CAPITOLO 5: CONVERGENZE E APPROSSIMAZIONI

LEGGE DEI GRANDI NUMERI (LEGGE DEBOLE)

Sia $\{X_n : n \geq 1\}$ una successione di v.a. i.i.d. (cioè indipendenti e identicamente distribuite).
Supponiamo che tutte abbiano speranza matematica finita; allora, poiché sono identicamente distribuite, hanno tutte le stesse speranze matematiche e indicheremo questo valore comune con μ.

Allora si ha

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

dove

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

è la media aritmetica delle prime n v.a., cioè $X_1, \ldots, X_n$.

NOTAZIONE CHE USEREMO ANCHE IN SEGUITO

LEGGE FORTE

Nelle stesse ipotesi si ha

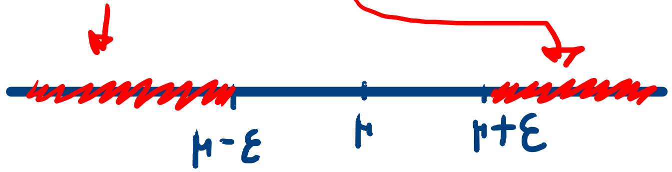
$$P(\underbrace{\{\omega \in \Omega : \lim_{n \to \infty} \bar{X}_n(\omega) = \mu\}}) = 1$$

In effetti si dimostra che questo sottoinsieme di Ω è un evento (va oltre gli obiettivi del corso).

COMMENTO

Quindi, anche se $\varepsilon > 0$ è molto vicino a zero, la probabilità che la v.a. $\overline{X}_n = \dfrac{X_1 + \dots + X_n}{n}$ assuma valori nell'insieme segnato

e quindi $(\mu-\varepsilon, \mu+\varepsilon)$ è molto piccolo



tende a zero per $n \to \infty$.

Dimostriamo la legge debole dei grandi numeri con l'ulteriore ipotesi restrittiva di v.a. $\{X_n : n \geq 1\}$ con varianze finite. Poiché le v.a. $\{X_n : n \geq 1\}$ sono identicamente distribuite, hanno tutte la stessa varianza e indicheremo questo valore comune con σ^2.

Si usa la disuguaglianza di Chebishev. Per fare questo osserviamo che:

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{X_1+\dots+X_n}{n}\right] \underset{\text{linearità}}{=} \frac{\mathbb{E}[X_1]+\dots+\mathbb{E}[X_n]}{n} = \frac{\overbrace{\mu+\dots+\mu}^{n \text{ volte}}}{n} = \frac{\cancel{n}\mu}{\cancel{n}} = \mu$$

$$\mathrm{Var}[\bar{X}_n] = \mathrm{Var}\left[\frac{X_1+\dots+X_n}{n}\right] =$$

$$= \frac{1}{n^2}\mathrm{Var}[X_1+\dots+X_n] \overset{\text{indip.}}{=} \frac{1}{n^2}\left\{\mathrm{Var}[X_1]+\dots+\mathrm{Var}[X_n]\right\} = \frac{1}{n^2}\{\underbrace{\sigma^2+\dots+\sigma^2}_{n \text{ volte}}\} = \frac{1}{n^{\cancel{2}}}\cdot\cancel{n}\sigma^2 = \frac{\sigma^2}{n}$$

Allora $\forall \varepsilon > 0$

$$P(|\bar{X}_n - \mathbb{E}[\bar{X}_n]| \geq \varepsilon) \leq \frac{\mathrm{Var}[\bar{X}_n]}{\varepsilon^2} \implies P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2/n}{\varepsilon^2}$$

$\boxed{\text{DIS. DI CHEBYSHEV PER } \bar{X}_n}$

$$\implies 0 \leq P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \underbrace{\frac{\sigma^2}{n\varepsilon^2}}_{\to 0 \text{ per } n\to\infty}.$$

In conclusione, per il Teorema sul confronto sui limiti, si ha

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0.$$

Con questo risultato si possono risolvere i seguenti esercizi.

1) Sia $\{X_n : n \geq 1\}$ una successione di v.a., e in generale X_n indica il numero che nel lancio n-simo del dado. Sia $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$. Dire per quale valore di μ si ha

$$\lim_{n \to \infty} P\left(|\bar{X}_n - \mu| \geq \varepsilon\right) = 0 \qquad \forall \varepsilon > 0.$$

RISPOSTA. Il valore di μ richiesto è $\mu = \sum_{k=1}^{6} k\, P(X_i = k) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = \frac{7}{2}$.

2) Sia $\{N(t) : t \geq 0\}$, dove $N(t) = \sum_{n \geq 1} 1_{\{T_n \leq t\}}$, un Processo di Poisson di intensità $\lambda > 0$.
Dire per quale valore di μ si ha

$$\lim_{n \to \infty} P\left(\left|\frac{T_n}{n} - \mu\right| \geq \varepsilon\right) = 0 \qquad \forall \varepsilon > 0.$$

RISPOSTA: In generale si ha $T_n = S_1 + \dots + S_n$ dove $\{S_n : n \geq 1\}$ sono v.a. i.i.d. $Exp(\lambda)$.
Quindi possiamo fare riferimento alla legge dei grandi numeri per $\frac{T_n}{n} = \frac{S_1 + \dots + S_n}{n} = \bar{S}_n$, e in corrispondenza si ha

$$\mu = \mathbb{E}[S_n] = \int_0^\infty x \lambda e^{-\lambda x}\, dx = \underset{\text{calcoli già fatti}}{\dots} = \frac{1}{\lambda}.$$

TEOREMA LIMITE CENTRALE

Sia $\{X_n : n \geq 1\}$ una successione di v.a. i.i.d. con media comune finita μ, e varianza comune finita σ^2. Inoltre supponiamo che $\sigma^2 > 0$, e quindi le v.a. $\{X_n : n \geq 1\}$ non sono costanti. Allora

$$\lim_{n\to\infty} P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \qquad \forall x \in \mathbb{R}.$$

oss. Qui si può anche avere $<$

Inoltre si capisce l'importanza della Normale standard a cui deve fare riferimento qualunque sia la distribuzione delle v.a. $\{X_n : n \geq 1\}$

COMMENTI

- Il termine "centrale" è da considerare come sinonimo di "importante".
- $\mathbb{E}[X_1 + \dots + X_n] = n\mu$ e $\mathrm{Var}[X_1 + \dots + X_n] = n\sigma^2$ per le ipotesi; quindi

$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

è una v.a. standardizzata (cioè di media zero e varianza 1).

- Nel caso particolare in cui $\{X_n : n \geq 1\}$ sono v.a. $N(\mu, \sigma^2)$, per quanto abbiamo detto sulle combinazioni lineari di Normali indipendenti il risultato è vero per ogni n (e non solo come limite per $n \to \infty$), cioè

$$P\left(\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \qquad \forall x \in \mathbb{R} \text{ e } \forall n \geq 1.$$

FORMULAZIONE DEL TEOREMA "CON LE MEDIE AL POSTO DELLE SOMME".

Nelle ipotesi del TLC abbiamo detto che si ha

$$\lim_{n\to\infty} P\left(\frac{\overbrace{X_1+\dots+X_n}^{\text{somme}} - n\mu}{\sigma\sqrt{n}} \le x\right) = \Phi(x) \qquad \forall x\in\mathbb{R}.$$

Quindi, se dividiamo numeratore e denominatore per n, si ha

$$\frac{X_1+\dots+X_n-n\mu}{\sigma\sqrt{n}} = \frac{\frac{X_1+\dots+X_n}{n} - \frac{\cancel{n}\mu}{\cancel{n}}}{\frac{\sigma\cancel{\sqrt{n}}}{\cancel{n}\sqrt{n}}} = \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$$

questa è la media che abbiamo indicato con $\bar{X}_n$

In conclusione si ha lo stesso evento (e quindi la stessa probabilità) espresso con le medie al posto delle somme e quindi

$$\lim_{n\to\infty} P\left(\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \le x\right) = \Phi(x) \qquad \forall x\in\mathbb{R}.$$

Altre formulazioni (per le somme) equivalenti e utili per gli esercizi:

$$\lim_{n\to\infty} P\left(\frac{X_1+\dots+X_n-n\mu}{\sigma\sqrt{n}} \geq x\right) = 1-\Phi(x) \qquad \forall x\in\mathbb{R}$$

$$\lim_{n\to\infty} P\left(a \leq \frac{X_1+\dots+X_n-n\mu}{\sigma\sqrt{n}} \leq b\right) = \Phi(b)-\Phi(a) \qquad \forall a,b\in\mathbb{R} \text{ t.c. } a<b.$$

<u>OSS.</u>

Anche qui le disuguaglianze negli eventi possono essere strette ("<" al posto di "≤" e/o ">" al posto di "≥").

<u>OSS.</u>

Formulazioni analoghe possono essere fatte "con le medie al posto delle somme" anche in questo caso (si usa lo stesso trucco di dividere numeratore e denominatore per n).

ESERCIZIO

Sia $\{X_n : n \geq 1\}$ una successione di v.a. i.i.d. con distribuzione $U(0, 2a)$, per $a > 0$.

Calcolare

$$\lim_{n \to \infty} P(X_1 + \dots + X_n > na + x\sqrt{n}), \text{ per ogni } x \in \mathbb{R}.$$

Inoltre trovare x (in funzione di a) affinché tale limite sia uguale a $1 - \Phi(1/2)$.

SVOLGIMENTO.

Per le notazioni introdotte, e per quanto sappiamo sulla distribuzione uniforme, si ha

$$\mu = \mathbb{E}[X_n] = \frac{0+2a}{2} = \frac{\cancel{2}a}{\cancel{2}} = a$$

$$\sigma^2 = \mathrm{Var}[X_n] = \frac{(2a-0)^2}{12} = \frac{\cancel{4}a^2}{\cancel{12}_3} = \frac{a^2}{3}$$

$$\Rightarrow \sigma = \frac{a}{\sqrt{3}}.$$

Quindi il limite richiesto si calcola come segue:

$$P(X_1 + \dots + X_n > na + x\sqrt{n}) = P\left(\frac{X_1 + \dots + X_n - na}{\sqrt{n}} > x\right)$$

$$= P\left(\frac{X_1 + \dots + X_n - na}{\frac{a}{\sqrt{3}}\sqrt{n}} > \frac{x}{a/\sqrt{3}}\right) =$$

$$= P\left(\frac{X_1 + \dots + X_n - na}{\frac{a}{\sqrt{3}}\sqrt{n}} > \sqrt{3}\frac{x}{a}\right) \xrightarrow[n \to \infty]{\text{per il TLC}} 1 - \Phi\left(\sqrt{3}\frac{x}{a}\right).$$

Infine si deve avere $\sqrt{3}\frac{x}{a} = \frac{1}{2} \Rightarrow x = \frac{1}{2}\frac{a}{\sqrt{3}} = \frac{a}{2\sqrt{3}}$.

ESERCIZIO

Sia $\{X_n : n \geq 1\}$ una successione di v.a. i.i.d.

Supponiamo che la loro media comune sia uguale a 1, e che la loro varianza comune sia uguale a 16.

1) Calcolare

$$\lim_{n\to\infty} P\left(\frac{1}{2} \leq \frac{X_1 + \cdots + X_n - n}{\sqrt{n}} \leq 1\right).$$

2) Trovare il valore di $y \in \mathbb{R}$ tale che

$$\lim_{n\to\infty} P\left(\frac{X_1 + \cdots + X_n - n}{\sqrt{n}} \geq y\right) = \Phi(-2).$$

SVOLGIMENTO

Si deve fare riferimento al TLC con $\mu = 1$ e $\sigma = \sqrt{16} = 4$. Quindi

1) $P\left(\frac{1}{2} \le \frac{X_1 + \cdots + X_n - n}{\sqrt{n}} \le 1\right) = P\left(\frac{1}{2} \le \frac{X_1 + \cdots + X_n - n \cdot 1}{\sqrt{n}} \le 1\right)$

$= P\left(\frac{1}{4 \cdot 2} \le \underbrace{\frac{X_1 + \cdots + X_n - n \cdot 1}{4 \cdot \sqrt{n}}}_{= \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}} \le \frac{1}{4}\right) \xrightarrow[n \to \infty]{} \Phi\left(\frac{1}{4}\right) - \Phi\left(\frac{1}{8}\right)$

per il TLC

2) $P\left(\frac{X_1 + \cdots + X_n - n}{\sqrt{n}} \ge y\right) = P\left(\frac{X_1 + \cdots + X_n - n \cdot 1}{\sqrt{n}} \ge y\right) = P\left(\underbrace{\frac{X_1 + \cdots + X_n - n \cdot 1}{4\sqrt{n}}}_{= \frac{X_1 + \cdots + X_n - n\mu}{6\sqrt{n}}} \ge \frac{y}{4}\right)$

$\xrightarrow[n \to \infty]{} 1 - \Phi\left(\frac{y}{4}\right).$

per il TLC

Allora si deve avere $1 - \Phi\left(\frac{y}{4}\right) = \Phi(-2)$, da cui segue

$\cancel{1} - \Phi\left(\frac{y}{4}\right) = \cancel{1} - \Phi(2)$, $\Phi\left(\frac{y}{4}\right) = \Phi(2)$, ($\Phi$ è strett. crescente e quindi è invertibile) $\frac{y}{4} = 2$, $y = 8$

ESERCIZIO

Sia $\{X_n : n \geq 1\}$ una successione di v.a. i.i.d. $\sim Exp(\lambda = 4)$.

1) Calcolare

$$\lim_{n\to\infty} P\left(-\frac{1}{8} \leq \frac{X_1 + \dots + X_n - \frac{n}{4}}{\sqrt{n}} \leq \frac{1}{16}\right)$$

esprimendo il risultato con Φ con argomento positivo.

2) Trovare il valore di $z>0$ tale che

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \dots + X_n - \frac{n}{4}}{\sqrt{n}}\right| \leq \frac{2}{3}\right) = 2\Phi(z) - 1.$$

SVOLGIMENTO

Per quanto visto in lezioni precedenti si deve fare riferimento al TLC con $\mu = \frac{1}{\lambda}$ e $\sigma^2 = \frac{1}{\lambda^2}$

$\Updownarrow$

$\sigma = \frac{1}{\lambda}$.

Quindi nel caso specifico $\mu = \frac{1}{4}$ e $\sigma = \frac{1}{4}$. Allora si ha

1) $$P\left(-\frac{1}{8} \le \frac{X_1+\dots+X_n-\frac{n}{4}}{\sqrt{n}} < \frac{1}{16}\right) = P\left(\frac{-1/8}{1/4} \le \underbrace{\frac{X_1+\dots+X_n-\frac{n}{4}}{\frac{1}{4}\sqrt{n}}}_{=\frac{X_1+\dots+X_n-n\mu}{\sigma\sqrt{n}}} \le \frac{1/16}{1/4}\right)$$

$$\xrightarrow[n\to\infty]{} \Phi\left(\frac{1/16}{1/4}\right) - \Phi\left(-\frac{1/8}{1/4}\right) =$$

per il TLC

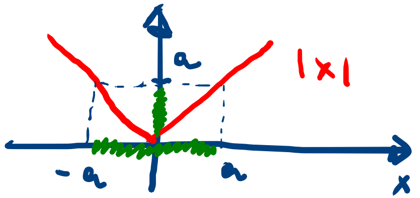
$$= \Phi\left(\frac{4}{16}\right) - \Phi\left(-\frac{4}{8}\right) = \Phi\left(\frac{1}{4}\right) - \Phi\left(-\frac{1}{2}\right) =$$

bisogna avere Φ con argomenti positivi.....

$$= \Phi\left(\frac{1}{4}\right) - \left(1 - \Phi\left(\frac{1}{2}\right)\right) = \Phi\left(\frac{1}{4}\right) + \Phi\left(\frac{1}{2}\right) - 1.$$

2) $$P\left(\left|\frac{X_1+\dots+X_n-\frac{n}{4}}{\sqrt{n}}\right| \le \frac{2}{3}\right) = P\left(-\frac{2}{3} \le \frac{X_1+\dots+X_n-\frac{n}{4}}{\sqrt{n}} \le \frac{2}{3}\right) =$$

$|x| \le a \iff -a \le x \le a$
$(a>0)$



$$= P\left(\frac{-\frac{2}{3}}{\frac{1}{4}} \le \underbrace{\frac{X_1+\dots+X_n-n\cdot\frac{1}{4}}{\frac{1}{4}\sqrt{n}}}_{=\frac{X_1+\dots+X_n-n\mu}{\sigma\sqrt{n}}} \le \frac{2/3}{1/4}\right) \xrightarrow[n\to\infty]{} \Phi\left(\frac{2/3}{1/4}\right) - \Phi\left(-\frac{2/3}{1/4}\right)$$

per il TLC

$$= \Phi\left(\frac{8}{3}\right) - \Phi\left(-\frac{8}{3}\right)$$

Quindi dobbiamo trovare il valore di z per cui si ha

$$\Phi\left(\frac{8}{3}\right) - \Phi\left(-\frac{8}{3}\right) = 2\,\Phi(z) - 1 .$$

Procediamo con alcuni calcoli per il 1° membro:

$$\Phi\left(\frac{8}{3}\right) - \Phi\left(-\frac{8}{3}\right) = \Phi\left(\frac{8}{3}\right) - \left(1 - \Phi\left(\frac{8}{3}\right)\right) = \Phi\left(\frac{8}{3}\right) - 1 + \Phi\left(\frac{8}{3}\right) = 2\,\Phi\left(\frac{8}{3}\right) - 1 .$$

Allora si deve avere

$$2\,\Phi\left(\frac{8}{3}\right) - \cancel{1} = 2\,\Phi(z) - \cancel{1} , \quad \cancel{2}\,\Phi\left(\frac{8}{3}\right) = \cancel{2}\,\Phi(z) ,$$

$$\Phi\left(\frac{8}{3}\right) = \Phi(z)$$

ed essendo Φ strettamente crescente e quindi invertibile

$$\boxed{\frac{8}{3} = z}$$

APPROSSIMAZIONE NORMALE

Vogliamo presentare un'applicazione del TLC per calcolare in maniera approssimata le probabilità di certi eventi.

Siano $X_1, \dots, X_n$ v.a. i.i.d. con media μ e varianza σ^2 finite (supponiamo che $\sigma^2 > 0$ per evitare casi banali). Allora, se n è grande, si può supporre che:

→ $X_1 + \dots + X_n$ ha distribuzione $N(n\mu, n\sigma^2)$;

→ $\overline{X}_n = \dfrac{X_1 + \dots + X_n}{n}$ ha distribuzione $N\left(\mu, \dfrac{\sigma^2}{n}\right)$.

Queste due affermazioni non sono esatte ma sono approssimazioni.

Solo in un caso le affermazioni "→" sono esatte; è il caso in cui le v.a. $X_1, \dots, X_n$ hanno distribuzione $N(\mu, \sigma^2)$. In tal caso è vera per ogni $n \geq 1$ (anche non grande).

ESERCIZIO

Un calcolatore somma un milione di numeri e, ad ogni somma, genera un errore aleatorio con distribuzione $U(-0{,}5\cdot 10^{-10}, 0{,}5\cdot 10^{-10})$. Gli errori sono indipendenti. Calcolare la probabilità che l'errore totale sia in valore assoluto sia minore di $0{,}5\cdot 10^{-7}$ sfruttando l'approssimazione Normale.

RISPOSTA

L'errore totale è la somma $X_1 + \dots + X_{10^6}$ di v.a. i.i.d. con distribuzione uniforme indicata. Si ha

$$P\left(|X_1 + \dots + X_{10^6}| < 0.5\cdot 10^{-7}\right) = P\left(-0.5\cdot 10^{-7} < X_1 + \dots + X_{10^6} < 0.5\cdot 10^{-7}\right).$$

Sfrutteremo l'approssimazione Normale di $X_1 + \dots + X_{10^6}$ perché 10^6 è grande.

(il numero degli addendi).

Quindi $X_1 + \dots + X_{10^6}$ ha (approssimativamente) distribuzione $N(10^6\mu, 10^6\sigma^2)$

dove (per quanto abbiamo visto in lezioni passate sulla distribuzione uniforme) si ha

$$\mu = \frac{-0.5\cdot 10^{-10} + 0.5\cdot 10^{-10}}{2} = 0 \qquad \sigma^2 = \frac{\left(0.5\cdot 10^{-10} - (-0.5\cdot 10^{-10})\right)^2}{12} = \frac{(10^{-10})^2}{12}$$

$$\Rightarrow \sigma = \sqrt{\frac{(10^{-10})^2}{12}} = \frac{10^{-10}}{\sqrt{12}} = \frac{10^{-10}}{2\sqrt{3}}$$

Quindi

$$P\left(-0.5\cdot 10^{-7} < X_1 + \dots + X_{10^6} < 0.5\cdot 10^{-7}\right) =$$

$$= P\left(\frac{-0.5\cdot 10^{-7} - 10^6\cdot 0}{\frac{10^{-10}}{2\sqrt{3}}\sqrt{10^6}} < \boxed{\frac{X_1 + \dots + X_{10^6} - 10^6\cdot 0}{\frac{10^{-10}}{2\sqrt{3}}\sqrt{10^6}}} < \frac{0.5\cdot 10^{-7} - 10^6\cdot 0}{\frac{10^{-10}}{2\sqrt{3}}\sqrt{10^6}}\right) =$$

$$10^{-10}\cdot\sqrt{10^6} = 10^{-10}\cdot 10^3 = 10^{-7}$$

APPROSSIMAZIONE NORMALE QUI

$$= P\left(\frac{-0.5\cdot \cancel{10^{-7}}}{\frac{\cancel{10^{-7}}}{2\sqrt{3}}} < \dots < \frac{0.5\cdot \cancel{10^{-7}}}{\frac{\cancel{10^{-7}}}{2\sqrt{3}}}\right) = P\left(-\underbrace{0.5\cdot 2}_{=1}\sqrt{3} < \dots < \underbrace{0.5\cdot 2}_{=1}\cdot\sqrt{3}\right) \simeq$$

$$= \Phi(\sqrt{3}) - \Phi(-\sqrt{3}) = \Phi(\sqrt{3}) - (1 - \Phi(\sqrt{3})) = 2\,\Phi(\sqrt{3}) - 1 = 2\cdot 0.95819 - 1 = 0.91638.$$

con le tavole

ESERCIZIO

Siano $X_1, \dots, X_{100}$ v.a. i.i.d. con distribuzione Gamma di parametri $\alpha = 2$, $\beta = 4$.

Calcolare $P(X_1 + \dots + X_{100} > 60)$ e $P(42 < X_1 + \dots + X_{100} < 48)$ sfruttando l'approssimazione Normale.

RISPOSTA

Per l'approssimazione Normale (il numero di addendi 100 è grande) possiamo dire che

$X_1 + \dots + X_{100}$ ha (approssimativamente) distribuzione $N(100 \cdot \mu, 100 \cdot \sigma^2)$

dove, per quanto visto in passato sulla distribuzione Gamma, si ha

$$\mu = \frac{\alpha}{\beta} = \frac{2}{4} = \frac{1}{2} \quad e \quad \sigma^2 = \frac{\alpha}{\beta^2} = \frac{2}{4^2} = \frac{2}{16} = \frac{1}{8} \Rightarrow \sigma = \sqrt{\frac{1}{8}} = \frac{1}{\sqrt{8}} = \frac{1}{2\sqrt{2}}.$$

Quindi

1) $P(X_1 + \dots + X_{100} > 60) = P\left(\frac{X_1 + \dots + X_{100} - 100 \cdot \mu}{\sigma \sqrt{100}} > \underbrace{\frac{60 - 100 \cdot \mu}{\sigma\sqrt{100}}} \right) \cong$ APPROSSIMAZIONE NORMALE QUI

$= \frac{60 - 100 \cdot \frac{1}{2}}{\frac{1}{2\sqrt{2}}\sqrt{100}} = \frac{60-50}{\frac{1}{2\sqrt{2}} \cdot 10} = \frac{\cancel{10}}{\cancel{10}} 2\sqrt{2} = 2\sqrt{2}$

$\cong 1 - \Phi(2\sqrt{2})$.

2) $P(42 < X_1 + \dots + X_{100} < 48) = P\left(\underbrace{\frac{42 - 100\cdot\mu}{\sigma\sqrt{100}}} < \frac{X_1 + \dots + X_{100} - 100 \cdot \mu}{\sigma\sqrt{100}} < \underbrace{\frac{48 - 100 \cdot \mu}{\sigma \sqrt{100}}} \right) \cong$

$= \frac{42 - 100 \cdot \frac{1}{2}}{\frac{1}{2\sqrt{2}} \cdot \sqrt{100}} = \frac{42-50}{\frac{1}{2\sqrt{2}} \cdot 10} = -\frac{8}{\cancel{10}_5} \cancel{2}\sqrt{2} = -\frac{8}{5}\sqrt{2}$

$= \frac{48 - 100 \cdot \frac{1}{2}}{\frac{1}{2\sqrt{2}} \cdot \sqrt{100}} = \frac{48-50}{\frac{1}{2\sqrt{2}} \cdot 10} = -\frac{2}{\cancel{10}_5} \cdot \cancel{2}\sqrt{2} = -\frac{2}{5}\sqrt{2}$

$\cong \Phi\left(-\frac{2}{5}\sqrt{2}\right) - \Phi\left(-\frac{8}{5}\sqrt{2}\right) =$ se si vogliono avere argomenti positivi

$= 1 - \Phi\left(\frac{2}{5}\sqrt{2}\right) - \left(1 - \Phi\left(\frac{8}{5}\sqrt{2}\right)\right) = \cancel{1} - \Phi\left(\frac{2}{5}\sqrt{2}\right) - \cancel{1} + \Phi\left(\frac{8}{5}\sqrt{2}\right) = \Phi\left(\frac{8}{5}\sqrt{2}\right) - \Phi\left(\frac{2}{5}\sqrt{2}\right)$

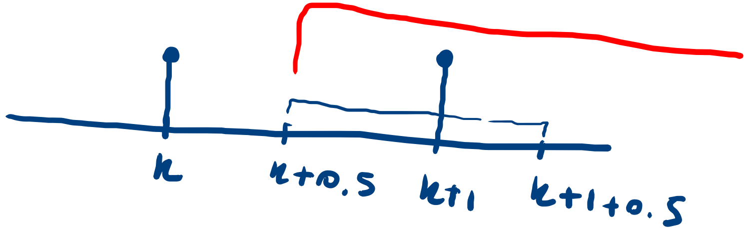
CORREZIONE DI CONTINUITÀ

Si tratta di una correzione che si fa quando si considera l'approssimazione Normale con variabili aleatorie $X_1, \dots, X_n$ (con n grande) $\boxed{\text{a valori interi}}$.

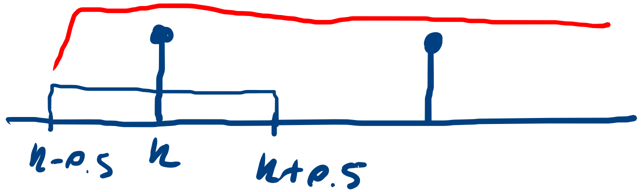
In questo caso al valore intero k corrisponde l'intervallo $(k-0.5, k+0.5)$.

ESEMPI

1) $\{X_1 + \dots + X_n > k\} = \{X_1 + \dots + X_n \geq k+1\} =$
$= \{X_1 + \dots + X_n > k + 0.5\}$



2) $\{X_1 + \dots + X_n \geq k\} = \{X_1 + \dots + X_n > k - 0.5\}$



Si possono fare ragionamenti analoghi per eventi con le disuguaglianze inverse, o con intervalli limitati.

ESERCIZIO

Si lancia 900 volte un dado equo.

1) Calcolare la probabilità che esca almeno 180 volte il numero 6.
2) Calcolare la stessa probabilità nel caso di dado truccato per il quale la probabilità che esca 6 in ogni lancio è $\frac{2}{9}$ (anziché $\frac{1}{6}$; quindi un po' più alta).

In entrambe le domande si usi l'approssimazione Normale e la correzione di continuità.

SVOLGIMENTO

In entrambi i casi si ha $P(X \geq 180)$ dove $X \sim Bin(n=900, p)$ dove $\begin{cases} p = 1/6 & \text{DOMANDA 1} \\ p = 2/9 & \text{DOMANDA 2} \end{cases}$

Quindi senza approssimazione Normale si potrebbe solo dire che

$$P(X \geq 180) = \sum_{k=180}^{900} \binom{900}{k} p^k (1-p)^{900-k}.$$

Si tratta di un valore esatto ma, per quantificare, bisogna approssimare...

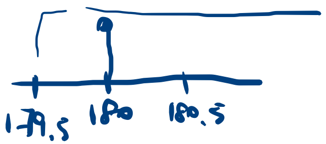
Con l'approssimazione Normale invece dobbiamo pensare che

$$\underbrace{X = X_1 + \dots + X_{900}}_{\substack{\text{Indipendenti} \\ \text{Bernoulli}(p)}} \quad \text{e quindi } X \text{ ha (approssimativamente) distribuzione } N(900 \cdot p,\ 900\,p(1-p))$$

perché $\begin{cases} \mu = \mathbb{E}[X_n] = p \\ \sigma^2 = \mathrm{Var}[X_n] = p(1-p). \end{cases}$

Allora

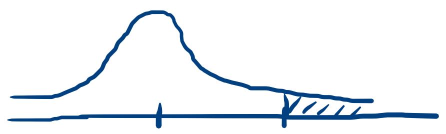
$$P(X \geq 180) \overset{\text{correzione continuità}}{=} P(X \geq 179.5) = P\left(\frac{X - 900p}{\sqrt{p(1-p)}\,\sqrt{900}} \geq \frac{179.5 - 900 \cdot p}{\underbrace{\sqrt{p(1-p)} \cdot 30}_{*}} \right).$$



Ora calcoliamo $*$ nei due casi $p = \frac{1}{6}$ e $p = \frac{2}{9}$.

Caso 1) $\displaystyle * = \frac{179.5 - 900\cdot\frac{1}{6}}{\sqrt{\frac{1}{6}\left(1-\frac{1}{6}\right)}\cdot 30} = \frac{179.5-150}{\sqrt{\frac{5}{36}}\cdot 30} = \frac{29.5}{5\sqrt{5}} \cong 2.63$

da cui segue $\displaystyle P(X\geq 180) \underset{\substack{\uparrow\\ \text{app. Normale}}}{\cong} 1-\Phi(2.63) \underset{\substack{\uparrow\\ \text{tavola}}}{=} 1-0.99573 = 0.00427.$



Caso 2) $\displaystyle * = \frac{179.5 - 900\cdot\frac{2}{9}}{\sqrt{\frac{2}{9}\left(1-\frac{2}{9}\right)}\sqrt{900}} = \frac{179.5-200}{\sqrt{\frac{14}{81}}\cdot 30} = -\frac{20.5}{\frac{\sqrt{14}}{9}\cdot 30} \cong -1.64$

da cui segue $\displaystyle P(X\geq 180) \underset{\substack{\uparrow\\ \text{app. Normale}}}{\cong} 1-\Phi(-1.64) = 1-(1-\Phi(1.64)) = \cancel{1}-\cancel{1}+\Phi(1.64) = 0.9495.$

ESERCIZIO

Si lancia una moneta equa 400 volte.

Calcolare la probabilità che il numero di teste ottenute sia compreso tra 195 e 210.
Usare l'approssimazione Normale e la correzione di continuità.

RISPOSTA

Sia X la v.a. che conta il numero di teste ottenute. Allora $X \sim Bin\left(n=400, p=\frac{1}{2}\right)$
e il risultato esatto è

$$P(195 \leq X \leq 210) = \sum_{k=195}^{210} \binom{400}{k} \left(\frac{1}{2}\right)^{400}.$$

Per l'approssimazione Normale possiamo supporre che $X \sim N\Big(\underbrace{400 \cdot \tfrac{1}{2}}_{=200}, \underbrace{400 \cdot \tfrac{1}{2}\left(1-\tfrac{1}{2}\right)}_{=100}\Big)$.

Quindi

correzione continuità

$$P(195 \leq X \leq 210) = P(194{,}5 < X < 210{,}5) =$$

APPROSSIMAZIONE NORMALE QUI

$$= P\left(\frac{194.5 - 400 \cdot \frac{1}{2}}{\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)}\sqrt{400}} \leq \frac{X - 400 \cdot \frac{1}{2}}{\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)}\sqrt{400}} \leq \frac{210.5 - 400 \cdot \frac{1}{2}}{\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)}\sqrt{400}} \right) \simeq \Phi\left(\frac{210.5-200}{10}\right) - \Phi\left(\frac{194.5-200}{10}\right) = \Phi(1.05) - \Phi(-0.55).$$

ESERCIZIO

Solita convenzione $\overline{X}_n = \dfrac{X_1 + \dots + X_n}{n}$

Sia $\{X_n : n \geq 1\}$ una successione di v.a. i.i.d. $\sim U(-1,3)$.

1) Calcolare $\lim_{n \to \infty} P\left(1 \leq \dfrac{\overline{X}_n - 1}{1/\sqrt{n}} < 2\right)$

2) Calcolare con l'approssimazione Normale $P\left(\overline{X}_{900} - 1 < \dfrac{1}{15}\right)$.

SVOLGIMENTO

1) Si usa il TLC con $\mu = \dfrac{-1+3}{2} = 1$ e $\sigma^2 = \dfrac{(3-(-1))^2}{12} = \dfrac{4^2}{12} = \dfrac{16}{12} = \dfrac{4}{3}$ (e quindi $\sigma = \sqrt{\frac{4}{3}} = \frac{2}{\sqrt{3}}$).

$$P\left(1 \leq \frac{\overline{X}_n - 1}{1/\sqrt{n}} < 2\right) = P\left(\frac{1}{\frac{2}{\sqrt{3}}} \leq \boxed{\frac{\overline{X}_n - 1}{\frac{2}{\sqrt{3}} \cdot \frac{1}{\sqrt{n}}}} < \frac{2}{\frac{2}{\sqrt{3}}}\right) \xrightarrow[n \to \infty]{} \Phi\left(\frac{2}{2/\sqrt{3}}\right) - \Phi\left(\frac{1}{2/\sqrt{3}}\right) = \Phi(\sqrt{3}) - \Phi(\sqrt{3}/2).$$

questo è $\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$

TLC CON MEDIE

METODO ALTERNATIVO CON SOMME AL POSTO DELLE MEDIE

Si ha $P\left(1 \leq \dfrac{\overline{X}_n - 1}{1/\sqrt{n}} < 2\right) = P\left(\dfrac{1}{\frac{2}{\sqrt{3}}} \leq \dfrac{X_1 + \dots + X_n - n}{\frac{2}{\sqrt{3}}\sqrt{n}} < \dfrac{2}{\frac{2}{\sqrt{3}}}\right) \xrightarrow[n \to \infty]{} \Phi\left(\dfrac{2}{2/\sqrt{3}}\right) - \Phi\left(\dfrac{1}{2/\sqrt{3}}\right) = \Phi(\sqrt{3}) - \Phi(\sqrt{3}/2)$

TLC CON SOMME

2) $$P\left(\bar{X}_{900}-1<\frac{1}{15}\right)=P\Bigg(\underbrace{\frac{\bar{X}_{900}-1}{\frac{2}{\sqrt{3}}/\sqrt{900}}}_{=\frac{\bar{X}_{900}-\mu}{\sigma/\sqrt{n}}}<\frac{1/15}{\frac{2}{\sqrt{3}}/\sqrt{900}}\Bigg)\underset{\substack{\uparrow\\ \text{APPROSSIMAZIONE}\\ \text{NORMALE QUI}}}{\simeq}\Phi\left(\frac{1/15}{\frac{2}{\sqrt{3}}\cdot\frac{1}{\sqrt{900}}}\right)=\Phi\left(\frac{1/15}{\frac{2}{\sqrt{3}}\cdot\frac{1}{30}}\right)=$$

$$=\Phi\left(\frac{\sqrt{3}}{\cancel{2}}\cdot\frac{\cancel{30}}{\cancel{15}}\right)=\Phi(\sqrt{3})$$

METODO ALTERNATIVO CON LE SOMME AL POSTO DELLE MEDIE

Si ha $$P\left(\bar{X}_{900}-1<\frac{1}{15}\right)=P\left(X_1+\dots+X_{900}-900<\frac{1}{\cancel{15}}\cdot\overset{60}{\cancel{900}}\right)$$

$$=P\Bigg(\underbrace{\frac{X_1+\dots+X_{900}-900}{\frac{2}{\sqrt{3}}\cdot\sqrt{900}}}_{=\frac{X_1+\dots+X_{900}-900\cdot\mu}{\sigma\cdot\sqrt{n}}}<\underbrace{\frac{60}{\frac{2}{\sqrt{3}}\sqrt{900}}}_{=\frac{\cancel{60}}{\frac{\cancel{2}}{\sqrt{3}}\cdot\cancel{30}}=\sqrt{3}}\Bigg)\simeq\Phi(\sqrt{3})$$