

Compito principale della statistica è trarre conclusioni da un insieme di dati

Def: Se $X_1, \dots, X_n$ sono v.z. indipendenti con la stessa distribuzione, diciamo che costituiscono un campione della distribuzione

Media Campionaria

$X_1, \dots, X_n$ campione da distribuzione f , con $E[X_1] = \int x f(x) dx = \mu$
e $\text{Var}[X_1] = \int (x - \mu)^2 f(x) dx = \sigma^2$.

Media campionaria $\bar{X} = \frac{X_1 + \dots + X_n}{n}$

$$\square E[\bar{X}] = \mu \quad \square \text{Var}[\bar{X}] = \frac{\sigma^2}{n} \quad \square \text{DS}[\bar{X}] = \sqrt{\text{Var}[\bar{X}]} = \frac{\sigma}{\sqrt{n}}$$

$\bar{X}$ ha lo stesso valore atteso di ogni X_i , ma varianza più piccola di un fattore $1/n$. Se $n=100$, la deviazione standard di $\bar{X}$ è 10 volte più piccola di quella di X_1 , se $n=10000$, di 100 volte. (8

Per approssimare la distribuzione di $\bar{X}$ usiamo il teorema del limite centrale: $X_1, \dots, X_n$ v.z. i.i.d., $\mu = E[X]$, $\sigma^2 = \text{Var}(X)$.

$$\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n} \sigma} \xrightarrow{n \rightarrow \infty} Z \sim N(0, 1)$$

Essendo $\bar{X} = \frac{X_1 + \dots + X_n}{n}$, ho $\bar{X} - \mu = \frac{1}{n} (X_1 + \dots + X_n - n\mu)$

quindi $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \dots + X_n - n\mu}{n \cdot \sigma/\sqrt{n}} = \frac{X_1 + \dots + X_n - n\mu}{\sqrt{n} \sigma} \rightarrow Z$

Per esempio: $P[\bar{X} < a] = P\left[\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{a - \mu}{\sigma/\sqrt{n}}\right] \approx$

$$\approx P\left[Z < \frac{a - \mu}{\sigma/\sqrt{n}}\right] = \Phi\left(\frac{a - \mu}{\sigma/\sqrt{n}}\right)$$

Per che valori di n , si può applicare il TLC?

Nei libri di solito scrivono $n=30$ oppure $n=50$ come regola. In realtà la bontà dell'approssimazione dipende dalla distribuzione delle X_i ; se sono molto simmetriche già $n=5$, 10 può dare buoni risultati, se invece sono fortemente asimmetriche, $n=50$ può ancora essere troppo piccolo.

Esempio Qual è la prob. di ottenere almeno 29 "Teste" in 50 lanci di una moneta equilibrata?

$$X_i = \begin{cases} 1 & \text{TESTE} \\ 0 & \text{CROCE} \end{cases} \quad X_i \sim \text{Bi}(1, \frac{1}{2}), \quad E[X_i] = p = \frac{1}{2}, \quad \text{Var}[X_i] = p(1-p) = \frac{1}{4}$$

$$\begin{aligned} P[X_1 + \dots + X_{50} \geq 29] &= 1 - P[X_1 + \dots + X_{50} \leq 28] \approx 1 - \Phi\left(\frac{28 - 50 \cdot \frac{1}{2}}{\sqrt{50} \cdot \sqrt{\frac{1}{4}}}\right) \\ &= 1 - \Phi(0.85) = 0.2 \end{aligned}$$

Il valore esatto $[X_1 + \dots + X_{50} \sim \text{Bi}(50, 1/2)]$ è 0,16.

Si può migliorare l'approssimazione (nel caso in cui la v.v. X_i sia discreta) notando che

$$\begin{aligned} \mathbb{P}[X_1 + \dots + X_{50} > 28] &= \mathbb{P}[X_1 + \dots + X_{50} \geq 29] \\ &= \mathbb{P}[X_1 + \dots + X_{50} > 28.5] \approx 1 - \Phi\left(\frac{28.5 - 25}{\sqrt{50 \cdot 1/2}}\right) \\ &= 1 - \Phi(0.99) = 0,16. \end{aligned}$$

Questa operazione è detta correzione di continuità

Varianza Campionaria

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

$X_1, \dots, X_n$ campione
con $X_i \sim N(\mu, \sigma^2)$

Teorema :

$$\frac{n-1}{\sigma^2} S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$$

si distribuisce con legge chi-quadrato con $n-1$ gradi di libertà, dove Y è chi-quadrato $(n-1) \Leftrightarrow Y = Z_1^2 + \dots + Z_{n-1}^2$ con $Z_i \sim N(0,1)$ indipendenti