Introduzione all'inferenza statistica, IV lezione

Carla Rampichini

Dipartimento di Statistica "Giuseppe Parenti" - Firenze - Italia

rampichini@ds.unifi.it - www.ds.unifi.it/rampi/

Proprietà asintotiche

- $T(\mathbf{X})$ è asintoticamente corretto per θ se:

$$\boxed{\lim_{n\to\infty} \mathbb{E}(T(\mathbf{X})) = \theta \Leftrightarrow \lim_{n\to\infty} b(T(\mathbf{X})) = 0.}$$

- $T(\mathbf{X})$ consistente in media quadratica (EQM) per θ se:

$$\boxed{\lim_{n\to\infty} \mathbb{E}(T(\mathbf{X}) - \theta)^2 = 0}$$

  $\Rightarrow \lim_{n\to\infty} V(T(\mathbf{X})) = 0, \lim_{n\to\infty} b^2(T(\mathbf{X})) = 0$
  Ossia, per $n \to \infty$ $P(T(\mathbf{X}) = \theta) = 1$.
  La consistenza in EQM $\Rightarrow$ la non distorsione asintotica.
  Viceversa, se vale $\lim_{n\to\infty} \mathbb{E}(T(\mathbf{X})) = \theta$ perchè $T(\mathbf{X})$ sia consistente deve valere anche $\lim_{n\to\infty} V(T(\mathbf{X})) = 0$

- $T(\mathbf{X})$ è consistente in probabilità per θ se per $\forall\varepsilon > 0$ si ha:

$$\boxed{\lim_{n\to\infty} P(\mid T(\mathbf{X}) - \theta \mid < \varepsilon) = 1 \Leftrightarrow \text{Plim}\ T(\mathbf{X}) = 1}$$

- Efficienza asintotica

Metodi di costruzione degli stimatori

Esitono molti metodi pratici utilizzati per costruire stimatori che godono di determinate proprietà.

- metodo dei momenti
- metodo dei minimi quadrati
- stimatori di massima verosimiglianza
- stimatori bayesiani
- algoritmo EM

Metodo della massima verosimiglianza

- E' uno dei metodi più utilizzati. Proposto da Fisher nel 1920;
- gli stimatori di ML godono di proprietà asintotiche ottimali;
- la stima di θ è il valore $\hat{\theta}$ che rende massima $\mathcal{L}(\theta; \mathbf{x})$.

► *Intuitivamente*: il valore del parametro per il quale i *dati osservati* hanno la più alta probabilità di presentarsi è il miglior stimatore di θ.

$$\boxed{\hat{\theta} : S(\theta) = \frac{\partial \log \mathcal{L}(\theta; \mathbf{x})}{\partial \theta} = 0}$$

sotto il vincolo $\frac{\partial^2 \log \mathcal{L}(\theta; \mathbf{x})}{\partial \theta}|_{\hat{\theta}} < 0$ per assicurare che il punto di stazionarietà sia un punto di massimo.

► Attenzione: tale interpretazione non può essere rovesciata! $\mathcal{L}(\theta; \mathbf{x})$ **non è** una funzione di probabilità per i possibili valori di $\theta \epsilon \Theta$!!

Proprietà degli stimatori ML

- sotto condizioni piuttosto generali, $\hat{\theta}_{ML}$ è consistente. Quindi è *asintoticamente corretto*, ma *distorto* in piccoli campioni;
- se esiste una statistica sufficiente $\hat{\theta}_{ML}$ è una sua funzione;
- se esite lo stimatore a varianza minima (MVE), questo coincide con $\hat{\theta}_{ML}$; se MVE non esiste, non si può dire nulla sulla $V(\hat{\theta}_{ML})$ in piccoli campioni
- *invarianza*: se $\hat{\theta}_{ML}$ è lo stimatore ML di θ, allora $\gamma(\hat{\theta}_{ML})$ è lo stimatore ML di $\gamma(\theta)$;
- $\hat{\theta}_{ML}$ è asintoticamente efficiente, cioè $V(\hat{\theta}_{ML}) \to [I_n(\theta)]^{-1} = [nI(\theta)]^{-1}$, per X_i i.i.d.;
- $\hat{\theta}_{ML}$ è asintoticamente Normale: $\hat{\theta}_{ML} \sim N(\theta, [I_n(\theta)]^{-1})$, asintoticamente anche per la funzione *score* si ha: $S(\theta) \sim N(0, I_n(\theta))$

Problemi degli stimatori ML

▸A volte si incontrano difficoltà computazionali

- Può non esistere una soluzione esplicita ⇒ ricorso a metodi numerici iterativi;
- la log-verosimiglianza può essere discontinua o avere una derivata prima discontinua o ammetere il massima in un punto estremo.

Metodo dei minimi quadrati

▸ Dato un campione casuale $\mathbf{X}$, con X_i i.i.d. e $\mathbf{X} = \mathbf{A}\boldsymbol{\theta} + \varepsilon$, dove $\mathbf{A}$ è una matrice nota a pieno rango di dimensione $(n \times p)$ e ε è un vettore casuale con media nulla e matrice di covariana $V(\varepsilon) = \mathbb{E}(\varepsilon\varepsilon') = \mathbf{V}\sigma^2$, $\mathbf{V}$ simmetrica e nota in base al **Principio dei minimi quadrati**: possiamo stimare θ minimizzando la forma quadratica:

$$\boxed{(\mathbf{x} - \mathbf{A}\boldsymbol{\theta})'\mathbf{V}^{-1}(\mathbf{x} - \mathbf{A}\boldsymbol{\theta})}$$

▸ Si ottiene lo stimatore dei MQ:

$$\boxed{\hat{\boldsymbol{\theta}}_{MQ} = (\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}\mathbf{X}}$$

Con matrice di covarianza:

$$\boxed{\hat{\boldsymbol{\theta}}_{MQ} = (\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\sigma^2}$$

▸ Utile in modelli lineari, per i quali fornisce stimatori **lineari**, **corretti** e a **varianza minima**.

Approfondimenti

- generalizzazione dei MQ ordinari (GLS, Greene, cap. 11)
- metodo dei momenti e sue generalizzazioni (Piccolo, 16.2, pp. 581-585; Greene, cap. 11)
- metodi di simulazione (p.e. Simulated makimum likelihood)
- metodi bayesiani (p.e. MCMC)

Test delle ipotesi

Genericamente diamo le seguenti **definizioni**:

- un'*ipotesi* è un'affermazione che specifica completamente o parzialmente la distribuzione di probabilità di una v.c. X.
- un *TEST statistico* è una procedura che consente di rispondere a domande del tipo: è ragionevole pensare che $\mu = \mu_0$?
- In maniera più formale: un test statistico è una **regola di decisione** istituita sullo spazio campionario $\mathcal{X}$ mediante la quale, in funzione del campione osservato $\mathbf{x}$, si decide se rifiutare o meno un'ipotesi statistica riferta alla popolazione.

Indice argomenti:

- tipi di test
- criteri per rappresentare e valutare un test statistico: è possibile trovare un test *ottimo*?
- metodi generali per la costruzione di test e loro proprietà

Come si costruisce un TEST statistico

- si formula un'ipotesi su θ, H_0, che può essere:
  - un valore specifico $\theta = \theta_0$
  - un insieme di valori $\theta \epsilon \omega, \omega \subset \Theta$, p.e. $\theta > \theta_0$
- I dati supportano questa ipotesi o suggeriscono di abbandonarla?

I vari **tipi di test** statistici si distinguono per:

1. utilizzo dell'informazione campionaria:
   (a) per misurare l'evidenza statistica contro H_0
   (b) (a) $+$ se l'evidenza empirica contro H_0 è sufficientemente forte, *si rifiuta* H_0 come inappropriata
2. si prevede un'ipotesi alternativa H_1 da adottare se H_0 viene rifiutata
3. si costruisce il test e se ne verificano le proprietà in relazione ad H_1

Tipi di test

- **Test di significatività puro**: procedura che si limita a valutare l'evidenza contro H_0. La misura dell'evidenza contro H_0 è detta *livello di significatività* del test (Cox e Hinkley, 1974).
- **Test di significatività**: procedura che individua una regola per rifiutare H_0, dove per *livello di significatività* si intende la probabilità di rifiutare H_0 quando questa è vera applicando la regola data più volte in circostanze simili. L'evidenza contro H_0 nel caso in esame è invece detta *livello critico* del test.
- **Test delle ipotesi**: regola per riufiutare H_0 in favore di H_1 (Neyman e Pearson, 1933). Considera, oltre al livello di significatività, anche il rischio di accettare H_0 quando questa è falsa e la *potenza* del test. Il test delle ipotesi ha sia una funzione inferenziale che di tipo decisionale.

Ipotesi nulla e ipotesi alternativa

▶ L'affermazione da verificare tramite il test statistico è detta **ipotesi nulla** H_0.

▶ Il test è costruito in modo da valutare quanto sia forte l'evidenza fornita dai dati contro H_0.

L'ipotesi nulla può essere:

- **semplice**: specifica completamente la distribuzione di $\mathbf{X}$ su $\mathcal{X}$.
- **composta**: non specifica completamente la distribuzione.

Esempio

Sia $\mathbf{X}$ un campione casuale di dimensione n, con $X \sim N(\mu, \sigma^2)$:

(i) H_0 semplice: $\mu = \mu_0, \sigma = \sigma_0$

(ii) H_0 composta: $\mu = \mu_0$, $\sigma \epsilon (0, \infty)$.

▶ L'ipotesi da contrastare ad H_0 è detta **ipotesi alternativa** H_1.

Test delle ipotesi

In un problema di test delle ipotesi, dopo aver osservato $\mathbf{X} = \mathbf{x}$, lo sperimentatore deve decidere se accettare H_0 come vera o rifiutarla come falsa decidendo che H_1 è vera.

▶Un **test delle ipotesi** è una **regola** che specifica:

(i) per quali valori campionari si prende la decisione di accettare H_0, individuando la *regione di accetazione* del test

(ii) per quali valori campionari si rifiuta H_0 e si accetta H_1 come vera, individuando la *regione di rifiuto* o regione critica del test

▶Un test delle ipotesi è di solito specificato in termini di una **statistica test** $W(\mathbf{X})$ funzione dei dati campionari.

Esempio

Sia $H_0 : \mu = \mu_0$. La procedura di test specifica che H_0 viene rifiutata se la media campionaria è maggiore di 3: $W(\mathbf{X}) = \overline{X}$, *regione di rifiuto* $\{(x_1, x_2, \ldots, x_n) : \overline{x} > 3\}$.

Tipi di errore

In un test delle ipotesi

$$H_0 : \theta \epsilon \Theta_0 \qquad \textit{contro} \qquad H_1 : \theta \epsilon \Theta_0^c$$

$\Theta = \Theta_0 \cup \Theta_0^c$

si possono commettere **due tipi di errore**:

- *Errore di I tipo*: rifiutare H_0 quando è vera, cioè quando $\theta \epsilon \Theta_0$
- *Errore di II tipo*: accettare H_0 quando è falsa, cioè quando $\theta \epsilon \Theta_0^c$

		Decisione	
		Accetto H_0	Rifiuto H_0
Verità	H_0	decisione corretta	errore I tipo
	H_1	errore II tipo	decisione corretta

Le diverse procedure di test delle ipotesi vengono valutate in base alla probabilità di commettere questi due tipi di errore.

Probabilità di errore

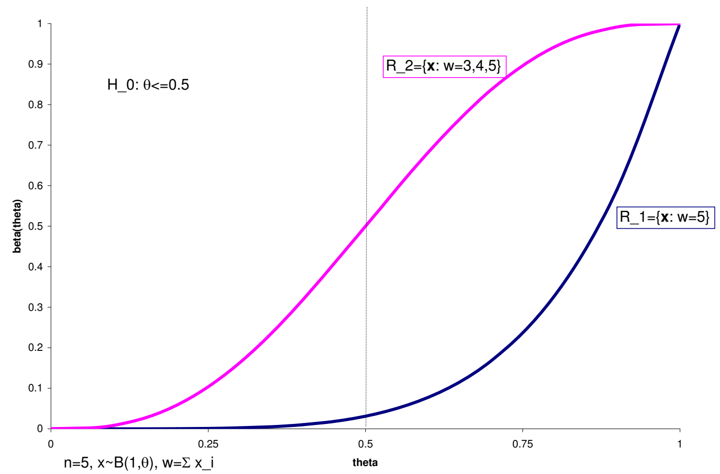
Indichiamo con R la *regione di rifiuto* del test e con R^c la *regione di accettazione*.

- Per $\theta \in \Theta_0$, il test sbaglia quando $\mathbf{x} \epsilon R$: la probabilità di commettere un *errore di I tipo* (rifiuto H_0 quando è vera) è pari a $P(\mathbf{x} \epsilon R \mid H_0)$.
- La probabiltà di commettere un *errore di II tipo* (accetto H_0 quando è falsa) è invece pari a $P(\mathbf{x} \in R^c \mid H_1)$.
- non si commette errore se si rifiuta H_0 quando questa è falsa. La probabilità di rifiutare H_0 quando è falsa è pari a: $P(\mathbf{x} \in R \mid H_1)$ ed è detta *potenza del test*

Per un test con regione di rifiuto R, la funzione $\beta(\theta) = P_\theta(\mathbf{x} \in R)$ è detta *funzione potenza del test*. Il test ideale ha una funzione potenza pari a 0 per $\theta \in \Theta_0$ e pari a 1 per $\theta \in \Theta_0^c$. In generale, $\beta(\theta)$ dipende dalla dimensione campionaria n.

Esempio: $X \sim (1, \theta) \Rightarrow \mathbf{X} \sim (n, \theta)$ Binomiale

Funzione potenza del test



(*si veda potenza_test.xls*)

Quale test scegliere?

▶Per n fissato è impossibile minimizzare sia l'errore di I tipo che l'errore di II tipo (per ridurre la prob di errore di I tipo dobbiamo rifiutare meno spesso, cioè ridurre l'ampiezza di R, questo induce ad accettare H_0 con maggiore frequenza, per cui aumenta la prob di errore di II tipo).

▶ In generale, si fissa l'errore di I tipo e si cerca il test che ha l'errore di II tipo più piccolo.

Definizioni:

- Un test con funzione potenza $\beta(\theta)$ è un test di dimensione α se $\sup_{\theta\epsilon\Theta_0} \beta(\theta) = \alpha$, con $0 \leq \alpha \leq 1$.
- Un test con funzione potenza $\beta(\theta)$ è un test di livello α se $\sup_{\theta\epsilon\Theta_0} \beta(\theta) \leq \alpha$, con $0 \leq \alpha \leq 1$.
- Si dice *p-value* il livello di significatività più piccolo osservato dato il campione **x**: $P(\mathbf{x}\epsilon R \mid \theta\epsilon\Theta_0) = \widehat{\alpha}(\mathbf{x})$.

p-value

- La probabilità, calcolata per H_0 vera, che la statistica test assuma un valore uguale o più *estremo* di quello osservato è detta **p-value**.
- Più è piccolo il p-value, più forte è l'evidenza empirica contro H_0.
- Per calcolare il **p-value** è necessario conoscere la distribuzione campionaria della statistica test.

Esempio: $X \sim N(\mu, \sigma^2 = 9)$, $n = 4$, $\mathbf{x} = \{190.5\ 189.0\ 195.5\ 187.0\}$, $\overline{x} = 190.5$

$$H_0 : \mu = 187 \text{ vs } H_1 : \mu > 187$$

▶ $\overline{X} \mid H_0 \sim N(187, 9/4)$, *p-value*: $P(\overline{x} \geq 190.5 \mid \mu = 187)$

▶ sotto H_0, possiamo standardizzare $\overline{X}$, $Z \sim N(0,1)$:
$z = \frac{190.5-187}{3/2} = 2.3$, $P(z \geq 2.3) = 1 - 0.9901 = 0.01$

Test uniformemente più potente

Regione critica ottimale Dato il livello del test α, si definisce *ottimale* una regione critica R che possiede la potenza più elevata fra tutte quelle di pari ampiezza α. Cioè, data $P(\mathbf{X} \in R \mid H_0) = \alpha$, per qualsiasi altra regione critica R' di uguale ampiezza α si ha:

$$\boxed{P(\mathbf{X} \in R \mid H_1) > P(\mathbf{X} \in R' \mid H_1)}$$

★Essendo preoccupati dell'errato rifiuto di H_0, la *regione critica ottimale* cautela ad un **livello noto** contro l'**errore di I tipo** (rifiuto H_0 quando è vera) e, tra le regioni critiche che soddisfano tale requisito, sceglie quella **più potente** (rifiuta H_0 quando è falsa).

★Di solito si fissa α in base alla *cautela* contro l'errore di I tipo che si ritiene più opportuna:

modesta $\alpha = 0.05$ pari al 5% (1 su 20)

elevata $\alpha = 0.01$ pari al 1%

notevole $\alpha = 0.001$ pari al 1 per mille

Lemma di Neyman-Pearson

Sia $\mathbf{X} = (X_1, \ldots, X_n)$ un campione casuale generato da $X \sim f(x;\theta)$ e si voglia verificare

$$H_0 : \theta = \theta_0 \textit{ contro } H_1 : \theta = \theta_1$$

Se $\mathcal{L}(\theta; \mathbf{X})$ è la funzione di verosimiglianza di $\mathbf{X}$, allora la *regione critica ottimale* al livello α per H_0 contro H_1 è quella regione R dello spazio campionario $\mathcal{X}$ che soddisfa le due seguenti condizioni :

$$\boxed{\text{(i) } \frac{\mathcal{L}(\theta_1;\mathbf{X})}{\mathcal{L}(\theta_0;\mathbf{X})} \geq c \qquad \text{(ii) } P(\mathbf{X} \in R \mid H_0) = \alpha}$$

- la regione ottimale può non essere unica, il lemma indivua *una* delle possibili R;
- il valore di c è determinato dall'ampiezza α della R;
- il lemma risponde al *principio di verosimiglianza*.

Statistica sufficiente e lemma di N-P

La R individuata dal lemma di Neyman-Pearson è funzione del campione casuale *solo* attraverso la statistica sufficiente minimale $T(\mathbf{X})$ per θ quando essa esiste. Per il teorema di fattorizzazione si ha:

$$\boxed{\frac{\mathcal{L}(\theta_1;\mathbf{X})}{\mathcal{L}(\theta_0;\mathbf{X})} = \frac{g(T(\mathbf{X});\theta_1)h(\mathbf{X})}{g(T(\mathbf{X});\theta_0)h(\mathbf{X})} = \frac{g(T(\mathbf{X});\theta_1)}{g(T(\mathbf{X});\theta_0)} = \eta(\mathbf{X};\theta_0,\theta_1)}$$

Esempio

Test per la media di una distribuzione normale con varianza nota $X \sim N(\theta, \sigma^2)$

$$\boxed{H_0 : \theta = \theta_0 \text{ contro } H_1 : \theta = \theta_1, \text{ con } \theta_0 < \theta_1}$$

$$\boxed{R = \{\mathbf{x} : \overline{X} > \theta_0 + z_\alpha \tfrac{\sigma}{\sqrt{n}}\}}$$

Test del rapporto di verosimiglianza

►Il Lemma di Neyman-Pearson consente di costruire test più potenti di ampiezza α prefissata per un'ipotesi nulla semplice contro un'alternativa semplice o unidirezionale.

►Il **test del rapporto di verosimiglianza** (LRT) consente di costruire test per ipotesi H_0 e H_1 completamente generali.

Definizione. *LRT* per $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_0^c$ è

$$\boxed{\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})}.}$$

Un **LRT** è un qualunque test la cui *regione di rifiuto* ha la forma

$$\boxed{R = \lambda(\mathbf{x}) \leq c, \forall c : 0 \leq c \leq 1.}$$

► $\lambda(\mathbf{x})$ è piccolo se ci sono punti di θ in Θ_0^c che rendono $\mathbf{x}$ più verosimile che per ogni altro punto in Θ_0. In questo caso il criterio LRT suggerisce di rifiutare H_0 in favore di H_1.

LRT e statistica sufficiente

Teorema Se $T(\mathbf{X})$ è una statistica sufficiente per θ e $\lambda(\mathbf{x})$ e $\lambda^*(t)$ sono le statistiche LRT basate rispettivamente su $\mathbf{X}$ e T allora $\lambda^*(T(\mathbf{X})) = \lambda(\mathbf{x})$ per ogni $\mathbf{x} \in \mathcal{X}$.

Esempio:

$X_1, \ldots, X_n$, con $X \sim N(\mu, \sigma^2)$, σ^2 nota

► $H_0 : \mu = \mu_0$, di dimensione α: $P(x \in R \mid H_0) = \alpha$

►Statistica test $\overline{X}$

► La regione critica *R* varia in base all'ipotesi alternativa:

- $H_1 : \mu > \mu_0,\ R = \{\mathbf{x} : \overline{X} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}\}$
- $H_1 : \mu < \mu_0,\ R = \{\mathbf{x} : \overline{X} \leq \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}\}$
- $H_1 : \mu \neq \mu_0,\ R = \{\mathbf{x} : \overline{X} \leq \mu_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \overline{X} \geq \mu_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\}$

Test sulla media di una v.c. Normale

$X_1, \ldots, X_n$, con $X \sim N(\mu, \sigma^2)$, σ^2 incognita.

► $H_0 : \mu = \mu_0$, di dimensione α: $P(x \in R \mid H_0) = \alpha$

►Statistica test $\overline{X}$, $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2$.

► La regione critica *R* varia in base all'ipotesi alternativa:

- $H_1 : \mu > \mu_0,\ R = \{\mathbf{x} : \overline{X} \geq \mu_0 + t_{(\alpha,g)} \frac{S}{\sqrt{n}}\}$
- $H_1 : \mu < \mu_0,\ R = \{\mathbf{x} : \overline{X} \leq \mu_0 - t_{(\alpha,g)} \frac{S}{\sqrt{n}}\}$
- $H_1 : \mu \neq \mu_0,\ R = \{\mathbf{x} : \overline{X} \leq \mu_0 - t_{(\alpha/2,g)} \frac{S}{\sqrt{n}}, \overline{X} \geq \mu_0 + t_{(\alpha/2,g)} \frac{S}{\sqrt{n}}\}$

► I quantili $t_{(\alpha,g)}$ sono relativi ad una v.c. $\mathcal{T}$ di Student con $g = n - 1$ gradi di libertà.

► Per n grande ($n > 100$) la distribuzione della v.c. $\mathcal{T}$ di Student è approssimata dalla $N(\mu, \sigma^2) \Rightarrow$ i quantili $t_{(\alpha,g)}$ possono essere sostituiti dai quantili z_α.