

## Capitolo 6

### Teoremi limite classici

**Abstract** 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 della molti algoritmi che utilizzano metodi probabilistici (metodi di Monte Carlo). Il teorema limite centrale giustifica il ruolo centrale che le variabili casuali Gaussiane hanno nella modellistica e nella statistica. Inoltre consente di effettuare numerosi calcoli approssimati di probabilità di interesse applicativo (approssimazione normale).

#### 6.1 La legge dei grandi numeri

Sia data una successione  $(X_n)_{n \geq 1}$  di variabili casuali scalari 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

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

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

**Definizione 6.1.** Si assuma che le  $X_n$  ammettano tutte la stessa media  $\mu \in (0, \infty)$ . Diremo che la successione  $(X_n)$  soddisfa alla legge debole dei grandi numeri se per ogni  $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0. \quad (6.1)$$

In altre parole, la legge dei grandi numeri afferma che per  $n \rightarrow \infty$  la media campionaria  $\bar{X}_n$  converge (nel senso descritto da (6.1)) verso media probabilistica, fornendo così una giustificazione *a posteriori* della nozione di valor medio.

Resta naturalmente da stabilire sotto quali condizioni sulla successione  $(X_n)$  sia valida la legge dei grandi numeri. L'ipotesi più comunemente assunta è quella in cui le variabili  $X_n$  sono tra loro indipendenti e che abbiano tutte la stessa distribuzione. Diremo, in tal caso, che le variabili  $X_n$  sono *indipendenti e identicamente distribuite (i.i.d.)* o, più semplicemente, che  $(X_n)$  è una successione i.i.d.. Facendo l'ipotesi aggiuntiva che le variabili  $X_n$  ammettano momento secondo, è facile dare una dimostrazione elementare della legge dei grandi numeri.

**Proposizione 6.1.** *Sia  $(X_n)$  una successione di variabili i.i.d. che ammettono momento secondo. Allora la successione  $(X_n)$  soddisfa alla legge debole dei grandi numeri.*

*Dimostrazione.* Poniamo  $\mu = E(X_n)$ , e  $\sigma^2 = \text{Var}(X_n)$ . Dalla linearità del valor medio si vede che

$$E(\bar{X}_n) = \mu. \quad (6.2)$$

Inoltre, usando il Corollario 3.3 si ha

$$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}. \quad (6.3)$$

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

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}, \quad (6.4)$$

da cui la tesi segue immediatamente.

*Osservazione 6.1.* Nella Proposizione 6.1, l'ipotesi che le variabili siano i.i.d. può essere notevolmente indebolita. Infatti è sufficiente assumere che le variabili  $X_n$  abbiano tutte la stessa media e la stessa varianza, cioè  $E(X_n) = \mu$  e  $\text{Var}(X_n) = \sigma^2$  per ogni  $n \in \mathbb{N}$ , e che siano scorrelate a due a due, cioè  $\text{Cov}(X_i, X_j) = 0$  per  $i \neq j$ , e le relazioni (6.2) e (6.3) continuano a valere (si veda il Corollario 3.3). Di conseguenza anche la relazione (6.4) resta valida e si ottiene la legge debole dei grandi numeri sotto queste ipotesi più generali.

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. Riportiamo di seguito l'enunciato (la dimostrazione è qui omessa).

**Teorema 6.1.** *Sia  $(X_n)$  una successione di variabili i.i.d. che ammettono valor medio. Allora la successione  $(X_n)$  soddisfa alla legge debole dei grandi numeri.*

*Esempio 6.1.* Per avere un'idea della portata applicativa della legge dei grandi numeri, consideriamo il seguente problema. Sia  $f$  una funzione di  $\mathbb{R}$  in  $\mathbb{R}$  Riemann integrabile nell'intervallo  $[a, b]$ . Le funzioni il cui integrale è calcolabile esattamente

con metodi analitici sono, in realtà, abbastanza poche, e perciò sono importanti metodi numerici per il calcolo approssimato di tale integrale. I metodi più comuni consistono nel “discretizzare” l’intervallo  $[a, b]$ , e, ad esempio, approssimare  $\int_a^b f(x)dx$  con la somma di Riemann

$$\frac{1}{N} \sum_{i=1}^N f\left(a + \frac{i}{N}(b-a)\right),$$

dove  $1/N$  è il “passo” della discretizzazione. Tale approssimazione “ingenua”, basata direttamente sulla definizione di integrale di Riemann, può essere migliorata, in particolare quando siano note ulteriori informazioni su  $f$  come, ad esempio, la sua derivata. Per una  $f$  sufficientemente “buona” è possibile, una volta fissato il massimo errore di approssimazione “tollerabile”, determinare *esattamente* quanto grande dev’essere  $N$  per garantire di non superare tale soglia di errore.

Gli algoritmi *stocastici*, ossia quelli che utilizzano generazione di numeri casuali, sono basati su un principio diverso. Vengono fissati due numeri: la soglia di errore e la massima probabilità tollerabile di commettere un errore maggiore della soglia data. In altre parole, non si pretende 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, siano  $X_1, \dots, X_N \sim U(a, b)$  indipendenti. In altre parole,  $X_1, \dots, X_N$  sono  $N$  numeri casuali generati con distribuzione uniforme in  $[a, b]$ . Si noti che, per ogni  $i$ ,

$$E[f(X_i)] = \frac{1}{b-a} \int_a^b f(x)dx.$$

La legge dei grandi numeri applicata alle variabili casuali  $f(X_1), \dots, f(X_N)$ , ci dice che

$$\lim_{N \rightarrow +\infty} P \left[ \left| \frac{1}{N} \sum_{i=1}^N f(X_i) - \frac{1}{b-a} \int_a^b f(x)dx \right| > \varepsilon \right] = 0,$$

dove  $\varepsilon > 0$ . Dunque, se  $N$  è sufficientemente grande, la quantità aleatoria  $\frac{1}{N} \sum_{i=1}^N f(X_i)$  è una buona approssimazione dell’integrale “normalizzato”  $\frac{1}{b-a} \int_a^b f(x)dx$  con probabilità elevata. Si può dire di più. Sia  $\varepsilon$  la soglia di errore nel calcolo di  $\frac{1}{b-a} \int_a^b f(x)dx$ , e  $\delta > 0$  la probabilità con cui si accetta di compiere un errore maggiore di  $\varepsilon$ . Vogliamo determinare quanti numeri casuali dobbiamo generare affinché

$$P \left[ \left| \frac{1}{N} \sum_{i=1}^N f(X_i) - \frac{1}{b-a} \int_a^b f(x)dx \right| > \varepsilon \right] \leq \delta. \quad (6.5)$$

Dalla dimostrazione della Proposizione 6.1, sappiamo che la probabilità in (6.5) è minore o uguale a  $\frac{\text{Var}[f(X_1)]}{\varepsilon^2 N}$ . Supponiamo sia nota una costante  $M > 0$  tale che  $|f(x)| \leq M$  per ogni  $x \in [a, b]$ . Allora

$$\text{Var}[f(X_1)] \leq E[f^2(X_1)] \leq M^2.$$

Ne segue che la disuguaglianza (6.5) vale se

$$\frac{M^2}{\varepsilon^2 N} \leq \delta \iff N \geq \frac{M^2}{\delta \varepsilon^2}. \quad (6.6)$$

Dunque, se generiamo almeno  $\frac{M^2}{\delta \varepsilon^2}$  numeri casuali, sappiamo che con probabilità maggiore o uguale a  $1 - \delta$  la quantità  $\frac{1}{N} \sum_{i=1}^N f(X_i)$  dista non più di  $\varepsilon$  da  $\frac{1}{b-a} \int_a^b f(x) dx$ .

Questo metodo per il calcolo approssimato di integrali definiti ha il vantaggio di essere molto facile da implementare, in quanto richiede solo un generatore di numeri casuali con distribuzione uniforme. Tuttavia, benchè le disuguaglianze in (6.6) possano essere migliorate, per ottenere una precisione accettabile è necessario generare molti numeri casuali, il che rende questo metodo meno efficiente degli algoritmi “deterministici”. Questo discorso cambia radicalmente quando si tratta di calcolare integrali di funzioni di molte variabili. In tal caso esistono varianti multidimensionali dell’algoritmo appena descritto, che risultano, in dimensione elevata, assai più efficienti degli algoritmi deterministici.

### 6.1.1 Il teorema di approssimazione di Weierstrass

Tra le conseguenze più interessanti della legge dei grandi numeri (o, più semplicemente, della disuguaglianza di Chebyshev) è la seguente dimostrazione *costruttiva* del teorema di approssimazione di Weierstrass. In ciò che segue, per  $f \in \mathcal{C}([0, 1])$ , consideriamo la seguente successione di polinomi, detti i *Polinomi di Bernstein* di  $f$ :

$$p_n(x) = \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

**Teorema 6.2.** *Se  $f \in \mathcal{C}([0, 1])$ , la successione dei Polinomi di Bernstein di  $f$  converge uniformemente a  $f$ , cioè*

$$\lim_{n \rightarrow +\infty} \sup_{x \in [0, 1]} |f(x) - p_n(x)| = 0.$$

*Dimostrazione.* Ricordiamo che una funzione continua su un intervallo chiuso e limitato di  $\mathbb{R}$  è limitata e uniformemente continua, cioè  $M := \sup_{x \in [0, 1]} |f(x)| < +\infty$ , e per ogni  $\varepsilon > 0$  esiste  $\delta > 0$  tale che  $|y - x| < \delta \Rightarrow |f(y) - f(x)| < \varepsilon$ .

Fissiamo dunque  $\varepsilon > 0$ , e sia  $\delta$  tale che  $|y - x| < \delta \Rightarrow |f(y) - f(x)| < \varepsilon/2$ . Siano inoltre, per  $x \in [0, 1]$ ,  $X_1, X_2, \dots, X_n \sim \text{Be}(x)$  indipendenti. Sia inoltre

$$A_{n,\varepsilon} := \{|\bar{X}_n - x| \geq \varepsilon\}.$$

Per lo stesso argomento usato nella dimostrazione della legge dei grandi numeri

$$P(A_{n,\varepsilon}) \leq \frac{x(1-x)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}, \quad (6.7)$$

dove si è usato il fatto che la funzione  $x \mapsto x(1-x)$  ha massimo in  $x = 1/2$ . A questo punto, osservando che

$$p_n(x) = E(f(\bar{X}_n)),$$

e usando (6.7), si ha

$$\begin{aligned} |f(x) - p_n(x)| &= |f(x) - E(f(\bar{X}_n))| = |E(f(x) - f(\bar{X}_n))| \leq E[|f(x) - f(\bar{X}_n)|] \\ &= E[1_{A_{n,\varepsilon}} |f(x) - f(\bar{X}_n)|] + E[1_{A_{n,\varepsilon}^c} |f(x) - f(\bar{X}_n)|] \\ &\leq 2MP(A_{n,\varepsilon}) + \frac{\varepsilon}{2}P(A_{n,\varepsilon}^c) \leq \frac{M}{2n\varepsilon^2} + \frac{\varepsilon}{2} < \varepsilon \end{aligned}$$

se  $n \geq n_0$ , per un opportuno  $n_0$ . Poichè quest'ultima stima è indipendente da  $x$ , la conclusione segue facilmente.

## 6.2 Il teorema limite centrale: enunciato e metodo dell'approssimazione normale

La dimostrazione del Teorema 6.1, come abbiamo visto, è basata sul fatto che  $\bar{X}_n$  ha varianza che tende a 0 per  $n \rightarrow +\infty$ . Se la differenza  $\bar{X}_n - \mu$ , dove  $\mu = E(X_i)$ , viene *amplificata* di una quantità proporzionale a  $\sqrt{n}$ , si ottiene una nuova successione di variabili casuali,

$$Y_n = \sqrt{n}[\bar{X}_n - \mu], \quad (6.8)$$

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

Prima di enunciare il teorema, osserviamo che le variabili casuali Gaussiane hanno un comportamento peculiare rispetto l'espressione (6.8). Infatti, applicando ricorsivamente la Proposizione 5.4, si ha che se  $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$  indipendenti, allora  $X_1 + X_2 + \dots + X_n \sim N(n\mu, n\sigma^2)$ , e quindi

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

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)  $Z_n = \frac{1}{\sigma}Y_n$ , anzichè  $Y_n$ .

**Teorema 6.3.** (teorema limite centrale). Sia  $(X_n)$  una successione i.i.d. di variabili casuali che ammettono momento secondo e con varianza non nulla. Posto  $\mu = E(X_n)$ ,  $\sigma^2 = \text{Var}(X_n)$  e

$$Z_n \equiv \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n},$$

allora per ogni  $x \in \mathbb{R}$

$$\lim_{n \rightarrow +\infty} P(Z_n \leq x) = P(Z \leq x),$$

dove  $Z \sim N(0, 1)$ .

Il Teorema 6.3, la cui dimostrazione verrà data nel paragrafo successivo, 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 6.2.** Si lancia  $n$  volte un dado equilibrato.

- Se  $n = 1000$ , qual è la probabilità che il punteggio totale sia minore o uguale di 3400?
- Quanto grande deve essere  $n$  affinché con probabilità maggiore o uguale a 0.99 il punteggio totale sia almeno  $3.3n$ ?
- Quanto grande deve essere  $n$  affinché 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

$$Z_n \equiv \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n}.$$

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

- Vogliamo calcolare

$$P(X_1 + \dots + X_{1000} \leq 3400)$$

o, equivalentemente,

$$P(\bar{X}_{1000} \leq 3.4). \quad (6.9)$$

Si noti anzitutto che le variabili casuali  $X_i$  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

$$\text{Var}(X_i) = \sigma^2 \simeq 2.917.$$

La probabilità in (6.9) si può riscrivere nella forma, con  $n = 1000$ ,

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

Se  $n$  è “sufficientemente grande”, la probabilità precedente è approssimativamente uguale a

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

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

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

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

$$P(X_1 + \dots + X_n \geq 3.3n) \geq 0.99$$

o, equivalentemente,

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

In altre parole vogliamo trovare per quali valori di  $n$

$$\Phi(0.117\sqrt{n}) \geq 0.99. \quad (6.10)$$

Dalle tavole per  $\Phi$  si vede che

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

Essendo  $\Phi$  strettamente crescente, (6.10) è equivalente a

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

(c) Vogliamo che sia

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

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 6.2, 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 casuali  $X_i \sim Be(p)$ . In tal caso l'approssimazione è buona se  $np \geq 5$ . Se  $p = 1/2$ , per cui la distribuzione di  $X_i$  è esattamente simmetrica rispetto alla media, è 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 6.2.* Nelle parti a. e c. dell'esempio 6.2 abbiamo visto istanze del seguente problema: date  $n$  variabili casuali i.i.d. a valori interi  $X_1, X_2, \dots, X_n$  e  $m \in \mathbb{N}$  calcolare, usando l'approssimazione normale,

$$P(X_1 + \dots + X_n \leq m). \quad (6.11)$$

Posto  $\mu = E(X_i)$  e  $\sigma^2 = Var(X_i)$ , la probabilità in (6.11) è uguale a

$$\Phi\left(\frac{\frac{m}{n}-\mu}{\sigma}\sqrt{n}\right). \quad (6.12)$$

Tuttavia, usando il fatto che le  $X_i$  sono a valori interi, la probabilità in (6.11) è uguale a

$$P(X_1 + \dots + X_n < m + 1)$$

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

$$\Phi\left(\frac{\frac{m+1}{n} - \mu}{\sigma} \sqrt{n}\right). \quad (6.13)$$

Nell'esempio 6.2 la differenza tra (6.11) e (6.13) 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 cosiddetta *correzione di continuità*, che consiste nel rimpiazzare  $m$  in (6.11) con  $m + \frac{1}{2}$ . Tale "mediazione" tra (6.11) e (6.13), 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.

### 6.3 Teorema limite centrale: dimostrazione

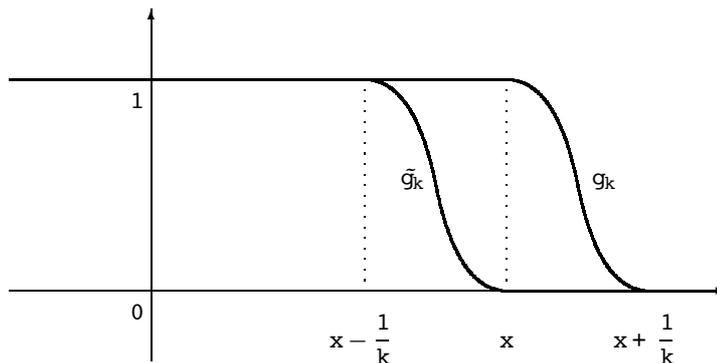
La dimostrazione del Teorema 6.3 verrà divisa in più passi. Inoltre, dimostreremo prima il teorema sotto l'ipotesi aggiuntiva che le variabili casuali  $X_i$  ammettano momento terzo. Il caso generale necessita di un po' più di attenzione, e verrà trattato alla fine di questo paragrafo. Iniziamo con alcuni lemmi ausiliari.

**Lemma 6.1.** Denotiamo con  $\mathcal{C}_b^3$  l'insieme delle funzioni di  $\mathbb{R}$  in  $\mathbb{R}$  le cui prime tre derivate esistono continue e limitate. Sia  $(Z_n)_{n \geq 1}$  una successione di variabili casuali, e  $Z$  una variabile casuale la cui funzione di ripartizione  $F_Z$  è continua. Supponiamo che per ogni  $g \in \mathcal{C}_b^3$  si abbia che

$$\lim_{n \rightarrow +\infty} E[g(Z_n)] = E[g(Z)].$$

Allora per ogni  $x \in \mathbb{R}$

$$\lim_{n \rightarrow +\infty} F_{Z_n}(x) = F_Z(x).$$



**Figura 6.1** Una rappresentazione grafica delle funzioni  $g_k$  e  $\tilde{g}_k$ , che approssimano dall'alto e dal basso la funzione indicatrice  $1_{(-\infty, x]}(\cdot)$ .

*Dimostrazione.* Sia  $x \in \mathbb{R}$ , e  $k \geq 1$ . Denotiamo con  $g_k$  e  $\tilde{g}_k$  due elementi di  $\mathcal{C}_b^3$  tali che per ogni  $z \in \mathbb{R}$

$$1_{(-\infty, x - \frac{1}{k}]}(z) \leq \tilde{g}_k(z) \leq 1_{(-\infty, x]}(z) \leq g_k(z) \leq 1_{(-\infty, x + \frac{1}{k}]}(z).$$

(vedi Figura 6.1).

Ne segue che, per ogni variabile casuale  $W$

$$\begin{aligned} F_W\left(x - \frac{1}{k}\right) &= E\left[1_{(-\infty, x - \frac{1}{k}]}(W)\right] \leq E[\tilde{g}_k(W)] \leq E[1_{(-\infty, x]}(W)] \\ &= F_W(x) \leq E[g_k(W)] \leq E[1_{(-\infty, x + \frac{1}{k}]}(W)] = F_W\left(x + \frac{1}{k}\right). \end{aligned} \quad (6.14)$$

Usando le disuguaglianze in (6.14) e l'ipotesi, abbiamo

$$\limsup_{n \rightarrow +\infty} F_{Z_n}(x) \leq \lim_{n \rightarrow +\infty} E[g_k(Z_n)] = E[g(Z_n)] \leq F_Z\left(x + \frac{1}{k}\right)$$

e

$$\liminf_{n \rightarrow +\infty} F_{Z_n}(x) \geq \lim_{n \rightarrow +\infty} E[\tilde{g}_k(Z_n)] = E[\tilde{g}(Z_n)] \geq F_Z\left(x - \frac{1}{k}\right).$$

Pertanto

$$F_Z\left(x - \frac{1}{k}\right) \leq \liminf_{n \rightarrow +\infty} F_{Z_n}(x) \leq \limsup_{n \rightarrow +\infty} F_{Z_n}(x) \leq F_Z\left(x + \frac{1}{k}\right).$$

Prendendo il limite per  $k \rightarrow +\infty$  nelle disuguaglianze precedenti e usando il fatto che, essendo  $F_Z$  continua,

$$\lim_{k \rightarrow +\infty} F_Z\left(x - \frac{1}{k}\right) = \lim_{k \rightarrow +\infty} F_Z\left(x + \frac{1}{k}\right) = F_Z(x),$$

troviamo

$$\liminf_{n \rightarrow +\infty} F_{Z_n}(x) = \limsup_{n \rightarrow +\infty} F_{Z_n}(x) = F_Z(x),$$

e si conclude.

**Lemma 6.2.** *Siano  $V, Y, Z$  tre variabili casuali indipendenti tali che  $Y, Z$  ammettono momento terzo,  $E(Y) = E(Z)$ ,  $E(Y^2) = E(Z^2)$ . Sia  $g \in \mathcal{C}_b^3$  e  $C = \sup_{x \in \mathbb{R}} |g^{(3)}(x)|$ . Allora*

$$|E[g(V+Y)] - E[g(V+Z)]| \leq \frac{C}{6} [E(|Y|^3) + E(|Z|^3)].$$

*Dimostrazione.* La formula di Taylor per funzioni di classe  $\mathcal{C}^3$  con resto integrale ci dà, per ogni  $x, h \in \mathbb{R}$

$$g(x+h) = g(x) + g'(x)h + \frac{1}{2}g''(x)h^2 + R_2(x, h),$$

dove

$$R_2(x, h) = \frac{1}{2} \int_x^{x+h} (x+h-t)^2 g^{(3)}(t) dt.$$

In particolare

$$|R_2(x, h)| \leq \frac{C}{6} |h|^3. \quad (6.15)$$

Si ricava facilmente che

$$g(x+h) - g(x+k) = g'(x)[h-k] + \frac{1}{2}g''(x)[h^2 - k^2] + R_2(x, h) - R_2(x, k). \quad (6.16)$$

A questo punto poniamo, in (6.15),  $x = V, h = Y, k = Z$  e prendiamo la media:

$$E[g(V+Y)] - E[g(V+Z)] = E[g'(V)(Y-Z)] + \frac{1}{2}E[g''(V)(Y^2 - Z^2)] + E[R_2(V, Y) - R_2(V, Z)].$$

Ma, essendo  $V, Y, Z$  indipendenti e  $E(Y) = E(Z)$ ,  $E(Y^2) = E(Z^2)$ , abbiamo

$$E[g'(V)(Y-Z)] = E[g'(V)]E[(Y-Z)] = 0 \quad E[g''(V)(Y^2 - Z^2)] = E[g''(V)]E[(Y^2 - Z^2)] = 0.$$

Ne segue, anche usando (6.15),

$$\begin{aligned} |E[g(V+Y)] - E[g(V+Z)]| &= |E[R_2(V,Y) - R_2(V,Z)]| \leq E[|R_2(V,Y)|] + E[|R_2(V,Z)|] \\ &\leq \frac{C}{6} [E(|Y|^3) + E(|Z|^3)], \end{aligned}$$

che completa la dimostrazione.

La seguente proposizione rappresenta il “cuore” della dimostrazione del Teorema 6.3.

**Proposizione 6.2.** *Siano  $Y_1, Y_2, \dots, Y_n$  variabili casuali i.i.d., che ammettono momento terzo, e tali che  $E(Y_1) = 0$ ,  $E(Y_1^2) = 1$ . Siano inoltre  $W_1, W_2, \dots, W_n$  variabili casuale con le medesime proprietà, cioè sono i.i.d.,  $E(W_1) = 0$ ,  $E(W_1^2) = 1$ . Sia inoltre  $g \in \mathcal{C}_b^3$  e  $C := \sup_{x \in \mathbb{R}} |g^{(3)}(x)|$ . Allora*

$$\left| E \left[ g \left( \frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[ g \left( \frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right| \leq \frac{C}{6} \frac{E(|Y_1|^3) + E(|W_1|^3)}{\sqrt{n}}.$$

*Dimostrazione.* Sia  $Y := (Y_1, Y_2, \dots, Y_n)$  e  $W := (W_1, W_2, \dots, W_n)$ . Il risultato da dimostrare dipende solo dalle distribuzioni marginali di  $Y$  e  $W$ , ma non dalla distribuzione congiunta di  $Y$  e  $W$ . Non è perciò restrittivo assumere che  $Y$  e  $W$  siano indipendenti, cioè le variabili casuali  $Y_1, Y_2, \dots, Y_n, W_1, W_2, \dots, W_n$  sono tutte indipendenti tra loro. L’idea chiave consiste nello scrivere la seguente somma telescopica:

$$\begin{aligned} &E \left[ g \left( \frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[ g \left( \frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \\ &= \sum_{k=0}^{n-1} \left\{ E \left[ g \left( \frac{Y_1 + \dots + Y_k + Y_{k+1} + W_{k+2} + \dots + W_n}{\sqrt{n}} \right) \right] \right. \\ &\quad \left. - E \left[ g \left( \frac{Y_1 + \dots + Y_k + W_{k+1} + W_{k+2} + \dots + W_n}{\sqrt{n}} \right) \right] \right\} \\ &= \sum_{k=0}^{n-1} \left\{ E \left[ g \left( V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[ g \left( V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right\}, \end{aligned}$$

dove abbiamo posto

$$V_k := \frac{Y_1 + \dots + Y_k + W_{k+2} + \dots + W_n}{\sqrt{n}}.$$

Per il Lemma 6.2

$$\left| E \left[ g \left( V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[ g \left( V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right| \leq \frac{C}{6} \frac{E(|Y_1|^3) + E(|W_1|^3)}{n\sqrt{n}}.$$

Pertanto

$$\begin{aligned}
& \left| E \left[ g \left( \frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[ g \left( \frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right| \\
& \leq \sum_{k=0}^{n-1} \left| E \left[ g \left( V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[ g \left( V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right| \\
& \leq n \frac{C E(|Y_1|^3) + E(|W_1|^3)}{6 n \sqrt{n}} = \frac{C E(|Y_1|^3) + E(|W_1|^3)}{6 \sqrt{n}},
\end{aligned}$$

che è quanto volevamo dimostrare.

Dalla proposizione precedente segue il fatto, assolutamente non banale e non intuitivo, che

$$\lim_{n \rightarrow +\infty} \left\{ E \left[ g \left( \frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[ g \left( \frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right\} = 0 \quad (6.17)$$

indipendentemente dalle distribuzioni delle  $Y_i$  e delle  $W_i$ .

*Dimostrazione (Dimostrazione del Teorema 6.3 con l'ipotesi aggiuntiva  $E(|X_1|^3) < +\infty$ ).*

Usando le notazioni nell'enunciato del Teorema 6.3, poniamo

$$Y_i := \frac{X_i - \mu}{\sigma}.$$

Sia inoltre  $(W_n)$  una successione di variabili casuali i.i.d. con distribuzione  $N(0, 1)$ . Per quanto visto nella Proposizione 5.4,

$$\frac{W_1 + \dots + W_n}{\sqrt{n}} \sim N(0, 1).$$

Quindi, se  $Z \sim N(0, 1)$  e  $g \in \mathcal{C}_b^3$ ,

$$E \left[ g \left( \frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] = E[g(Z)]$$

ed è quindi indipendente da  $n$ . Quindi possiamo applicare (6.17) e otteniamo

$$\lim_{n \rightarrow +\infty} E \left[ g \left( \frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] = E[g(Z)].$$

A questo punto, per completare la dimostrazione e dimostrare la convergenza delle funzioni di ripartizione, basta applicare il Lemma 6.1.

Nella parte restante di questo paragrafo vedremo come modificare la precedente dimostrazione al fine di dimostrare il Teorema 6.3 *senza* l'ipotesi aggiuntiva  $E(|X_1|^3) < +\infty$ . La modifica fondamentale è costituita da una versione più raffinata del Lemma 6.2

**Lemma 6.3.** *Siano  $V, Y, Z$  tre variabili casuali indipendenti tali che  $Z$  ammette momento terzo,  $E(Y) = E(Z)$ ,  $E(Y^2) = E(Z^2)$ . Sia  $g \in \mathcal{C}_b^3$  e  $C_3 = \sup_{x \in \mathbb{R}} |g^{(3)}(x)|$ ,  $C_2 := \sup_{x \in \mathbb{R}} |g''(x)|$ . Allora, per ogni  $\varepsilon > 0$*

$$|E[g(V+Y)] - E[g(V+Z)]| \leq \frac{C_3}{2} E[|Y|^3 1_{\{|Y| \leq \varepsilon\}}] + \frac{C_2}{2} E[|Y|^2 1_{\{|Y| > \varepsilon\}}] + \frac{C_3}{6} E[|Z|^3]$$

*Dimostrazione.* La formula di Taylor arrestata al primo ordine ci dà

$$\begin{aligned} g(x+h) &= g(x) + g'(x)h + \int_x^{x+h} (x+h-t)g''(t)dt \\ &= g(x) + g'(x)h + \frac{1}{2}g''(x)h^2 + \int_x^{x+h} (x+h-t)[g''(t) - g''(x)]dt \\ &= g(x) + g'(x)h + \frac{1}{2}g''(x)h^2 + \bar{R}_2(x, h), \end{aligned}$$

dove

$$\bar{R}_2(x, h) := \int_x^{x+h} (x+h-t)[g''(t) - g''(x)]dt,$$

e dove abbiamo usato il fatto che

$$\int_x^{x+h} (x+h-t)dt = \frac{h^2}{2}.$$

Per stimare  $\bar{R}_2(x, h)$  osserviamo anzitutto che

$$|\bar{R}_2(x, h)| \leq \sup_{t \in [x, x+h]} |g''(t) - g''(x)| \int_x^{x+h} (x+h-t)dt = \frac{h^2}{2} \sup_{t \in [x, x+h]} |g''(t) - g''(x)|.$$

Ora, usiamo due diverse stime della differenza  $|g''(t) - g''(x)|$ , per  $t \in [x, x+h]$ . In modo banale

$$|g''(t) - g''(x)| \leq 2C_2 \quad (6.18)$$

Inoltre, per il teorema del valor medio

$$|g''(t) - g''(x)| \leq C_3|t-x| \leq C_3h. \quad (6.19)$$

Ora, usiamo (6.18) se  $|h| > \varepsilon$  e (6.19) se  $|h| \leq \varepsilon$ . Otteniamo

$$|\bar{R}_2(x, h)| \leq C_2 h^2 1_{[-\varepsilon, \varepsilon]^c}(h) + \frac{C_3}{2} |h|^3 1_{[-\varepsilon, \varepsilon]}(h). \quad (6.20)$$

Applicando (6.16) esattamente come nel Lemma 6.2, ma usando il resto  $\bar{R}_2$  per  $g(V+Y)$  e  $R_2$  per  $g(V+Z)$ , otteniamo:

$$E[g(V+Y)] - E[g(V+Z)] = E[g'(V)(Y-Z)] + \frac{1}{2}E[g''(V)(Y^2-Z^2)] + E[\bar{R}_2(V, Y) - R_2(V, Z)].$$

Come nel Lemma 6.2, essendo  $V, Y, Z$  indipendenti e  $E(Y) = E(Z)$ ,  $E(Y^2) = E(Z^2)$ , abbiamo

$$E[g'(V)(Y-Z)] = E[g'(V)]E[(Y-Z)] = 0 \quad E[g''(V)(Y^2-Z^2)] = E[g''(V)]E[(Y^2-Z^2)] = 0.$$

Pertanto,

$$\begin{aligned} |E[g(V+Y)] - E[g(V+Z)]| &= |E[\bar{R}_2(V, Y) - R_2(V, Z)]| \leq E[|\bar{R}_2(V, Y)|] + E[|R_2(V, Z)|] \\ &\leq C_2 E[Y^2 1_{[-\varepsilon, \varepsilon]^c}(Y)] + \frac{C_3}{2} E[|Y|^3 1_{[-\varepsilon, \varepsilon]}(Y)] + \frac{C}{6} E(|Z|^3) \end{aligned}$$

che completa la dimostrazione.

Prima di completare la dimostrazione del Teorema 6.3 dimostriamo il seguente Lemma.

**Lemma 6.4.** *Sia  $(\Omega, \mathcal{A}, P)$  uno spazio di probabilità, e  $X \in L^1(\Omega, \mathcal{A}, P)$ . Allora*

$$\lim_{a \rightarrow +\infty} E[X 1_{[-a, a]^c}(X)] = 0.$$

*Dimostrazione.* Dimostriamo la tesi solo nel caso in cui  $(\Omega, \mathcal{A}, P) = (\Omega, P)$  è uno spazio di probabilità discreto. Il caso generale richiede argomenti un po' più avanzati, non trattati in questo corso. Usando la definizione di somma infinita

$$E[|X|] = \sum_{\omega \in \Omega} |X(\omega)| P(\{\omega\}) = \sup_{A \subseteq \Omega: |A| < +\infty} \sum_{\omega \in A} |X(\omega)| P(\{\omega\}).$$

Pertanto, per ogni  $\varepsilon > 0$ , esiste un sottoinsieme finito  $A_\varepsilon$  di  $\Omega$  tale che

$$\sum_{\omega \in A_\varepsilon} |X(\omega)| P(\{\omega\}) \geq E[|X|] - \varepsilon.$$

Sia  $a_\varepsilon := \max\{|X(\omega)| : \omega \in A_\varepsilon\}$ , che ovviamente è finito poichè  $A_\varepsilon$  è un insieme finito. Abbiamo

$$\sum_{\omega: |X(\omega)| \leq a_\varepsilon} |X(\omega)| P(\{\omega\}) \geq \sum_{\omega \in A_\varepsilon} |X(\omega)| P(\{\omega\}) \geq E[|X|] - \varepsilon,$$

quindi

$$E[|X| 1_{[-a_\varepsilon, a_\varepsilon]^c}(X)] = E[|X|] - \sum_{\omega: |X(\omega)| \leq a_\varepsilon} |X(\omega)| P(\{\omega\}) \leq \varepsilon.$$

Da questo la tesi del Lemma segue immediatamente.

*Dimostrazione (Dimostrazione del Teorema 6.3).* Seguiamo l'argomento e le notazioni usate nella Proposizione 6.2. Poniamo  $Y_i := \frac{X_i - \mu}{\sigma}$ ,  $W_1, W_2, \dots, W_n \sim N(0, 1)$

indipendenti. Allora, come nella dimostrazione della Proposizione 6.2, per  $g \in \mathcal{C}_b^3$  si ha

$$\begin{aligned} & E \left[ g \left( \frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[ g \left( \frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \\ &= \sum_{k=0}^{n-1} \left\{ E \left[ g \left( \frac{Y_1 + \dots + Y_k + Y_{k+1} + W_{k+2} + \dots + W_n}{\sqrt{n}} \right) \right] \right. \\ &\quad \left. - E \left[ g \left( \frac{Y_1 + \dots + Y_k + W_{k+1} + W_{k+2} + \dots + W_n}{\sqrt{n}} \right) \right] \right\} \\ &= \sum_{k=0}^{n-1} \left\{ E \left[ g \left( V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[ g \left( V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right\}, \end{aligned}$$

con

$$V_k := \frac{Y_1 + \dots + Y_k + W_{k+2} + \dots + W_n}{\sqrt{n}}.$$

Per il Lemma 6.3

$$\begin{aligned} & \left| E \left[ g \left( V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[ g \left( V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right| \\ & \leq C_2 n^{-1} E \left[ Y_1^2 1_{[-\varepsilon\sqrt{n}, \sqrt{n}\varepsilon]^c}(Y_1) \right] + \frac{C_3}{2n\sqrt{n}} E \left[ |Y_1|^3 1_{[-\sqrt{n}\varepsilon, \sqrt{n}\varepsilon]}(Y) \right] + \frac{C}{6n\sqrt{n}} E(|W_1|^3). \end{aligned}$$

Perciò

$$\begin{aligned} & \left| E \left[ g \left( \frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[ g \left( \frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right| \\ & \leq C_2 E \left[ Y_1^2 1_{[-\varepsilon\sqrt{n}, \sqrt{n}\varepsilon]^c}(Y_1) \right] + \frac{C_3}{2\sqrt{n}} E \left[ |Y_1|^3 1_{[-\sqrt{n}\varepsilon, \sqrt{n}\varepsilon]}(Y_1) \right] + \frac{C}{6\sqrt{n}} E(|W_1|^3) \\ & \leq C_2 E \left[ Y_1^2 1_{[-\varepsilon\sqrt{n}, \sqrt{n}\varepsilon]^c}(Y_1) \right] + \frac{C_3}{2} \varepsilon + \frac{C}{6\sqrt{n}} E(|W_1|^3), \end{aligned}$$

dove abbiamo usato il fatto che  $|Y_1|^3 1_{[-\sqrt{n}\varepsilon, \sqrt{n}\varepsilon]}(Y_1) \leq \varepsilon\sqrt{n}|Y_1|^2$ . Mandando  $n \rightarrow +\infty$  e usando l'arbitrarietà di  $\varepsilon$ , concludiamo che

$$\lim_{n \rightarrow +\infty} \left\{ E \left[ g \left( \frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[ g \left( \frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right\} = 0.$$

Da qui in poi la dimostrazione è identica a quella vista in precedenza con l'ipotesi aggiuntiva.