Campione statistico

L'inferenza statistica ha lo scopo di ricavare informazioni su una popolazione tramite l'analisi di un campione, partendo dall'ipotesi che ci sia una distribuzione di probabilità nella popolazione, e che le osservazioni sul campione siano la realizzazione di variabili aleatorie indipendenti con quella distribuzione.

Un campione quindi è una famiglia $X_1, \dots, X_n$ di variabili aleatorie indipendenti e con stessa funzione di ripartizione F (*legge di probabilità del campione*), sconosciuta totalmente o in parte.

Una funzione di un campione statistico è detta *statistica campionaria.*

Stima di media e varianza

Dato il campione $X_1, \dots, X_n$, con momento secondo, $\mu = E[X_i]$ e $\sigma^2 = \mathrm{var}(X_i)$, allora:

$$E[\overline{X}] = \mu \qquad E[S^2] = \sigma^2$$

dove

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} \qquad S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$$

sono *media* e *varianza campionaria.*

Questo significa che media e varianza campionaria sono una stima corretta della media e varianza delle variabili.

Dimostrazione

Ovvio per $E[\overline{X}] = \mu$.
Per $E[S^2] = \sigma^2$: usiamo

$$\sum_{i=1}^n (X_i - \overline{X})^2 = \sum_{i=1}^n X_i^2 - n\overline{X}^2,$$

sapendo che:

$$\begin{aligned}
\mathrm{var}(\overline{X}) &= \frac{\sigma^2}{n} && (X_1, \dots, X_n \text{ indipendenti}) \\
E[X_i^2] &= \mathrm{var}(X_i) + E[X_i]^2 = \sigma^2 + \mu^2 \\
E[\overline{X}^2] &= \mathrm{var}(\overline{X}) + E[\overline{X}]^2 = \frac{\sigma^2}{n} + \mu^2.
\end{aligned}$$

Allora:

$$\begin{aligned}
E\left[\sum (X_i - \overline{X})^2\right] &= E\left[\sum X_i^2\right] - nE[\overline{X}^2] \\
&= n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) \\
&= (n-1)\sigma^2.
\end{aligned}$$

Dividendo per $(n-1)$ si ha $E[S^2] = \sigma^2$.