Media campionaria

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

Momento campionario di ordine K

$$M_n^{(k)} = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

Varianza campionaria

$$S_0^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - m)^2$$

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2$$

Proposizioni

$$S_0^2 = M_n^{(2)} - 2m\overline{X}_n + m^2$$

$$S_n^2 = \frac{1}{n-1} \left(n M_n^{(2)} - n \overline{X}_n^2\right)$$

Proposizione (Valore atteso e varianza della media campionaria)

Se la popolazione X è dotata di media e di varianza ($E[X] = m$, $Var[X] = v$), allora dato un campione i.i.d., $X_1, \dots, X_n$, abbiamo

$$E[\bar{X}_n] = m \qquad Var[\bar{X}_n] = \frac{1}{n} v$$

Proposizione (Valore atteso della varianza campionaria)

Sotto le ipotesi della proposizione precedente

$$E[S_0^2] = v \qquad E[S_n^2] = v$$

Proprietà (Distribuzione della media campionaria)

Sia $X_1, X_2, \dots, X_n$ un campione estratto da una popolazione $X \sim Nor(\mu = m, \sigma^2 = v)$, allora, detta S_0^2 la varianza campionaria a media nota

$$S_0^2 \sim \frac{v}{n} \chi^2(\nu = n)$$

Per la varianza campionaria a media incognita

$$S_n^2 \sim \frac{v}{n-1} \chi^2(\nu = n-1)$$

In sintesi:

$$\frac{\bar{X}_n - m}{\sqrt{\frac{v}{n}}} \sim Nor(\mu = 0, \sigma^2 = 1), \quad \frac{n}{v} S_0^2 \sim \chi^2(\nu = n)$$

$$\frac{n-1}{v} S_n^2 \sim \chi^2(\nu = n-1)$$

Proprietà: $\overline{X}_n$ è indipendente da S_n^2

Proposizione

Sia $X_1, \dots, X_n$ un campione estratto da una popolazione $X \sim Nor(\mu = m, \sigma^2 = v)$, allora

$$\frac{\overline{X}_n - m}{\sqrt{\frac{S_n^2}{n}}} \sim T(\nu = n-1)$$

Definizione (convergenza di una distribuzione)

Si dice che la successione di variabili aleatorie $\{X_n\}$ converge in distribuzione alla variabile aleatoria X, $X_n \xrightarrow{d} X$, se

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

per tutti i valori di x ove $F_X(x)$ è continua

Definizione (convergenza in probabilità

Si dice che la successione di variabili aleatorie $\{X_n\}$ converge in probabilità alla variabile aleatoria X se

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P\left[\, |X_n - X| < \varepsilon \,\right] = 1$$

$X_n \xrightarrow{P} X$ in simboli

Definizione (convergenza quasi certa)

Si dice che la successione di variabili aleatorie $\{X_n\}$ converge quasi certamente alla variabile aleatoria X, se

$$\forall \varepsilon > 0 \quad P\left[\, \lim_{n\to\infty} |X_n - X| < \varepsilon \,\right] = 1$$

$$X_n \xrightarrow{q.c} X$$

Definizione (convergenza in media quadratica)

$$X_n \xrightarrow{m.q.} X \qquad \lim_{n\to\infty} E\left[(X_n - X)^2 \right] = 0$$

$$\begin{matrix} X_n \xrightarrow{q.c.} X \\ X_n \xrightarrow{m.q.} X \end{matrix} \quad \Rightarrow \quad X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{d} X$$

Teorema (Legge debole dei grandi numeri)

Sia X_n una successione di variabili aleatorie i.i.d.. provenienti da una popolazione X dotata di media ($E[X]=m$), allora

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P\left[\,|\bar{X}_n - m| < \varepsilon\,\right] = 1$$

cioè, la successione delle medie campionarie converge in probabilità ad m, il valore atteso di tutte le X_n.

Teorema (Legge forte dei grandi numeri)

Stesse ipotesi del teorema precedente

$$P\left[\lim_{n\to\infty} |\bar{X}_n - m| < \varepsilon\right] = 1$$

Teorema (centrale del limite)

Sia X_n una succ di variabili aleatorie i.i.d. provenienti da una popolazione X dotata di media e di varianza ($E[X]=m$, $Var[X]=v$) allora per ogni $x \in \mathbb{R}$

$$\lim_{n \to \infty} P\left[\frac{\overline{X}_n - m}{\sqrt{\frac{V}{n}}} \leq x \right] = \phi(x)$$

distribuzione normale

cioè la successione delle medie campionarie standardizzate converge in legge ad una $Nor(\mu=0, \sigma^2=1)$

$$\overline{X}_n \sim Nor\left[\mu = m \,,\, \sigma^2 = \frac{V}{n}\right]$$

Osservazioni

1) $Bin(n,p) \approx Nor(\mu = np, \sigma^2 = np)$

se $n \geq 30$, $np \geq 5$, $nq \geq 5$

2) $\overline{Bin}(n,p) \approx Nor\left(\mu = \frac{n}{p}, \sigma^2 = \frac{n(1-p)}{p^2}\right)$

se $n \geq 30$

3) $Poi(\lambda) \approx Nor(\mu = \lambda, \sigma^2 = \lambda)$

se $\lambda \geq 30$

Definizione (Statistica)

Si dice statistica di una variabile aleatoria T, funzione nota delle variabili aleatorie $X_1, \dots, X_n$

$$T(X_1, \dots, X_n) = T(X)$$

Definizione (di stimatore)

Data una popolazione dipendente da un parametro ϑ, e un campione casuale $X_1, X_2, \dots, X_n$, si dice stimatore una statistica T che non dipende da ϑ e che serve per stimare ϑ

Definizione (correttezza, correttezza asintotica)

Uno stimatore si dice corretto (o non distorto) se il suo valore atteso coincide con la quantità da stimare

$$E[T_n] = \vartheta$$

Si dirà asintoticamente corretto se

$$E[T_n] \xrightarrow{n \to \infty} \vartheta$$

Definizione (consistenza)

Uno stimatore si dice consistente in media quadratica se

$$E\left[(T_n-\vartheta)^2\right] \xrightarrow{n\to\infty} 0$$

si dirà consistente in probabilità

se $$P\left[\,|T_n-\vartheta| < \varepsilon\,\right] \xrightarrow{n\to\infty} 0 \qquad \forall \varepsilon>0$$

Definizione (Asintotica normalità)

$$\frac{T_n - E[T_n]}{\sqrt{Var[T_n]}} \overset{n\to\infty}{\sim} Nor(0,1)$$

Esempio

Sia $\vartheta>0$. Supponiamo che X abbia

densità $$f_X(x) = \begin{cases} \vartheta\, x^{\vartheta-\ell} & 0\le x\le 1 \\ 0 & \text{altrove} \end{cases} \quad \text{con } \vartheta>0$$

$$E[X] = \int_0^1 x\left(\vartheta x^{\vartheta-1}\right)dx = \frac{\vartheta}{\vartheta+1}$$

$$E[X] = \frac{\vartheta}{\vartheta+1} \quad \Rightarrow \quad \vartheta = \frac{E[X]}{1-E[X]}$$

lo stimatore di ϑ è $\boxed{\hat{\vartheta} = \frac{\overline{X}_n}{1-\overline{X}_n}}$

Esempio

$$f_X(x) = \begin{cases} \frac{2\vartheta+1}{2}\left(x^2\right)^{\vartheta} & -1 \leq x \leq 1 \\ 0 & \text{altrove} \end{cases} \qquad \vartheta \geq 0$$

$$E[X] = 0 \qquad \int_{-1}^{1} x \, f_X(x) = 0$$

$$E[X^2] = \int_{-1}^{1} x^2 \, \frac{2\vartheta+1}{2} \left(x^2\right)^{\vartheta} dx = \frac{2\vartheta+1}{2\vartheta+3}$$

$$\vartheta = \frac{3E[X^2]-1}{2\left(1-E[X^2]\right)} \qquad M_n^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$\hat{\vartheta} = \frac{3M_n^2-1}{2\left(1-M_n^2\right)}$$

Esercizi:

1) $X \sim Ber(p)$

2) $Poi(\mu)$

Metodo della massima verosomiglianza

$X_1, X_2, \ldots, X_n$

$$\mathcal{L}(x_1, x_2, \ldots, x_n, \vartheta) = p_X(x_1, \vartheta) \cdot p_X(x_2, \vartheta) \ldots p_X(x_n, \vartheta) =$$

$$= \prod_{i=1}^{n} p_X(x_i, \vartheta)$$

Esempio

$$f_X(x) = \begin{cases} \vartheta x^{\vartheta - 1} & 0 \le x \le 1 \\ 0 & \text{altrove} \end{cases} \qquad \vartheta > 0$$

$$\mathcal{L}(\vartheta) = [\vartheta x_1^{\vartheta-1}][\vartheta x_2^{\vartheta-1}] \ldots [\vartheta x_n^{\vartheta-1}] =$$

$$= \vartheta^n \left[\prod_i x_i\right]^{\vartheta - 1}$$

$$\frac{d}{d\vartheta} \mathcal{L}(\vartheta) = \frac{n}{\vartheta} + \ln\left[\prod_i x_i\right] = 0$$

$$\vartheta = -\frac{n}{\ln\left[\prod_i x_i\right]} \qquad \vartheta = -\frac{n}{\sum_i \ln x_i}$$

Esercizi

1) $p_X(x) = \vartheta^x (1-\vartheta)^{1-x}$

2) $\mathrm{Poi}(\mu)$

3) $\mathrm{Exp}(\lambda)$

4) $\mathrm{Norm}(\mu, \sigma^2)$