Legge dei grandi numeri e approssimazione normale

martedì 26 settembre 2023 16:58

Premesse

Disuguaglianza di Markov

Sia X una v.a. (discreta e continua) a valori non negativi

Allora, per ogni $\alpha > 0$, si ha

$$\frac{E[X]}{\alpha} \geq P(X \geq \alpha)$$

Dim

Sia $\alpha > 0$. Definiamo la v.a.

$$I = \begin{cases} 1 & \text{se } X \geq \alpha \\ 0 & \text{se } X < \alpha \end{cases}$$

$$\Rightarrow \left.\begin{array}{l} \text{Se } X \geq \alpha \Rightarrow \frac{X}{\alpha} \geq 1 = I \geq 0 \\ \text{Se } X < \alpha \Rightarrow \frac{X}{\alpha} \geq 0 = I \end{array}\right\} \Rightarrow 0 \leq I \leq \frac{X}{\alpha}$$

perché $X \geq 0$ per ipotesi

Se $E[X]$ è infinita $\Rightarrow$ La tesi è vera

Supp. $E[X]$ finita

$$E[I] = 0 \cdot P(I=0) + 1 \cdot P(I=1) = P(I=1)$$
$$= P(X \geq \alpha) \qquad (1)$$

Ma poiché $0 \leq I \leq \frac{X}{\alpha}$

$$\Rightarrow E[I] \leq E\left[\frac{X}{\alpha}\right] = \frac{E[X]}{\alpha} \qquad (2)$$

Da (1) e (2) segue che

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

$\square$

<u>Disuguaglianza di Chebishev</u>

Sia X una v.a. (discreta o continua) che ammette media e varianza finite.
Allora, per ogni $\eta > 0$, si ha

$$P(|X - E[X]| \geq \eta) \leq \frac{VAR(X)}{\eta^2}$$

<u>Dim</u>

Introduciamo la v.a. $Y = (X - E[X])^2$ che è a valori non negativi.
Ora che

$$E[Y] = E[(X - E[X])^2] = VAR(X)$$

Inoltre, abbiamo che

$$P(|X - E[X]| \geq \eta) = P(|X - E[X]|^2 \geq \eta^2)$$

$$= P(Y \geq \eta^2) \leq \frac{E[Y]}{\eta^2} = \frac{VAR(X)}{\eta^2}$$

Per la disuguaglianza di Markov

$\square$

Vedremo alcuni teoremi utili nelle applicazioni.

$\boxed{\text{Def}}$ Sia $(X_n)_{n\in\mathbb{N}}$ una successione di v.a. def. su uno stesso spazio di probabilità $(\Omega, \mathcal{A}, P)$ e sia X una v.a. def. su $(\Omega, \mathcal{A}, P)$. Diremo che $(X_n)_{n\in\mathbb{N}}$ converge in probabilità a X se per ogni $\eta > 0$

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

e in tal caso scriveremo $X_n \xrightarrow{P} X$

Oss

Intuitivamente significa che i valori assunti da X_n hanno sempre meno probabilità di discostarsi da quelli di X al crescere di n.

Lemma

Sia $(X_n)_{n\in\mathbb{N}}$ una successione di v.a. indipendenti def. su uno stesso spazio di probabilità, aventi media μ e varianza $\sigma^2 > 0$ finite.

Definite, per ogni $n \in \mathbb{N}$, le v.a.

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \dots + X_n}{n}$$

si ha che

$$E[\overline{X}_n] = \mu \quad , \quad VAR(\overline{X}_n) = \frac{\sigma^2}{n}$$

Dim

$$E[\overline{X}_n] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{\cancel{n}}\cdot\cancel{n}\mu = \mu$$

Poiché le X_i sono indipendenti, si ha

$$VAR(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} VAR(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Teorema (Legge dei grandi numeri in forma debole)

Sia $(X_n)_{n\in\mathbb{N}}$ una succ. di v.a. indipendenti definite su di uno stesso spazio di probabilità, aventi media μ e varianza $\sigma^2 > 0$ finite. Allora

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| > \eta) = 0 \qquad \forall \eta > 0$$

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

In altre parole, la tesi è equivalente a

$$\overline{X}_n \xrightarrow{P} \mu$$

Dim

Per la disuguaglianza di Chebishev, fissato ad arbitrio $\eta > 0$, si ha

$$P\left(\left|\overline{X}_n - E[\overline{X}_n]\right| > \eta\right) \leq \frac{VAR(\overline{X}_n)}{\eta^2}$$

$$\Rightarrow \quad 0 \leq P\left(\left|\overline{X}_n - \mu\right| > \eta\right) \leq \frac{\sigma^2}{n\eta^2} \xrightarrow[n\to\infty]{} 0$$

$$\Rightarrow \quad P\left(\left|\overline{X}_n - \mu\right| > \eta\right) \xrightarrow[n\to\infty]{} 0$$

$\square$

<u>Esempio</u>

Un numero al lotto non esce da 324 estrazioni. Questo giustifica l'asserzione che tale ritardo dovrebbe comportare che il numero esce nelle imminenti estrazioni altrimenti si avrebbe una violazione della legge dei grandi numeri?

\- - - -

Per la legge dei grandi numeri le frequenze si devono stabilizzare.

Allora 342 volte crea un forte squilibrio?

Vediamo che questo numero non è poi così grande.

Calcoliamo la probabilità di estrazione.

Il numero di cinquine "buone" segue un'ipergeometrica

$$\Rightarrow \binom{1}{1}\binom{89}{4}$$

Il numero di cinquine totali è $\binom{90}{5}$

$$\Rightarrow p = \frac{\binom{1}{1}\binom{89}{4}}{\binom{90}{5}} = \frac{1}{18}$$

Abbiamo un campione di 324 estrazioni.
Se consideriamo "successo" l'estrazione del nostro numero
$\Rightarrow \overline{X}_{324} = 0$

$$\Rightarrow \left| \overline{X}_{324} - p \right| = \left| 0 - \frac{1}{18} \right| = \frac{1}{18} = \frac{1}{\sqrt{324}} = \frac{1}{\sqrt{n}} \xrightarrow[n \to \infty]{} 0$$

ciò è in accordo con la legge dei grandi numeri.

L'errore che si commette è che non si considera $\overline{X}_n$ bensì $\sum X_i = n \overline{X}_n$ perché dico "nelle ultime n estrazioni il numero non è uscito".

$$\Rightarrow \left| n \overline{X}_n - np \right| = \frac{n}{\sqrt{n}} = \sqrt{n}$$

e ciò può apparire in contrasto con la legge dei grandi numeri ma in realtà è una applicazione errata perché essa si applica a $\overline{X}_n$

Curiosità: Il numero più ritardatario della storia del Gioco del Lotto è il numero 53 della ruota Nazionale con un ritardo di 257 estrazioni.

Def Sia $(X_n)_{n\in\mathbb{N}}$ una succ di v.a. e sia $F_i(t)$ la f.r. di X_i.
Sia X una v.a. di f.r. $F(t)$
Diremo che X_n converge in legge a X
se $\lim_{n\to\infty} F_n(t) = F(t)$ $\forall t \in \mathbb{R}$ ove F è continua

In tal caso scriveremo $X_n \xrightarrow{\mathcal{L}} X$.

Teorema del limite centrale

Sia $(X_n)_{n\in\mathbb{N}}$ una succ. di v.a. indipendenti, equidistribuite (ossia di medesima legge), di media μ e varianza $\sigma^2 > 0$ finite.
Allora la variabile normale standardizzata, def. da

$$S_n^* = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n}$$

con $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, tende in legge ad una variabile $S \sim N(0,1)$

Oss

Nella pratica per $n \gg 1$ si approssima S_n^* con S di legge $N(0,1)$
Un criterio empirico per l'applicabilità di tale approssimazione, che viene detta approssimazione normale, è che n sia almeno 30.

Oss

Nel caso particolare di leggi binomiali $B(n,p)$ un criterio empirico è $np \geq 5$ e $n(1-p) \geq 5$

Esempio

Si consideri un esame di ammissione composto da 10 domande, ciascuna con 3 risposte di cui una sola è esatta.

Per essere ammessi occorre fornire almeno il 70% di risposte esatte.

1) Qual è la probabilità di superare il test rispondendo a caso?

2) Qual è la probabilità di superare il test rispondendo a caso se le domande fossero 100?

---

1) Sia X la v.a. che conta il numero di risposte esatte (successi in 10 prove)

Rispondendo a caso, la probabilità di successo è $\frac{1}{3}$

$$\Rightarrow X \sim B\left(10, \frac{1}{3}\right)$$

$$\Rightarrow p(X \geq 7) = \sum_{k=7}^{10} \binom{10}{k} \left(\frac{1}{3}\right)^{k} \left(\frac{2}{3}\right)^{10-k} \approx 1\%$$

2) Se le domande fossero 100 si potrebbe ripetere il ragionamento ma il calcolo del coefficiente binomiale sarebbe complicato.

Applichiamo la regola empirica

$$\begin{cases} np = 100 \cdot \frac{1}{3} > 5 \\ n(1-p) = 100 \cdot \frac{2}{3} > 5 \end{cases} \implies \text{ok!}$$

È possibile usare l'approssimazione normale

$$P(X \geq 70) \underset{\substack{\uparrow \\ \text{correzione per continuità}}}{=} P(X \geq 69.5) = 1 - P(X < 69.5)$$

$$= 1 - \phi\left(\frac{69.5 - \frac{100}{3}}{\sqrt{\frac{1}{3} \cdot \frac{2}{3}}\sqrt{100}}\right)$$

$$\approx 1 - \underbrace{\phi(7.67)}_{\approx 1} \approx 1 - 1 = 0$$

per la regola del 3σ