Calcolo delle Probabilità e Statistica Matematica – A.A. 2009/10

CAPITOLO 8 – Teoremi Limite

8.1 Introduzione

8.2 La disuguaglianza di Chebyshev e la legge debole dei grandi numeri

8.3 Il teorema del limite centrale

8.1 Introduzione

I risultati più significativi del calcolo delle probabilità sono certamente i teoremi limite.

Di questi, i più importanti sono quelli che vengono classificati come *leggi dei grandi numeri* e come *teoremi del limite centrale*.

In generale, un teorema viene considerato una legge dei grandi numeri quando stabilisce delle condizioni sotto le quali la media di una successione di variabili aleatorie converge (in qualche senso) alla media attesa.

D'altra parte, i teoremi del limite centrale sono caratterizzati dal determinare condizioni sotto le quali la somma di un grande numero di variabili aleatorie (opportunamente standardizzata) ha la distribuzione di probabilità che tende a essere normale.

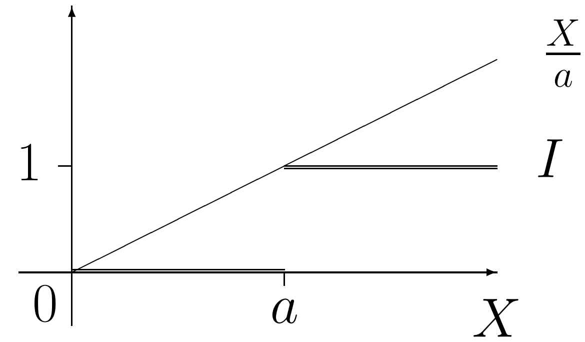
8.2 La disuguaglianza di Chebyshev e la legge debole dei grandi numeri

Proposizione. Disuguaglianza di Markov. Se X è una variabile aleatoria che assume solo valori non negativi, allora per ogni numero reale $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Dimostrazione. Per $a > 0$, sia

$$I = \mathbf{1}_{\{X \geq a\}} = \begin{cases} 1 & \text{se } X \geq a \\ 0 & \text{altrimenti.} \end{cases}$$



Ricordando che $X \geq 0$, si ha: $I \leq \dfrac{X}{a}$.

Dalle proprietà del valore atteso segue: $E[I] \leq \dfrac{E[X]}{a}$.

Ricordando che per la variabile di Bernoulli I risulta $E[I] = P(I = 1) = P(X \geq a)$, si ha

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Proposizione. Disuguaglianza di Chebyshev. Se X è una variabile aleatoria di media μ e varianza σ^2, entrambe finite, allora per ogni numero reale $k > 0$,

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}.$$

Dimostrazione. Essendo $(X - \mu)^2$ una variabile aleatoria non negativa, possiamo applicare la disuguaglianza di Markov, con $a = k^2$:

$$P(X \geq a) \leq \frac{E(X)}{a} \qquad \text{diventa} \qquad P\{(X - \mu)^2 \geq k^2\} \leq \frac{E[(X - \mu)^2]}{k^2}.$$

Essendo $(X - \mu)^2 \geq k^2$ se e solo se $|X - \mu| \geq k$, si ricava infine:

$$P\{|X - \mu| \geq k\} \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}.$$

Analogamente si può ottenere una limitazione inferiore per $P\{\mu - k < X < \mu + k\}$:

$$P\{|X - \mu| < k\} \geq 1 - \frac{\sigma^2}{k^2}.$$

Esempio. Si supponga che il numero di studenti che si iscrive ad un corso di lingua inglese è descritto da una variabile aleatoria X di media $\mu = 50$.
(a) Cosa si può dire su $P(X \geq 75)$?
(b) Cosa si può dire su $P(40 < X < 60)$ se $\sigma^2 = 25$?

Soluzione. (a) Facendo uso della disuguaglianza di Markov, con $a = 75$, si ha

$$P(X \geq 75) \leq \frac{E(X)}{75} = \frac{50}{75} = \frac{2}{3}.$$

(b) Ricordiamo che dalla disuguaglianza di Chebyshev segue

$$P\{|X - \mu| < k\} \geq 1 - \frac{\sigma^2}{k^2}.$$

Quindi, essendo $\mu = 50$ e $\sigma^2 = 25$, per $k = 10$ si ottiene

$$P(40 < X < 60) = P\{|X - 50| < 10\} = P\{|X - \mu| < k\} \geq 1 - \frac{\sigma^2}{k^2} = 1 - \frac{25}{100} = \frac{3}{4}.$$

Esempio. Sia X uniformemente distribuita su $(0, 10)$. Allora, poiché $E[X] = 5$ e $\text{Var}(X) = 25/3$, dalla disuguaglianza di Chebyshev segue

$$P\{|X - 5| > 4\} \leq \frac{25}{3 \cdot 16} = \frac{25}{48} \approx 0{,}52$$

mentre il valore esatto è

$$P\{|X - 5| > 4\} = P(X < 1) + P(X > 9) = \frac{1}{10} + \frac{1}{10} = 0{,}2.$$

Esempio. Sia X una variabile normale di media μ e varianza σ^2. Risulta

$$P\{\mu - h\,\sigma < X < \mu + h\,\sigma\} = P\Big(-h < \frac{X - \mu}{\sigma} < h\Big) = 2\,\Phi(h) - 1,$$

mentre dalla disuguaglianza di Chebyshev, per $k = h\,\sigma$, si ha

$$P\{\mu - h\,\sigma < X < \mu + h\,\sigma\} = P\{|X - \mu| < h\,\sigma\} \geq 1 - \frac{\sigma^2}{(h\,\sigma)^2} = 1 - \frac{1}{h^2}.$$

	$h = 1$	$h = 1{,}5$	$h = 2$	$h = 2{,}5$	$h = 3$
$2\,\Phi(h) - 1$	0,6826	0,8664	0,9544	0,9876	0,9974
$1 - 1/h^2$	0	0,5555	0,75	0,84	0,8888

Teorema. La legge debole dei grandi numeri. Sia $X_1, X_2, \ldots$ una successione di variabili aleatorie i.i.d., ognuna con media finita $E[X_i] = \mu$. Allora, posto $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, per ogni $\epsilon > 0$ si ha

$$\lim_{n\to\infty} P\{|\overline{X}_n - \mu| \geq \epsilon\} = 0.$$

Dimostrazione. Assumiamo inoltre che $\mathrm{Var}(X_i) = \sigma^2$ sia finita. Poiché $E[\overline{X}_n] = \mu$ e $\mathrm{Var}[\overline{X}_n] = \sigma^2/n$, dalla disuguaglianza di Chebyshev si ha, per ogni $\epsilon > 0$

$$P\{|\overline{X}_n - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{n}\frac{1}{\epsilon^2}.$$

Procedendo al limite per $n \to \infty$ segue immediatamente l'asserto.

8.3 Il teorema del limite centrale

Vedremo ora uno dei più importanti risultati della teoria della probabilità. Esso, in breve, asserisce che la somma di un grande numero di variabili aleatorie indipendenti ha distribuzione approssimativamente normale. La dimostrazione è omessa per brevità.

Teorema. Il teorema del limite centrale. Sia $X_1, X_2, \ldots$ una successione di variabili aleatorie i.i.d., ognuna di media μ e varianza $\sigma^2 > 0$ finite. Allora, posto

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}},$$

per ogni $a \in \mathbb{R}$ si ha: $\lim_{n\to\infty} P\{Z_n \le a\} = \Phi(a) \qquad \left(= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx \right).$

In altri termini, la somma standardizzata $Z_n = \frac{X_1+\cdots+X_n-E[X_1+\cdots+X_n]}{\sqrt{\mathrm{Var}(X_1+\cdots+X_n)}}$ tende ad avere distribuzione normale standard quando $n \to \infty$. Vale quindi l'approssimazione

$$P\left(\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le a\right) \approx \Phi(a) \qquad (n \text{ grande}).$$

Esempio. Si genera a caso una sequenza di n bit $X_1, \ldots, X_n$, ossia variabili aleatorie di Bernoulli indipendenti di parametro 1/2. Posto $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, determinare un limite inferiore e un'approssimazione per $P(0{,}49 < \overline{X}_n < 0{,}51)$ se $n = 10000$.

Soluzione. Ricordando che $E(\overline{X}_n) = \mu$ e $Var(\overline{X}_n) = \sigma^2/n$, dove $\mu = 1/2$ e $\sigma^2 = 1/4$, per $n = 10000$ dalla disuguaglianza di Chebyshev si trae il limite inferiore

$$P(0{,}49 < \overline{X}_n < 0{,}51) = P\{|\overline{X}_n - \mu| < 0{,}01\} \geq 1 - \frac{1/4}{n(0{,}01)^2} = \frac{3}{4} = 0{,}75.$$

L'approssimazione basata sul teorema del limite centrale, con Z normale standard, è:

$$\begin{aligned} P(0{,}49 < \overline{X}_n < 0{,}51) &= P(4900 < X_1 + \ldots + X_n < 5100) \\ &= P(4899{,}5 < X_1 + \ldots + X_n < 5100{,}5) \qquad \text{(la correzione di continuità)} \\ &\approx P\left(\frac{4899{,}5 - 5000}{50} < Z < \frac{5100{,}5 - 5000}{50}\right) = P(-2{,}01 < Z < 2{,}01) \\ &= \Phi(2{,}01) - \Phi(-2{,}01) = 2\,\Phi(2{,}01) - 1 = 2 \cdot 0{,}9778 - 1 = 0{,}9556, \end{aligned}$$

essendo $n\mu = 10000\,(1/2) = 5\,000$ e $\sigma\sqrt{n} = (1/2)\sqrt{10000} = 50$.

Esempio. Si supponga che il numero di studenti che si iscrive ad un corso universitario è descritto da una variabile aleatoria di Poisson X di media $\mu = 100$. Qual è la probabilità che si iscrivano almeno 120 studenti?

Soluzione. Non è agevole determinare il valore esatto, essendo

$$P(X \geq 120) = e^{-100} \sum_{k=120}^{\infty} \frac{(100)^k}{k!} = 1 - e^{-100} \sum_{k=0}^{119} \frac{(100)^k}{k!}.$$

Ricordando che X può essere riguardata come somma di $n = 100$ variabili aleatorie di Poisson indipendenti di media 1 (e quindi di varianza 1), possiamo adoperare l'approssimazione basata sul teorema del limite centrale:

$$\begin{aligned} P(X \geq 120) &= P(X \geq 119{,}5) \qquad \text{(la correzione di continuità)} \\ &= P\left(\frac{X - 100}{\sqrt{100}} \geq \frac{119{,}5 - 100}{\sqrt{100}}\right) \\ &\approx 1 - \Phi(1{,}95) \\ &= 1 - 0{,}9744 = 0{,}0256. \end{aligned}$$

Esempio. Determinare la probabilità approssimata che lanciando a caso 10 dadi la somma X dei valori ottenuti sia compresa tra 30 e 40, estremi inclusi.

Soluzione. Indicando l'esito del lancio i-esimo con X_i, questa è una variabile uniforme discreta con $E[X_i] = 7/2$ e $\mathrm{Var}(X_i) = 35/12$. Pertanto risulta

$$E[X] = 35, \qquad \mathrm{Var}(X) = \frac{350}{12} = 29{,}1666.$$

L'approssimazione basata sul teorema del limite centrale dà quindi:

$$\begin{aligned}
P(30 \leq X \leq 40) &= P(29{,}5 \leq X \leq 40{,}5) \qquad \text{(la correzione di continuità)} \\
&= P\left(\frac{29{,}5 - 35}{\sqrt{29{,}1666}} \leq \frac{X - 35}{\sqrt{29{,}1666}} \leq \frac{40{,}5 - 35}{\sqrt{29{,}1666}}\right) \\
&\approx \Phi(1{,}02) - \Phi(-1{,}02) \\
&= 2\,\Phi(1{,}02) - 1 \\
&= 2 \cdot 0{,}8461 - 1 = 0{,}6922.
\end{aligned}$$

Esempio. Siano X_i, $i = 1, \ldots, 24$, variabili aleatorie indipendenti, con distribuzione uniforme su $(0, 1)$. Determinare in modo approssimato i valori di

$$P\left(\sum_{i=1}^{24} X_i > 14\right), \qquad P\left(8 < \sum_{i=1}^{24} X_i < 16\right).$$

Soluzione. Poiché $E[X_i] = 1/2$ e $\mathrm{Var}(X_i) = 1/12$, posto $X = \sum_{i=1}^{24} X_i$ si ha $E[X] = 12$ e $\mathrm{Var}(X) = 2$. Quindi, dal teorema del limite centrale si trae:

$$P(X > 14) = P\left(\frac{X-12}{\sqrt{2}} > \frac{14-12}{\sqrt{2}}\right) \approx 1 - \Phi(1{,}41) = 1 - 0{,}9207 = 0{,}0793$$

ed inoltre

$$\begin{aligned} P(8 < X < 16) &= P\left(\frac{8-12}{\sqrt{2}} \le \frac{X-12}{\sqrt{2}} \le \frac{16-12}{\sqrt{2}}\right) \\ &\approx \Phi(2{,}83) - \Phi(-2{,}83) \\ &= 2\,\Phi(2{,}83) - 1 \\ &= 2 \cdot 0{,}9977 - 1 = 0{,}9954. \end{aligned}$$

Teorema di DeMoivre-Laplace. Sia S_n il numero di successi che si realizzano in n prove indipendenti, in ognuna delle quali il successo ha probabilità p, e sia Z una variabile normale standard. Per $a < b$ si ha:

$$\lim_{n\to+\infty} P\left\{a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right\} = P\,(a \le Z \le b) = \Phi\big(b\big) - \Phi\big(a\big).$$

Il teorema di DeMoivre-Laplace afferma che, se n è grande, la standardizzata di una variabile aleatoria binomiale di parametri n e p ha approssimativamente la stessa distribuzione di una variabile aleatoria normale standard.

Esempio. Sia X il numero di volte che esce testa lanciando $n = 40$ volte una moneta equa ($p = 0{,}5$). Determinare $P(X = 20)$. Ricavarne poi un'approssimazione normale e confrontare il risultato con la soluzione esatta. Cosa cambia se $p = 0{,}4$?

Soluzione. Per $p = 0{,}5$ il risultato esatto è

$$P(X = 20) = \binom{40}{20} (0{,}5)^{40} \simeq 0{,}1254.$$

Usiamo l'approssimazione normale. Mentre la variabile aleatoria binomiale è discreta a valori interi, la normale è una variabile continua ed è essenziale scrivere $P(X = i)$ come $P(i - 1/2 < X < i + 1/2)$ prima di applicare l'approssimazione normale. Si ha:

$$P(X = 20) = P\big(\,19{,}5 \le X < 20{,}5\big) = P\big(\frac{19{,}5 - 20}{\sqrt{10}} < \frac{X - 20}{\sqrt{10}} < \frac{20{,}5 - 20}{\sqrt{10}}\big)$$

$$\simeq P\big(-0{,}16 < \frac{X - 20}{\sqrt{10}} < 0{,}16\big) = \Phi\big(0{,}16\big) - \Phi\big(-0{,}16\big) = 2\,\Phi\big(0{,}16\big) - 1 = 0{,}1272.$$

Per $p = 0{,}4$ il risultato esatto e l'approssimazione normale sono:

$$P(X = 20) = \binom{40}{20} (0{,}4)^{20} (0{,}6)^{20} \simeq 0{,}0554,$$

$$P(X = 20) = P\big(\frac{19{,}5 - 16}{\sqrt{9{,}6}} < \frac{X - 16}{\sqrt{9{,}6}} < \frac{20{,}5 - 16}{\sqrt{9{,}6}}\big) \approx \Phi\big(1{,}45\big) - \Phi\big(1{,}13\big) = 0{,}0557.$$