Capitolo 4

Teoremi limite classici

I Teoremi limite classici, la Legge dei Grandi Numeri e il Teorema Limite Centrale, costituiscono il nucleo del Calcolo delle Probabilità, per la loro portata sia teorica che applicativa. La legge dei grandi numeri è, tra l'altro, alla base di molti algoritmi che utilizzano metodi probabilistici (metodi di Monte Carlo). Il Teorema limite centrale giustifica il ruolo centrale che le variabili aleatorie gaussiane hanno nella modellistica e nella statistica. Inoltre consente di effettuare numerosi calcoli approssimati di probabilità di interesse applicativo (approssimazione normale).

4.1 Convergenze di variabili aleatorie

Per comprendere meglio i due teoremi limite, introduciamo due diverse definizioni relative alla convergenza di successioni di variabili aleatorie $(X_n)_n$ ad una variabile aleatoria limite X.

Definizione 4.1.1 Diciamo che le variabili aleatorie reali $(X_n)_n$ convergono **in probabilità** verso X (scritto $X_n \xrightarrow{P} X$) se, per ogni $\epsilon > 0$, abbiamo

$$\lim_{n\to\infty} P\{|X_n - X| > \epsilon\} = 0$$

Diciamo che le variabili aleatorie reali $(X_n)_n$ convergono **in legge** verso X (scritto $X_n \rightharpoonup X$) se abbiamo

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$$

con F_{X_n} (risp. F_X) funzione di ripartizione di X_n (risp. X), per ogni t in cui F_X è continua.

Queste due convergenze hanno proprietà abbastanza diverse. Innanzitutto si può dimostrare (ma non lo faremo) che la convergenza in probabilità implica quella in legge. Inoltre, il limite in probabilità è unico, mentre quello in legge no: difatti, se $X_n \rightharpoonup X$ ed Y ha la stessa legge di X, allora si ha anche che $X_n \rightharpoonup Y$. Per questo motivo, se designamo con P_X la legge di X, si scrive anche $X_n \rightharpoonup P_X$ in luogo di $X_n \rightharpoonup X$.

Abbiamo già incontrato un esempio di convergenza in legge nel caso dell'approssimazione di Poisson: riformuliamo qui il risultato con questo nuovo linguaggio.

Proposizione 4.1.2 *Se $X_n \sim B(n, p_n)$, con $np_n \to \lambda > 0$, allora $X_n \rightharpoonup Po(\lambda)$.*

Dimostrazione. Dalla Proposizione 2.5.6 sappiamo che, per ogni $k \in \mathbb{N}$, $p_{X_n}(k) \to X$, dove X è una generica variabile aleatoria con legge $Po(\lambda)$. La funzione di ripartizione F_X è discontinua solo su $\mathbb{N}$, quindi per ogni $t \notin \mathbb{N}$ si ha

$$\lim_{n\to\infty} F_{X_n}(t) = \lim_{n\to\infty} \sum_{k\le t} p_{X_n}(k) = \lim_{n\to\infty} \sum_{k=0}^{[t]} p_{X_n}(k) = \sum_{k=0}^{[t]} \lim_{n\to\infty} p_{X_n}(k) = \sum_{k=0}^{[t]} p_X(k) = F_X(t)$$

e abbiamo la tesi. ∎

4.2 La Legge dei Grandi Numeri

Sia data una successione $(X_n)_{n\ge1}$ di variabili aleatorie reali definite nello stesso spazio di probabilità $(\Omega, \mathcal{A}, P)$. Per avere più chiaro il senso di ciò che segue, si può immaginare che le X_n rappresentino misurazioni successive di una grandezza, ad esempio una grandezza fisica, la cui aleatorietà è dovuta all'imprecisione degli strumenti di misura. Se si effettuano n misure successive, è assai naturale considerare la media aritmetica dei risultati ottenuti, cioè

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Nel linguaggio del calcolo delle probabilità, $\overline{X}_n$ è detta **media campionaria**. Una parte considerevole dei Teoremi limite del calcolo delle probabilità riguarda il comportamento asintotico, per $n \to +\infty$, della media campionaria.

Definizione 4.2.1 Si assuma che le X_n ammettano tutte la stessa madia μ. Diremo che la successione $(X_n)_n$ soddisfa la **Legge (debole) dei Grandi Numeri** se per ogni $\varepsilon > 0$

$$\lim_{n\to+\infty} P\{|\overline{X}_n - \mu| > \varepsilon\} = 0.$$

Con il linguaggio delle convergenze, $(X_n)_n$ soddisfa la Legge (debole) dei Grandi Numeri se $\bar{X}_n \xrightarrow{P} \mu$.

In altre parole, la Legge dei Grandi Numeri afferma che la media campionaria converge alla media probabilistica, fornendo così una giustificazione *a posteriori* della nozione di valor medio.

Resta, naturalmente, da stabilire sotto quali condizioni sulla successione $(X_n)_n$ sia valida la Legge dei Grandi Numeri. L'ipotesi più comunemente assunta è quella in cui le X_n sono i.i.d.

Il prossimo risultato è una versione "elementare" della legge dei grandi numeri.

Proposizione 4.2.2 *Sia (X_n) una successione i.i.d., e supponiamo che le X_n ammettano momento secondo. Allora la successione (X_n) soddisfa la Legge debole dei Grandi Numeri.*

Dimostrazione. Poniamo $\mu := E[X_n]$, e $\sigma^2 := \mathrm{Var}[X_n]$. Dalla linearità del valor medio si vede che

$$E[\overline{X}_n] = \mu.$$

Inoltre, usando il Corollario 2.9.5 si ha

$$\mathrm{Var}[\overline{X}_n] = \frac{\sigma^2}{n}.$$

Ma allora, applicando a $\overline{X}_n$ la Disuguaglianza di Chebischev, si ha

$$P\left\{|\overline{X}_n - \mu| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2},$$

da cui la tesi segue immediatamente. ∎

Usando una tecnica più sofisticata, è possibile dimostrare la Legge debole dei Grandi Numeri per successioni i.i.d., senza assumere l'esistenza del momento secondo. La dimostrazione è qui omessa.

Teorema 2 *Sia $(X_n)_n$ una successione i.i.d., e supponiamo che le X_n ammettano valor medio. Allora la successione $(X_n)_n$ soddisfa alla Legge debole dei Grandi Numeri.*

4.2.1 Il metodo Monte Carlo

La Legge dei Grandi Numeri ha parecchie applicazioni pratiche, molte delle quali sfruttano il fatto che il valore medio di una variabile aleatoria può essere calcolato attraverso la media campionaria di un opportuno numero n (generalmente grande) di simulazioni indipendenti della variabile aleatoria in esame. I problemi risolubili con questo metodo sono di due tipi principali:

1. problemi probabilistici, in cui bisogna calcolare la media $E[X]$ di una variabile aleatoria X di legge sconosciuta o poco maneggevole;

2. problemi puramente deterministici, in cui però gli algoritmi deterministici non sono efficienti o non sono neppure percorribili; in tal caso, può convenire ricondurre il problema deterministico ad un problema probabilistico.

Vediamo qui sotto due esempi di queste tipologie di problemi.

Esempio 4.2.1 Una compagnia di assicurazioni, per calcolare il premio da far pagare ad ogni assicurato, deve calcolare quanto è la somma totale che dovrà rimborsare in media per ogni anno per i sinistri dei propri assicurati. Considerando un modello molto semplice, questa somma può essere rappresentata dalla variabile aleatoria

$$X := \sum_{i=1}^{N_t} e^{Y_i} \tag{4.2.0}$$

dove N_t rappresenta il numero (aleatorio!) di sinistri fino all'istante t ed e^{Y_i} il rimborso (anch'esso aleatorio) pagato al sinistro i-esimo. Un modello molto semplice si ottiene imponendo che $N_t \sim Po(\lambda t)$ e $Y_i \sim N(\mu, \sigma^2)$, dove $\lambda > 0, \mu, \sigma$ sono delle opportune costanti. Resta da specificare la dipendenza delle variabili aleatorie N_t, $(Y_i)_i$. Se queste variabili aleatorie sono indipendenti tra di loro, si riesce a calcolare

$$E[X] = E[N_t]E[e^{Y_i}]$$

ma già la dimostrazione di questa formula non è fattibile senza introdurre nuovi strumenti matematici rispetto a quelli finora trattati. Il problema è che l'ipotesi di indipendenza è spesso troppo forte! Infatti, se la compagnia sta assicurando ad esempio un danno climatico (causato ad esempio da uno tsunami, o più semplicemente da una grandinata), tipicamente le variabili $(Y_i)_i$ saranno fortemente correlate, e pertanto la legge di X (e pertanto il suo valor medio) non è facilmente calcolabile con metodi analitici.

Negli algoritmi *stocastici*, ossia quelli che utilizzano generazione di numeri casuali, non si può essere *certi* che non si supererà un certo errore, e quindi sono basati su un principio diverso. Invece che uno solo, vengono fissati due numeri: la soglia di errore ammissibile e la massima probabilità tollerabile di commettere un errore maggiore della soglia data. In altre parole, non si può pretendere la *certezza* di commettere un errore piccolo, ma soltanto che sia estremamente improbabile commettere un errore maggiore della soglia fissata.

Tornando ai dettagli del problema in esame, se vogliamo calcolare il valor medio di una certa variabile aleatoria X, siano $X_1, \ldots, X_N$ indipendenti e con la stessa legge di X. Se si vuole implementare questo metodo al calcolatore, $X_1, \ldots, X_N$ sono N numeri casuali generati con distribuzione uguale a X. La legge dei grandi numeri applicata alle variabili aleatorie $X_1, \ldots, X_N$, ci dice che

$$\lim_{N \to +\infty} P\left\{\left|\frac{1}{N}\sum_{i=1}^{N} X_i - E[X]\right| > \varepsilon\right\} = 0,$$

dove $\varepsilon > 0$. Dunque, se N è sufficientemente grande, la quantità aleatoria $\frac{1}{N}\sum_{i=1}^{N} X_i$ è, con probabilità elevata, una buona approssimazione del valore medio cercato $E[X]$. Si può dire di più. Sia ε la soglia di errore nel calcolo di $E[X]$, e $\delta > 0$ la probabilità con cui si accetta di compiere un errore maggiore di ε. Vogliamo determinare quanti numeri casuali dobbiamo generare affinchè

$$P\left\{\left|\frac{1}{N}\sum_{i=1}^{N} X_i - E[X]\right| > \varepsilon\right\} \leq \delta. \tag{4.2.1}$$

Dalla dimostrazione della Proposizione 4.2.2, sappiamo che la probabilità in (4.2.1) è minore o uguale a $\frac{\mathrm{Var}[X_1]}{\varepsilon^2 N}$. Supponiamo sia nota una costante $M > 0$ tale che $|X| \leq M$ quasi certamente. Allora

$$\mathrm{Var}[X_1] \leq E[X_1^2] \leq M^2.$$

Ne segue che la disuguaglianza (4.2.1) vale se

$$\frac{M^2}{\varepsilon^2 N} \leq \delta \iff N \geq \frac{M^2}{\delta\varepsilon^2}. \tag{4.2.2}$$

Dunque, se generiamo almeno $\frac{M^2}{\delta\varepsilon^2}$ numeri casuali, sappiamo che con probabilità maggiore o uguale a $1 - \delta$ la quantità $\frac{1}{N}\sum_{i=1}^{N} X_i$ dista non più di ε da $E[X]$.

Questo metodo per il calcolo approssimato di integrali definiti ha il vantaggio di essere molto facile da implementare, in quanto richiede solo di generare una successione di variabili aleatorie con legge assegnata. Tuttavia, benchè le disuguaglianze in (4.2.2) possano essere migliorate, per ottenere una precisione accettabile è necessario generare molti numeri casuali.

Esempio 4.2.2 (continuazione dell'Esempio 4.2.1). Nel caso in cui si voglia calcolare la media dell'espressione (4.2.0), si procede nel modo seguente:

1. si fissa N pari al numero di realizzazioni di X da simulare;
2. per ogni $k = 1, \ldots, N$ si simula una variabile aleatoria $N_t^{(k)} \sim Po(\lambda t)$;
3. per $i = 1, \ldots, N_t$ si simulano le variabili aleatorie $Y_i^{(k)} \sim N(\mu, \sigma^2)$;
4. si calcola la somma $X_k := \sum_{i=1}^{N_t^{(k)}} e^{Y_i^{(k)}}$ e si aggiunge alla somma già trovata $\sum_{j=1}^{k-1} X_j$;

5. si torna al punto 2. finchè k non è uguale a N;

6. si divide $\sum_{k=1}^{N} X_i$ per N: questa è l'approssimazione di $E[X]$.

Quindi, per generare le N realizzazioni $X_1, \ldots, X_N$ bisogna generare un numero di variabili aleatorie pari a N variabili aleatorie di Poisson N_t più la somma di queste N variabili aleatorie. Dato che queste variabili aleatorie hanno tutte parametro λt, in media sarà necessario generare $N(1 + \lambda t)$ variabili aleatorie.

4.3 Il Teorema limite centrale

La dimostrazione del Teorema 4.2.2, come abbiamo visto, è basata sul fatto che $\overline{X}_n$ ha varianza che tende a 0 per $n \to +\infty$. Se la differenza $\overline{X}_n - \mu$, dove $\mu = E(X_i)$, viene *amplificata* di una quantità proporzionale a $\sqrt{n}$, si ottiene la nuova successione di variabili aleatorie

$$Y_n := \sqrt{n}[\overline{X}_n - \mu] = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}}, \tag{4.3.1}$$

che hanno media zero e varianza uguale a $\sigma^2 = \mathrm{Var}[X_i]$ (verificarlo!). Il Teorema limite centrale fornisce la distribuzione di Y_n nel limite per $n \to +\infty$.

Prima di enunciare il teorema, osserviamo che le variabili aleatorie gaussiane hanno un comportamento peculiare rispetto all'espressione (4.3.1). Difatti, applicando ricorsivamente la Proposizione 3.4.2, si ha che se $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ indipendenti, allora $X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2)$, e quindi

$$Y_n = \sqrt{n}[\overline{X}_n - \mu] \sim N(0, \sigma^2).$$

Dunque, se le X_i hanno distribuzione gaussiana, Y_n ha anch'essa distribuzione gaussiana, e tale distribuzione è indipendente da n. Il Teorema limite centrale afferma che, anche se le X_i non hanno distribuzione gaussiana, la distribuzione di Y_n è "vicina", in un senso opportuno, ad una gaussiana. Seguendo la tradizione più diffusa, enunciamo il Teorema limite centrale usando la successione *normalizzata* (con varianza 1) $S_n^* = \frac{1}{\sigma} Y_n$, anzichè Y_n.

Teorema 3 (Teorema limite centrale) *Sia $(X_n)_n$ una successione i.i.d. di variabili aleatorie che ammettono momento secondo e con varianza non nulla. Posto $\mu := E[X_n]$, $\sigma^2 := \mathrm{Var}[X_n]$ e*

$$S_n^* := \frac{\overline{X}_n - \mu}{\sigma} \sqrt{n} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma \sqrt{n}},$$

allora per ogni intervallo I di $\mathbb{R}$

$$\lim_{n \to +\infty} P\{S_n^* \in I\} = P\{Z \in I\},$$

dove $Z \sim N(0,1)$. In altre parole, $S_n^* \rightharpoonup N(0,1)$.

Il Teorema 3, la cui dimostrazione è omessa, può essere usato per effettuare calcoli approssimati di probabilità legate alla media campionaria. Lo schema, che chiameremo di **approssimazione normale**, è quello descritto nell'esempio seguente.

Esempio 4.3.1 Si lancia n volte un dado equilibrato.

a. Se $n = 1000$, qual è la probabilità che il punteggio totale sia minore o uguale di 3400?

b. Quanto grande deve essere n affinchè con probabilità maggiore o uguale a 0.99 il punteggio totale sia almeno $3.3n$?

c. Quanto grande deve essere n affinchè con probabilità maggiore o uguale a 0.99 il punteggio totale sia almeno 500?

La strategia che si segue è la seguente, in tutti e tre i casi. Sia X_i il punteggio ottenuto all'i-esimo lancio. Si esprime la probabilità in esame in termini di

$$Z_n := \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n}.$$

Successivamente, assumendo n sufficientemente grande, si sostituisce a Z_n il "limite" Z, ottenendo un valore approssimato, ma esplicitamente calcolabile.

a. Vogliamo calcolare

$$P\{X_1 + \cdots + X_{1000} \leq 3400\}$$

o, equivalentemente,

$$P\left\{\overline{X}_{1000} \leq 3.4\right\}. \tag{4.3.2}$$

Si noti anzitutto che le variabili aleatorie X_i assumono i valori $1, 2, 3, 4, 5, 6$ ognuno con probabilità $1/6$. Da ciò si trova facilmente che

$$E[X_i] = \mu = 3.5$$

e

$$\mathrm{Var}[X_i] = \sigma^2 \simeq 2.917.$$

Ponendo $n = 1000$, la probabilità in (4.3.2) si può riscrivere nella forma

$$P\left\{Z_n \leq \frac{3.4 - \mu}{\sigma}\sqrt{n}\right\} \simeq P\{Z_n \leq -1.85\}.$$

Se $n = 1000$ è "sufficientemente grande", la probabilità precedente è approssimativamente uguale a

$$P\{Z \leq -1.85\} = \Phi(-1.85),$$

dove con Φ denotiamo la funzione di ripartizione di una normale standard. I valori di Φ si possono trovare in apposite tavole, che tipicamente forniscono i valori di $\Phi(x)$ per $x > 0$. I rimanenti valori di Φ si ottengono osservando che, essendo la densità di Z una funzione pari,

$$\Phi(-x) = P\{Z \leq -x\} = P\{Z \geq x\} = 1 - \Phi(x).$$

Concludiamo allora che

$$P\{X_1 + \cdots + X_{1000} \leq 3400\} \simeq 1 - \Phi(1.85) \simeq 0.032.$$

b. Procediamo come sopra, ma lasciando incognito il valore di n. Vogliamo che sia

$$P\{X_1 + \cdots + X_n \geq 3.3n\} \geq 0.99$$

o, equivalentemente,

$$\begin{aligned} 0.99 &\leq P\{\overline{X}_n \geq 3.3\} = P\left\{Z_n \geq \frac{3.3-\mu}{\sigma}\sqrt{n}\right\} \simeq P\left\{Z_n \geq -0.117\sqrt{n}\right\} \\ &\simeq P\left\{Z \geq -0.117\sqrt{n}\right\} = \Phi(0.117\sqrt{n}). \end{aligned}$$

(nell'ultima uguaglianza abbiamo sfruttato il fatto che la densità normale è una funzione pari). Vogliamo quindi trovare per quali valori di n

$$\Phi(0.117\sqrt{n}) \geq 0.99. \tag{4.3.3}$$

Dalle tavole per Φ si vede che

$$q_{0.99} = \Phi^{-1}(0.99) \simeq 2.326.$$

dove indichiamo con q_α il cosiddetto **quantile di ordine** α della distribuzione normale (vedi appendice A; nel caso di una distribuzione assolutamente continua, il quantile è esattamente uguale alla funzione inversa della funzione di ripartizione). Essendo Φ strettamente crescente, (4.3.3) è equivalente a

$$0.117\sqrt{n} \geq 2.326 \;\Leftrightarrow\; n \geq 395,23 \;\Leftrightarrow\; n \geq 396.$$

c. Vogliamo che sia

$$P\{X_1 + \cdots + X_n \geq 500\} \geq 0.99 \;\Leftrightarrow\; P\left\{Z_n \geq \frac{\frac{500}{n}-\mu}{\sigma}\sqrt{n}\right\} \geq 0.99.$$

Come prima, approssimiamo la precedente probabilità con quella corrispondente rimpiazzando Z_n con Z, ottenendo

$$\Phi\left(-\frac{\frac{500}{n}-\mu}{\sigma}\sqrt{n}\right) \geq 0.99,$$

che equivale a

$$-\frac{\frac{500}{n}-\mu}{\sigma}\sqrt{n} \geq 2.326,$$

che riscriviamo nella forma

$$3.5n - 3.973\sqrt{n} - 500 \geq 0.$$

Risolvendo la precedente come disequazione di secondo grado in $\sqrt{n}$, si trova

$$\sqrt{n} \geq 12.53 \;\Leftrightarrow\; n \geq 158.$$

Naturalmente, i risultati ottenuti nell'esempio precedente si possono considerare affidabili se effettivamente Z_n è "molto vicino" al suo limite Z. Esistono risultati che forniscono stime esplicite per l'errore in tale approssimazione. Ci basta qui rimarcare che se la distribuzione di X_i non è troppo asimmetrica rispetto alla media, le probabilità del tipo $P\{Z_n \in I\}$, con I intervallo di $\mathbb{R}$, sono in termini pratici indistinguibili da $P\{Z \in I\}$ quando n è dell'ordine di alcune decine.

Nella gran parte dei casi $n \geq 30$ è sufficiente ad avere un'ottima approssimazione. Dunque, se nell'Esempio 4.3.1, con riferimento ai quesiti b. e c., avessimo ottenuto dei valori di n minori di 30, tali valori si sarebbero dovuti scartare, in quanto per essi il procedimento di approssimazione usato non è affidabile.

Una stima più precisa di quanto dev'essere grande n per poter usare l'approssimazione normale è nota nel caso in cui le variabili aleatorie $X_i \sim Be(p)$. In tal caso l'approssimazione è buona se $np \geq 5$ e $n(1-p) \geq 5$, oppure se si verifica la condizione (più forte) $np(1-p) \geq 5$. Se $p = 1/2$, per cui la distribuzione di X_i è esattamente simmetrica rispetto alla media, è quindi sufficiente che sia $n \geq 10$. Nel caso di p molto vicino a 0 o a 1, per cui la distribuzione è altamente asimmetrica, sono necessari valori più grandi di n: ad esempio, per $p = 0.01$, occorre avere $n \geq 500$!

Osservazione 4.3.2 (correzione di continuità) Nelle parti a. e c. dell'esempio 4.3.1 abbiamo visto istanze del seguente problema: date n variabili aleatorie i.i.d. a valori interi $X_1, X_2, \ldots, X_n$ e $m \in \mathbb{N}$ calcolare, usando l'approssimazione normale,

$$P\{X_1 + \cdots + X_n \leq m\}. \tag{4.3.4}$$

Posto $\mu := E[X_i]$ e $\sigma^2 := \mathrm{Var}[X_i]$, la probabilità in (4.3.4) è uguale a

$$\Phi\left(\frac{\frac{m}{n} - \mu}{\sigma}\sqrt{n}\right). \tag{4.3.5}$$

Tuttavia, usando il fatto che le X_i sono a valori interi, la probabilità in (4.3.4) è uguale a

$$P\{X_1 + \cdots + X_n < m + 1\}$$

che, usando di nuovo l'approssimazione normale e la continuità di Φ, è approssimativamente uguale a

$$\Phi\left(\frac{\frac{m+1}{n} - \mu}{\sigma}\sqrt{n}\right). \tag{4.3.6}$$

Nell'esempio 4.3.1 la differenza tra (4.3.4) e (4.3.6) è pressoché irrilevante, ma non è sempre così. Supponiamo, ad esempio, $n = 25$, $X_i = Be(1/2)$ e $m = 15$. In questo caso $\mu = 1/2$, $\sigma = 1/2$. Pertanto

$$\Phi\left(\frac{\frac{m}{n} - \mu}{\sigma}\sqrt{n}\right) = \Phi(1) \simeq 0.841,$$

mentre

$$\Phi\left(\frac{\frac{m+1}{n} - \mu}{\sigma}\sqrt{n}\right) = \Phi(1.4) \simeq 0.919.$$

La differenza è considerevole! Per avere comunque una buona approssimazione è opportuno usare la cosidetta **correzione di continuità**, che consiste nel rimpiazzare m in (4.3.4) con $m + \frac{1}{2}$. Tale "mediazione" tra (4.3.4) e (4.3.6), nel caso delle distribuzioni "usuali" (Binomiale, Geometrica, Poisson, che hanno un andamento sufficientemente regolare) migliora talvolta considerevolmente la precisione nell'approssimazione. Nell'esempio appena considerato

$$\Phi\left(\frac{\frac{m+1/2}{n} - \mu}{\sigma}\sqrt{n}\right) = \Phi(1.2) \simeq 0.8849.$$

Il valore esatto della probabilità stimata è

$$\frac{1}{2^{25}} \sum_{k=0}^{15} \binom{25}{k} \simeq 0.8852,$$

da cui si vede l'estrema accuratezza dell'approssimazione.

Esercizi

Esercizio 89 Vengono generati n numeri casuali tra 0 e 1, con distribuzione uniforme. Quanti numeri è necessario generare affinché la probabilità che la somma di essi sia compresa tra $0.49n$ e $0.51n$ sia maggiore o uguale a 0.99?

Esercizio 90 Un venditore porta a porta deve vendere 10 copie di un libro. Se ogni singolo cliente acquista il libro con probabilità 0.1, quanti clienti deve visitare il venditore affinché la probabilità di vendere tutti e dieci i libri sia almeno 0.99?

Esercizio 91 Calcolare approssimativamente la probabilità che una variabile aleatoria X con distribuzione di Poisson di parametro 100 assuma un valore minore di 95.

Esercizio 92 Un congegno è costituito da una componente elettrica che viene rimpiazzata non appena smette di funzionare. Dunque, se $T_1, T_2, \ldots, T_n$ sono i tempi di vita di n componenti che si hanno a disposizione, il tempo di vita totale del congegno è $T = T_1 + T_2 + \cdots + T_n$. Si supponga che $T_i \sim \text{Exp}(1)$, e che le T_i siano indipendenti. Utilizzando l'approssimazione normale calcolare:

a. se $n = 100$ la probabilità $P\{T < 90\}$;

b. il valore minimo di n per cui $P\{T < 90\} \leq 0.05$.

Esercizio 93 Un giocatore di pallacanestro ha una percentuale di successo nei tiri da tre punti del 20%. Calcolare:

a. la probabilità che in 100 tiri faccia non più di 54 punti;

b. il numero minimo di tiri che deve effettuare per realizzare almeno 57 punti con probabilità maggiore o uguale a 0.95.

Esercizio 94 Il numero giornaliero di passeggeri sui treni da Milano a Firenze è una variabile aleatoria di distribuzione incognita. Supponendo che il valore atteso sia pari a 3000 e la varianza pari a 10^6, si calcoli approssimativamente la probabilità che in 30 giorni il numero complessivo di viaggiatori sia almeno 10^5.

Esercizio 95 Sia $\{X_n\}$ una successione di variabili aleatorie i.i.d. con distribuzione di Poisson di parametro 1. Usando opportunamente il Teorema del Limite Centrale per tale successione, calcolare

$$\lim_{n \to +\infty} e^{-n} \sum_{k=0}^{n+\sqrt{n}} \frac{n^k}{k!}.$$

Esercizio 96 Il gruppo promotore di un referendum ritiene che il 60% della popolazione sia disposta a firmare per la relativa raccolta di firme. Si assuma che le persone a cui viene richiesto di firmare siano scelte a caso. Dovendo raccogliere 30.000 firme, quante persone è necessario interpellare affinchè la soglia delle 30.000 firme sia raggiunta con probabilità di almeno $0,95$?

Esercizio 97 Si assuma che, in un libro di 400 pagine, la probabilità che una pagina sia priva di errori sia 0.98, indipendentemente dalle altre pagine. Sia X il numero di pagine che contengono almeno un errore.

a. Qual è la distribuzione di X?

b. Usando l'approssimazione normale, calcolare approssimativamente la probabilità dell'evento $\{X \geq 4\}$.

c. Calcolare la probabilità al punto b. usando un altro tipo di approssimazione, visto a lezione.

Capitolo 5

Statistica

Una delle caratteristiche di quanto visto finora era che la misura di probabilità P si considerava nota, o comunque fissata. Questo spesso non succede in pratica: se difatti a volte può essere sensato rappresentare un fenomeno aleatorio con particolari classi di eventi e variabili aleatorie (esempi: variabili aleatorie di Poisson per eventi rari, esponenziali per vita di strumenti esenti da usura, ecc.), quasi sempre i parametri da cui queste distribuzioni dipendono non sono noti. Sorge quindi l'esigenza di ricavarsi da dati sperimentali quali potrebbero essere i valori plausibili dei parametri significativi. Questo è uno dei due problemi più frequenti della cosiddetta **statistica inferenziale**, e va sotto il nome di **stima dei parametri**: l'altro (che verrà trattato solo marginalmente) consiste nel confermare o rifiutare una data ipotesi sui parametri (esempio: la media è uguale ad un certo valore di riferimento), e va sotto il nome di **test di ipotesi**.

5.1 Modello statistico, campione statistico e stimatori

Passiamo ora alla formalizzazione che si usa più spesso in statistica inferenziale. Come già detto, si suppone spesso che la probabilità dipenda da uno o più parametri, che si rappresentano con la generica lettera θ, che può essere un numero, un vettore o un altro tipo di quantità, e si suppone che appartenga ad un insieme Θ, chiamato **spazio dei parametri** (ad esempio, se P è una legge gaussiana di parametri ignoti, si pone $\theta = (\mu, \sigma)$): spesso $\Theta \subseteq \mathbb{R}^k$, con k opportuno (nell'esempio appena visto, $\theta = (\mu, \sigma) \in \Theta := \mathbb{R} \times (0, +\infty)$).

Definizione 5.1.1 Definiamo **modello statistico** l'oggetto $(\Omega, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$, dove, per ogni $\theta \in \Theta$, P_θ è una misura di probabilità su $(\Omega, \mathcal{A})$.

Nell'ambito di questo modello statistico possono essere svolti degli esperimenti aleatori, i cui risultati possono come al solito essere rappresentati da variabili aleatorie. Questa volta però la loro legge dipenderà dalla particolare scelta di θ, poichè la stessa definizione di legge dipende dalla particolare probabilità P_θ presente sullo spazio $(\Omega, \mathcal{A})$. Allo scopo di stimare θ (o suoi componenti), un'idea usata spesso è quella di prendere delle variabili aleatorie i.i.d., ciascuna delle quali rappresenta un "esperimento" fatto sul fenomeno aleatorio. Formalizziamo ora questo concetto.

Definizione 5.1.2 Chiamiamo **campione statistico**, o più brevemente **campione**, una successione $(X_n)_n$ di variabili aleatorie che siano i.i.d. sotto ogni P_θ, $\theta \in \Theta$; la sequenza finita $X_1, \ldots, X_n$ viene chiamata **campione di taglia** n.

Una delle situazioni più comuni è quella in cui il fenomeno aleatorio sotto esame è relativo ad una "popolazione" che non si può osservare nella sua interezza: in questo caso, ognuna delle variabili aleatorie rappresenta un'osservazione fatta su un diverso individuo estratto dalla popolazione, e il loro spazio di arrivo E dipenderà da quello che vogliamo misurare nella popolazione. La legge che hanno queste variabili aleatorie (che è la stessa per tutte dato che sono i.d., e dipenderà da θ) rappresenta la distribuzione che ha nella popolazione la quantità che vogliamo misurare. Da questo segue che le uniche informazioni sulla popolazione che possiamo avere sono quelle dateci dal campione. Possiamo quindi costruire delle stime di θ dai campioni.

Definizione 5.1.3 Uno **stimatore**, o **statistica campionaria**, è una variabile aleatoria della forma $Y_n := f_n(X_1, \dots, X_n)$, con $f_n : E^n \to \mathbb{R}$ misurabile (in modo che Y_n sia una variabile aleatoria).

I seguenti stimatori sono molto importanti in Statistica.

Esempio 5.1.1 Definiamo **media campionaria** la variabile aleatoria

$$\bar{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

Esempio 5.1.2 Supponiamo che le variabili aleatorie $(X_n)_n$ abbiano media nota e uguale a $m_X := E_\theta[X_i]$ e definiamo **varianza campionaria** nel caso di media nota la variabile aleatoria

$$s_X^2 := \frac{1}{n} \sum_{i=1}^{n} (X_i - m_X)^2$$

Esempio 5.1.3 Nel caso in cui invece la media delle $(X_n)_n$ non sia nota, definiamo **varianza campionaria** nel caso di media non nota la variabile aleatoria

$$s_X^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

A volte è utile fornire una espressione alternativa per s_X^2: notando che

$$\begin{aligned} \sum_{i=1}^{n} (X_i - \bar{X})^2 &= \sum_{i=1}^{n} (X_i^2 - 2X_i\bar{X} + \bar{X}^2) = \sum_{i=1}^{n} X_i^2 - 2\bar{X} \sum_{i=1}^{n} X_i + n\bar{X}^2 = \\ &= \sum_{i=1}^{n} X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \end{aligned}$$

possiamo scrivere lo stimatore in questo modo:

$$s_X^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}$$

Esempio 5.1.4 In entrambi i casi precedenti possiamo definire **deviazione standard campionaria** la variabile aleatoria

$$s_X = \sqrt{s_X^2}$$

La Definizione 5.1.3 è tuttavia piuttosto vaga: mentre nei precedenti esempi è plausibile pensare che le statistiche campionarie definite siano indicative delle quantità che vogliono stimare, sarebbe bene definire delle proprietà che individuino dei "buoni" stimatori. Ci occuperemo di questo nella prossima Sezione.

5.2 Proprietà degli stimatori

Definizione 5.2.1 Diciamo che uno stimatore Y_n di un parametro θ_i, $i = 1, \ldots, k$, é **corretto** o **non distorto** se la sua speranza è uguale alla quantità che deve stimare: più precisamente, se per ogni $\theta \in \Theta$ si ha $E_\theta[Y_n] = \theta_i$, dove E_θ è la speranza sotto la probabilità P_θ.

Vediamo ora che quasi tutti gli stimatori dei precedenti esempi sono corretti.

Esempio 5.2.1 (continuazione dell'Esempio 5.1.1) Supponiamo che uno dei parametri da stimare, cioè una delle componenti di θ, sia $E_\theta[X_i] = m_X$, allora $\bar{X}$ è uno stimatore corretto di m_X: difatti,

$$E_\theta[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} E_\theta[X_i] = \frac{1}{n}\sum_{i=1}^{n} m_X = m_X$$

Esempio 5.2.2 (continuazione dell'Esempio 5.1.2) Supponiamo che uno dei parametri da stimare sia $\mathrm{Var}_\theta[X_i] = \sigma^2$, mentre supponiamo di conoscere $E_\theta[X_i] = m_X$; allora s_X^2 definito nell'Esempio 5.1.2 è uno stimatore corretto di σ^2: difatti

$$E_\theta[s_X^2] = \frac{1}{n}\sum_{i=1}^{n} E_\theta[(X_i^2 - m_X)^2] = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}_\theta[X_i] = \frac{1}{n} \cdot n\sigma^2 = \sigma^2$$

Esempio 5.2.3 (continuazione dell'Esempio 5.1.3) Supponiamo che due dei parametri da stimare siano $E_\theta[X_i] = m_X$ e $\mathrm{Var}_\theta[X_i] = \sigma^2$; allora s_X^2 definito nell'Esempio 5.1.3 è uno stimatore corretto di σ^2: difatti

$$\begin{aligned}
E_\theta[s_X^2] &= \frac{E_\theta[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2]}{n-1} = \frac{\sum_{i=1}^{n} E_\theta[X_i^2] - nE_\theta[\bar{X}^2]}{n-1} = \\
&= \frac{\sum_{i=1}^{n}(\mathrm{Var}_\theta[X_i] + E_\theta[X_i]^2) - n(\mathrm{Var}_\theta[\bar{X}] + E_\theta[\bar{X}]^2)}{n-1} = \\
&= \frac{\sum_{i=1}^{n}(\sigma^2 + m_X^2) - n(\sigma^2/n + m_X^2)}{n-1} = \frac{n\sigma^2 - \sigma^2}{n-1} = \sigma^2
\end{aligned}$$

Questo ci fornisce la giustificazione per cui nello stimatore si divide per $n-1$ invece che per n.

Esempio 5.2.4 (continuazione dell'Esempio 5.1.4) Abbiamo appena visto che gli stimatori $\bar{X}$ e s_X^2 sono non distorti, quest'ultimo sia nel caso di media nota che nel caso di media non nota. Se invece esaminiamo lo stimatore σ_X, notiamo che questo stimatore è distorto sia nel caso di media nota che nel caso di media non nota. Infatti, da

$$\mathrm{Var}_\theta[s_X] = E_\theta[s_X^2] - E_\theta[s_X]^2 \geq 0$$

ricaviamo che

$$E_\theta[s_X] \leq \sqrt{E_\theta[s_X^2]} = \sqrt{\sigma^2} = \sigma$$

In generale quindi la speranza dello stimatore della deviazione standard é minore della deviazione standard vera, ed è uguale ad essa se e solo se $P_\theta\{s_X = \sigma_X\} = 1$, il che non succede praticamente mai.

Vediamo ora un'altra importante proprietà degli stimatori.

Definizione 5.2.2 Uno stimatore Y_n di un parametro θ_i, $i = 1, \ldots, k$, si dice **consistente** se all'aumentare della taglia n del campione, lo stimatore converge in probabilità al parametro che stima: più precisamente, se $Y_n \xrightarrow{P_\theta} \theta_i$ per ogni $\theta \in \Theta$.

Possiamo dimostrare che gli stimatori di media e varianza sono consistenti; in più, anche lo stimatore della deviazione standard è consistente.

Esempio 5.2.5 (continuazione dell'Esempio 5.1.1) Per la media campionaria, ricordando la Legge dei Grandi Numeri abbiamo che

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathrm{P}_\theta} m_X$$

quindi $\bar{X}$ è consistente.

Esempio 5.2.6 (continuazione dell'Esempio 5.1.2) Per quanto riguarda s_X^2 nel caso di media nota, siccome le variabili aleatorie $((X_i - m_X)^2)_i$ sono i.i.d., tutte di media uguale a $E_\theta[(X_i - m_X)^2] = \mathrm{Var}[X_i] = \sigma^2$, si ha

$$s_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m_X)^2 \xrightarrow{\mathrm{P}_\theta} E_\theta[(X_i - m_X)^2] = \sigma^2$$

e quindi in questo caso s_X^2 è consistente.

Esempio 5.2.7 (continuazione dell'Esempio 5.1.3) Per quanto riguarda s_X^2 nel caso di media non nota, siccome le variabili aleatorie $(X_i^2)_i$ sono i.i.d., tutte di media uguale a $E_\theta[X_i^2] = \sigma^2 + m_X^2$, si ha

$$s_X^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1} = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\bar{X}^2 \xrightarrow{\mathrm{P}_\theta} E_\theta[X_i^2] - m_X^2 = \sigma^2 + m_X^2 - m_X^2 = \sigma^2$$

e quindi anche in questo caso s_X^2 è consistente.

Esempio 5.2.8 (continuazione dell'Esempio 5.1.4) Per quanto riguarda s_X sia nel caso di media nota che nel caso di media non nota, per le proprietà della convergenza in probabilità si ha

$$s_X = \sqrt{s_X^2} \xrightarrow{\mathrm{P}_\theta} \sqrt{\sigma^2} = \sigma$$

e quindi s_X è consistente.

Possiamo quindi notare che lo stimatore della deviazione standard, pur essendo distorto, converge alla deviazione standard della popolazione; il fatto che sia distorto implicherà che lo stimatore tenderà ad avere valori più bassi della deviazione standard, mentre per media e varianza non si avrà questo effetto.

L'ultima proprietà degli stimatori che trattiamo è quella della normalità asintotica.

Definizione 5.2.3 Uno stimatore Y_n di un parametro θ_i, $i = 1, \ldots, k$, si dice **approssimativamente normale** se per ogni $\theta \in \Theta$ esiste una costante positiva $\sigma(\theta)$ (che può quindi dipendere da θ) tale che

$$\sqrt{n}\frac{Y_n - \theta_i}{\sigma(\theta)} \rightharpoonup N(0,1)$$

Esempio 5.2.9 (continuazione dell'Esempio 5.1.1) È abbastanza facile mostrare che, se $\mathrm{Var}_\theta[X_i] < +\infty$ per ogni $\theta \in \Theta$, allora la media campionaria è uno stimatore asintoticamente normale. Difatti, posto $\sigma(\theta) := \mathrm{Var}_\theta[X_i]$, per il Teorema Limite Centrale, abbiamo

$$\frac{\sqrt{n}(\bar{X} - m_X)}{\sigma(\theta)} \overset{\text{legge}}{\to} N(0,1)$$

Questo risultato è indipendente dalla distribuzione delle X_i (sempre a patto che ammettano varianza σ^2 finita!), e quindi è indipendente dalla particolare distribuzione della popolazione iniziale. Questo significa che per n grande la funzione di distribuzione di $\sqrt{n}(\bar{X} - \mu)/\sigma$ è vicina a quella di $N(0,1)$, e quindi quella di $\bar{X}$ è vicina a quella di $N(\mu, \sigma^2/n)$, e quindi si può approssimare $\bar{X}$ con una variabile aleatoria gaussiana di legge $N(\mu, \sigma^2/n)$.

5.3 Intervalli di confidenza asintotici e test di ipotesi

La normalità asintotica di uno stimatore permette di quantificare lerrore di stima che si commette nello stimare θ_i con Y_n. Per $0 < \alpha << 1$, cerchiamo un $x > 0$ tale che

$$\lim_{n\to\infty} P_\theta \left\{ \left| \frac{\sqrt{n}(Y_n - \theta_i)}{\sigma(\theta)} \right| \le x \right\} = 1 - \alpha$$

Se Y_n è asintoticamente normale, allora si ha

$$1 - \alpha = P_\theta\{|Z| \le x\} = P_\theta\{-x \le |Z| \le x\} = F_Z(x) - F_Z(-x)$$

con $Z \sim N(0,1)$, e quindi

$$1 - \alpha = 2F_Z(x) - 1$$

e infine $F_Z(x) = 1 - \alpha/2$, e infine $x = q_{1-\alpha/2}$, dove $q_\alpha := \Phi^{-1}(\alpha)$ è il quantile di una legge $N(0,1)$. Si ha quindi, per "n grande",

$$1 - \alpha \simeq P_\theta \left\{ \left| \frac{\sqrt{n}(Y_n - \theta_i)}{\sigma(\theta)} \right| \le q_{1-\alpha/2} \right\} = P_\theta \left\{ \theta_i \in \left[Y_n - \frac{\sigma(\theta)}{\sqrt{n}} q_{1-\alpha/2}, Y_n + \frac{\sigma(\theta)}{\sqrt{n}} q_{1-\alpha/2} \right] \right\}$$

Questo ragionamento giustifica la seguente definizione.

Definizione 5.3.1 Se Y_n è uno stimatore asintoticamente normale di θ_i, chiamiamo **intervallo di confidenza asintotico** di livello 1 - α per il parametro θ_i l'intervallo

$$\left[Y_n - \frac{\sigma(\theta)}{\sqrt{n}} q_{1-\alpha/2}, Y_n + \frac{\sigma(\theta)}{\sqrt{n}} q_{1-\alpha/2} \right]$$

Questa definizione non è però pienamente "operativa": difatti in essa compare la quantità $\sigma(\theta)$, che potrebbe dipendere da θ, e che quindi in generale è sconosciuta. Ci sono due possibili approcci a questo problema.

1. Il primo si può utilizzare quando si sappia che $\sup_{\theta\in\Theta} \sigma(\theta) = \bar{\sigma} < +\infty$: in questo caso, infatti, l'intervallo di confidenza è contenuto sicuramente in

$$\left[Y_n - \frac{\bar{\sigma}}{\sqrt{n}} q_{1-\alpha/2}, Y_n + \frac{\bar{\sigma}}{\sqrt{n}} q_{1-\alpha/2} \right] \tag{5.3.0}$$

e quindi

$$P_\theta\left\{\theta_i \in \left[Y_n - \frac{\bar\sigma}{\sqrt{n}}q_{1-\alpha/2}, Y_n + \frac{\bar\sigma}{\sqrt{n}}q_{1-\alpha/2}\right]\right\} \geq P_\theta\left\{\theta_i \in \left[Y_n - \frac{\sigma(\theta)}{\sqrt{n}}q_{1-\alpha/2}, Y_n + \frac{\sigma(\theta)}{\sqrt{n}}q_{1-\alpha/2}\right]\right\} = 1-\alpha$$

di conseguenza si può usare l'intervallo in (5.3.0) come approssimazione per l'intervallo di confidenza asintotico.

2. Se però accadesse che $\sup_{\theta\in\Theta}\sigma(\theta) = +\infty$, non si potrebbe più utilizzare l'approccio precedente. È allora lecito usare una stima per $\sigma(\theta)$ in luogo di questa quantità sconosciuta.

Esempio 5.3.1 Supponiamo di avere un campione $X_1, \dots, X_n$ di variabili aleatorie $Be(p)$ e di volere stimare $\theta := p$: siccome questo parametro è anche la speranza matematica delle $X_1, \dots, X_n$, un "buon" stimatore potrebbe essere quello della media campionaria:

$$\hat p := \bar X_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Per il Teorema Limite Centrale, sappiamo che

$$\sqrt{n}\frac{\bar X_n - p}{\sqrt{p(1-p)}} \rightharpoonup N(0,1)$$

e quindi $\hat p = \bar X_n$ è asintoticamente normale con $\sigma(p) := \sqrt{p(1-p)}$, che non è noto. Si ha tuttavia che $\sup_{p\in(0,1)}\sigma(p) = \frac{1}{2}$, e quindi l'intervallo di confidenza asintotico è sicuramente contenuto in

$$\left[\hat p - \frac{1}{2\sqrt{n}}q_{1-\alpha/2}, \hat p + \frac{1}{2\sqrt{n}}q_{1-\alpha/2}\right]$$

Se invece decidiamo di usare il secondo approccio, possiamo approssimare l'intervallo di confidenza asintotico con

$$\left[\hat p - \sqrt{\frac{\hat p(1-\hat p)}{n}}q_{1-\alpha/2}, \hat p + \sqrt{\frac{\hat p(1-\hat p)}{n}}q_{1-\alpha/2}\right]$$

Utilizzando gli intervalli di confidenza, è possibile rispondere alla tipica domanda di un cosiddetto **test di ipotesi**. Tipicamente questi test si presentano nella forma di domande sui parametri (ad esempio: "la media è nulla?" oppure "è uguale ad un valore che abbiamo in mente?"), e sono riconducibili a due regioni disgiunte dello spazio dei parametri Θ_0, chiamata **ipotesi**, e Θ_1, chiamata **alternativa**, tali che $\Theta_0 \cup \Theta_1 = \Theta$.

Esempio 5.3.2 Supponiamo di avere un campione $X_1, \dots, X_n$ di legge $N(m, \sigma^2)$, con entrambi i parametri incogniti: dunque $\Theta := \mathbb{R} \times (0, +\infty)$. Ci chiediamo se la media possa essere uguale ad una media di riferimento $\bar m$ (magari proveniente da osservazioni passate su una popolazione "di base"): questa ipotesi da testare si traduce nel fissare l'ipotesi $\Theta_0 := \{m = \bar m\}$ e l'alternativa $\Theta_1 := \{m \neq \bar m\}$, dove le due scritture stanno per

$$\Theta_0 := \{(m,\sigma^2) \in \Theta \mid m = \bar m\}, \qquad \Theta_1 := \{(m,\sigma^2) \in \Theta \mid m \neq \bar m\}$$

In una situazione come quella dell'Esempio 5.3.2, è possibile usare gli intervalli di confidenza per dare una risposta alla domanda del test. Supponiamo infatti di avere uno stimatore asintoticamente normale Y_n per la media m, e di poter calcolare (o approssimare con uno dei due metodi visti) un intervallo di confidenza asintotico per m. Sappiamo dunque che, dato un parametro della forma $\theta = (m, \sigma^2)$, per definizione di intervallo di confidenza asintotico si ha

$$1 - \alpha \simeq P_\theta \left\{ m \in \left[Y_n - \frac{\sigma(\theta)}{\sqrt{n}} q_{1-\alpha/2}, Y_n + \frac{\sigma(\theta)}{\sqrt{n}} q_{1-\alpha/2} \right] \right\}$$

e quindi in particolare se $m = \bar{m}$, chiamato $\bar{\theta} := (\bar{m}, \sigma^2)$, si ha

$$1 - \alpha \simeq P_{\bar{\theta}} \left\{ \bar{m} \in \left[Y_n - \frac{\sigma(\bar{\theta})}{\sqrt{n}} q_{1-\alpha/2}, Y_n + \frac{\sigma(\bar{\theta})}{\sqrt{n}} q_{1-\alpha/2} \right] \right\}$$

quindi, con probabilità $1 - \alpha$ sotto $P_{\bar{\theta}}$, il valore $\bar{m}$ apparterrà all'intervallo di confidenza di livello $1 - \alpha$. Il ragionamento che a questo punto si fa in Statistica è il seguente:

- se $\bar{m}$ appartiene all'intervallo di confidenza, allora non ci sono evidenze statistiche che "qualcosa è andato storto", e si **accetta** l'ipotesi Θ_0;
- se invece $\bar{m}$ non appartiene all'intervallo di confidenza, allora è successo qualcosa di improbabile sotto $P_{\bar{\theta}}$, oppure semplicemente non è quella la probabilità "giusta" per il nostro fenomeno: siccome non è dato sapere quale di queste cose sia vera, si **rifiuta** l'ipotesi Θ_0 e si accetta l'alternativa Θ_1.

Vediamo quindi che la risposta ad un test di ipotesi non viene data in termini di "vero" o "falso". Questo è giustificato dal fatto che l'intervallo di confidenza asintotico, essendo funzione di stimatori, è a sua volta una variabile aleatoria i cui valori dipendono (pur tramite leggi di probabilità) dal caso, e quindi le risposte prodotte tramite esso (ad esempio se un dato valore vi appartiene o meno) possono cambiare a seconda del risultato dell'esperimento aleatorio: pertanto tramite questo metodo (tipico di tutta la Statistica inferenziale) non è possibile dimostrare "verità" o "falsità" di affermazioni, ma solo quanto siano o meno plausibili. Ci si accontenta quindi di dire "l'ipotesi è accettata (rifiutata)", in luogo del più forte ma indimostrabile "l'ipotesi è vera (falsa)".