Legge dei grandi numeri

Come succede frequentemente in statistica, consideriamo una successione $\{X_n\}_{n\in\mathbb{N}}$ di v.a. indipendenti, tutte dotate della stessa legge sconosciuta. Se ne vuole stimare la speranza. A questo problema viene incontro la cosiddetta legge dei grandi numeri.

Prima però occorrono due risultati preliminari.

Teorema (disuguaglianza di Markov) Sia X una v.a. strettamente positiva, e sia $\varepsilon>0$. Allora

$$P(X>\varepsilon)\leq \frac{1}{\varepsilon}\mathbb{E}[X].$$

dim Essendo $X\geqslant \varepsilon I_{\{X>\varepsilon\}}$, per isotonia della speranza si ha

$$\mathbb{E}[X]\geqslant \mathbb{E}[\varepsilon I_{\{X>\varepsilon\}}]=\varepsilon\,\mathbb{E}[I_{\{X>\varepsilon\}}]=\varepsilon P(X>\varepsilon).\ \square$$

Teorema (disuguaglianza di Chebyshev) Sia X una v.a. con speranza μ e varianza σ^2 (finite). Allora per ogni $\varepsilon>0$ si ha

$$P(|X-\mu|\geqslant\varepsilon)\leq\frac{\sigma^2}{\varepsilon^2}.$$

dim. Si ha $P(|X-\mu|\geqslant\varepsilon)=P(|X-\mu|^2\geqslant\varepsilon^2)$ e, per la disuguaglianza di Markov,

$$P(|X-\mu|\geqslant\varepsilon)\leq\frac{1}{\varepsilon^2}\mathbb{E}[(X-\mu)^2]=\frac{1}{\varepsilon}\mathrm{Var}[X]=\frac{\sigma^2}{\varepsilon^2}.\ \square$$

Queste due disuguaglianze sono utili per stimare la probabilità di eventi relativi a v.a. di cui non è nota la legge, ma delle quali è nota la media, oppure sono note media e varianza.

Esempio Una fabbrica produce mediamente 50 pezzi ogni settimana. Si può stimare che occasionalmente la fabbrica superi i 75 pezzi in una settimana?

Se X è la v.a. discreta che descrive i pezzi prodotti per settimana, la legge di X è ignota, ma si sa che $\mathbb{E}[X] = 50$. Vogliamo conoscere $P(X \geq 75)$: usando la disuguaglianza di Markov si ha

$$P(X \geq 75) \leq \frac{1}{75}\,\mathbb{E}[X] = \frac{2}{3}.$$

Dunque la risposta è sì, con buona probabilità.

Osservazione. Se nella disuguaglianza di Chebyshev poniamo $\varepsilon = k\sigma$, con $k>0$, essa si può riscrivere come

$$P(|X-\mu| > k\sigma) \leq \frac{1}{k^2},$$

ovvero: la probabilità che una v.a. differisca dalla sua media per più di k volte la sua deviazione standard non può mai superare il valore $\frac{1}{k^2}$.

Diamo ora una definizione.

Definizione Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di v.a. reali. Diciamo che X_n converge in probabilità verso una v.a. reale X se si ha
$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0 \qquad \forall \varepsilon > 0.$$

Possiamo ora enunciare il teorema principale.

Teorema (legge dei grandi numeri) Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di v.a. indipendenti, tutte dotate della stessa legge. Dette μ e σ^2 la speranza e la varianza (finite) comuni a tutte le X_n, posto $S_n = X_1 + \dots + X_n$, la successione $\left\{\frac{S_n}{n}\right\}_{n\in\mathbb{N}^+}$ converge in probabilità verso la costante μ.

dim. Si ha anzitutto
$$\mathbb{E}\left[\frac{S_n}{n}\right] = \mathbb{E}\left[\frac{X_1 + \dots + X_n}{n}\right] = \mu,$$
$$\mathrm{Var}\left[\frac{S_n}{n}\right] = \frac{1}{n^2}\mathrm{Var}[X_1 + \dots + X_n] = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}[X_k] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Dalla disuguaglianza di Chebyshev segue perciò
$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \le \frac{\sigma^2}{n\varepsilon^2}$$
e dunque per $n\to\infty$ si ha la tesi. $\square$

Esempio Se le X_n sono v.a. discrete bernoulliane, allora S_n è il numero di successi nelle prime n prove. Perciò $\frac{S_n}{n}$ è la frequenza dei successi relativi alle prime n prove. La legge dei grandi numeri ci dice allora che

$$\frac{S_n}{n} \longrightarrow \mu \text{ in probabilità per } n \to \infty.$$

Ma nel caso delle v.a. bernoulliane si ha $\mu = p$, ove p è la probabilità di successo in una prova. Dunque la frequenza dei successi nelle prime n prove tende alla probabilità di successo in una singola prova: $\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) = 0 \quad \forall \varepsilon > 0$.

Esempio (metodo Montecarlo) Sia $f : [0,1] \to \mathbb{R}$ una funzione integrabile secondo Riemann. Se $\{X_n\}_{n\in\mathbb{N}}$ è una successione di v.a. indipendenti, tutte dotate della legge uniforme in $[0,1]$, allora la successione $\{f(X_n)\}$ è ancora formata da v.a. indipendenti, tutte con speranza $\mathbb{E}[f(X_1)]$. Allora, per la legge dei grandi numeri, si ha in probabilità

$$\frac{f(X_1) + \dots + f(X_n)}{n} \longrightarrow \mathbb{E}[f(X_1)] = \int_0^1 f(x)\,dx.$$

Questo fatto suggerisce un metodo di calcolo approssimato per l'integrale di f in $[0,1]$: basta aver a disposizione un generatore casuale di numeri $X_1, \dots, X_n$ con legge uniforme in $[0,1]$, e calcolarne la media. Per n grande, si approssimerà $\int_0^1 f(x)dx$.

Questo metodo di approssimazione, detto metodo Montecarlo non è velocissimo ma è molto affidabile e molto utilizzato nella pratica.

Il teorema certamente più importante del calcolo delle probabilità, quantomeno dal punto di vista delle applicazioni, è certamente il "teorema limite centrale", traduzione grossolana del tedesco "zentraler Grenzwertsatz" (teorema centrale del limite), ormai entrata nell'uso: ci limitiamo ad enunciarlo.

Teorema limite centrale Sia $\{X_n\}_{n\in\mathbb{N}}$ una successione di v.a. indipendenti, tutte dotate della stessa legge, aventi media μ e varianza σ^2 (entrambe finite). Poniamo

$$S_n = X_1 + \dots + X_n, \qquad T_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Detta poi $F_n(x)$ la funzione di ripartizione di T_n e posto

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx,$$

si ha, per ogni $t\in\mathbb{R}$,

$$\lim_{n\to\infty} F_n(t) = \Phi(t).$$

In altre parole, la somma S_n di n v.a. indipendenti di cui si sa solo che hanno tutte media μ e varianza σ^2, per n

grande ha approssimativamente legge normale $N(n\mu, n\sigma^2)$.

Una tipica applicazione del teorema limite centrale è la seguente: volendo calcolare $P(S_n \leq c)$, ove $c \in \mathbb{R}$ e S_n è la somma di n v.a. indipendenti X_j, tutte dotate della stessa legge sconosciuta, con media μ e varianza σ^2 (finita), possiamo approssimare il numero

$$P(S_n \leq c) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{c - n\mu}{\sigma\sqrt{n}}\right)$$

con la quantità

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

Questa approssimazione è detta <u>approssimazione normale</u>, visto che $\Phi(x)$ è la funzione di ripartizione di una v.a con legge $N(0,1)$.

<u>Esempio</u> (Misure di una grandezza fisica). Quando si effettua la misura di una grandezza fisica, il risultato della misura sarà il "valore vero" più un termine casuale (l'errore di misura) che è dovuto alla risultante di molti effetti che perturbano gli strumenti utilizzati e le operazioni di lettura. Il risultato della misura è dunque una v.a.; in assenza di un errore sistematico, si può pensare che tale v.a. sia della forma $\mu + Y$, dove μ è il vero valore, mentre Y denota l'errore assoluto, il quale, secondo

il teorema limite centrale, avrà approssimativamente una legge normale di media 0 (dato che stiamo supponendo che non vi sia un errore sistematico). Dunque il risultato della misura è una v.a. $X=\mu+Y$, ed è naturale supporre che X abbia legge normale $\mathcal{N}(\mu,\sigma^2)$. I parametri μ, σ^2 saranno l'oggetto della ricerca, e ciò sarà il compito della statistica. Il teorema limite centrale giustifica l'idea che un effetto casuale che sia la risultante di molti piccoli effetti, segua necessariamente una legge normale: qui sta la sua fondamentale importanza.

Osservazione. Sia X una v.a. con legge normale $\mathcal{N}(\mu,\sigma^2)$. Supponiamo di dover calcolare le due probabilità

$$P(X\le c), \quad P(a\le X\le b).$$

Osservato che la v.a. $Z=\frac{X-\mu}{\sigma}$ ha legge $\mathcal{N}(0,1)$, si ha

$$P(X\le c)=P\left(Z\le \frac{c-\mu}{\sigma}\right)=\Phi\left(\frac{c-\mu}{\sigma}\right),$$

mentre

$$\begin{aligned} P(a\le X\le b) &= P\left(\frac{a-\mu}{\sigma}\le Z\le \frac{b-\mu}{\sigma}\right)= \\ &= P\left(Z\le \frac{b-\mu}{\sigma}\right)-P\left(Z\le \frac{a-\mu}{\sigma}\right)= \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right), \end{aligned}$$

e i valori richiesti della funzione Φ si ricavano dalla tavola dei valori della funzione Φ.

Tali tavole, tuttavia, riporta solo i valori $\Phi(z)$ per $z>0$: infatti, se $z<0$ si può utilizzare la formula

$$\Phi(-x) = 1-\Phi(x), \quad \forall x\in\mathbb{R},$$

che segue dal fatto, già menzionato, che la densità della legge normale è una funzione pari.

Esempio Sia X una v.a. con legge $\mathcal{N}(3,16)$. Si vuole trovare la probabilità $P(X<11)$. Si ha

$$P(X<11) = P\left(\frac{X-3}{4} < \frac{11-3}{4}\right) = P\left(\frac{X-3}{4} < 2\right) = \Phi(2) \simeq 0.9772.$$

Se invece si vuole calcolare $P(X>-1)$, abbiamo

$$P(X>-1) = P\left(\frac{X-3}{4} > \frac{-1-3}{4}\right) = P\left(\frac{X-3}{4} > -1\right) = P\left(\frac{X-3}{4} < 1\right) =$$
$$= \Phi(1) \simeq 0.8413.$$

Se, infine, vogliamo trovare $P(2<X<7)$, si ha

$$P(2<X<7) = P\left(\frac{2-3}{4} < \frac{X-3}{4} < \frac{7-3}{4}\right) = P\left(-\frac{1}{4} < \frac{X-3}{4} < 1\right) =$$
$$= \Phi(1) - \Phi\left(-\frac{1}{4}\right) = \Phi(1) - [1-\Phi(0,25)] =$$
$$\simeq 0.8413 - 1 + 0.5987 = 0.4400.$$

In molti problemi non si deve calcolare una probabilità del tipo $P(X \leq c)$, ma si chiede l'inverso: fissato $\lambda \in]0,1[$, come si deve scegliere c affinché risulti

$$P(X \leq c) \geq \lambda \ ?$$

A questo scopo è fondamentale la seguente

<u>Definizione</u> Sia X una v.a. reale. Per $\alpha \in]0,1[$, il <u>quantile di ordine α</u> di X è il numero

$$x_\alpha = \inf\{x \in \mathbb{R} : P(X \leq x) \geq \alpha\}.$$

Ricordiamo che $P(X \leq x) = F_X(x)$ è la funzione di ripartizione di X: essa è crescente e continua a destra. Però se X ha una densità f_X continua su $\mathbb{R}$, allora X è diffusa, F_X è strettamente crescente e di classe C^1 e in particolare

$$x_\alpha = \min\{x \in \mathbb{R} : P(X \leq x) \geq \alpha\}$$

cosicché

$$F_X(x_\alpha) = P(X \leq x_\alpha) = \alpha.$$

La risposta alla domanda precedente, "come occorre scegliere c?" è dunque: dev'essere $c \geq x_\lambda$.

I quantili della legge normale ridotta $\mathcal{N}(0,1)$, quelli della legge di Student $t(n)$ e quelli della legge del chi-quadro $\chi^2(n)$ si indicano con i simboli

$$\phi_\alpha\,,\quad t_\alpha(n)\,,\quad \chi^2_\alpha(n)\,.$$

Dunque

$$P(X\le\phi_\alpha)=\alpha,\quad P(Y\le\chi^2_\alpha(n))=\alpha,\quad P(T\le t_\alpha(n))=\alpha$$

ove X, Y, T sono v.a. con leggi $\mathcal{N}(0,1)$, $\chi^2(n)$, $t(n)$.

<u>Osservazione</u> Se X ha legge $\mathcal{N}(0,1)$, sappiamo già che, essendo la densità f_X una funzione pari, si ha, scelto $x=\phi_\alpha$,

$$P(X\le-\phi_\alpha)=P(X\ge\phi_\alpha)=1-P(X\le\phi_\alpha)=1-\alpha$$

Ne segue

$$P(|X|\ge\phi_{1-\alpha/2})=P(X\le-\phi_{1-\alpha/2})+P(X\ge\phi_{1-\alpha/2})=$$

$$=2\left[1-\left(1-\tfrac{\alpha}{2}\right)\right]=2\,\frac{\alpha}{2}=\alpha.$$

Analogamente, se T ha legge di Student $t(n)$,

$$P(T\le-t_\alpha(n))=1-\alpha,\qquad P\left(|T|\ge-t_{1-\frac{\alpha}{2}}(n)\right)=\alpha.$$

<u>Esempi</u> (1) In una fabbrica di lampadine il titolare ha stimato che il tempo medio di vita di una lampadina è di 100 giorni, con una deviazione standard di 30 giorni. Qual è la probabilità che 36 lampadine siano sufficienti per 10 anni?

Consideriamo una sequenza $X_1, \dots, X_{36}$ di $n=36$ v.a. indipendenti, tutte dotate della stessa legge con speranza $\mu = 100$ e deviazione standard $\sigma = 30$, che rappresentano la durata della i-esima lampadina. Posto $S = X_1 + \dots + X_{36}$, S rappresenta la durata del sistema delle 36 lampadine. Dovremo dunque calcolare la probabilità $P(S > 3650)$: 10 anni misurati in giorni. (39)

Per il teorema limite centrale, ossia sfruttando l'approssimazione normale, S avrà approssimativamente legge normale di media $n\mu = 3600$ e deviazione standard $\sqrt{n}\sigma = 180$. Si ha allora

$$P(S > 3650) = P\left(\frac{S-3600}{180} > \frac{3650-3600}{180}\right) = P\left(\frac{S-3600}{180} > 0.28\right) =$$

$$1 - P\left(\frac{S-3600}{180} \le 0.28\right) \cong 1 - \Phi(0.28) = 1 - 0.6103 = 0.3897.$$

[2] Una popolazione contiene in proporzioni uguali due tipi di individui (tipo A e tipo B). Si estrae un campione di 100 individui: qual è la probabilità che il campione contenga almeno 65 individui di tipo B?

Se $X_i = \begin{cases} 1 & \text{se l'i-esimo individuo è di tipo A} \\ 0 & \text{altrimenti,} \end{cases}$ allora X_i è una v.a. bernoulliana con $p = \frac{1}{2}$. Quindi $S = \sum_{i=1}^{100} X_i$ ha legge binomiale $B(100, \frac{1}{2})$. La probabilità richiesta è

$$P(S \geq 65) = \sum_{k=65}^{100} \binom{100}{k} \frac{1}{2^{100}},$$

non facile da calcolare. Possiamo usare la legge di Poisson

$P(np)$, cioè $P(50)$? No, perchè $n=100$ non è eccessivamente grande e $p=\frac{1}{2}$ non è molto piccolo. Usando invece l'approssimazione normale, con $\mu=np=50$, e $\sigma=\sqrt{npq}=$ $=5$, si ha

$$P(S\geq 65)=1-P(S<65)\simeq 1-\Phi\left(\frac{64-50}{5}\right)=1-\Phi(2.8)\simeq$$
$$\simeq 0.0026.$$

Si può anzi scrivere, essendo $S<65$ se e solo se $S<64.5$,

$$1-P(S<65)=1-P(S<64.5)=1-\Phi\left(\frac{64.5-50}{5}\right)=1-\Phi(2.9)=0.00186$$

Si noti che, come si sa dall'esperienza empirica, per una v.a. S a valori in $\mathbb{N}$, se $c\in\mathbb{N}$, si ottiene una miglior approssimazione di $P(S\leq c)$ e $P(S\geq c)$ prendendo invece $P(S\leq c+\frac{1}{2})$ e $P(S\geq c-\frac{1}{2})$: gli eventi sono gli stessi, ma l'approssimazione normale dei secondi è più precisa. Questa scelta si chiama <u>correzione di continuità</u>.

<u>Esercizio</u> Tra i 1400 studenti di una scuola, si è calcolato che all'incirca i $\frac{2}{7}$ del totale consumano il pranzo alla mensa scolastica. Quanti pasti occorre preparare per avere una probabilità del 99% che essi siano sufficienti per tutte le richieste?

Se X è il numero di studenti che va a mensa in un certo giorno, e N è il numero dei pasti da preparare, deve essere $P(X\leq N)=0.99$; X ha legge $B(1400,\frac{2}{7})$, quindi $\mu=np=400$, $\sigma=\sqrt{npq}=\sqrt{285.7}\simeq$

$P(np)$, cioè $P(50)$? No, perché $n=100$ non è eccessivamente grande e $p=\frac{1}{2}$ non è molto piccolo. Usando invece l'approssimazione normale, con $\mu=np=50$, e $\sigma=\sqrt{npq}=$ $=5$, si ha

$$P(S\geq 65)=1-P(S<65)\simeq 1-\Phi\left(\frac{64-50}{5}\right)=1-\Phi(2.8)\simeq$$
$$\simeq 0.0026.$$

Si può anzi scrivere, essendo $S<65$ se e solo se $S<64.5$,

$$1-P(S<65)=1-P(S<64.5)=1-\Phi\left(\frac{64.5-50}{5}\right)=1-\Phi(2.9)=0.00186$$

Si noti che, come si sa dall'esperienza empirica, per una v.a. S a valori in $\mathbb{N}$, se $c\in\mathbb{N}$, si ottiene una miglior approssimazione di $P(S\leq c)$ e $P(S>c)$ prendendo invece $P(S\leq c+\frac{1}{2})$ e $P(S\geq c+\frac{1}{2})$: gli eventi sono gli stessi, ma l'approssimazione normale dei secondi è più precisa. Questa scelta si chiama <u>correzione di continuità</u>.

<u>Esercizio</u> Tra i 1400 studenti di una scuola, si è calcolato che all'incirca i $\frac{2}{7}$ del totale consumano il pranzo alla mensa scolastica. Quanti pasti occorre preparare per avere una probabilità del 99% che essi siano sufficienti per tutte le richieste?

Se X è il numero di studenti che va a mensa in un certo giorno, e N è il numero dei pasti da preparare, deve essere $P(X\leq N)=0.99$; X ha legge $B(1400,\frac{2}{7})$, quindi $\mu=np=400$, $\sigma=\sqrt{npq}=\sqrt{285.7}\simeq$

$\simeq 16.90$. Usando l'approssimazione normale,

$$0.99 = P(X \le N) = P\left(\frac{X-400}{16.90} \le \frac{N-400}{16.90}\right) \simeq \Phi\left(\frac{N-400}{16.90}\right);$$

dunque deve essere $\frac{N-400}{16.90} = \phi_{0.99}$, cioè il quantile di ordine 0.99. Dalle tavole segue

$$\phi_{0.99} = 2.33,$$

quindi dovrà essere

$$N = (2.33)(16.90) + 400 = 439.77 \simeq 440.$$

<u>Esempio</u>. In una certa università, per problemi di capienza delle aule, il numero ideale di studenti per il primo anno di corso è di 150. Sapendo che soltanto il 30% degli studenti ammessi, mediamente, segue le lezioni, l'università adotta la politica di accettare 450 iscrizioni. Qual è la probabilità che più di 150 studenti del 1° anno frequentino le lezioni?

Sia X il numero di studenti che frequentano. Se si ipotizza che ogni studente decida indipendentemente dagli altri di frequentare o no, si può supporre che X abbia legge $B(150, 0.3)$.

Usando l'approssimazione normale e la correzione di continuità,

$$P(X>150) = P(X \geq 150.5) = P\left(\frac{X-450\cdot 0.3}{\sqrt{450\cdot 0.3\cdot 0.7}} > \frac{150.5-135}{\sqrt{94.5}}\right) =$$

$$= P\left(\frac{X-135}{\sqrt{94.5}} > 1.59\right) \simeq 1-\Phi(1.59) \cong 0.0559.$$