STIMATORI

<u>modello statistico</u>: modellizza un esperimento aleatorio in cui non abbiamo ancora attribuito univocamente la probabilità

costituito da:

- spazio campionario Ω
- famiglia di probabilità $(\mathbb{P}_\vartheta)_\vartheta$, con $\vartheta \in \Theta$

ϑ <u>parametro</u>, Θ <u>spazio dei parametri</u>

<u>esempio</u>: supponiamo di voler effettuare un esp. al. dove siano definite $X_1, \dots, X_n$ i.i.d. $\sim Be(p)$, con p sconosciuto

Allora $\vartheta = p \in \Theta = (0,1)$, e $\mathbb{P}_\vartheta$ è la probabilità t.c. $X_1, \dots, X_n$ i.i.d. $\sim Be(\vartheta)$ sotto $\mathbb{P}_\vartheta$

<u>esempio</u>: se nell'esp. al. abbiamo bisogno invece di

$X_1, \dots, X_n$ i.i.d. $\sim N(\mu, \sigma^2)$, con μ, σ^2 sconosciute,

allora $\vartheta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0, +\infty)$

Diciamo che $X_1, \dots, X_n$ sono un <u>campione</u> (<u>statistico</u>) <u>di taglia n</u> se le $X_1, \dots, X_n$ sono i.i.d. sotto $\mathbb{P}_\vartheta$, $\forall \vartheta \in \Theta$

A partire da un campione, possiamo definire lo <u>stimatore</u> di un parametro come una funzione delle v.al. del campione.

esempio: stimatore della media

$$\bar{X} := \frac{1}{n} \sum_{1}^{n} {}_i X_i$$

esempio: stimatore della varianza

$$S_X^2 = \frac{1}{n-1} \sum_{1}^{n} {}_i \left(X_i - \bar{X}\right)^2$$

esempio: stimatore della deviazione standard

$$S_X = \sqrt{S_X^2}$$

Gli stimatori possono godere (o meno) di alcune proprietà.

Uno stimatore si dice corretto, o non distorto, se la sua speranza sotto $\mathbb{P}_\vartheta$ è uguale al parametro che sta stimando

esempio: media

$X_1, \dots, X_n$ i.i.d. con $E_\vartheta[X_i] = \vartheta_1$ $\left(\vartheta = (\vartheta_1, \dots, \vartheta_K)\right)$

speranza rispetto a $\mathbb{P}_\vartheta$

Allora

$$E_\vartheta\left[\bar{X}\right] = E_\vartheta\left[\frac{1}{n}\sum_{1}^{n} {}_i X_i\right] = \frac{1}{n}\sum_{1}^{n} {}_i E_\vartheta[X_i] = \frac{1}{n}\sum_{1}^{n} {}_i \vartheta_1 =$$

$$= \frac{1}{n} \cdot n\vartheta_1 = \vartheta_1$$

esempio: varianza

Innanzitutto si può scrivere

$$\sum_{1}^{n} {}_i (X_i - \bar{X})^2 = \sum_{1}^{n} {}_i \left(X_i^2 - 2X_i\bar{X} + \bar{X}^2\right) =$$

$$\sum_1^n {}_i (X_i - \bar{X})^2 = \sum_1^n {}_i \left(X_i^2 - 2X_i\bar{X} + \bar{X}^2\right) =$$

$$= \sum_1^n {}_i X_i^2 - 2\bar{X}\sum_1^n {}_i X_i + \sum_1^n {}_i \bar{X}^2 =$$

$$= \sum_1^n {}_i X_i^2 - 2\bar{X} \cdot n\bar{X} + n\bar{X}^2 =$$

$$= \sum_1^n {}_i X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 = \sum_1^n {}_i X_i^2 - n\bar{X}^2$$

Allora

$$S_X^2 = \frac{1}{n-1}\left(\sum_1^n {}_i X_i^2 - n\bar{X}^2\right)$$

Supponiamo che $\vartheta = (\mu, \sigma^2)$, con $E_\vartheta[X_i] = \mu$, $Var_\vartheta[X_i] = \sigma^2$

calcolate rispetto a $\mathbb{P}_\vartheta$

Allora

$$E_\vartheta[S_X^2] = E_\vartheta\left[\frac{1}{n-1}\left(\sum_1^n {}_i X_i^2 - n\bar{X}^2\right)\right] =$$

$$= \frac{1}{n-1}\left(\sum_1^n {}_i E_\vartheta[X_i^2] - n E_\vartheta[\bar{X}^2]\right)$$

Ma $Var[X] = E[X^2] - E[X]^2 \Rightarrow$

$$E[X^2] = Var[X] + E[X]^2$$

$$E_\vartheta[X_i^2] = Var_\vartheta[X_i] + E_\vartheta[X_i]^2 = \sigma^2 + \mu^2$$

$$E_\vartheta[\bar{X}^2] = Var_\vartheta[\bar{X}] + E_\vartheta[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$$

$$Var_\vartheta[\bar{X}] = Var_\vartheta\left[\frac{1}{n}\sum_1^n {}_i X_i\right] = \frac{1}{n^2} Var_\vartheta\left[\sum_1^n {}_i X_i\right] =$$

$$= \frac{1}{n^2}\sum_1^n {}_i Var_\vartheta[X_i] = \frac{1}{n^2}\sum_1^n {}_i \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$E\left[S_X^2\right] = \frac{1}{n-1}\left(\sum_{i=1}^{n} E_\theta\left[X_i^2\right] - n\, E_\theta\left[\bar{X}^2\right]\right) =$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}\left(\sigma^2 + \mu^2\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) =$$

$$= \frac{1}{n-1}\left(n\sigma^2 + \cancel{n\mu^2} - \sigma^2 - \cancel{n\mu^2}\right) =$$

$$= \frac{(n-1)\sigma^2}{n-1} = \sigma^2$$

<u>esempio</u>: dev. standard

$$E_\theta\left[\overbrace{\sqrt{S_X^2}}^{=S_X}\right] \neq \sigma \quad (!)$$

Difatti

$$Var_\theta\left[S_X\right] = E_\theta\left[S_X^2\right] - E_\theta\left[S_X\right]^2 > 0$$

$$E_\theta\left[S_X\right]^2 < E_\theta\left[S_X^2\right] = \sigma^2$$

$$E_\theta\left[S_X\right] < \sigma$$

Una stimatore si dice <u>consistente</u> se, per $n \to \infty$, converge in probabilità al parametro che sta stimando.

<u>esempio</u>: media $\left(E_\theta[X_i] = \mu\right)$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}_\theta} E_\theta\left[X_i\right]^{=\mu} \quad (\text{l.g.n.})$$

<u>esempio</u>: varianza $\left(E_\theta[X_i] = \mu,\ Var_\theta[X] = \sigma^2\right)$

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

$$S_X^2 = \frac{1}{n-1}\left(\sum_1^n {}_i X_i^2 - n\bar{X}^2\right) =$$

$$= \frac{1}{n-1}\left(n \cdot \frac{1}{n}\sum_1^n {}_i X_i^2 - n\bar{X}^2\right) =$$

$$= \frac{n}{n-1}\left(\underbrace{\frac{1}{n}\sum_1^n {}_i X_i^2}_{} - \bar{X}^2\right) \xrightarrow{\mathbb{P}_\theta} 1 \cdot \left(E_\theta[X_i^2] - \mu^2\right) =$$

$$= \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

<u>esempio</u>: dev. st.

$$S_X = \sqrt{S_X^2} \xrightarrow{\mathbb{P}_\theta} \sqrt{\sigma^2} = \sigma$$

Uno stimatore si dice <u>asintoticamente normale</u> se, sottratta la media e diviso per la dev. st., converge in legge a $N(0,1)$

Per il t.l.c.,

$$\frac{\bar{X} - E_\theta[\bar{X}]}{\sqrt{Var_\theta[\bar{X}]}} = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\frac{1}{n}\sum_1^n {}_i X_i - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{legge} N(0,1)$$

la media campionaria è asint. normale

La dev. st. di $\bar{X}$ è $\frac{\sigma}{\sqrt{n}}$; se è sconosciuta, usiamo lo stimatore

$$S_{\bar{X}} = \frac{S_X}{\sqrt{n}}$$ <u>errore standard della media</u>

$\boxed{\text{Es. } 3.5}$ $(X_i)_i$ i.i.d. $\sim Be(p)$

$\boxed{\text{Es. 3.5}}$ $(X_i)_i$ i.i.d. $\sim Be(p)$

1) stimatore per $p = E[X_i]$:

$$\hat{p} = \frac{1}{n} \sum_{1}^{n} {}_i X_i = \bar{X}$$

è uno stimatore di p

2) se $p = \frac{1}{2}$:

$\mathbb{P}\{S_{15} \geq 11\}$, con $S_{15} = \sum_{1}^{15} {}_i X_i \sim B(15; \frac{1}{2})$

o si fa il calcolo con la legge esatta ($\sim$ 20 cifre d.)

oppure: $np = n(1-p) = 15 \times \frac{1}{2} = 7,5 > 5$

appr. normale! (nota: $np(1-p) = 15 \cdot \frac{1}{2} \cdot \frac{1}{2} = 3,75 < 5$)

$$\mathbb{P}\{S_{15} \geq 11\} = \mathbb{P}\left\{ \underbrace{\frac{S_{15} - np}{\sqrt{np(1-p)}}}_{= Z \approx N(0,1)} \geq \frac{10,5 - 7,5}{\sqrt{3,75}} \right\} \simeq$$

$F_Z(1,55)$

$$\simeq \mathbb{P}\left\{Z \geq \frac{3}{1,93}\right\} = \mathbb{P}\{Z \geq 1,55\} = 1 - \mathbb{P}\{Z \leq 1,55\} =$$

$$= 1 - 0,93943 = 0,06057$$

3) se $S_n = 11$, stimare p:

$$\hat{p} = \frac{1}{n} \sum_{1}^{n} {}_i X_i = \frac{1}{n} S_n = \frac{11}{15} = 0,73$$