

LA FUNZIONE DI VEROSIMIGLIANZA

HA UN RUOLO IMPORTANTE NELLA PROCEDURE DI
INFERENZA STATISTICA COME:

- 1) METODO DI COSTRUZIONE DI STIMATORI (IN
SITUAZIONI COMPLESSE)
- 2) METODO DI INDIVIDUAZIONE DI TEST
UNIFORMEMENTE PIU' POTENTI

Funzione di verosimiglianza

Osservato un determinato campione $X_1, \dots, X_n$ estratto da una popolazione X la cui distribuzione dipende da un parametro θ

La **funzione di verosimiglianza** $L(\theta)$ indica la probabilità di osservare il campione al variare del parametro θ

Poiché le osservazioni campionarie sono **indipendenti** e **identicamente distribuite** possiamo scrivere la funzione di verosimiglianza come prodotto delle probabilità delle singole osservazioni campionarie:

$$\begin{aligned} L(\theta) &= P(\text{dati osservati}; \theta) = P(X_1 = x_1, \dots, X_n = x_n; \theta) = \\ &= P(X_1 = x_1; \theta) \dots P(X_n = x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \end{aligned}$$

dove $f(x_i; \theta)$ corrisponde alla probabilità o alla densità a seconda se X è discreta o continua.

Metodo di massima verosimiglianza

Osservazione:

Nella funzione di verosimiglianza i dati campionari sono fissati, mentre il valore del parametro incognito può variare.

Quindi la $L(\theta)$ deve essere vista come una funzione del solo parametro θ che ci indica quanto è plausibile uno specifico valore del parametro ignoto, una volta che sia stato osservato il campione.

Se θ_1 e θ_2 sono due valori del parametro ignoto e se, dato il campione, $L(\theta_1) > L(\theta_2)$, diremo che il valore θ_1 è **più verosimile** di θ_2 , alla luce del campione osservato.

Metodo di massima verosimiglianza

Il **metodo di massima verosimiglianza** per ottenere una stima del parametro ignoto, consiste nel prendere il valore di θ che massimizza la funzione di verosimiglianza $L(\theta)$.

Il valore che massimizza la $L(\theta)$ è la **stima di massima verosimiglianza** di θ ed è indicato con $\hat{\theta}$.

Per convenienza si utilizza massimizzare il logaritmo della $L(\theta)$ detta **log-verosimiglianza**, $\log L(\theta)$.

La stima di massima verosimiglianza di θ è il valore $\hat{\theta}$ che massimizza la $\log L(\theta)$, ossia

$$\log L(\hat{\theta}) = \sup_{\theta} \log L(\theta)$$

Stimatore di massima verosimiglianza

La **stima di massima verosimiglianza** di θ è la soluzione dell'equazione di verosimiglianza: $\partial \log L(\theta) / \partial \theta = 0$ che nel punto $\theta = \hat{\theta}$ soddisfa $\partial^2 \log L(\theta) / \partial \theta^2 = 0$

Al variare del campione osservato si avrà in generale una diversa stima di massima verosimiglianza del parametro. Si ottiene perciò una v.c. detta **stimatore di massima verosimiglianza** del parametro θ .

Se la popolazione è Normale con media μ e varianza σ^2 gli stimatori di massima verosimiglianza per μ e σ^2 sono:

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

Il livello del corso non permette una trattazione approfondita del metodo di costruzione della massima verosimiglianza

Ma se:

$$X \approx P(\lambda) \quad \Rightarrow \quad \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$X \approx \text{Exp}(\lambda) \quad \Rightarrow \quad \hat{\lambda} = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}}$$

in fine se:

$$X \approx \text{Ber}(\pi) \quad \Rightarrow \quad \hat{\pi} = \frac{1}{n} \sum_{i=1}^n X_i$$

Esempio

Popolazione Bernoulliana con parametro π

Campione osservato $x_1, \dots, x_n$

$$L(\pi) = \prod_{i=1}^n p(X_i = x_i; \pi) = \prod_{i=1}^n \pi^{x_i} (1 - \pi)^{1-x_i} =$$

$$= \pi^{x_1 + \dots + x_n} (1 - \pi)^{(1-x_1) + \dots + (1-x_n)} =$$

$$= \pi^{x_1 + \dots + x_n} (1 - \pi)^{n - (x_1 + \dots + x_n)}$$

$$\partial \log L(\pi) / \partial \pi = \frac{\partial}{\partial \pi} \left[\log \left(\pi^{x_1 + \dots + x_n} (1 - \pi)^{n - (x_1 + \dots + x_n)} \right) \right] =$$

$$= \frac{x_1 + \dots + x_n - n\pi}{\pi(1 - \pi)} = 0 \quad \Rightarrow \quad x_1 + \dots + x_n - n\pi = 0$$

$$\hat{\pi} = \bar{x}$$

Test del rapporto delle massime verosimiglianze

Un test con livello di significatività pari ad α e una funzione di potenza $\pi^*(\theta)$ è detto **uniformemente più potente** a livello α se:

$$\pi^*(\theta) \geq \pi(\theta), \quad \theta \in \Theta_1$$

per ogni altro test con uguale livello di significatività α e funzione di potenza $\pi(\theta)$.

Test uniformemente più potenti possono essere individuati mediante l'approccio basato sul **rapporto delle massime verosimiglianze**.

Test uniformemente più potenti possono essere individuati mediante l'approccio basato sul rapporto delle massime verosimiglianze

Dato un problema di verifica d'ipotesi:

la statistica rapporto delle massime verosimiglianze è: $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_1$

$$\lambda(\mathbf{x}) = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$$

dove $\hat{\theta}_0$ è la stima di max verosimiglianza di θ con il vincolo $\theta \in \Theta_0$

mentre $\hat{\theta}$ è la stima di max verosimiglianza non vincolata.

$\Lambda(x)$ varia tra 0 e 1 : più è vicino a 0 più è grande l'evidenza fornita dai dati contro l'ipotesi nulla. Quindi è logico considerare un test che porti a rifiutare H_0 a favore di H_1 quando $\Lambda(x)$ assume valori piccoli:

$$\text{R.C.: } \Lambda(x) \leq c$$

Per un opportuno c tale che $0 < c < 1$.

In genere è scelto tale che il livello di significatività del test sia pari ad un α prefissato.

Test chi-quadrato sull'adattamento

Tale test permette di verificare se un modello teorico è adeguato a rappresentare un fenomeno di interesse per il quale è stata osservata la distribuzione di frequenza. Il test può essere applicato sia su variabili quantitative discrete o continue (suddivise in classi) sia su variabili qualitative.

Data una variabile X con K modalità

$A_1, A_2, \dots, A_k$ alle quali siano associate nella popolazione di riferimento le probabilità $P(A_i)$, $i = 1, 2, \dots, k$.

Si supponga di avere osservato le frequenze

$n_1, n_2, \dots, n_k$ (frequenze osservate) in corrispondenza di $A_1, A_2, \dots, A_k$.

L'obiettivo è verificare sulla base di una distribuzione di frequenze osservata, se un modello teorico è idoneo a descrivere il fenomeno.

Un modello teorico consiste in una distribuzione di probabilità:

$$P(A_1) = p_1, P(A_2) = p_2, \dots, P(A_k) = p_k \text{ con } p_i \geq 0 \text{ e } \sum_{i=1}^n p_i = 1.$$

L'ipotesi nulla che si vuole sottoporre a test è:

$$H_o : P(A_i) = p_i, \text{ per } i = 1, 2, \dots, k.$$

l'ipotesi alternativa, sottintesa assume che invece la distribuzione delle p_i non è adeguata.

Per effettuare il test si debbono confrontare le frequenze osservate con le frequenze teoriche, ossia quelle attese sotto l'ipotesi nulla

Se l'ipotesi nulla è vera le frequenze attese in un campione di n osservazioni sono pari a :

$$\hat{n}_1 = np_1, \hat{n}_2 = np_2, \dots, \hat{n}_k = np_k.$$

Il test si esegue confrontando le frequenze teoriche con quelle osservate:

quanto più è grande la distanza tra esse tanto meno verosimile è H_0 .

Si utilizza la statistica test:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i}$$

che sotto l'ipotesi nulla è nota ed al crescere di n si distribuisce come una v.c. chi-quadrato con $k-1$ g.d.l.

Il valore della statistica test aumenta all'aumentare del numeratore per cui è ragionevole individuare la regione di accettazione per valori piccoli della statistica test: quanto più è piccolo il suo valore osservato tanto più plausibile appare l'ipotesi nulla:

$$\text{R.A.: } \chi^2 \leq \chi^2_{k-1, \alpha} \quad \text{e} \quad \text{R.C.: } \chi^2 > \chi^2_{k-1, \alpha}$$

sono le regioni di accettazione e rifiuto al livello di significatività α

Nota: Sul test di omoschedasticità

Nel caso di test di confronto tra medie per due campioni indipendenti provenienti da due popolazioni normali con σ_1^2 e σ_2^2 incognite essendo $n_1 \leq 30$ e/o $n_2 \leq 30$ è necessario effettuare il test di omoschedasticità ossia di uguaglianza delle varianze.

Se si accetta l'ipotesi di omoschedasticità si utilizza nella statistica test "differenza tra le medie campionarie" studentizzata come stimatore della varianza (unica e pari a σ^2) della popolazione la varianza *pooled*.

Si noti che se si rifiuta l'ipotesi di omoschedasticità la statistica test diventa:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx t_g \quad \text{con} \quad g = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{\left(\frac{S_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2} \right)^2}{n_2 - 1}}$$