normale alla binomiale

(continua)

Si noti che la forma della distribuzione binomiale ($b(k;\ n = 100,\ p = 0,3)$) è **simile** alla forma della distribuzione normale.

Per confrontare in maniera più diretta queste due distribuzioni, consideriamole nella loro forma standardizzata.

Approssimazione normale alla binomiale
(continua)

Sia S_n il numero di successi in n prove bernoulliane.

La somma standardizzata S_n sarà:

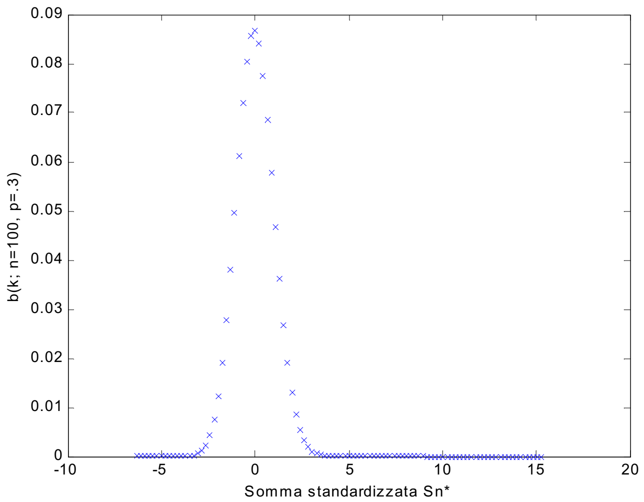
$$S^*_n = \frac{S_n - np}{\sqrt{npq}}$$

Approssimazione normale alla binomiale
(continua)

Creiamo un altro grafico in cui i valori della distribuzione binomiale calcolati in precedenza sono associati a S^*_n anziché a S_n . In altre parole, ciascuna probabilità $b(k;\ n,\ p)$ sarà associata a:

$$x_k = \frac{k - np}{\sqrt{npq}}$$

Approssimazione normale alla binomiale
(continua)

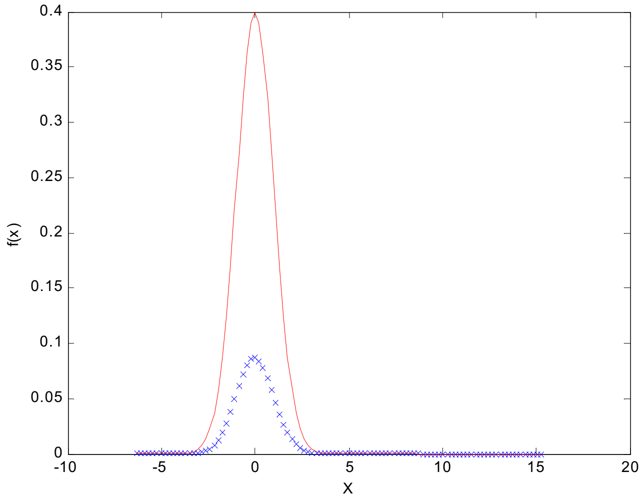


Approssimazione normale alla binomiale
(continua)

Esaminiamo insieme in un unico grafico i valori della distribuzione binomiale associati a S^*_n e i valori della distribuzione normale standardizzata.

Approssimazione normale alla binomiale

(continua)



Approssimazione normale alla binomiale

(continua)

La ragione per le diverse altezze dei due grafici deriva dal fatto che, nel caso della distribuzione binomiale, la somma di tutti i valori di probabilità rappresentati dai simboli "$\times$" è uguale a 1 mentre, nel caso della distribuzione normale standardizzata è **l'area** sottesa alla funzione di densità ad essere uguale a 1.

Approssimazione normale alla binomiale

(continua)

Immaginiamo di collegare i punti segnati dai simboli "$\times$" con una spezzata e di calcolare l'area racchiusa da questa spezzata.

Moltiplichiamo dunque i valori rappresentati dai simboli "$\times$" per una costante ε in modo tale da rendere l'area sottesa alla spezzata uguale a 1.

Approssimazione normale alla binomiale

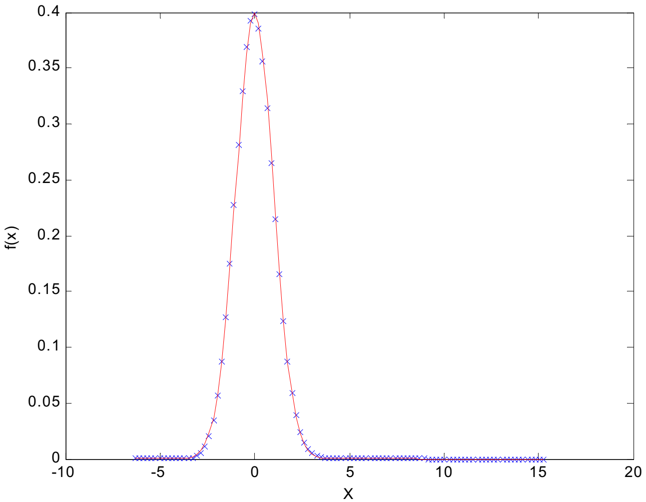
(continua)

La costante cercata ε è uguale a:

$$\varepsilon = \sqrt{npq}$$

Approssimazione normale alla binomiale

(continua)



Approssimazione normale alla binomiale

(continua)

Per valori grandi di *n* possiamo dunque vedere che la distribuzione binomiale è molto simile alla distribuzione normale.

Questa conclusione è formulata più precisamente nel seguente teorema di De Moivre-Laplace che rappresenta una forma particolare del TLC.

Teorema di De Moivre-Laplace

$\forall$ *a* e *b*, $-\infty < a < b + \infty$,

$$\lim_{n\to\infty} P\left(a < \frac{S_n - np}{\sqrt{npq}} \le b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-z^2/2} dx$$

Esempio*

La variabile aleatoria *Y* ha una distribuzione binomiale con $n = 25$ e $p = 0{,}4$. Si trovino le probabilità esatte per gli eventi $Y \le 8$ e $Y = 8$ e si confrontino queste probabilità con i valori ottenuti con l'approssimazione normale.

Esempio*

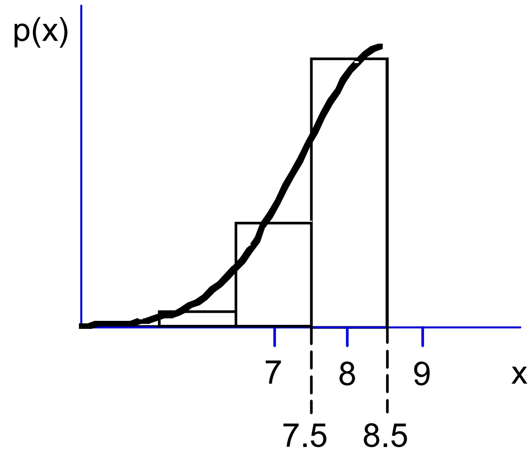
Il fatto di avere $n = 25$ e $p = 0{,}4$ (ovvero, un valore p non troppo diverso da 0,5) giustifica l'approssimazione alla normale in base al TLC.

Esempio*

Valori esatti della distribuzione binomiale:

$P(Y \leq 8) = 0{,}2735$

$P(Y = 8) = 0{,}1200$



Esempio*

Possiamo pensare a Y come se fosse una variabile Z normalmente distribuita con media np e varianza npq.

$$P(Y \leq 8) \approx P(Z \leq 8{,}5)$$

$$= P\left(\frac{Z - np}{\sqrt{npq}} \leq \frac{8{,}5 - (25 \times 0{,}4)}{\sqrt{25 \times 0{,}4 \times 0{,}6}}\right)$$

$$= P(Z \leq -0{,}6124) = 0{,}2701$$

Esempio*

Il valore ottenuto usando l'approssimazione normale alla binomiale (0,2701) è molto simile al valore esatto della distribuzione binomiale [$P(Y \leq 8) = 0{,}2735$] anche per un piccolo valore di n.

Esempio*

Per trovare l'approssimazione normale alla binomiale nel caso di $Y = 8$ dobbiamo trovare l'area sottesa alla curva normale tra i punti 7,5 e 8,5 perché questo intervallo contiene la barra dell'istogramma che rappresenta la probabilità associata a 8 successi in 25 prove con probabilità di successo $p = 0,4$.

Esempio*

Per calcolare la probabilità compresa nell'intervallo [7,5   8,5] calcoliamo prima i due punti z corrispondenti a questi due valori e poi facciamo la differenza tra i valori della distribuzione normale cumulativa associati a questi due punti z.

Esempio*

$$z_1 = \frac{8{,}5 - (25 \times 0{,}4)}{\sqrt{25 \times 0{,}4 \times 0{,}6}} = \text{-}0{,}6124$$

$$z_2 = \frac{7{,}5 - (25 \times 0{,}4)}{\sqrt{25 \times 0{,}4 \times 0{,}6}} = \text{-}1{,}0206$$

$$P(-1{,}0206 \leq Z \leq \text{-}0{,}6124) = 0{,}1164$$

Esempio*

Anche in questo caso, l'approssimazione normale alla binomiale produce un valore molto simile a quello cercato [$P(Y = 8) = 0,12$].

Esempio

Il 10% degli item prodotti da una certa macchina è difettoso. Trovate la probabilità che, in un campione casuale di 400 items prodotti da quella macchina, ci siano al massimo 30 item difettosi.

Esempio

Usiamo l'approssimazione normale alla binomiale:

$$z = \frac{30{,}5-(400\times 0{,}1)}{\sqrt{400\times 0{,}1\times 0{,}9}} = \text{-}1{,}5833$$

$$P(Z\leq -1{,}5833)=0{,}0567$$

Esempio

Un sondaggio ha rivelato che le intenzioni di voto a favore del candidato *A* ammontano al 50%. Assumendo che i risultati del sondaggio siano attendibili, calcolare la probabilità che *A* ottenga almeno il 55% dei voti nella circoscrizione 1 se 100 votanti si recano a votare in quella circoscrizione.

Esempio

Consideriamo i 100 votanti della circoscrizione 1 come 100 prove di un processo bernoulliano con probabilità di "successo" uguale a 0,5. Questa infatti è la probabilità stimata dal sondaggio.

Esempio

Il problema richiede $P(M_n \geq 0,55)$ nel caso di una distribuzione bernoulliana con parametri n = 100 e p = 0,5.

Esempio

Sia M_n la proporzione di "successi" in n prove bernoulliane. Il valore M_n standardizzato è:

$$M_n^* = \frac{M_n - p}{\sqrt{\frac{pq}{n}}} = \frac{0,55 - 0,5}{\sqrt{\frac{0,5 \times 0,5}{100}}} = 1,0$$

Esempio

Resta da trovare la probabilità che una variabile standardizzata M_n assuma un valore maggiore di 1,0. Usando l'approssimazione normale alla binomiale otteniamo:

$$P(Z \geq 1,0) = 0,1587$$

Esempio

Usando l'approssimazione normale alla binomiale si trovi la probabilità che, in 10 lanci di una moneta onesta, l'esito testa sia osservato da 4 a 6 volte (inclusi).

Esempio

$$z_1 = \frac{3{,}5-(10\times 0{,}5)}{\sqrt{10\times 0{,}5\times 0{,}5}} = \text{-}0{,}9487$$

$$z_2 = \frac{6{,}5-(10\times 0{,}5)}{\sqrt{10\times 0{,}5\times 0{,}5}} = 0{,}9487$$

$$P(-0{,}9487 \le Z \le 0{,}9487) = 0{,}6572$$

Esempio

Una linea aerea ha determinato che il 5% delle persone che effettuano una prenotazione su un certo volo in seguito non acquistano il biglietto. Se la linea aerea effettua 160 prenotazioni per un volo che ha solo 155 posti disponibili, si trovi la probabilità che vi sarà un posto disponibile per ciascuna persona dotata di prenotazione che effettivamente acquista il biglietto.

Esempio

$$z = \frac{155{,}5-(160\times 0{,}95)}{\sqrt{160\times 0{,}95\times 0{,}05}} = 1{,}2696$$

$$P(Z \le 1.2696) = 0{,}8979$$

Esercizio 8.1

Sia X una variabile binomiale con parametri $n = 50$ e $p = 0{,}3$.

Si calcoli $P(12 \le X \le 16)$ usando l'approssimazione normale.

Esercizio 8.2

Un dado viene lanciato 24 volte. Si usi il TLC per stimare la probabilità che la somma dei punti prodotti in ciascun lancio sia (a) uguale a 84, (b) maggiore di 84.