



UNIVERSITÀ
DEGLI STUDI DI TRIESTE



Dipartimento di scienze economiche,
aziendali, matematiche e statistiche
"Bruno de Finetti"

Statistica (c.p.)

3. Verifica d'ipotesi

Francesco Pauli

DEAMS

Università di Trieste

A.A. 2017/2018

Indice

- 1 Generalità
- 2 Test connessi alla verosimiglianza
- 3 Regioni di confidenza
- 4 Verosimiglianza profilo
- 5 Esempi vari

Verifica d'ipotesi: il contesto

In un contesto sperimentale “classico” si può confutare una teoria osservando un risultato che è incompatibile con essa.

Ad esempio:

- ho la teoria secondo cui la velocità con cui un oggetto cade a terra non dipende dalla sua massa
- metto alla prova la teoria facendo cadere dalla stessa altezza oggetti di massa diversa (nel vuoto) e misurando il tempo che impiegano per arrivare a terra
- (supponiamo che le misure siano esatte)
- se i tempi sono diversi, la teoria è confutata

Verifica d'ipotesi: il contesto

Quando l'esperimento su cui si fonda la verifica non ha un risultato deterministico, è possibile che tutti i risultati osservabili siano compatibili con la teoria.

Ad esempio nell'esperimento illustrato la mia misura del tempo impiegato ad arrivare a terra è soggetta a errore casuale, allora i tempi saranno certamente diversi per effetto dell'errore.

La teoria però associa a ciascuno di essi una diversa probabilità.

Il risultato sperimentale non consente allora di confutare con certezza una teoria ma possiamo

- misurare il grado di incompatibilità (quanto l'osservazione è sorprendente alla luce della teoria)
- decidere di accettare o rifiutare la teoria tenendo sotto controllo la probabilità di errore

Verifica d'ipotesi statistica parametrica

Si ha un modello statistico parametrico

$$(\mathcal{Y}, p_{\theta}, \Theta)$$

si formula un'ipotesi su quale sia il meccanismo generatore dei dati, cioè il vero valore del parametro

$$\text{Ipotesi nulla: } H_0 : \theta = \theta_0$$

In generale, qualunque campione sarà compatibile con θ_0 , nel senso che

$$p_{\theta_0}(y) > 0 \quad \forall y \in \mathcal{Y}$$

Esistono allora due approcci

- approccio di Fisher: valore p , misura il grado di incompatibilità/lontananza del campione con θ_0
- approccio di Neyman-Fisher: regioni di rifiuto, è una regola che prescrive per quali campioni y si rifiuta la teoria, formulata in modo sia bassa la probabilità di rifiutare quando la teoria è vera

Approccio di Fisher

Fisher suggerì di misurare quanto compatibile è il dato sperimentale con la teoria con la **probabilità di osservare un campione almeno altrettanto estremo se è vera l'ipotesi nulla**, questa probabilità è detta valore p (p -value).

Più specificamente la procedura di Fisher è

- 1 specificare lo spazio campionario ed individuare una statistica sufficiente $t(\cdot)$ ed il relativo spazio;
- 2 calcolare la distribuzione di $t(\cdot)$ nell'ipotesi da testare: $P_0(t)$;
- 3 una volta osservato il risultato y^* , corrispondente al riassunto t^* si considera l'insieme dei risultati meno probabili di t^*

$$\mathcal{T} = \{t | P_0(t) \leq P_0(t^*)\}$$

e si calcola il **valore** p

$$p^* = P_0(t \in \mathcal{T})$$

Approccio di Neyman Pearson

Oltre all'ipotesi nulla si ha un'ipotesi alternativa

$$\text{Ipotesi alternativa: } H_1 : \theta = \theta_1$$

Neyman e Pearson suggerirono di determinare una regione dello spazio campionario \mathcal{R} in corrispondenza alla quale si rifiuta H_0 tenendo sotto controllo la probabilità di errore.

La procedura è

- 1 determinare lo spazio campionario (eventualmente una statistica $t(\cdot)$);
- 2 determinare la distribuzione di Y nelle due ipotesi: P_0 e P_1 ;
- 3 fissare un **livello di significatività** α (basso) e determinare \mathcal{R} t.c.

$$P_0(Y \in \mathcal{R}) \leq \alpha; \quad P_1(Y \in \mathcal{R}) \text{ grande}$$

in questo modo si tiene sotto controllo la probabilità (condizionata alle ipotesi) di errore.

Approccio di Neyman Pearson

Oltre all'ipotesi nulla si ha un'ipotesi alternativa

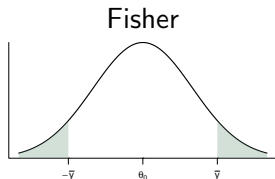
Ipotesi alternativa: $H_1 : \theta = \theta_1$

Neyman e Pearson suggerirono di determinare una regione dello spazio campionario \mathcal{R} in corrispondenza alla quale si rifiuta H_0 tenendo sotto controllo la probabilità di errore.

θ y	$y \notin \mathcal{R}$	$y \in \mathcal{R}$
H_0	✓	errore I tipo $\alpha = P_0(Y \in \mathcal{R})$
H_1	errore II tipo $1 - \beta = P_1(Y \notin \mathcal{R})$	✓

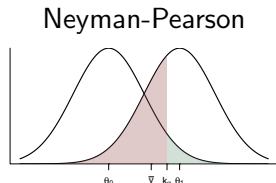
Approcci in parallelo

Modello: $Y_1, \dots, Y_n \sim IID(\mathcal{N}(\theta, 1))$, statistica $\bar{Y} \sim \mathcal{N}(\theta, n^{-1})$



$$H_0 : \theta = \theta_0$$

$$p = 2(1 - \Phi_{\theta_0, n^{-1}}(|\bar{y}|))$$



$$H_0 : \theta = \theta_0; \quad H_1 : \theta = \theta_1$$

$$\mathcal{R}_\alpha = \{\bar{Y} > k_\alpha = \Phi_{\theta_0, n^{-1}}^{-1}(1 - \alpha)\}$$

$$P_0(\bar{Y} \in \mathcal{R}_\alpha) = \alpha$$

$$\beta := P_1(\bar{Y} \in \mathcal{R}_\alpha) = 1 - \Phi_{\theta_1, n^{-1}}(k_\alpha)$$

Approcci in parallelo

Fisher

Fornisce una misura (tra 0 e 1) di quanto sorprendente è un risultato.

—

Fa riferimento alla sola ipotesi nulla.

—

Non c'è controllo dell'errore.

Neyman-Pearson

Fornisce un risultato dicotomico (accetto/rifiuto) l'ipotesi nulla.

—

Si basa sul confronto ipotesi nulla/alternativa.

—

Tiene sotto controllo la frequenza (condizionata) di errori in una sequenza di test.

Incompatibilità tra i due approcci

C'è un legame tecnico tra i due approcci: se i test sono egualmente formulati infatti il valore p di Fisher è il minimo livello di significatività con cui si rifiuta l'ipotesi nulla secondo la procedura di Neyman-Pearson.

I due approcci rimangono incompatibili, in particolare

- è scorretto attribuire a un valore p un significato in termini di probabilità di errore
- è scorretto dedurre, in una procedura di Neyman-Pearson, una misura dell'evidenza contro l'ipotesi nulla in un particolare esperimento.

Ruolo della verifica d'ipotesi

Abbiamo presentato il test statistico come metodo di falsificazione di una teoria (ipotesi scientifica): esso è in tal caso il risultato finale dell'analisi statistica.

Lo strumento del test di ipotesi può essere impiegato anche in fasi intermedie dell'analisi, ad esempio nelle fasi di selezione del modello: ci si basa sui test, ad esempio, nella selezione delle variabili esplicative nella regressione/GLM.

Ancora, i test possono essere usati per il controllo dell'adeguatezza di un modello, l'ipotesi da verificare in tal caso è un assunto alla base del modello stesso, in caso di rifiuto/valore p basso si conclude che il modello non è adeguato (ad es. test di normalità sui residui di una regressione).

Definizione di sistema d'ipotesi

Sia dato modello parametrico $(\mathcal{Y}, p_\theta, \Theta)$, un **sistema d'ipotesi** è una partizione dello spazio parametrico Θ in due costituenti Θ_0 e $\Theta_1 = \Theta \setminus \Theta_0$.

Schematicamente un sistema d'ipotesi si indica con

$$\begin{cases} H_0 : \theta \in \Theta_0 & \text{(ipotesi nulla)} \\ H_1 : \theta \in \Theta_1 & \text{(ipotesi alternativa)} \end{cases}$$

Un'ipotesi si dice “semplice” se specifica completamente una distribuzione in \mathcal{F} (Θ_0 è un singoletto), composta altrimenti.

Con questa definizione, l'ipotesi alternativa è sempre implicitamente definita.

Test di Neyman-Pearson

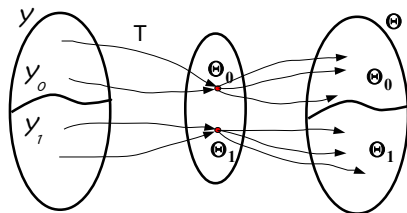
Un **test** d'ipotesi alla Neyman-Pearson è una funzione

$$T(y) : \mathcal{Y} \rightarrow \{\Theta_0, \Theta_1\},$$

Il test induce una partizione dello spazio campionario in

$$\mathcal{Y}_0 = T^{-1}(\Theta_0) \quad \text{regione di accettazione}$$

$$\mathcal{Y}_1 = T^{-1}(\Theta_1) \quad \text{regione di rifiuto}$$



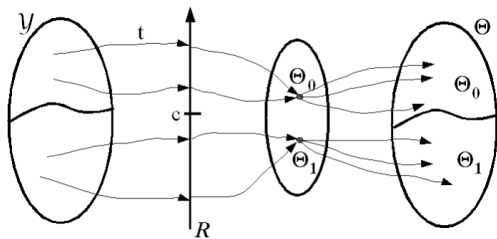
Test di Neyman-Pearson

Spesso il test si basa su una statistica, o funzione test,

$$t : \mathcal{Y} \rightarrow \mathbb{R}$$

il cui valore determina il valore del test, ossia è tale che

$$T(y) = \Theta_0 \Leftrightarrow t(y) \in A \subset \mathbb{R}$$



Esempio delle due urne (proporzione)

Supponiamo di avere due urne con N palline bianche e nere, la proporzione delle bianche è 0.5 in una, 0.7 nell'altra.

Estraiamo, con rimpiazzo, n palline da una delle urne e sia

$$t = \frac{1}{n} \# \{\text{bianche nel campione}\}$$

si ha $nt \sim \text{Binom}(n, \theta)$.

In termini di verifica d'ipotesi si ha il sistema

$$H_0 : \theta = 0.5; \quad H_1 : \theta = 0.7$$

Intuitivamente risulta naturale definire il test

$$T(y) = \{0.7\} \Leftrightarrow t(y) > k$$

dove k è un reale di $[0, 1]$ scelto opportunamente. Vedremo poi come come questa scelta intuitiva possa essere giustificata formalmente.

Funzione potenza

Si dice **funzione potenza** la probabilità di rifiutare l'ipotesi in funzione di θ

$$\beta(\theta) = P_{\theta}(T(Y) \in \mathcal{Y}_1)$$

Il livello di significatività del test è allora

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$$

La probabilità di errore di II tipo è, al variare di $\theta \in \Theta_1$,

$$1 - \beta(\theta)$$

Proprietà di un test

Tra le proprietà desiderabili di un test di ipotesi si definiscono

- **non distorsione**: un test si dice **non distorto** se

$$P_{\theta}(Y \in \mathcal{Y}_1) \leq \alpha \quad \forall \theta \in \Theta_0$$

$$P_{\theta}(Y \in \mathcal{Y}_1) \geq \alpha \quad \forall \theta \in \Theta_1$$

- **consistenza**: se $\beta_n(\theta)$ è la funzione potenza di un test di ipotesi basato su un campione IID di numerosità n diremo che il test è consistente sse

$$\lim_{n \rightarrow \infty} \beta_n(\theta) = 1; \quad \forall \theta \in \Theta_1$$

Esempio delle due urne (proporzione)

Consideriamo l'esempio delle due urne con $n = 10$, allora

$$\beta(\theta) = P_{\theta} \left(\frac{S(Y)}{10} > k \right) = 1 - F_{\theta}(10k)$$

se F_{θ} è la f.d.r. della Binom($n = 10, \theta$).

Nella tabella che segue riportiamo le funzioni potenza relative ai test che possiamo definire variando k :

k	1	2	3	4	5	6	7	8	9	10
$\beta(0.5)$	0.999	0.989	0.945	0.828	0.623	0.377	0.172	0.055	0.011	0.001
$\beta(0.7)$	1.000	0.999	0.998	0.989	0.953	0.850	0.650	0.383	0.150	0.028

Esempio: media di una popolazione normale

Sia $Y = (Y_1, \dots, Y_n) \sim IID(\mathcal{N}(\theta, 1))$ e consideriamo il sistema di ipotesi

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1 (> \theta_0)$$

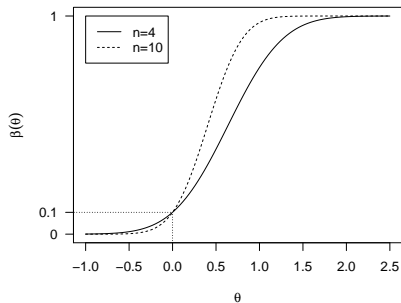
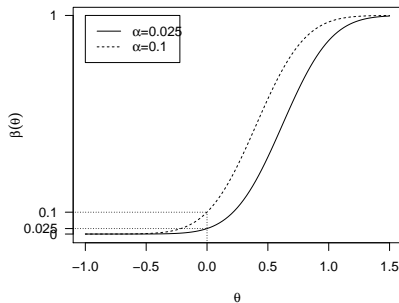
Consideriamo il test definito da

$$T(Y) = \{\theta_1\} \Leftrightarrow \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} > k$$

Essendo $\bar{Y} \sim \mathcal{N}(\theta, \frac{1}{n})$ la funzione potenza del test è

$$\beta(\theta) = 1 - \Phi(\sqrt{n}(k - \theta))$$

Esempio: media di una popolazione normale



Potenza del test per la media della normale al variare del livello di significatività e della numerosità campionaria.

Obiettivo e compromesso

In linea generale, si vorrebbe $\beta(\theta)$ grande per $\theta \in \Theta_1$ e bassa per $\theta \in \Theta_0$.

Le due esigenze sono in contrasto, quindi si procede

- fissando il livello di significatività
- massimizzando (ove possibile) $\beta(\theta)$ in Θ_1 col vincolo del livello di significatività.

Si parla di test uniformemente più potente di livello α quando si individua un test con associata $\beta^*(\theta)$ tale che

$$\beta^*(\theta) > \beta(\theta) \quad \forall \theta \in \Theta_1$$

dove $\beta(\cdot)$ è la funzione potenza di qualunque test di livello α .

Lemma di Neyman-Pearson

Lemma di Neyman-Pearson

Dato il modello parametrico $(\mathcal{Y}, p_\theta; \Theta)$ e il sistema d'ipotesi

$$\begin{cases} H_0 : \theta = \theta_0 & (\text{ipotesi nulla}) \\ H_1 : \theta = \theta_1 & (\text{ipotesi alternativa}). \end{cases}$$

il test avente come regione di rifiuto

$$\mathcal{Y}_1 = \left\{ y \mid \lambda^*(y) = \frac{L(\theta_0; y)}{L(\theta_1; y)} \leq \lambda_\alpha \right\}$$

dove $P(\lambda^*(y) \leq \lambda_\alpha; \theta_0) = \alpha$ è quello più potente tra i test di livello non superiore ad α .

Dimostrazione del lemma di Neyman-Pearson

Sia \mathcal{Y}_1 la zona di rifiuto del test di Neyman-Pearson, e \mathcal{Y}_A quella di un generico test di livello α , allora

- (i) $\lambda^*(y) = \frac{L(\theta_0; y)}{L(\theta_1; y)} \leq \lambda_\alpha \quad \forall y \in \mathcal{Y}_1$;
- (ii) $\alpha = \int_{\mathcal{Y}_1} f(y; \theta_0) d\nu(y) \geq \int_{\mathcal{Y}_A} f(y; \theta_0) d\nu(y) \quad \forall \mathcal{Y}_A \neq \mathcal{Y}_1$

La tesi che vogliamo dimostrare è che sotto queste ipotesi

$$\int_{\mathcal{Y}_1} f(y; \theta_1) d\nu(y) \geq \int_{\mathcal{Y}_A} f(y; \theta_1) d\nu(y) \quad \forall \mathcal{Y}_A \neq \mathcal{Y}_1$$

che equivale a

$$\int_{\mathcal{Y}_1 \cap \mathcal{Y}_A} f(y; \theta_1) d\nu(y) + \int_{\mathcal{Y}_1 \cap \bar{\mathcal{Y}}_A} f(y; \theta_1) d\nu(y) \geq \int_{\mathcal{Y}_1 \cap \mathcal{Y}_A} f(y; \theta_1) d\nu(y) + \int_{\bar{\mathcal{Y}}_1 \cap \mathcal{Y}_A} f(y; \theta_1) d\nu(y)$$

che equivale, togliendo a entrambi i membri il primo addendo,

$$\int_{\mathcal{Y}_1 \cap \bar{\mathcal{Y}}_A} f(y; \theta_1) d\nu(y) \geq \int_{\bar{\mathcal{Y}}_1 \cap \mathcal{Y}_A} f(y; \theta_1) d\nu(y)$$

Dimostrazione del lemma di Neyman-Pearson

Sia \mathcal{Y}_1 la zona di rifiuto del test di Neyman-Pearson, e \mathcal{Y}_A quella di un generico test di livello α , allora

$$(i) \quad \lambda^*(y) = \frac{L(\theta_0; y)}{L(\theta_1; y)} \leq \lambda_\alpha \quad \forall y \in \mathcal{Y}_1;$$

$$(ii) \quad \alpha = \int_{\mathcal{Y}_1} f(y; \theta_0) d\nu(y) \geq \int_{\mathcal{Y}_A} f(y; \theta_0) d\nu(y) \quad \forall \mathcal{Y}_A \neq \mathcal{Y}_1$$

La tesi che vogliamo dimostrare è che sotto queste ipotesi

$$\int_{\mathcal{Y}_1 \cap \bar{\mathcal{Y}}_A} f(y; \theta_1) d\nu(y) \geq \int_{\bar{\mathcal{Y}}_1 \cap \mathcal{Y}_A} f(y; \theta_1) d\nu(y)$$

Dimostrazione del lemma di Neyman-Pearson

Sia \mathcal{Y}_1 la zona di rifiuto del test di Neyman-Pearson, e \mathcal{Y}_A quella di un generico test di livello α , allora

- (i) $\lambda^*(y) = \frac{L(\theta_0; y)}{L(\theta_1; y)} \leq \lambda_\alpha \quad \forall y \in \mathcal{Y}_1;$
- (ii) $\alpha = \int_{\mathcal{Y}_1} f(y; \theta_0) d\nu(y) \geq \int_{\mathcal{Y}_A} f(y; \theta_0) d\nu(y) \quad \forall \mathcal{Y}_A \neq \mathcal{Y}_1$

La tesi che vogliamo dimostrare è che sotto queste ipotesi

$$\int_{\mathcal{Y}_1 \cap \bar{\mathcal{Y}}_A} f(y; \theta_1) d\nu(y) \geq \int_{\bar{\mathcal{Y}}_1 \cap \mathcal{Y}_A} f(y; \theta_1) d\nu(y)$$

Si ha poi

- (i) $\Rightarrow \int_{\mathcal{Y}_1 \cap \bar{\mathcal{Y}}_A} f(y; \theta_1) d\nu(y) \geq \lambda_\alpha^{-1} \int_{\mathcal{Y}_1 \cap \bar{\mathcal{Y}}_A} f(y; \theta_0) d\nu(y)$
- $\int_{\bar{\mathcal{Y}}_1 \cap \mathcal{Y}_A} f(y; \theta_1) d\nu(y) \leq \lambda_\alpha^{-1} \int_{\bar{\mathcal{Y}}_1 \cap \mathcal{Y}_A} f(y; \theta_0) d\nu(y)$

condizione sufficiente per la tesi è dunque che

$$\lambda_\alpha^{-1} \int_{\mathcal{Y}_1 \cap \bar{\mathcal{Y}}_A} f(y; \theta_0) d\nu(y) \geq \lambda_\alpha^{-1} \int_{\bar{\mathcal{Y}}_1 \cap \mathcal{Y}_A} f(y; \theta_0) d\nu(y)$$

che equivale a

$$\int_{\mathcal{Y}_1 \cap \mathcal{Y}_A} f(y; \theta_0) d\nu(y) + \int_{\mathcal{Y}_1 \cap \bar{\mathcal{Y}}_A} f(y; \theta_0) d\nu(y) \geq \int_{\bar{\mathcal{Y}}_1 \cap \mathcal{Y}_A} f(y; \theta_0) d\nu(y) + \int_{\bar{\mathcal{Y}}_1 \cap \bar{\mathcal{Y}}_A} f(y; \theta_0) d\nu(y)$$

che è poi l'ipotesi (ii) CVD.

Esempio delle due urne

Il rapporto λ^* risulta pari a

$$\lambda^* = \frac{0.5^{S(Y)} 0.5^{n-S(Y)}}{0.7^{S(Y)} 0.3^{n-S(Y)}}$$

e pertanto, secondo il lemma di Neyman Pearsonn, il test ottimale è del tipo

$$\frac{0.5^{S(Y)} 0.5^{n-S(Y)}}{0.7^{S(Y)} 0.3^{n-S(Y)}} < k$$

il che equivale, passando ai logaritmi e raccogliendo $S(Y)$, a

$$S(Y) \log \left(\frac{0.5 \times 0.3}{0.7 \times 0.3} \right) < k'$$

essendo $\log \left(\frac{0.5 \times 0.3}{0.7 \times 0.3} \right) < 0$ la forma finale della zona di rifiuto è $\{y | S(y) > k''\}$, cioè il test già proposto su base intuitiva nello sviluppare l'esempio.

Esempio delle due urne

In generale, se il sistema di ipotesi è del tipo

$$\begin{cases} H_0 : \theta = \theta_0 & \text{(ipotesi nulla)} \\ H_1 : \theta = \theta_1 & \text{(ipotesi alternativa)}. \end{cases}$$

si ottiene

$$\lambda^* = \frac{\theta_0^{S(Y)}(1 - \theta_0)^{n-S(Y)}}{\theta_1^{S(Y)}(1 - \theta_1)^{n-S(Y)}}$$

e quindi

$$S(Y) \log \left(\frac{\theta_0(1 - \theta_1)}{\theta_1(1 - \theta_0)} \right) < k'$$

dove notiamo che se $\theta_1 > \theta_0$ allora $\log \left(\frac{\theta_0(1 - \theta_1)}{\theta_1(1 - \theta_0)} \right) < 0$ e quindi la zona di rifiuto è $S(Y) > k''$, viceversa se $\theta_1 < \theta_0$

Esempio: media per popolazione normale

Nel caso di un campione iid da una $\mathcal{N}(\theta, 1)$ il rapporto λ^* è pari a

$$\lambda^* = \frac{(2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (y_i - \theta_0)^2 \right\}}{(2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (y_i - \theta_1)^2 \right\}}$$

con alcuni passaggi algebrici possiamo riscrivere il rapporto, e quindi la zona di rifiuto, come

$$\lambda^* = \left\{ -\frac{1}{2} \left(2(\theta_1 - \theta_0) \sum_{i=1}^n y_i + n(\theta_0^2 - \theta_1^2) \right) \right\} < k$$

equivalentemente, la zona di rifiuto può essere scritta

$$\bar{y}(\theta_0 - \theta_1) < k$$

e quindi

- se $\theta_0 < \theta_1$ rifiuto H_0 per valori “grandi” di \bar{y}
- viceversa se $\theta_0 > \theta_1$.

Test e statistiche sufficienti

Teorema: Test e statistiche sufficienti

Se S è una statistica sufficiente e ci si trova nelle ipotesi del Lemma di Neyman-Pearson, allora il test più potente ha regione di rifiuto

$$g(s; \theta_0) < k_\alpha g(s; \theta_1)$$

dove g è la distribuzione di S e k_α è tale che $P_{\theta_0}(g(s; \theta_0) < k_\alpha g(s; \theta_1)) = \alpha$

La dimostrazione deriva semplicemente dal fatto che $f(y; \theta) = h(y)g(s; \theta)$ per il teorema di fattorizzazione.

Test per sistemi di ipotesi più generali

Se l'ipotesi alternativa non è atomica la probabilità di errore di II tipo è una funzione su Θ_1 , il teorema di Neyman Pearson si può estendere con alcune premesse.

Definizione: rapporto di verosimiglianza monotono

Sia θ unidimensionale e $T \sim g(t; \theta)$ univariata, T ha rapporto di verosimiglianza monotono sse

$$\theta_2 > \theta_1 \Rightarrow \frac{g(t; \theta_2)}{g(t; \theta_1)} \text{ monotona in } t$$

Una conseguenza che è utile scrivere è

Teorema

Se T ha r.v.m. crescente allora, qualunque sia c ,

$$\theta_2 > \theta_1 \Rightarrow P_{\theta_1}(T > c) \leq P_{\theta_2}(T > c)$$

Risultato su RVM

Teorema

Se T ha r.v.m. crescente allora, qualunque sia c ,

$$\theta_2 > \theta_1 \Rightarrow P_{\theta_1}(T > c) \leq P_{\theta_2}(T > c)$$

Si ha

$$\begin{aligned} P_{\theta_1}(T > c) &= \int_c^{+\infty} g(t; \theta_1) dt = \int_c^{+\infty} \frac{g(t; \theta_1)}{g(t; \theta_2)} g(t; \theta_2) dt \\ &\leq \frac{g(c; \theta_1)}{g(c; \theta_2)} \int_c^{+\infty} g(t; \theta_2) dt = \frac{g(c; \theta_1)}{g(c; \theta_2)} P_{\theta_2}(T > c) \end{aligned}$$

in maniera analoga si mostra che

$$\frac{1 - P_{\theta_2}(T > c)}{1 - P_{\theta_1}(T > c)} \leq \frac{g(c; \theta_2)}{g(c; \theta_1)}$$

ne segue

$$\frac{P_{\theta_2}(T > c)}{P_{\theta_1}(T > c)} \geq \frac{1 - P_{\theta_2}(T > c)}{1 - P_{\theta_1}(T > c)} \Rightarrow \frac{1 - P_{\theta_1}(T > c)}{P_{\theta_1}(T > c)} \geq \frac{1 - P_{\theta_2}(T > c)}{P_{\theta_2}(T > c)}$$

Esempio normale

La media campionaria per un campione casuale semplice da una normale di media θ e varianza nota σ^2 ha rapporto di verosimiglianza

$$\begin{aligned}\frac{g(t; \theta_2)}{g(t; \theta_1)} &= \exp \left\{ -\frac{n}{2\sigma^2} \left((t - \theta_2)^2 - (t - \theta_1)^2 \right) \right\} \\ &= \exp \left\{ -\frac{n}{2\sigma^2} \left(2t(\theta_1 - \theta_2) + \theta_2^2 - \theta_1^2 \right) \right\}\end{aligned}$$

che è crescente (decescente) se $\theta_2 > \theta_1$ ($\theta_2 < \theta_1$).

Estensione del lemma di Neyman-Pearson

Teorema: Test UPP

Per il sistema d'ipotesi $H_0 : \theta \leq \theta_0$, $H_1 : \theta > \theta_0$, se T è sufficiente e ha r.v.m., il test con regione di rifiuto $T > t_0$ è UPP tra i test di livello α dove $\alpha = P_{\theta_0}(T > t_0)$.

Estensione del lemma di Neyman-Pearson, dimostrazione

Essendo T r.v.m., supponiamo crescente, si ha $\beta(\theta) = P_\theta(T > t_0)$ crescente. Allora il test $T > t_0$ è di livello $\alpha = P_{\theta_0}(T > t_0)$ in quanto per ogni $\theta \in \Theta_0$

$$\beta(\theta) \leq \sup_{\theta \leq \theta_0} \beta(\theta) = \beta(\theta_0) = \alpha$$

Tale test è il test del rapporto di verosimiglianza per qualunque sistema di ipotesi $H_0 : \theta = \theta_0, H_1 : \theta = \theta'$ con $\theta' > \theta_0$, infatti detto

$$k' = \inf_{t \in \mathcal{T}} \frac{g(t; \theta')}{g(t; \theta_0)}$$

dove $\mathcal{T} = \{t > t_0, g(t; \theta') > 0 \text{ o } g(t; \theta_0) > 0\}$ allora, in virtù del fatto che T ha r.v.m. crescente,

$$T > t_0 \Leftrightarrow \frac{g(t; \theta')}{g(t; \theta_0)} > k'$$

Cioè la regione di rifiuto ha la forma prevista dal lemma di Neyman-Pearson.

Indice

- 1 Generalità
- 2 Test connessi alla verosimiglianza**
- 3 Regioni di confidenza
- 4 Verosimiglianza profilo
- 5 Esempi vari

Indice

- 1 Generalità
- 2 Test connessi alla verosimiglianza
 - Test basati sulla quantità pivotale
 - Test del rapporto di verosimiglianza
 - Distribuzione (asintotica) del TRV e formulazioni equivalenti
 - Esempi di test
- 3 Regioni di confidenza
- 4 Verosimiglianza profilo
- 5 Esempi vari

Test basato su una quantità pivotale

Un modo per costruire un test è partire da una quantità pivotale, cioè una funzione del campione e del parametro la cui distribuzione non dipende dal parametro

$$q(y, \theta) \sim F$$

ad esempio nel caso di un campione IID da una normale con varianza σ^2 nota si usa, per l'inferenza sulla media θ

$$\frac{y - \theta}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

Test basato sulla quantità pivotale asintotica

A partire da $\hat{\theta} \sim \mathcal{N}(\theta, \mathcal{I}(\hat{\theta})^{-1})$ si ha, per θ_r , la quantità pivotale asintotica

$$\frac{\hat{\theta}_r - \theta_r}{\sqrt{[\mathcal{I}(\hat{\theta})^{-1}]_{rr}}} \sim \mathcal{N}(0, 1)$$

Considerato il sistema d'ipotesi

$$H_0 : \theta_r = \theta_{r0}, \quad H_1 : \theta_r \neq \theta_{r0}$$

se è vera l'ipotesi H_0 si ha

$$\frac{\hat{\theta}_r - \theta_{r0}}{\sqrt{[\mathcal{I}(\hat{\theta})^{-1}]_{rr}}} \sim \mathcal{N}(0, 1)$$

si ha la regione di rifiuto di livello α

$$\{y : |\theta_{r0} - \hat{\theta}_r| > \Phi^{-1}(1 - \alpha/2) \sqrt{[\mathcal{I}(\hat{\theta})^{-1}]_{rr}}\}$$

Test basato sulla quantità pivotale asintotica

A partire da $\hat{\theta} \sim \mathcal{N}(\theta, \mathcal{I}(\hat{\theta})^{-1})$ si ha, per θ_r , la quantità pivotale asintotica

$$\frac{\hat{\theta}_r - \theta_r}{\sqrt{[\mathcal{I}(\hat{\theta})^{-1}]_{rr}}} \sim \mathcal{N}(0, 1)$$

Considerato il sistema d'ipotesi

$$H_0 : \theta_r = \theta_{r0}, \quad H_1 : \theta_r \neq \theta_{r0}$$

se è vera l'ipotesi H_0 si ha

$$\frac{\hat{\theta}_r - \theta_{r0}}{\sqrt{[\mathcal{I}(\hat{\theta})^{-1}]_{rr}}} \sim \mathcal{N}(0, 1)$$

si ha il valore p

$$2 \left(1 - \left| \frac{\hat{\theta}_r - \theta_{r0}}{\sqrt{[\mathcal{I}(\hat{\theta})^{-1}]_{rr}}} \right| \right)$$

Indice

- 1 Generalità
- 2 Test connessi alla verosimiglianza
 - Test basati sulla quantità pivotale
 - **Test del rapporto di verosimiglianza**
 - Distribuzione (asintotica) del TRV e formulazioni equivalenti
 - Esempi di test
- 3 Regioni di confidenza
- 4 Verosimiglianza profilo
- 5 Esempi vari

Test del rapporto di verosimiglianza

Dato il modello $(\mathcal{Y}, p_\theta, \Theta)$ con verosimiglianza $L(\theta; y)$ e dato

$$H_0 : \theta \in \Theta_0; \quad H_1 : \theta \in \Theta_1$$

definiamo il **rapporto di verosimiglianza**

$$\lambda(y) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{L(\hat{\theta})} \in [0, 1]$$

- Se $\lambda(y) = 0$ i dati sono incompatibili con l'ipotesi.
- Se $\lambda(y) = 1$ l'ipotesi è massimalmente supportata dai dati.
- Se $\lambda(y) \in (0, 1)$ è ragionevole un criterio del tipo “i dati supportano l'ipotesi $\theta = \theta_0$ se $\lambda(y) \geq c$ ”.

Un test con regione di rifiuto del tipo

$$\{y \in \mathcal{Y} \mid \lambda(y) < c\}$$

si dice **test del rapporto di verosimiglianza**.

Log-rapporto di verosimiglianza

Si definisce **log-rapporto di verosimiglianza** la quantità

$$W(y) = -2 \log(\lambda(y)).$$

Essendo una trasformazione monotona la condizione $W(y) > k$ definisce un test equivalente

$$\{y \in \mathcal{Y} | \lambda(y) < c\} = \{y \in \mathcal{Y} | W(y) > k\}$$

(dove $k = -2 \log c$)

Rapporto di verosimiglianza e valore p , ipotesi nulla puntuale

Conviene poi scrivere le quantità nel caso particolare in cui $\Theta_0 = \{\theta_0\}$,

$$\lambda(y) = \frac{L(\theta_0)}{\sup_{\theta \in \Theta} L(\theta)} = \frac{L(\theta_0)}{L(\hat{\theta})} \in [0, 1]$$

e

$$W(Y) = -2 \log \lambda(Y) = -2(l(\theta_0) - l(\hat{\theta}))$$

Si definisce allora, dati i valori osservati λ^* e w^* , il valore p

$$p = P_{\theta_0}(\lambda(Y) < \lambda^*) = P_{\theta_0}(W(Y) > w^*)$$

Esempio. Media di una popolazione normale

Se $Y_1, \dots, Y_n \sim IID(\mathcal{N}(\theta, 1))$ e

$$H_0 : \theta = \theta_0 ; H_1 : \theta \neq \theta_0$$

ricordando che la SMV è $\hat{\theta} = \bar{y}$, il RV è dato da

$$\lambda(Y) = \frac{(2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (y_i - \theta_0)^2 \right\}}{(2\pi)^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (y_i - \bar{y})^2 \right\}} = \exp \left\{ -\frac{n}{2} (\bar{y} - \theta_0)^2 \right\}$$

e quindi

$$W(y) = n(\bar{y} - \theta_0)^2$$

per il quale, nell'ipotesi nulla, si ha

$$W(y) = n(\bar{y} - \theta_0)^2 \sim \chi_1^2$$

Esempio. Media di una popolazione normale

Dato il LRV

$$W(y) = n(\bar{y} - \theta_0)^2 \sim \chi_1^2$$

si ha il valore p del test secondo Fisher come

$$p = 1 - F_{\chi_1^2}(w^*)$$

Si mostri che è equivalente a

$$p = 2(1 - \Phi(\sqrt{n}|\bar{y} - \theta_0|))$$

Esempio. Media di una popolazione normale

Dato il LRV

$$W(y) = n(\bar{y} - \theta_0)^2 \sim \chi_1^2$$

La regione di rifiuto di livello α è

$$\{y : W(y) \geq k\} = \{y : \sqrt{n}|\bar{y} - \theta_0| \geq \Phi^{-1}(1 - \alpha/2)\}$$

La potenza del test è data da

$$\begin{aligned} \beta(\theta) &= P\left(|\bar{y} - \theta_0| \geq \frac{1}{\sqrt{n}}\Phi^{-1}(\alpha/2)\right) \\ &= P\left(\bar{y} < \theta_0 - \frac{1}{\sqrt{n}}\Phi^{-1}(\alpha/2)\right) + P\left(\bar{y} > \theta_0 + \frac{1}{\sqrt{n}}\Phi^{-1}(\alpha/2)\right) \\ &= \Phi(\sqrt{n}(\theta_0 - \theta_1) - z_{\alpha/2}) + 1 - \Phi(\sqrt{n}(\theta_0 - \theta_1) + z_{\alpha/2}) \end{aligned}$$

TRV e statistiche sufficienti

Il RV si può valutare sostituendo al campione la statistica sufficiente $T \sim g(t; \theta)$, il RV basato su T è infatti

$$\lambda^*(T(y)) = \frac{\sup_{\theta \in \Theta_0} L(\theta; T(y))}{\sup_{\theta \in \Theta} L(\theta; T(y))} = \frac{\sup_{\theta \in \Theta_0} g(T(y); \theta)}{\sup_{\theta \in \Theta} g(T(y); \theta)}$$

essendo però T sufficiente si ha $g(T(y); \theta)h(y) = f(y; \theta)$, sicché

$$\lambda^*(T(y)) = \frac{\sup_{\theta \in \Theta_0} g(T(y); \theta)h(y)}{\sup_{\theta \in \Theta} g(T(y); \theta)h(y)} = \frac{\sup_{\theta \in \Theta_0} f(y; \theta)}{\sup_{\theta \in \Theta} f(y; \theta)} = \lambda(y)$$

quindi se T è una statistica sufficiente è indifferente calcolare il TRV su T o su y .

Indice

- 1 Generalità
- 2 Test connessi alla verosimiglianza
 - Test basati sulla quantità pivotale
 - Test del rapporto di verosimiglianza
 - Distribuzione (asintotica) del TRV e formulazioni equivalenti
 - Esempi di test
- 3 Regioni di confidenza
- 4 Verosimiglianza profilo
- 5 Esempi vari

Distribuzione asintotica del TRV

Distribuzione asintotica del TRV

Sia $Y_1, \dots, Y_n \sim IID(f(x, \theta))$, sotto condizioni di regolarità sul modello, se $\theta \in \Theta_0$ allora

$$W(y) \sim \chi_d^2$$

al divergere della numerosità campionaria.

I gradi di libertà d sono dati dalla differenza tra il numero di parametri liberi specificati da $\theta \in \Theta_0$ e il numero di parametri liberi specificati da $\theta \in \Theta$.

Sviluppo di $W()$

Consideriamo il LRV

$$W(y) = -2 \log \lambda(y) = -2(l(\theta_0) - l(\hat{\theta}))$$

sviluppando $l(\theta)$ in serie di Taylor in $\hat{\theta}$ si ha

$$\begin{aligned} W(y) &= -2 \left(l'(\hat{\theta})(\theta_0 - \hat{\theta}) + \frac{1}{2} l''(\tilde{\theta})(\theta_0 - \hat{\theta})^2 \right) \\ &= -n(\theta_0 - \hat{\theta})^2 \frac{l''(\tilde{\theta})}{n} + o_p(1) \\ &= n(\hat{\theta} - \theta_0)^2 i(\theta_0) + o_p(1) \\ &= n(\hat{\theta} - \theta_0)^2 i(\hat{\theta}) + o_p(1) \end{aligned}$$

poiché

- $\tilde{\theta}$ è un valore compreso tra $\hat{\theta}$ e θ_0 e quindi se $\hat{\theta}$ è consistente, $\tilde{\theta} \xrightarrow{P} \theta_0$
- $\frac{l''(\tilde{\theta})}{n} = -i(\theta_0) + o_p(1)$
- $i(\hat{\theta}) = i(\theta_0) + o_p(1)$

Test di Wald

Si è così mostrata l'equivalenza asintotica tra $W(y)$ e

$$W_e(y) = n(\hat{\theta} - \theta_0)^2 i(\hat{\theta})$$

detto **test di Wald**.

Si noti che il test di Wald si basa sull'approssimazione quadratica della log-verosimiglianza.

Il test di Wald è equivalente a quello basato sulla quantità

$$\frac{\hat{\theta} - \theta_0}{\sqrt{(ni(\hat{\theta}))^{-1}}},$$

(di cui è il quadrato).

Test di Wald

Si è così mostrata l'equivalenza asintotica tra $W(y)$ e

$$W_e(y) = n(\hat{\theta} - \theta_0)^2 i(\hat{\theta})$$

detto **test di Wald**.

Se θ è multidimensionale si ha

$$W_e(Y) = (\hat{\theta} - \theta_0)^T \mathcal{I}(\hat{\theta})(\hat{\theta} - \theta_0)$$

Test del punteggio

Ancora,

$$\begin{aligned} W(y) &= n(\hat{\theta} - \theta_0)^2 i(\theta_0) + o_p(1) \\ &= \frac{l'(\theta_0)^2}{ni(\theta_0)} + o_p(1) \end{aligned}$$

poiché $\sqrt{n}(\hat{\theta} - \theta_0) = \frac{l'(\theta_0)}{\sqrt{ni(\theta_0)}} + o_p(1)$, si ha

$$W(y) = \frac{l'(\theta_0)^2}{ni(\theta_0)} + o_p(1)$$

e si ha l'equivalenza asintotica con il **test del punteggio**

$$W_u(y) = \frac{l'(\theta_0)^2}{ni(\theta_0)}$$

che, si noti, non richiede la conoscenza di $\hat{\theta}$.

Test del punteggio

Ancora,

$$W(y) = \frac{l'(\theta_0)^2}{ni(\theta_0)} + o_p(1)$$

poiché $\sqrt{n}(\hat{\theta} - \theta_0) = \frac{l'(\theta_0)}{\sqrt{ni(\theta_0)}} + o_p(1)$, si ha

$$W(y) = \frac{l'(\theta_0)^2}{ni(\theta_0)} + o_p(1)$$

e si ha l'equivalenza asintotica con il **test del punteggio**

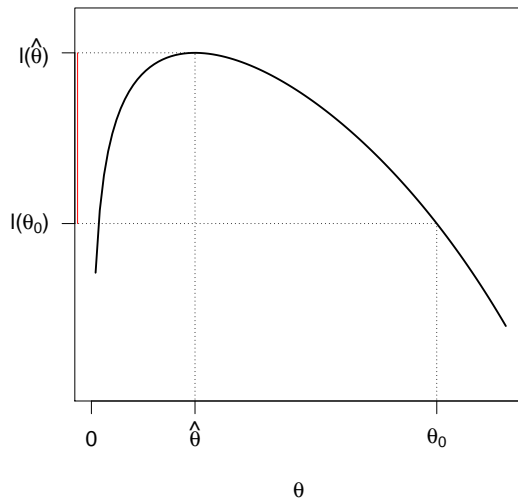
$$W_u(y) = \frac{l'(\theta_0)^2}{ni(\theta_0)}$$

che, si noti, non richiede la conoscenza di $\hat{\theta}$.

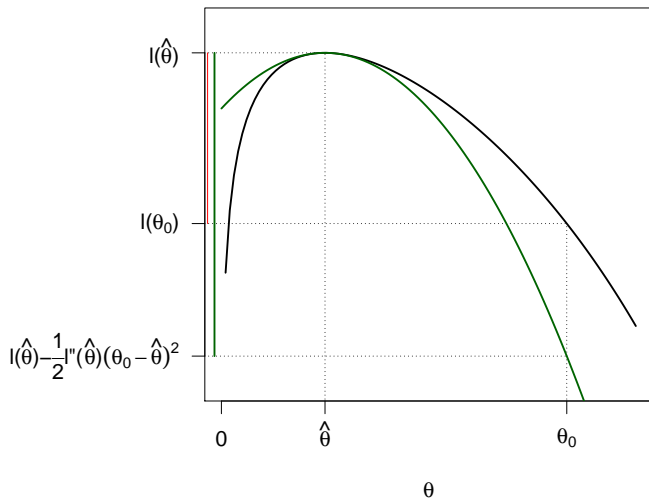
Se θ è multidimensionale si ha

$$W_u(Y) = l_*(\theta_0)^T \mathcal{I}(\theta_0)^{-1} l_*(\theta_0)$$

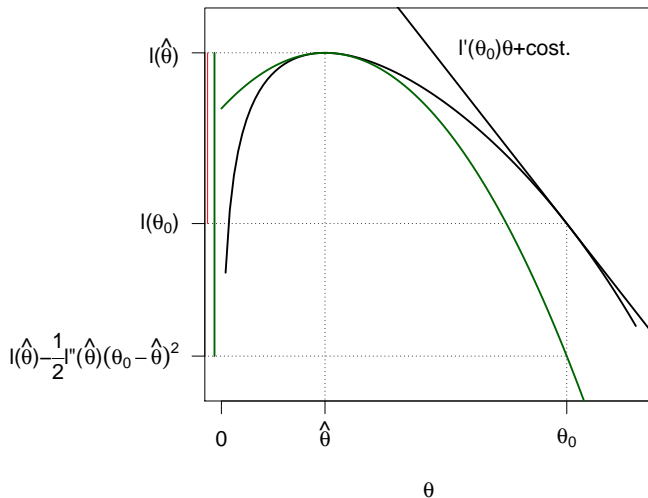
TRV, Wald e punteggio: rappresentazione grafica



TRV, Wald e punteggio: rappresentazione grafica



TRV, Wald e punteggi: rappresentazione grafica



Indice

- 1 Generalità
- 2 **Test connessi alla verosimiglianza**
 - Test basati sulla quantità pivotale
 - Test del rapporto di verosimiglianza
 - Distribuzione (asintotica) del TRV e formulazioni equivalenti
 - **Esempi di test**
- 3 Regioni di confidenza
- 4 Verosimiglianza profilo
- 5 Esempi vari

TRV per l'esponenziale traslata

Sia Y_1, \dots, Y_n IID e

$$Y_i \sim p_\theta(y) = e^{-(y-\theta)}$$

per $y \geq \theta$. Si ha allora

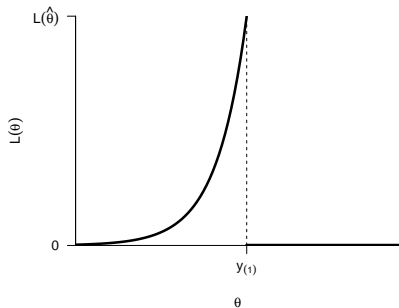
$$L(\theta) = \begin{cases} \exp\{-\sum_i y_i + n\theta\} & \text{se } \theta \leq y_{(1)} \\ 0 & \text{se } \theta > y_{(1)} \end{cases}$$

La SMV è

$$\hat{\theta} = y_{(1)}$$

e

$$L(\hat{\theta}) = \exp\left\{-\sum_i y_i + ny_{(1)}\right\}$$



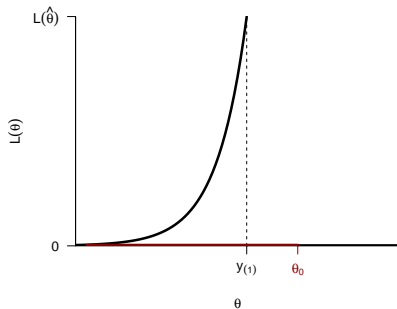
TRV per l'esponenziale traslata

Con riferimento al sistema d'ipotesi

$$H_0 : \theta \leq \theta_0; H_1 : \theta > \theta_0$$

si ha

$$\max_{\theta \in \Theta_0} L(\theta) = \begin{cases} L(\hat{\theta}) & \text{se } y_{(1)} \leq \theta_0 \end{cases}$$



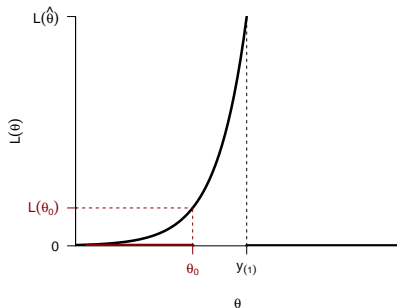
TRV per l'esponenziale traslata

Con riferimento al sistema d'ipotesi

$$H_0 : \theta \leq \theta_0; H_1 : \theta > \theta_0$$

si ha

$$\max_{\theta \in \Theta_0} L(\theta) = \begin{cases} L(\hat{\theta}) & \text{se } y_{(1)} \leq \theta_0 \\ L(\theta_0) & \text{se } y_{(1)} > \theta_0 \end{cases}$$



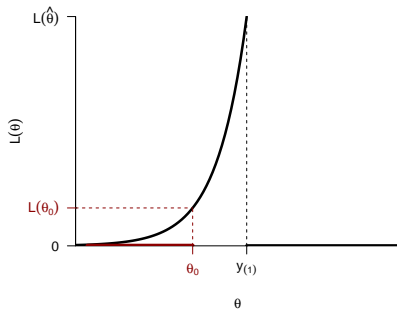
TRV per l'esponenziale traslata

Con riferimento al sistema d'ipotesi

$$H_0 : \theta \leq \theta_0; H_1 : \theta > \theta_0$$

si ha

$$\max_{\theta \in \Theta_0} L(\theta) = \begin{cases} L(\hat{\theta}) & \text{se } y_{(1)} \leq \theta_0 \\ L(\theta_0) & \text{se } y_{(1)} > \theta_0 \end{cases}$$



e si ha il rapporto di verosimiglianza

$$\lambda(y) = \begin{cases} 1 & \text{se } y_{(1)} \leq \theta_0 \\ \exp \{ -n(y_{(1)} - \theta_0) \} & \text{se } y_{(1)} > \theta_0 \end{cases}$$

e quindi la regione di rifiuto

$$\{y \text{ t.c. } n(y_{(1)} - \theta_0) > k\}$$

Media della normale: i tre test

Sia $Y_1, \dots, Y_n \sim_{iid} \mathcal{N}(\theta, \sigma^2)$ con σ^2 noto e sia $H_0 : \theta = \theta_0$, allora

$$l(\theta) = -\frac{n}{2\sigma^2}(\bar{Y} - \theta)^2$$

e si hanno dunque le quantità

$$W(Y) = 2(l(\hat{\theta}) - l(\theta_0)) = \frac{n}{\sigma^2}(\bar{Y} - \theta)^2$$

$$W_e(Y) = (\bar{Y} - \theta_0)^2 \mathcal{I}(\hat{\theta}) = (\bar{Y} - \theta_0)^2 \frac{n}{\sigma^2}$$

$$W_u(Y) = \frac{l_*(\theta_0)}{\mathcal{I}(\theta_0)} = \frac{(n(\bar{Y} - \theta_0)/\sigma^2)^2}{n/\sigma^2}$$

si ha cioè $W(Y) = W_e(Y) = W_u(Y)$.

Media della normale: i tre test

La regione di rifiuto $\mathcal{Y}_1 = \{Y \text{ t.c. } W(Y) > w_\alpha\}$ è quindi del tipo

$$\mathcal{Y}_1 = \left\{ Y \text{ t.c. } \frac{n}{\sigma^2}(\bar{Y} - \theta_0)^2 > w_\alpha \right\}$$

con w_α tale che $P_{\theta_0} \left(\frac{n}{\sigma^2}(\bar{Y} - \theta_0)^2 > w_\alpha \right) = \alpha$ siccome poi, nell'ipotesi H_0 $\bar{Y} \sim \mathcal{N}(\theta_0, \sigma^2/n)$ si ha

$$\frac{n}{\sigma^2}(\bar{Y} - \theta_0) \sim \chi_1^2$$

e quindi la regione di rifiuto è

$$\frac{n}{\sigma^2}(\bar{Y} - \theta_0) > \chi_{1,1-\alpha}^2$$

che equivale a

$$\frac{\bar{Y} - \theta_0}{\sigma/\sqrt{n}} < z_{\alpha/2} \circ \frac{\bar{Y} - \theta_0}{\sigma/\sqrt{n}} > z_{1-\alpha/2}$$

Media della normale: i tre test

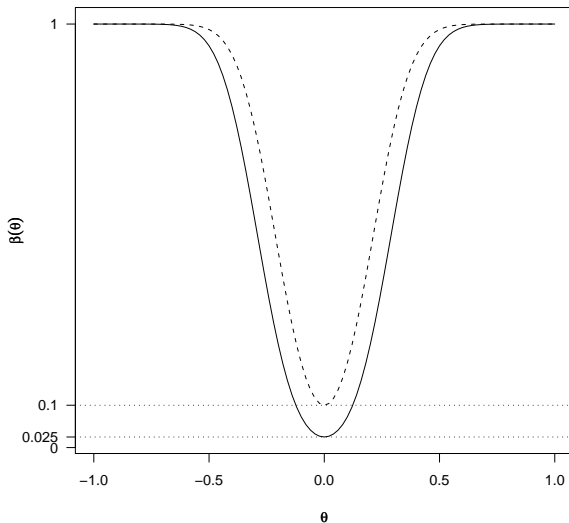
cioè

$$\bar{Y} < \theta_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \text{ o } \bar{Y} > \theta_0 + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

La funzione potenza vale

$$\begin{aligned} \beta(\theta) &= 1 - P_{\theta} (z_{\alpha/2} < \sqrt{n}(\bar{Y} - \theta_0)/\sigma < -z_{\alpha/2}) \\ &= 1 - P_{\theta} (z_{\alpha/2} < \sqrt{n}(\bar{Y} - \theta)/\sigma + \sqrt{n}(\theta - \theta_0)/\sigma < -z_{\alpha/2}) \\ &= 1 - P_{\theta} (z_{\alpha/2} - \sqrt{n}(\theta - \theta_0)/\sigma < \sqrt{n}(\bar{Y} - \theta)/\sigma < -z_{\alpha/2} - \sqrt{n}(\theta - \theta_0)) \\ &= 1 - \Phi(-z_{\alpha/2} - \sqrt{n}(\theta - \theta_0)/\sigma) + \Phi(z_{\alpha/2} - \sqrt{n}(\theta - \theta_0)/\sigma) \end{aligned}$$

Media della normale: i tre test



Esempio: esponenziale, TRV

Sia Y_1, \dots, Y_n i.i.d. da $p_\theta(y) = \theta e^{-\theta y}$, sicché

$$l(\theta) = n \log \theta - \theta n \bar{Y}$$

con $\hat{\theta} = \bar{Y}^{-1}$, quindi

$$\begin{aligned} W(Y) &= 2(l(\hat{\theta}) - l(\theta_0)) \\ &= 2(-n \log \bar{Y} - n - n \log \theta_0 + n\theta_0 \bar{Y}) \\ &= -2n(1 + \log(\bar{Y}\theta_0) - \theta_0 \bar{Y}) \end{aligned}$$

Usando i risultati asintotici la regione di rifiuto è dunque

$$\{\bar{Y} : 2n(\theta_0 \bar{Y} - \log(\bar{Y}\theta_0) - 1) > \chi_{1,1-\alpha}^2\}$$

mentre il valore p è

$$1 - F_{\chi_1^2}(2n(\theta_0 \bar{Y} - \log(\bar{Y}\theta_0) - 1))$$

Esempio: esponenziale, test di Wald

Il test di Wald d'altro canto ha espressione

$$\begin{aligned}
 W_e(Y) &= (\hat{\theta} - \theta_0)^2 \mathcal{I}(\hat{\theta}) \\
 &= (\bar{Y}^{-1} - \theta_0)^2 \frac{n}{\hat{\theta}^2} \\
 &= (\bar{Y}^{-2} + \theta_0^2 - 2\theta_0 \bar{Y}^{-1}) n \bar{Y}^2 \\
 &= n + n \bar{Y}^2 \theta_0^2 - 2n\theta_0 \bar{Y} \\
 &= n(\bar{Y}\theta_0 - 1)^2
 \end{aligned}$$

La regione di rifiuto è dunque

$$\{\bar{Y} : n(\bar{Y}\theta_0 - 1)^2 > \chi_{1,1-\alpha}^2\}$$

mentre il valore p è

$$1 - F_{\chi_1^2}(n(\bar{Y}\theta_0 - 1)^2)$$

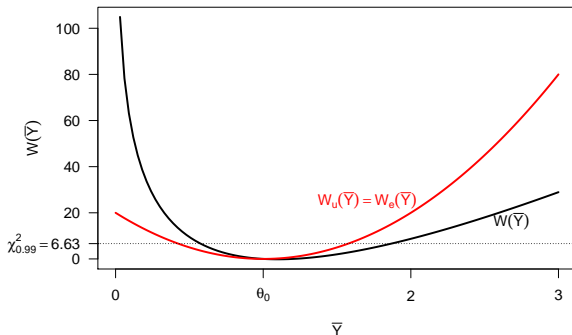
Esempio: esponenziale, test del punteggio

Infine, l'espressione per il test del punteggio è

$$\begin{aligned}
 W_u(Y) &= \frac{l'(\theta_0)^2}{\mathcal{I}(\theta_0)} \\
 &= (n\theta_0^{-1} - n\bar{Y})^2 \left(\frac{n}{\theta_0^2}\right)^{-1} \\
 &= \theta_0^2 n(\theta_0^{-2} + \bar{Y}^2 - 2\bar{Y}\theta_0^{-1}) \\
 &= n + n\bar{Y}^2\theta_0^2 - 2n\theta_0\bar{Y} \\
 &= n(\bar{Y}\theta_0 - 1)^2
 \end{aligned}$$

si ottiene quindi lo stesso risultato del test di Wald.

Esempio: esponenziale, rappresentazione grafica



test W e W_u/W_e per l'esponenziale

Indice

- 1 Generalità
- 2 Test connessi alla verosimiglianza
- 3 Regioni di confidenza**
- 4 Verosimiglianza profilo
- 5 Esempi vari

Regioni di confidenza

Regione di confidenza

Dato un modello parametrico $(\mathcal{Y}, \theta, p_\theta)$ una regione di confidenza di livello $1 - \alpha$ è un'applicazione

$$C : \mathcal{Y} \rightarrow \mathcal{P}(\Theta)$$

(con $P(\Theta)$ insieme delle parti di θ) tale che

$$P_\theta(\theta \in C(Y)) \geq 1 - \alpha \quad \forall \theta \in \Theta$$

Si definisce probabilità di copertura la quantità

$$c(\theta) = P_\theta(\theta \in C(Y))$$

Va sottolineato che è $C(Y)$ a essere casuale, non θ .

Intervallo di confidenza

Intervallo di confidenza

Dato un modello parametrico $(\mathcal{Y}, \theta, p_\theta)$ con $\theta \in \Theta \subset \mathbb{R}$ un intervallo di confidenza di livello $1 - \alpha$ è una coppia di funzioni

$$L, U : \mathcal{Y} \rightarrow \mathbb{R}$$

con $L \leq U$ tali che

$$P_\theta(L(Y) \leq \theta \leq U(Y)) \geq 1 - \alpha \quad \forall \theta \in \Theta$$

Quantità pivotali

Un modo per ottenere regioni di confidenza è utilizzare quantità pivotali

Quantità pivotali

Si dice **quantità pivotale** una funzione

$$T(Y, \theta) : \mathcal{Y} \times \Theta \rightarrow \mathbb{R}^m$$

tale che la distribuzione di T non dipende da θ .

Si noti che T non è una statistica in quanto dipende anche dal parametro. Se T è una quantità pivotale e $B \subset \mathbb{R}^m$ si ha

$$P_\theta(T(Y, \theta) \in B) = P(T(Y, \theta) \in B)$$

e quindi, posto $P(T(Y, \theta) \in B) =: 1 - \alpha$ la regione

$$\{y \text{ t.c. } T(y, \theta) \in B\}$$

è una regione di confidenza di livello $1 - \alpha$.

Esempio: normale

Dato un campione Y_1, \dots, Y_n da una $\mathcal{N}(\mu, \sigma^2)$ si ha la quantità pivotale

$$T(Y, \theta) = \frac{\sqrt{n}(y - \mu)}{\bar{S}} \sim t_{n-1}$$

per la quale si ha

$$P_{\theta}(-t_{n-1, 1-\alpha/2} \leq T(Y, \theta) \leq t_{n-1, 1-\alpha/2}) = 1 - \alpha$$

e da cui si ottengono gli intervalli

$$\left[\bar{Y} - t_{n-1, 1-\alpha/2} \frac{\bar{S}}{\sqrt{n}}, \bar{Y} + t_{n-1, 1-\alpha/2} \frac{\bar{S}}{\sqrt{n}} \right]$$

I.C. basato sulla quantità pivotale asintotica

È chiaro dunque che si possono ottenere regioni di confidenza approssimate sfruttando la quantità pivotale asintotica ottenibile da

$$\hat{\theta} \sim \mathcal{N}(\theta, \mathcal{I}(\theta)^{-1})$$

si ha infatti, per l'elemento r -esimo di θ

$$\frac{\hat{\theta}_r - \theta_r}{\sqrt{[\mathcal{I}(\theta)^{-1}]_{rr}}} \sim \mathcal{N}(0, 1)$$

e quindi l'intervallo di confidenza

$$\hat{\theta}_r \pm \Phi^{-1}(1 - \alpha/2) \sqrt{[\mathcal{I}(\theta)^{-1}]_{rr}}$$

I.C. basato sulla quantità pivotale asintotica

A partire da

$$\hat{\theta}_r \pm \Phi^{-1}(1 - \alpha/2) \sqrt{[\mathcal{I}(\theta)^{-1}]_{rr}}$$

concretamente, l'intervallo si ottiene sostituendo la stima

$$\hat{\theta}_r \pm \Phi^{-1}(1 - \alpha/2) \sqrt{[\mathcal{I}(\hat{\theta})^{-1}]_{rr}}$$

Ed è anche possibile usare l'informazione osservata invece dell'informazione attesa

$$\hat{\theta}_r \pm \Phi^{-1}(1 - \alpha/2) \sqrt{[J(\hat{\theta})^{-1}]_{rr}}$$

Test \rightarrow Regioni di confidenza

Per il modello $(\mathcal{Y}, \theta, p_\theta)$ si consideri

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

e sia $A(\theta_0)$ la regione di accettazione al livello α

$$A(\theta_0) = \{Y \text{ t.c. accetto } H_0\} \subset \mathcal{Y}$$

$A(\theta_0)$ è, per ogni fissato θ_0 , un sottoinsieme di \mathcal{Y} tale che

$$P_{\theta_0}(Y \in A(\theta_0)) \geq 1 - \alpha$$

definiamo allora, fissato il campione Y , l'insieme dei valori θ_0 che verrebbero accettati in base ai test della famiglia individuata sopra, cioè

$$C(Y) = \{\theta_0 \text{ t.c. } Y \in A(\theta_0)\}$$

si ha allora

$$P_\theta(\theta \in C(Y)) = P_\theta(Y \in A(\theta)) \geq 1 - \alpha$$

quindi $C(Y)$ è una regione di confidenza di livello $1 - \alpha$.

Regioni di confidenza \rightarrow Test

Viceversa, se $C(Y)$ definisce una famiglia di regioni di confidenza di livello $1 - \alpha$, cioè

$$P_{\theta}(\theta \in C(Y)) \geq 1 - \alpha$$

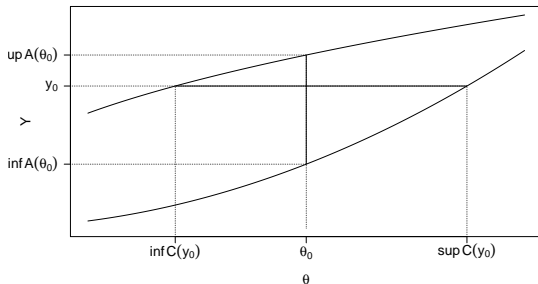
allora

$$A(\theta) = \{Y \text{ t.c. } \theta \in C(Y)\}$$

è una regione di accettazione per un test di livello α .

Regioni di confidenza \leftrightarrow Test

In figura si dà un'ipotetica rappresentazione del passaggio da intervalli di confidenza a verifica d'ipotesi.



Esempio: normale

Si ha, per la media di una normale con varianza nota pari a σ^2 la famiglia di test con regioni di accettazione

$$A(\theta) = \left[\theta - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \theta + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

da cui si deriva la famiglia di intervalli di confidenza

$$C(Y) = \left[\bar{Y} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{Y} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

Il grafico corrispondente alla figura vedrebbe in luogo delle curve due rette.

Regioni di confidenza basate sul RV

Usando il test del rapporto di verosimiglianza, la regione di rifiuto per $H_0 : \theta = \theta_0$ è

$$\{y : W(\theta_0) = 2(l(\hat{\theta}) - l(\theta_0)) > F_W^{-1}(1 - \alpha)\}$$

dove $W \sim F_W$.

Via il paradigma appena illustrato si ha quindi la regione di confidenza è

$$\{\theta_0 : W(\theta_0) = 2(l(\hat{\theta}) - l(\theta_0)) < F_W^{-1}(1 - \alpha)\}$$

cioè l'insieme dei θ_0 che soddisfa

$$l(\theta_0) > l(\hat{\theta}) - \frac{1}{2}F_W^{-1}(1 - \alpha)$$

Regioni di confidenza basate sul test di Wald

Con lo stesso ragionamento partendo dal test di Wald

$$W_e(Y) = n(\hat{\theta} - \theta_0)^T i(\hat{\theta})(\hat{\theta} - \theta_0)$$

otteniamo la regione di confidenza

$$\{\theta_0 : W_e(Y) = n(\hat{\theta} - \theta_0)^T i(\hat{\theta})(\hat{\theta} - \theta_0) < F_{W_e}^{-1}(1 - \alpha)\}$$

cioè, se θ è unidimensionale, l'insieme dei θ_0 che soddisfa

$$n(\hat{\theta} - \theta_0)^2 i(\hat{\theta}) < F_{W_e}^{-1}(1 - \alpha)$$

vale a dire

$$\frac{|\hat{\theta} - \theta_0|}{\sqrt{ni(\hat{\theta})}} < \sqrt{F_{W_e}^{-1}(1 - \alpha)}$$

Regioni di confidenza basate sul test del punteggio

Infine, partendo dal test del punteggio

$$W_u(Y) = I_*(\theta_0)^T \mathcal{I}(\theta_0)^{-1} I_*(\theta_0)$$

otteniamo la regione di confidenza

$$\{\theta_0 : W_u(Y) = I_*(\theta_0)^T \mathcal{I}(\theta_0)^{-1} I_*(\theta_0) < F_{W_u}^{-1}(1 - \alpha)\}$$

Esempio: esponenziale

Si ricordi che per Y_1, \dots, Y_n i.i.d. da $p_\theta(y) = \theta e^{-\theta y}$ si ha

$$W(Y) = -2n(1 + \log(\bar{Y}\theta_0) - \theta_0 \bar{Y})$$

con regione di rifiuto

$$\{\bar{Y} : 2n(\theta_0 \bar{Y} - \log(\bar{Y}\theta_0) - 1) > \chi_{1,1-\alpha}^2\}$$

la regione di confidenza è quindi data dai valori θ_0 che soddisfano la condizione

$$2n(\theta_0 \bar{Y} - \log(\bar{Y}\theta_0) - 1) < \chi_{1,1-\alpha}^2$$

$$\theta_0 \bar{Y} - \log \theta_0 < \frac{\chi_{1,1-\alpha}^2}{2n} + 1 + \log \bar{Y}$$

Esempio: esponenziale, test di Wald

Il test di Wald e del punteggio hanno espressione

$$W_e(Y) = W_u(y) = n(\bar{Y}\theta_0 - 1)^2$$

La regione di rifiuto è dunque

$$\{\bar{Y} : n(\bar{Y}\theta_0 - 1)^2 > \chi_{1,1-\alpha}^2\}$$

ovvero la regione di accettazione è

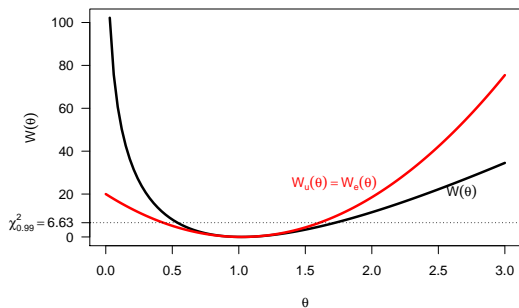
$$\frac{1 - \sqrt{\chi_{1-\alpha}^2/n}}{\theta_0} < \bar{Y} < \frac{1 + \sqrt{\chi_{1-\alpha}^2/n}}{\theta_0}$$

La regione di confidenza è dunque

$$\frac{1 - \sqrt{\chi_{1-\alpha}^2/n}}{\bar{Y}} < \theta_0 < \frac{1 + \sqrt{\chi_{1-\alpha}^2/n}}{\bar{Y}}$$

Regioni di confidenza \leftrightarrow Test

Quantità W e W_e per l'esponenziale.



Esempio: normale, σ^2

Si ha la quantità pivotale $T = n\hat{\sigma}^2/\sigma^2 \sim \chi_{n-1}^2$, sulla base di questa si ha

$$1-\alpha = P(c_1 < T < c_2) = P\left(\frac{c_1}{n\hat{\sigma}^2} < \frac{1}{\sigma^2} < \frac{c_2}{n\hat{\sigma}^2}\right) = P\left(\frac{n\hat{\sigma}^2}{c_2} < \sigma^2 < \frac{n\hat{\sigma}^2}{c_1}\right)$$

Esempio: Poisson

Sia $y_1, \dots, y_n \sim IID(\text{poi}(\lambda))$

$$L(\lambda) = \lambda^{n\bar{y}} e^{-n\lambda} \Rightarrow l(\lambda) = n\bar{y} \log \lambda - n\lambda$$

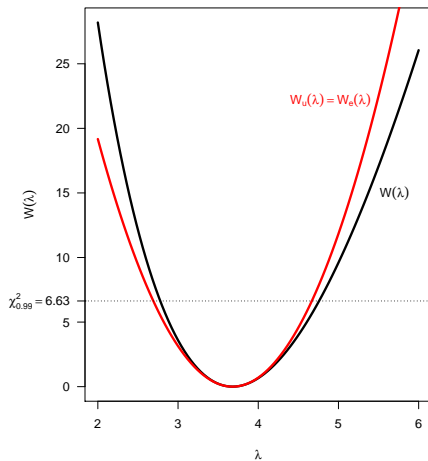
quindi si ha il rapporto di verosimiglianza

$$W = 2n(\bar{y} \log \bar{y} - \bar{y} - \bar{y} \log \lambda + \lambda)$$

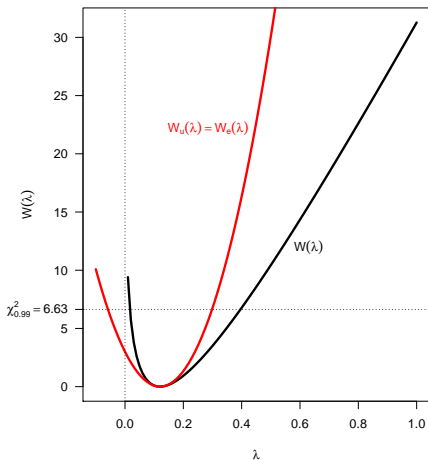
e si ha $l_*(\lambda) = n\bar{y}\lambda^{-1} - n$ e $l''(\lambda) = -n\bar{y}\lambda^{-2}$, quindi

$$W_e = W_u = n\bar{Y}^{-1}(\lambda - \bar{y})^2$$

Esempio: Poisson



Esempio: Poisson



Esempio: Poisson, riparametrizzazione

Sia $y_1, \dots, y_n \sim IID(\text{poi}(\theta))$ con $\theta = \log \lambda$

$$p(y; \theta) = \frac{e^{y\theta}}{y!} e^{-e^\theta}$$

$$L(\theta) = e^{\theta n \bar{y}} e^{-ne^\theta} \Rightarrow l(\theta) = \theta n \bar{y} - ne^\theta$$

si ha $\hat{\theta} = \log \bar{y}$ quindi si ha il rapporto di verosimiglianza

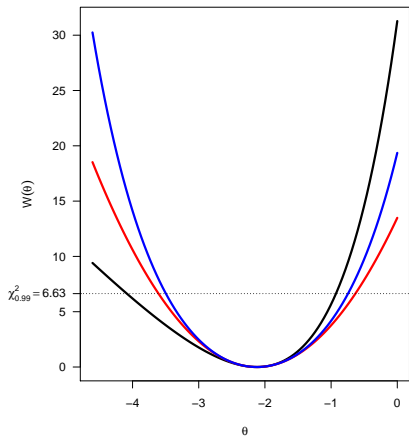
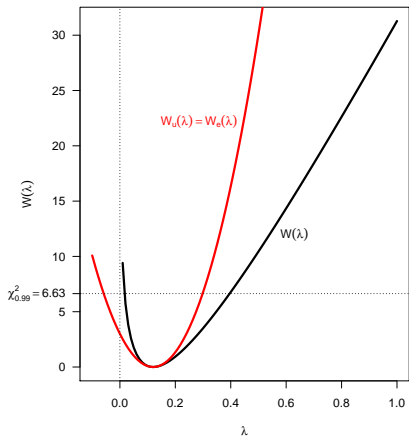
$$W = 2n(\bar{y} \log \bar{y} - \bar{y} - \bar{y}\theta + e^\theta)$$

e si ha $l_*(\theta) = n\bar{y} - ne^\theta$ e $l''(\theta) = -ne^\theta$, quindi

$$W_e = ne^\theta (\hat{\theta} - \theta)^2$$

$$W_u = ne^{-\theta} (\bar{y} - e^\theta)^2$$

Esempio: Poisson



Esempio: Poisson, riparametrizzazione

Possiamo anche, arrivati alle statistiche W , riscriverle in funzione di λ per tornare alla scala originaria.

Se lo facciamo per W ritroviamo la stessa funzione vista prima della riparametrizzazione

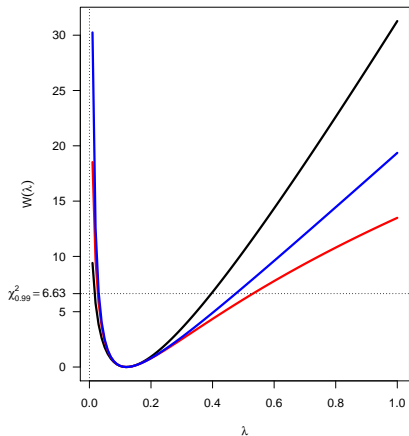
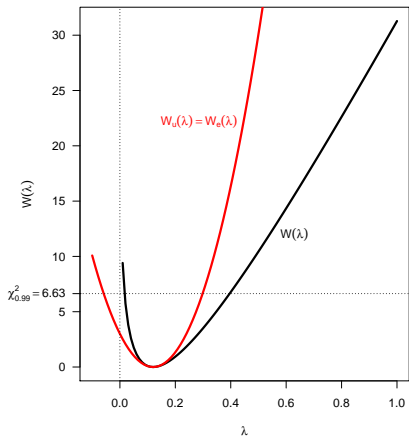
$$W = 2n(\bar{y} \log \bar{y} - \bar{y} - \bar{y} \log \lambda + \lambda)$$

mentre per le altre si ha

$$W_e = ne^{\theta}(\hat{\theta} - \theta)^2 \rightarrow n\lambda(\log \bar{y} - \log \lambda)^2$$

$$W_u = ne^{-\theta}(\bar{y} - e^{\theta})^2 \rightarrow n\lambda^{-1}(\bar{y} - \lambda)^2$$

Esempio: Poisson



Morale

Approssimazioni e parametrizzazioni

Una diversa parametrizzazione (che significa con lo stesso modello) può comportare una migliore approssimazione asintotica per i test.

Si noti che questo è vero anche per la quantità pivotale.

Esempio: binomiale

Nel caso della binomiale, sia $y \sim \text{Binom}(n, \theta)$

$$l(\theta) = y \log \theta + (n - y) \log(1 - \theta)$$

$$l_*(\theta) = \frac{y}{\theta} - \frac{n - y}{1 - \theta} = \frac{y - n\theta}{\theta(1 - \theta)}$$

$$l''(\theta) = -\frac{y}{\theta^2} - \frac{n - y}{(1 - \theta)^2}$$

$$\mathcal{I}(\theta) = \frac{n\theta}{\theta^2} - \frac{n(1 - \theta)}{(1 - \theta)^2} = \frac{n}{\theta(1 - \theta)}$$

e quindi

$$W_e(Y) = n(\hat{\theta} - \theta_0)^2 i(\hat{\theta}) = n \left(\frac{y}{n} - \theta_0 \right)^2 \frac{n}{\hat{\theta}(1 - \hat{\theta})} = \frac{(y - n\theta_0)^2}{n\hat{\theta}(1 - \hat{\theta})}$$

e

$$W_u(Y) = \frac{l'(\theta_0)^2}{ni(\theta_0)} = \frac{\left(\frac{y - n\theta_0}{\theta_0(1 - \theta_0)} \right)^2}{\frac{n}{\theta_0(1 - \theta_0)}} = \frac{(y - n\theta_0)^2}{n\theta_0(1 - \theta_0)}$$

Esempio: binomiale

Nel caso della binomiale, sia $y \sim \text{Binom}(n, \theta)$

$$W_e(Y) = n(\hat{\theta} - \theta_0)^2 i(\hat{\theta}) = n \left(\frac{y}{n} - \theta_0 \right)^2 \frac{n}{\hat{\theta}(1 - \hat{\theta})} = \frac{(y - n\theta_0)^2}{n\hat{\theta}(1 - \hat{\theta})}$$

e

$$W_u(Y) = \frac{l'(\theta_0)^2}{ni(\theta_0)} = \frac{\left(\frac{y - n\theta_0}{\theta_0(1 - \theta_0)} \right)^2}{\frac{n}{\theta_0(1 - \theta_0)}} = \frac{(y - n\theta_0)^2}{n\theta_0(1 - \theta_0)}$$

Entrambi nella forma facilmente interpretabile (molto più di quanto non si possa dire per la forma originale del RV)

$$\frac{(\text{osservato} - \text{atteso})^2}{\text{varianza}(\text{osservato})}$$

con una differente valutazione per la varianza, in W_e troviamo la stima della varianza, in W_u la varianza nell'ipotesi nulla.

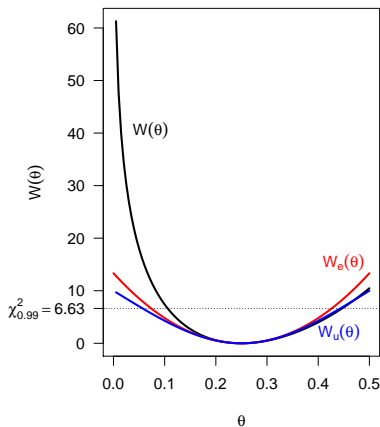
Esempio: binomiale

Nel caso della binomiale, sia $y \sim \text{Binom}(n, \theta)$

$$W = 2n \left(\hat{\theta} \log \frac{\hat{\theta}}{\theta} + (1 - \hat{\theta}) \log \frac{1 - \hat{\theta}}{1 - \theta} \right)$$

$$W_e = \frac{n}{\hat{\theta}(1 - \hat{\theta})} (\theta - \hat{\theta})^2$$

$$W_u = \frac{n}{\theta(1 - \theta)} (\theta - \hat{\theta})^2$$



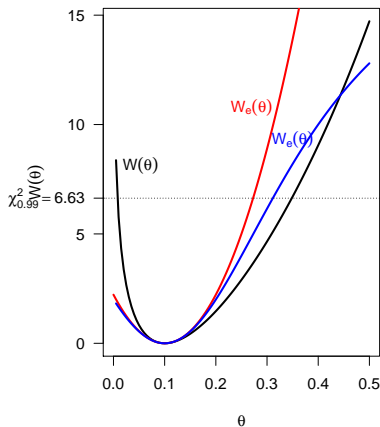
Esempio: binomiale

Nel caso della binomiale, sia $y \sim \text{Binom}(n, \theta)$

$$W = 2n \left(\hat{\theta} \log \frac{\hat{\theta}}{\theta} + (1 - \hat{\theta}) \log \frac{1 - \hat{\theta}}{1 - \theta} \right)$$

$$W_e = \frac{n}{\hat{\theta}(1 - \hat{\theta})} (\theta - \hat{\theta})^2$$

$$W_u = \frac{n}{\theta(1 - \theta)} (\theta - \hat{\theta})^2$$



Esempio: binomiale

Nel caso della binomiale, sia $y \sim \text{Binom}(n, \theta)$

$$W = 2n \left(\hat{\theta} \log \frac{\hat{\theta}}{\theta} + (1 - \hat{\theta}) \log \frac{1 - \hat{\theta}}{1 - \theta} \right)$$

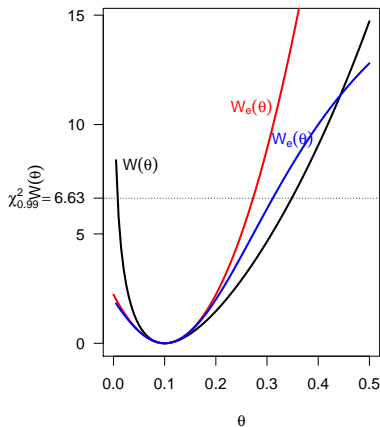
$$W_e = \frac{n}{\hat{\theta}(1 - \hat{\theta})} (\theta - \hat{\theta})^2$$

$$W_u = \frac{n}{\theta(1 - \theta)} (\theta - \hat{\theta})^2$$

Riparametrizzazioni consigliabili:

$$\psi = \log \frac{\theta}{1 - \theta}$$

$$\zeta = \Phi^{-1}(\theta)$$



Esempio numerico binomiale

Sia $n = 8$ e $\theta_0 = 0.7$, la regione di rifiuto è

$$\mathcal{Y}_1 = \{y \text{ t.c. } y < y_l \text{ o } y > y_u\}$$

Per determinare y_l e y_u occorre fare riferimento alla distribuzione di Y nell'ipotesi nulla, la cui tabulazione è

y	0	1	2	3	4	5	6	7	8
$\lambda(y)$	$6 \cdot 10^{-5}$	0.003	0.032	0.166	0.498	0.902	0.952	0.503	0.058
$P(Y = y; \theta_0)$	$6 \cdot 10^{-5}$	0.001	0.010	0.047	0.136	0.254	0.296	0.198	0.058

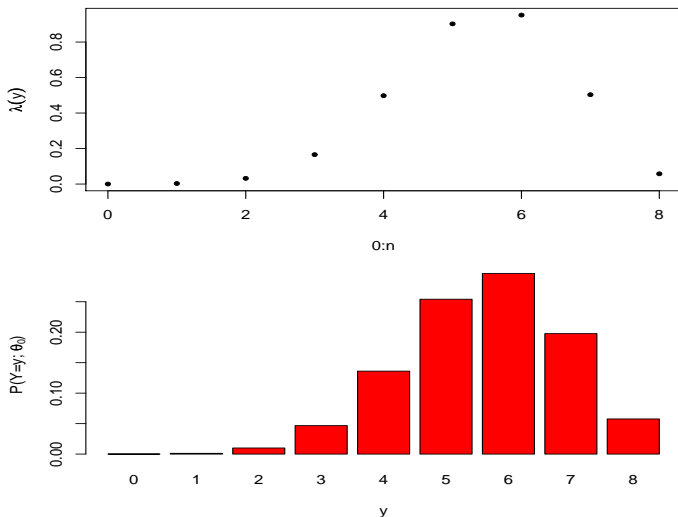
Conviene ordinare i valori ottenuti (tutta la tabella) sulla base del RV $\lambda(y)$

y	0	1	2	8	3	4	7	5	6
$\lambda(y)$	$6 \cdot 10^{-5}$	0.003	0.032	0.058	0.166	0.498	0.503	0.902	0.952
$P(Y = y; \theta_0)$	$6 \cdot 10^{-5}$	0.001	0.010	0.058	0.047	0.136	0.198	0.254	0.296

- reg. rifiuto al livello dell'1% : $\lambda(y) \leq 0.032$ ossia $\mathcal{Y}_1 = \{0, 1, 2\}$.
- reg. rifiuto al livello del 6.8% : $\lambda(y) \leq 0.058$ ossia $\mathcal{Y}_1 = \{0, 1, 2, 8\}$.

Questi sono i l.d.s. ottenibili più vicini al 5%, in un caso di v.a. discreta come questo non ottengo generalmente il test al livello cercato.

Esempio numerico binomiale



Una soluzione alternativa al problema di costruzione del test è nell'utilizzo

Indice

- 1 Generalità
- 2 Test connessi alla verosimiglianza
- 3 Regioni di confidenza
- 4 Verosimiglianza profilo**
- 5 Esempi vari

Verosimiglianza profilo

Consideriamo un modello (Y, p_θ, Θ) con

$$\theta = (\psi, \lambda)$$

(in generale le due componenti sono anch'esse multidimensionali.)

Definizione: verosimiglianza profilo

Data la verosimiglianza $L(\theta) = L(\psi, \lambda)$ si dice verosimiglianza profilo di ψ la funzione

$$L_p(\psi) = \max_{\lambda} L(\psi, \lambda)$$

In altre parole, per ogni ψ si calcola

$$\hat{\lambda}_\psi = \operatorname{argmax}_{\lambda} L(\psi, \lambda)$$

e si sostituisce questa al parametro λ in L ottenendo

$$L_p(\psi) = L(\psi, \hat{\lambda}_\psi)$$

Verosimiglianza profilo

Consideriamo un modello (Y, p_θ, Θ) con

$$\theta = (\psi, \lambda)$$

(in generale le due componenti sono anch'esse multidimensionali.)

Definizione: verosimiglianza profilo

Data la verosimiglianza $L(\theta) = L(\psi, \lambda)$ si dice verosimiglianza profilo di ψ la funzione

$$L_p(\psi) = \max_{\lambda} L(\psi, \lambda)$$

In maniera analoga si definisce la log-verosimiglianza profilo.

$$l_p(\psi) = l(\psi, \hat{\lambda}_\psi) = \max_{\lambda} l(\psi, \lambda)$$

Verosimiglianza profilo

Consideriamo un modello (Y, p_{θ}, Θ) con

$$\theta = (\psi, \lambda)$$

(in generale le due componenti sono anch'esse multidimensionali.)

Definizione: verosimiglianza profilo

Data la verosimiglianza $L(\theta) = L(\psi, \lambda)$ si dice verosimiglianza profilo di ψ la funzione

$$L_p(\psi) = \max_{\lambda} L(\psi, \lambda)$$

L'utilità è di isolare una parte del parametro, e lo si fa a volte per semplificare l'analisi, a volte perché solo una parte del parametro è di interesse, mentre l'altra è di disturbo.

Profilo per la media della normale

Consideriamo la verosimiglianza

$$L(\mu, \sigma^2) = (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right\}$$

Fissato μ , si ottiene $\hat{\sigma}_\mu^2$ massimizzando

$$f(\sigma^2) = l(\mu, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2$$

la cui derivata è

$$\frac{df}{d\sigma^2} = -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2$$

e si ha quindi

$$\hat{\sigma}_\mu^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2$$

Profilo per la media della normale

la verosimiglianza profilo è dunque

$$L_p(\mu) = (\hat{\sigma}_\mu^2)^{-n/2} \exp \left\{ -\frac{n}{2} \right\} \propto (\hat{\sigma}_\mu^2)^{-n/2}$$

La verosimiglianza profilo per σ è, molto più semplicemente

$$L_p(\sigma^2) = L(\bar{y}, \sigma^2)$$

dato che fissato σ^2 si ha $\hat{\mu}_{\sigma^2} = \hat{\mu} = \bar{y}$.

Profilo per la media della normale

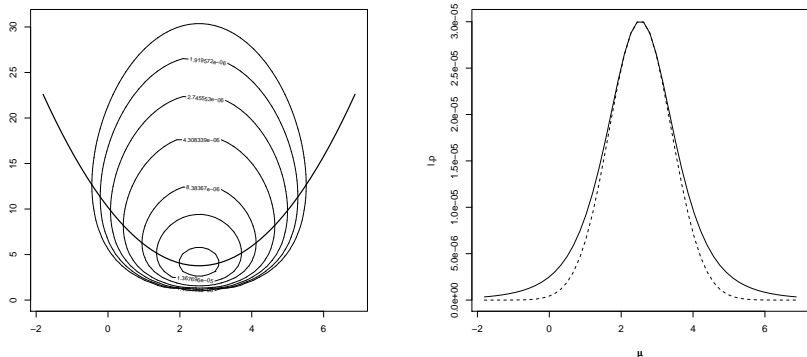


Figura: A sinistra: linee di livello della verosimiglianza della normale e luogo dei punti $(\mu, \hat{\sigma}_\mu^2)$; a destra: verosimiglianza profilo (linea continua) e sezione $L(\mu, \hat{\sigma}_\mu^2)$ (linea tratteggiata).

Caratteristiche della verosimiglianza profilo: il massimo

La funzione di verosimiglianza profilo viene poi trattata come una verosimiglianza, si ottengono perciò intervalli di confidenza profilo o si effettuano test profilo.

Massimo della verosimiglianza profilo e massimo della verosimiglianza

Si osserva facilmente che se si massimizza la verosimiglianza profilo il valore di massimo $\hat{\psi}_p$ che si ottiene è tale per cui $(\hat{\psi}_p, \hat{\lambda}_{\hat{\psi}})$ coincide con lo stimatore di massima verosimiglianza.

Caratteristiche della verosimiglianza profilo: il RVP

Data una verosimiglianza profilo si parla poi di rapporto di verosimiglianza profilo

$$\begin{aligned} W_p(\psi_0) &= 2(l_p(\hat{\psi}) - l_p(\psi_0)) \\ &= 2(l(\hat{\psi}, \hat{\lambda}) - l(\psi, \hat{\lambda}_{\psi_0})) \end{aligned}$$

Questo, nell'ipotesi $\psi = \psi_0$, ha distribuzione asintotica χ_q^2 con $q = \dim(\psi)$.

Indice

- 1 Generalità
- 2 Test connessi alla verosimiglianza
- 3 Regioni di confidenza
- 4 Verosimiglianza profilo
- 5 Esempi vari**

Indice

- 1 Generalità
- 2 Test connessi alla verosimiglianza
- 3 Regioni di confidenza
- 4 Verosimiglianza profilo
- 5 **Esempi vari**
 - **Test t**
 - Analisi della varianza a un criterio
 - Esempio: frequenze (multinomiale)

Esempio: media della normale

Sia $Y_1, \dots, Y_n \sim IID \mathcal{N}(\mu, \sigma^2)$ e

$$H_0 : \mu \in \Theta_0; \quad H_1 : \mu \in \Theta - \Theta_0$$

il rapporto di verosimiglianza è

$$\lambda(y) = \frac{\max_{\mu \in \Theta_0} L(\mu, \sigma^2)}{L(\hat{\mu}, \hat{\sigma}^2)} = \frac{L(\hat{\mu}_0, \hat{\sigma}_0^2)}{L(\hat{\mu}, \hat{\sigma}^2)}$$

dove

$$(\hat{\mu}_0, \hat{\sigma}_0^2) = \underset{\mu \in \Theta_0}{\operatorname{argmax}} L(\mu, \sigma^2)$$

notiamo che $\hat{\sigma}_0^2$ è la soluzione di

$$0 = \frac{dl(\hat{\mu}_0, \sigma^2)}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \hat{\mu}_0)^2 \Rightarrow \hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}_0)^2$$

Esempio: media della normale, RV

$$\begin{aligned}
 \lambda(y) &= \frac{(\hat{\sigma}_0^2)^{-n/2} \exp \left\{ -\frac{1}{2\hat{\sigma}_0^2} \sum_i (y_i - \hat{\mu}_0)^2 \right\}}{(\hat{\sigma}^2)^{-n/2} \exp \left\{ -\frac{1}{2\hat{\sigma}^2} \sum_i (y_i - \hat{\mu})^2 \right\}} \\
 &= \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} \right)^{-n/2} = \left(\frac{\sum (y_i - \hat{\mu}_0)^2}{\sum (y_i - \hat{\mu})^2} \right)^{-n/2} \\
 &= \left(\frac{\sum (y_i - \bar{y})^2 + n(\bar{y} - \hat{\mu}_0)^2}{\sum (y_i - \bar{y})^2} \right)^{-n/2} \\
 &= \left(1 + \frac{d^2}{n-1} \right)^{-n/2}
 \end{aligned}$$

con

$$d = \frac{\sqrt{n}(\bar{y} - \hat{\mu}_0)}{\sqrt{\frac{1}{n-1} \sum (y_i - \bar{y})^2}} \quad (1)$$

Esempio: media della normale, RV

Si ha dunque

$$W(y) = -2 \log \lambda(y) = n \log \left(1 + \frac{d^2}{n-1} \right).$$

e la regione di rifiuto ha forma

$$|d| = \frac{\sqrt{n}|\bar{y} - \hat{\mu}_0|}{\sqrt{\frac{1}{n-1} \sum (y_i - \bar{y})^2}} = \frac{\sqrt{n}|\bar{y} - \hat{\mu}_0|}{S} > k.$$

Per proseguire oltre è necessario precisare l'ipotesi H_0 .

Test bilaterale

Sia $\Theta_0 = \{\mu_0\}$, cioè

$$H_0 : \mu = \mu_0$$

quindi $\hat{\mu}_0 = \mu_0$ e

$$d = \frac{\sqrt{n}(\bar{y} - \mu_0)}{S} \sim_{H_0} t_{n-1}$$

dove questa quantità è nota come statistica t . (Si noti che

$$W(y) = n \log \left(1 + \frac{t^2}{n-1} \right) = n(t^2/(n-1) + o(1/n)) = t^2 + o(1) \rightarrow_d \chi_1^2.)$$

La regione di rifiuto di livello α è allora

$$t = \frac{\sqrt{n}|\bar{y} - \mu_0|}{S} > t_{n-1, 1-\alpha/2}$$

cioè si ottiene il familiare test

$$|\bar{y} - \mu_0| > \frac{S}{\sqrt{n}} t_{n-1, 1-\alpha/2}.$$

Test unilaterale

Sia $\Theta_0 = [-\infty, \mu_0]$, cioè

$$H_0 : \mu \leq \mu_0$$

quindi

$$\hat{\mu}_0 = \min\{\mu_0, \bar{y}\}$$

$$\hat{\sigma}_0^2 = \frac{1}{n} \sum_i (y_i - \hat{\mu}_0)^2 = \begin{cases} \hat{\sigma}^2 & \text{se } \bar{y} \leq \mu_0 \\ \frac{1}{n} \sum_i (y_i - \mu_0)^2 & \text{se } \bar{y} > \mu_0 \end{cases}$$

La statistica d è pari a

$$d = \max\{0, t\} = \begin{cases} 0 & \text{se } \bar{y} \leq \mu_0 \\ t & \text{se } \bar{y} > \mu_0 \end{cases}$$

e la regione di rifiuto è $t > t_{1-\alpha, n-1}$.

Test t a due campioni

Si hanno i due campioni

$$Y_1, \dots, Y_n \sim IID(\mathcal{N}(\mu_y, \sigma^2)) \quad Z_1, \dots, Z_m \sim IID(\mathcal{N}(\mu_z, \sigma^2))$$

con parametro

$$\theta = (\mu_y, \mu_z, \sigma^2) \in \Theta = \mathbb{R}^2 \mathbb{R}$$

Sia poi $\Theta_0 = \{(\mu, \mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+\}$ cioè

$$H_0 : \mu_y = \mu_z; \quad H_1 : \mu_y \neq \mu_z$$

Lo SMV è

$$\hat{\mu}_y = \bar{y}; \quad \hat{\mu}_z = \bar{z}; \quad \hat{\sigma}^2 = \frac{1}{n+m} \left(\sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^m (z_i - \bar{z})^2 \right)$$

mentre massimizzando limitatamente a Θ_0 si ha

$$\hat{\mu} = \frac{n\bar{y} + m\bar{z}}{n+m}; \quad \hat{\sigma}_0^2 = \frac{1}{n+m} \left(\sum_{i=1}^n (y_i - \hat{\mu})^2 + \sum_{i=1}^m (z_i - \hat{\mu})^2 \right)$$

Test t a due campioni, TRV

Il rapporto di verosimiglianza è allora

$$\begin{aligned} \lambda(y, z) &= \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} \right)^{-(n+m)/2} = \left(\frac{\sum_{i=1}^n (y_i - \hat{\mu})^2 + \sum_{i=1}^m (z_i - \hat{\mu})^2}{\sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^m (z_i - \bar{z})^2} \right)^{-(n+m)/2} \\ &= \left(\frac{\sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \hat{\mu})^2 + \sum_{i=1}^m (z_i - \bar{z})^2 + m(\bar{z} - \hat{\mu})^2}{\sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^m (z_i - \bar{z})^2} \right)^{-(n+m)/2} \\ &= \left(1 + \frac{n(\bar{y} - \hat{\mu})^2 + m(\bar{z} - \hat{\mu})^2}{\sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^m (z_i - \bar{z})^2} \right)^{-(n+m)/2} \end{aligned}$$

sostituendo $\hat{\mu} = (n\bar{y} + m\bar{z})/(n + m)$ si ottiene

$$= \left(1 + \frac{\frac{nm}{n+m}(\bar{y} - \bar{z})^2}{\sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^m (z_i - \bar{z})^2} \right)^{-(n+m)/2} = \left(1 + \frac{t^2}{n + m - 2} \right)^{-(n+m)/2}$$

con

$$t = \frac{\frac{\bar{y} - \bar{z}}{\sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^m (z_i - \bar{z})^2}{n + m - 2}}}$$

da cui la regione di rifiuto $|t| > k_\alpha$, ritroviamo l'usuale test.

Indice

- 1 Generalità
- 2 Test connessi alla verosimiglianza
- 3 Regioni di confidenza
- 4 Verosimiglianza profilo
- 5 **Esempi vari**
 - Test t
 - **Analisi della varianza a un criterio**
 - Esempio: frequenze (multinomiale)

Analisi della varianza a un criterio

Sia $Y_{ij} \sim \mathcal{N}(\mu_i, \sigma^2)$ per $i = 1, \dots, m$ e $j = 1, \dots, n$, indipendenti.

$H_0 : \mu_1 = \mu_2 = \dots = \mu_m$; $H_1 : \mu_h \neq \mu_k$ per qualche h, k

il massimo della verosimiglianza in Θ_0 si ha per

$$\hat{\mu}_0 = \frac{\sum_i \sum_j y_{ij}}{nm} = \bar{y}$$

$$\hat{\sigma}_0^2 = \frac{1}{nm} \sum_i \sum_j (y_{ij} - \bar{y})^2$$

mentre il massimo complessivo è per

$$\hat{\mu}_i = \frac{\sum_j y_{ij}}{m} = \bar{y}_i$$

$$\hat{\sigma}^2 = \frac{\sum_i \hat{\sigma}_i^2}{m} = \frac{1}{m} \sum_i \left(\sum_j (y_{ij} - \bar{y}_i)^2 \right)$$

Analisi della varianza a un criterio, RV

e quindi

$$\begin{aligned}
 \lambda(y) &= \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} \right)^{-nm/2} = \left(\frac{\sum_i \sum_j (y_{ij} - \bar{y})^2}{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2} \right)^{-nm/2} \\
 &= \left(\frac{\sum_i \sum_j (y_{ij} - \bar{y}_i + \bar{y}_i - \bar{y})^2}{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2} \right)^{-nm/2} \\
 &= \left(1 + \frac{n \sum_i (\bar{y}_i - \bar{y})^2}{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2} \right)^{-nm/2}
 \end{aligned}$$

e la regione critica si ha per

$$\frac{D_0}{D} = \frac{n \sum_i (\bar{y}_i - \bar{y})^2}{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2} > k$$

Analisi della varianza a un criterio, RV

e quindi

$$\lambda(y) = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} \right)^{-nm/2} = \left(1 + \frac{n \sum_i (\bar{y}_i - \bar{y})^2}{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2} \right)^{-nm/2}$$

e la regione critica si ha per

$$\frac{D_0}{D} = \frac{n \sum_i (\bar{y}_i - \bar{y})^2}{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2} > k$$

dove $D_0/\sigma^2 \sim \chi_{m-1}^2$ e $D/\sigma^2 \sim \chi_{m(n-1)}^2$ e D_0 e D sono indipendenti, si ha allora

$$F = \frac{D_0/(m-1)}{D/(m(n-1))} \sim F_{m-1, m(n-1)}$$

Indice

- 1 Generalità
- 2 Test connessi alla verosimiglianza
- 3 Regioni di confidenza
- 4 Verosimiglianza profilo
- 5 **Esempi vari**
 - Test t
 - Analisi della varianza a un criterio
 - **Esempio: frequenze (multinomiale)**

Esempio multinomiale

Si ha un campione Y_1, \dots, Y_n dove Y_i ha $r + 1$ possibili valori h_j e

$$P(Y_i = h_j) = \pi_j \quad j = 0, \dots, r$$

con $\sum_{j=0}^r \pi_j = 1$. La verosimiglianza è

$$L(\pi) = \prod_{i=1}^n \prod_{j=0}^r \pi_j^{|y_i=h_j|}$$

posto $n_j = \sum_i |Y_i = h_j|$, (si ha $\sum_{j=0}^r n_j = n$) la verosimiglianza è

$$L(\pi) = \prod_{j=1}^r \pi_j^{n_j}$$

$((n_1, \dots, n_r) \sim \text{multinomiale}(n, \pi))$

Esempio multinomiale: SMV

Con ciò si ha

$$l(\pi) = \sum_{j=0}^r n_j \log \pi_j = \sum_{j=1}^r n_j \log \pi_j + n_0 \log \left(1 - \sum_{j=1}^r \pi_j \right)$$

e quindi si hanno le equazioni di verosimiglianza

$$\frac{\partial l(\pi)}{\partial \pi_j} = \frac{n_j}{\pi_j} - \frac{n_0}{\pi_0} = 0, \quad j = 1, \dots, r$$

per $\hat{\pi}_0$ si ha

$$\pi_j = \frac{n_j}{n_0} \pi_0 \Rightarrow 1 = \sum_{j=0}^r \pi_j = \sum_{j=0}^r \frac{n_j}{n_0} \pi_0 = \frac{n}{n_0} \pi_0 \Rightarrow \hat{\pi}_0 = n_0/n$$

da cui

$$\hat{\pi}_j = \frac{n_j}{n} \quad j = 0, \dots, r$$

Esempio multinomiale: TRV

Sia l'ipotesi nulla rappresentata da una particolare distribuzione per Y

$$H_0 : \pi_j = p_j \quad j = 1, \dots, r$$

contro

$$H_1 : \pi_j \neq p_j \quad \text{per qualche } j$$

Si ha quindi

$$W(n_1, \dots, n_r) = 2(l(\hat{\pi}) - l(p)) = 2 \sum_{j=0}^r n_j \log \frac{\hat{\pi}_j}{p_j}$$

con distribuzione asintotica χ_r^2 . Questo è noto anche come test G .

Esempio multinomiale: informazione

Si ha

$$[\mathcal{I}(\pi)]_{ij} = E \left(-\frac{\partial^2 l(\pi)}{\partial \pi_i \partial \pi_j} \right) = E \left(|i = j| \frac{n_j}{\pi_j^2} - \frac{n_0}{\pi_0^2} \right) = |i = j| \frac{n}{\pi_j} + \frac{n}{\pi_0}$$

quindi posto $D = \text{diag} \left(\frac{\pi_0}{\pi_1}, \dots, \frac{\pi_0}{\pi_r} \right)$

$$\mathcal{I}(\pi) = (D + \mathbf{1}\mathbf{1}^T) \frac{n}{\pi_0}$$

si ha allora

$$\mathcal{I}(\pi)^{-1} = \frac{\pi_0}{n} (D + \mathbf{1}\mathbf{1}^T)^{-1} = \frac{\pi_0}{n} \left(D^{-1} - \frac{1}{1 + \mathbf{1}^T D^{-1} \mathbf{1}} D^{-1} \mathbf{1}\mathbf{1}^T D^{-1} \right)$$

Esempio multinomiale: informazione

Si ha $D^{-1} = \text{diag} \left(\frac{\pi_1}{\pi_0}, \dots, \frac{\pi_r}{\pi_0} \right)$ e quindi

$$\mathbf{1}^T D^{-1} \mathbf{1} = \sum_{j=1}^r \frac{\pi_j}{\pi_0} = \frac{1 - \pi_0}{\pi_0}$$

da cui

$$(1 + \mathbf{1}^T D^{-1} \mathbf{1})^{-1} = \pi_0$$

e

$$[D^{-1} \mathbf{1} \mathbf{1}^T D^{-1}]_{ij} = \pi_i \pi_j / \pi_0^2$$

Si ha perciò

$$\begin{aligned} [\mathcal{I}(\pi)^{-1}]_{ij} &= \frac{\pi_0}{n} \left(|i = j| \frac{\pi_j}{\pi_0} - \frac{\pi_i \pi_j}{\pi_0} \right) \\ &= n (|i = j| \pi_j - \pi_i \pi_j) \end{aligned}$$

Esempio multinomiale: test del punteggio

Essendo

$$[l_*(\pi)]_j = n_j/\pi_j - n_0/\pi_0$$

il test del punteggio è dato da

$$W_u = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^r \left(\frac{n_i}{p_i} - \frac{n_0}{p_0} \right) (|i=j|p_j - p_i p_j) \left(\frac{n_j}{p_j} - \frac{n_0}{p_0} \right)$$

$$= \frac{1}{n} \left(\sum_{i=1}^r \left(\frac{n_i}{p_i} - \frac{n_0}{p_0} \right)^2 p_i + \underbrace{\sum_{i=1}^r \sum_{j=1}^r \left(\frac{n_i}{p_i} - \frac{n_0}{p_0} \right) (|i=j|p_j - p_i p_j) \left(\frac{n_j}{p_j} - \frac{n_0}{p_0} \right)}_S \right)$$

dove la seconda sommatoria è pari a

$$S = \sum_{i=1}^r \left(\frac{n_i}{p_i} - \frac{n_0}{p_0} \right) p_i \sum_{j=1}^r \left(n_j - \frac{n_0 p_j}{p_0} \right) = \sum_{i=1}^r \left(\frac{n_i}{p_i} - \frac{n_0}{p_0} \right) p_i \left(n - n_0 - \frac{n_0}{p_0} (1 - p_0) \right)$$

$$= \sum_{i=1}^r \left(\frac{n_i}{p_i} - \frac{n_0}{p_0} \right) p_i \left(n - \frac{n_0}{p_0} \right) = \left(n - \frac{n_0}{p_0} \right)^2$$

Esempio multinomiale: test del punteggio

sostituendo perciò si ottiene

$$\begin{aligned}
 nW_u &= \sum_{i=1}^r \left(\frac{n_i}{p_i} - \frac{n_0}{p_0} \right)^2 p_i + \left(n - \frac{n_0}{p_0} \right)^2 \\
 &= \sum_{i=1}^r \left(\frac{n_i^2}{p_i} + \frac{n_0^2 p_i}{p_0^2} - 2 \frac{n_0 n_i}{p_0} \right) - n^2 - \frac{n_0^2}{p_0^2} - 2 \frac{nn_0}{p_0} \\
 &= \sum_{i=1}^r \frac{n_i^2}{p_i} + \frac{n_0^2}{p_0^2} (1 - p_0) - 2 \frac{n_0}{p_0} (n - n_0) - n^2 - \frac{n_0^2}{p_0^2} - 2 \frac{nn_0}{p_0} \\
 &= \sum_{i=1}^r \frac{n_i^2}{p_i} + \frac{n_0^2}{p_0} - n^2 = \sum_{i=0}^r \frac{n_i^2}{p_i} - 2n^2 + n^2 = \sum_{i=0}^r \left(\frac{n_i^2}{p_i} - 2nn_i + n^2 p_i \right) \\
 &= \sum_{i=0}^r \frac{1}{p_i} (n_i - 2nn_i p_i + n^2 p_i^2) = \sum_{i=0}^r \frac{1}{p_i} (n_i n_i p_i)^2
 \end{aligned}$$

cioè si ottiene

$$W_u(N_1, \dots, N_r) = \sum_{j=0}^r \frac{(n_j - np_j)^2}{np_j}$$

Esempio multinomiale: TRV e test del punteggio

Si hanno dunque le due statistiche test

$$G = W(n_1, \dots, n_r) = 2(l(\hat{\pi}) - l(p)) = 2 \sum_{j=0}^r n_j \log \frac{\hat{\pi}_j}{p_j}$$

$$\chi^2 = W_u(N_1, \dots, N_r) = \sum_{j=0}^r \frac{(n_j - np_j)^2}{np_j}$$

Quale si sceglie? Il test noto come χ^2 (cioè W_u) ha migliore approssimazione asintotica anche per frequenze attese relativamente basse ($np_j > 2$) mentre per G si richiede $np_j > 5$.