

Università degli studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Triennale in  
STATISTICA E TECNOLOGIE INFORMATICHE



RELAZIONE FINALE

**TEST BASATI SULLA VEROSIMIGLIANZA IN UN MODELLO  
PER LA CLASSIFICAZIONE SCORRETTA DI DATI BINARI**

Relatore Prof. Nicola Sartori  
Dipartimento di Scienze Statistiche

Laureando Edoardo Vignotto  
Matricola N 1077495

Anno Accademico 2015/2016

# Indice

<b>Elenco delle figure</b>	<b>v</b>
<b>Elenco delle tabelle</b>	<b>vii</b>
<b>Introduzione</b>	<b>ix</b>
<b>1 Concetti preliminari</b>	<b>1</b>
1.1 L'inferenza statistica . . . . .	1
1.2 Modello statistico . . . . .	2
1.3 Verosimiglianza . . . . .	3
1.3.1 Stima di massima verosimiglianza . . . . .	3
1.3.2 Funzione di verosimiglianza . . . . .	4
1.3.3 Condizioni di regolarità . . . . .	5
1.3.4 Quantità notevoli . . . . .	6
1.3.5 Proprietà legate alla verosimiglianza . . . . .	6
1.3.6 Distribuzione asintotica . . . . .	8
1.3.7 Test statistico . . . . .	9
1.3.8 Regioni di confidenza . . . . .	14
1.3.9 Verosimiglianza profilo . . . . .	16
1.3.10 Modelli a variabili indipendenti . . . . .	16
1.4 Statistiche sufficienti, ancillari e completezza . . . . .	17
1.5 Modello per dati dicotomici . . . . .	19
1.6 Simboli di Landau . . . . .	21

<b>2</b>	<b>Classificazione scorretta di dati binari</b>	<b>23</b>
2.1	Il modello . . . . .	23
2.2	Verifica di ipotesi . . . . .	26
<b>3</b>	<b>Radice con segno modificata del log-rapporto di verosiglianza</b>	<b>29</b>
3.1	Formula $p^*$ . . . . .	29
3.2	Statistica test $r_p^*$ . . . . .	30
3.3	Package <code>likelihoodAsy</code> . . . . .	31
3.4	Utilizzo di <code>likelihoohAsy</code> . . . . .	32
<b>4</b>	<b>Bootstrap prepivoting di <math>r</math></b>	<b>37</b>
4.1	Il bootstrap . . . . .	37
4.2	Bootstrap prepivoting di $r_p$ . . . . .	38
4.3	Intervalli di confidenza col bootstrap . . . . .	40
<b>5</b>	<b>Studi di simulazione</b>	<b>43</b>
5.1	Struttura della simulazione . . . . .	43
5.2	Test su $\beta_1$ . . . . .	44
5.2.1	$n = 100$ con $\eta = 0.95$ e $\omega = 0.85$ . . . . .	44
5.2.2	$n = 250$ con $\eta = 0.9$ e $\omega = 0.8$ . . . . .	46
5.2.3	$n = 250$ con $\eta = 0.95$ e $\omega = 0.85$ . . . . .	48
5.3	Test su $\beta_2$ . . . . .	50
5.3.1	$n = 100$ con $\eta = 0.95$ e $\omega = 0.85$ . . . . .	50
5.3.2	$n = 250$ con $\eta = 0.9$ e $\omega = 0.8$ . . . . .	52
5.3.3	$n = 250$ con $\eta = 0.95$ e $\omega = 0.85$ . . . . .	54
5.4	Test su $\beta_3$ . . . . .	56
5.4.1	$n = 100$ con $\eta = 0.95$ e $\omega = 0.85$ . . . . .	56
5.4.2	$n = 250$ con $\eta = 0.9$ e $\omega = 0.8$ . . . . .	58
5.4.3	$n = 250$ con $\eta = 0.95$ e $\omega = 0.85$ . . . . .	60
5.5	Commenti . . . . .	62

<i>INDICE</i>	iii
<b>6 Conclusioni</b>	<b>65</b>
<b>A Codice R</b>	<b>67</b>
<b>Bibliografia</b>	<b>95</b>



# Elenco delle figure

3.1	Valori di $r_p$ e $r_p^*$ per $\psi$ con $n = 100$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_1$	35
4.1	Grafico per la costruzione degli intervalli di confidenza con le procedure bootstrap con $n = 100$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_1$	41
5.1	qq-plot ottenuto con $r$ e $r^*$ per $n = 100$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_1$	45
5.2	qq-plot ottenuto col bootstrap di secondo e terzo ordine per $n = 100$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_1$	45
5.3	qq-plot ottenuto con $r$ e $r^*$ per $n = 250$ , $\eta = 0.9$ e $\omega = 0.8$ per $\beta_1$	47
5.4	qq-plot ottenuto col bootstrap di secondo e terzo ordine per $n = 250$ , $\eta = 0.9$ e $\omega = 0.8$ per $\beta_1$	47
5.5	qq-plot ottenuto con $r$ e $r^*$ per $n = 250$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_1$	49
5.6	qq-plot ottenuto col bootstrap di secondo e terzo ordine per $n = 250$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_1$	49
5.7	qq-plot ottenuto con $r$ e $r^*$ per $n = 100$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_2$	51
5.8	qq-plot ottenuto col bootstrap di secondo e terzo ordine per $n = 100$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_2$	51
5.9	qq-plot ottenuto con $r$ e $r^*$ per $n = 250$ , $\eta = 0.9$ e $\omega = 0.8$ per $\beta_2$	53

5.10 qq-plot ottenuto col bootstrap di secondo e terzo ordine per $n = 250, \eta = 0.9$ e $\omega = 0.8$ per $\beta_2$ . . . . .	53
5.11 qq-plot ottenuto con $r$ e $r^*$ per $n = 250, \eta = 0.95$ e $\omega = 0.85$ per $\beta_2$ . . . . .	55
5.12 qq-plot ottenuto col bootstrap di secondo e terzo ordine per $n = 250, \eta = 0.95$ e $\omega = 0.85$ per $\beta_2$ . . . . .	55
5.13 qq-plot ottenuto con $r$ e $r^*$ per $n = 100, \eta = 0.95$ e $\omega = 0.85$ per $\beta_3$ . . . . .	57
5.14 qq-plot ottenuto col bootstrap di secondo e terzo ordine per $n = 100, \eta = 0.95$ e $\omega = 0.85$ per $\beta_3$ . . . . .	57
5.15 qq-plot ottenuto con $r$ e $r^*$ per $n = 250, \eta = 0.9$ e $\omega = 0.8$ per $\beta_3$ . . . . .	59
5.16 qq-plot ottenuto col bootstrap di secondo e terzo ordine per $n = 250, \eta = 0.9$ e $\omega = 0.8$ per $\beta_3$ . . . . .	59
5.17 qq-plot ottenuto con $r$ e $r^*$ per $n = 250, \eta = 0.95$ e $\omega = 0.85$ per $\beta_3$ . . . . .	61
5.18 qq-plot ottenuto col bootstrap di secondo e terzo ordine per $n = 250, \eta = 0.95$ e $\omega = 0.85$ per $\beta_3$ . . . . .	61

# Elenco delle tabelle

3.1	Intervalli di confidenza per $\psi$ basati su $r_p^*$ e $r_p$ , con $n = 100$ , $\eta = 0.95$ , $\omega = 0.85$ e $x^* = (0, 1, 1)$ . . . . .	35
5.1	Risultati del test di Kolmogorov-Smirnov per $n = 100$ , $\eta =$ $0.95$ e $\omega = 0.85$ per $\beta_1$ . . . . .	44
5.2	$\alpha$ percentuali osservati confrontati con quelli teorici con $n =$ $100$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_1$ . . . . .	46
5.3	Livelli di copertura con $n = 100$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_1$	46
5.4	Risultati del test di Kolmogorov-Smirnov per $n = 250$ , $\eta = 0.9$ e $\omega = 0.8$ per $\beta_1$ . . . . .	46
5.5	$\alpha$ percentuali osservati confrontati con quelli teorici per $n =$ $250$ , $\eta = 0.9$ e $\omega = 0.8$ per $\beta_1$ . . . . .	48
5.6	Livelli di copertura con $n = 250$ , $\eta = 0.9$ e $\omega = 0.8$ per $\beta_1$ . .	48
5.7	Risultati del test di Kolmogorov-Smirnov per $n = 250$ , $\eta =$ $0.95$ e $\omega = 0.85$ per $\beta_1$ . . . . .	48
5.8	$\alpha$ percentuali osservati confrontati con quelli teorici per $n =$ $250$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_1$ . . . . .	50
5.9	Livelli di copertura con $n = 250$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_1$	50
5.10	Risultati del test di Kolmogorov-Smirnov per $n = 100$ , $\eta =$ $0.95$ e $\omega = 0.85$ per $\beta_2$ . . . . .	50
5.11	$\alpha$ percentuali osservati confrontati con quelli teorici per $n =$ $100$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_2$ . . . . .	52
5.12	Livelli di copertura con $n = 100$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_2$	52



5.13	Risultati del test di Kolmogorov-Smirnov per $n = 250$ , $\eta = 0.9$ e $\omega = 0.8$ per $\beta_2$ . . . . .	52
5.14	$\alpha$ percentuali osservati confrontati con quelli teorici per $n =$ $250$ , $\eta = 0.9$ e $\omega = 0.8$ per $\beta_2$ . . . . .	54
5.15	Livelli di copertura con $n = 250$ , $\eta = 0.9$ e $\omega = 0.8$ per $\beta_2$ . .	54
5.16	Risultati del test di Kolmogorov-Smirnov per $n = 250$ , $\eta =$ $0.95$ e $\omega = 0.85$ per $\beta_2$ . . . . .	54
5.17	$\alpha$ percentuali osservati confrontati con quelli teorici per $n =$ $250$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_2$ . . . . .	56
5.18	Livelli di copertura con $n = 250$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_2$	56
5.19	Risultati del test di Kolmogorov-Smirnov per $n = 100$ , $\eta =$ $0.95$ e $\omega = 0.85$ per $\beta_3$ . . . . .	56
5.20	$\alpha$ percentuali osservati confrontati con quelli teorici per $n =$ $100$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_3$ . . . . .	58
5.21	Livelli di copertura con $n = 100$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_3$	58
5.22	Risultati del test di Kolmogorov-Smirnov per $n = 250$ , $\eta = 0.9$ e $\omega = 0.8$ per $\beta_3$ . . . . .	58
5.23	$\alpha$ percentuali osservati confrontati con quelli teorici per $n =$ $250$ , $\eta = 0.9$ e $\omega = 0.8$ per $\beta_3$ . . . . .	60
5.24	Livelli di copertura con $n = 250$ , $\eta = 0.9$ e $\omega = 0.8$ per $\beta_3$ . .	60
5.25	Risultati del test di Kolmogorov-Smirnov per $n = 250$ , $\eta =$ $0.95$ e $\omega = 0.85$ per $\beta_3$ . . . . .	60
5.26	$\alpha$ percentuali osservati confrontati con quelli teorici per $n =$ $250$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_3$ . . . . .	62
5.27	Livelli di copertura con $n = 250$ , $\eta = 0.95$ e $\omega = 0.85$ per $\beta_3$	62

# Introduzione

Obiettivo di questo elaborato è di presentare un modello per la classificazione scorretta di dati binari e valutarne il comportamento in relazione a diverse procedure statistiche, utilizzando il software R.

In molte realtà è facile imbattersi in dati binari classificati scorrettamente. Ciò avviene in tutti quei casi in cui per la classificazione ci si affida a un qualche criterio non completamente corretto, come può avvenire in medicina quando si usa un test diagnostico con possibilità di errore per individuare gli individui affetti da una qualche malattia.

Rifacendosi a tale esempio, nel caso in cui si volesse costruire un modello di regressione per prevedere la probabilità di essere malato di un soggetto con certe caratteristiche, usare come risposta i risultati di tale test porterebbe a delle ovvie distorsioni. Un modello nato per risolvere questo tipo di problemi è quello presentato da P. McInturff et al. (2003), di cui ci si occuperà.

In particolare si studierà il suo comportamento in relazione a quattro diversi test per la verifica di ipotesi, che verranno presentati nei primi capitoli. Il più semplice test valutato sarà l'usuale statistica radice con segno del log-rapporto di log-verosimiglianza. Si vedrà poi una sua modificazione, implementata da Biello e Pierce (2015) nel pacchetto `likelihoodAsy`, che, sempre fondandosi sulla teoria della verosimiglianza, raggiunge un'accuratezza inferenziale maggiore. Infine si vedranno due alternative basate sul bootstrap parametrico. La bontà di tali metodologie verrà confrontata mediante simulazioni.

Si mostrerà, inoltre, come è possibile costruire intervalli di confidenza

con ognuna delle statistiche discusse.

# Capitolo 1

## Concetti preliminari

In questo capitolo ci si occuperà di introdurre alcuni concetti di base riguardanti l'inferenza statistica e il calcolo delle probabilità, che torneranno poi utili nel prosieguo di questo elaborato.

### 1.1 L'inferenza statistica

Per inferenza statistica si intende quell'insieme di metodi e di tecniche atte a trarre conclusioni di carattere generale su un fenomeno oggetto di studio, a partire da un'osservazione parziale di esso. In particolare, l'interesse è spesso rivolto ad uno o più *caratteri* di tale fenomeno e ai distinti modi in cui essi possono manifestarsi, ovvero alle loro *modalità*. L'insieme degli elementi oggetto di indagine formerà la *popolazione statistica* e ognuno di essi verrà detto *unità statistica*. Per vari motivi (costo di ispezione, tempestività, popolazione non ben definita, campionamento distruttivo etc.) spesso non è possibile osservare l'intera popolazione e si è quindi costretti ad accontentarsi di un'osservazione parziale, limitata solo ad alcune unità statistiche. Da questo *campione* di riferimento si vorrà, poi, inferire risultati di carattere generale sull'intera popolazione, con grande importanza riservata al controllo dell'errore che questa procedura induttiva può portare con sé (Azzalini, 2001, Capitolo 1). Ci sono diversi modi per fare tutto ciò: in

particolare si possono pensare le varie modalità osservate come realizzazioni di una *variabile casuale* latente e quindi usare risultati propri del calcolo delle probabilità per giungere a tali obiettivi. Oppure, soprattutto quando ciò non è analiticamente agevole, possiamo ricorrere a (pseudo)simulazioni di tali variabili o, per meglio dire, approssimazioni di esse suggerite dai dati osservati, al fine di coglierne il comportamento. I metodi di verosimiglianza che presenteremo appartengono al primo di questi casi, i metodi bootstrap.

Verranno usati, spesso implicitamente, principi propri del *paradigma fisheriano* e di quello *decisionale frequentista*: punti cardine del primo sono il *principio del campionamento ripetuto*, secondo il quale il campione osservato va considerato come una particolare realizzazione delle infinite che si potevano ottenere, e che avrebbero condotto quindi a conclusioni differenti, suggerendo quindi di modellare in questo senso la stima dell'errore commesso. Altri importanti principi sono il *principio di verosimiglianza*, che poi verrà trattato nel dettaglio, e quello di *pertinenza*, che afferma che la probabilità descrittiva di un evento deve essere considerata condizionata a tutto ciò che è noto. Il secondo paradigma, che è di fatto un'estensione del primo, si preoccupa di applicare principi medesimi guardando però all'inferenza statistica soprattutto come mezzo per risolvere problemi di decisione, cercando di trovare principi generali, non legati allo specifico problema, per affrontarli (Pace e Salvan, 1996, Capitolo 1). Tali principi sono la base logica su cui si basa in ultima istanza la correttezza del ragionamento induttivo che è poi l'inferenza statistica stessa.

## 1.2 Modello statistico

Specificare il modello statistico di riferimento significa considerare l'insieme delle funzioni di ripartizione (o, analogamente, delle funzioni di densità) a cui appartiene quella della variabile casuale che descrive il fenomeno in questione come un insieme di funzioni tutte dello stesso tipo. Nel caso di

modelli parametrici, tali distribuzioni sono distinte unicamente dal valore di un *parametro*  $\theta$  (all'occorrenza multidimensionale), libero di variare entro l'insieme  $\Theta \subseteq \mathbb{R}^d$ , con  $\dim(\theta) = d$ , chiamato *spazio parametrico*; in simboli, tale modello parametrico si può denotare come

$$\mathcal{F} = \{F(\cdot; \theta) : \theta \in \Theta \subseteq \mathbb{R}^d\},$$

dove  $F(\cdot; \theta)$  indica la funzione di ripartizione per un dato valore  $\theta$  (Azzalini, 2001, Capitolo 2). Se vi è una corrispondenza biunivoca tra lo spazio parametrico  $\Theta$  e l'insieme delle distribuzioni considerate dal modello si dirà che esso è *identificabile*, condizione assolutamente basilare se si vogliono ottenere buoni risultati inferenziali. Ci si occuperà quindi di fare inferenza riguardo a  $\theta_0$ , vero valore del parametro, associato a  $F_0$ , vera distribuzione della variabile casuale che descrive il fenomeno studiato. L'insieme  $\mathcal{Y}$  di tutti i possibili valori  $y$  che il campione può assumere, compatibilmente con il modello specificato, è chiamato *spazio campionario*. Il campione effettivamente osservato sarà indicato con  $y^{oss}$ .

## 1.3 Verosimiglianza

### 1.3.1 Stima di massima verosimiglianza

In una prima fase, quindi, si individuerà il modello statistico (anche detto *classe parametrica*) adeguato a descrivere il fenomeno oggetto di studio, nell'interesse di trarre informazioni sul parametro  $\theta$  che indicizza tale modello, che per questo elaborato sarà un parametro continuo libero di variare in tutto  $\mathbb{R}^d$ . Il primissimo problema che si pone è, poi, quello di calcolare una *stima puntuale* di tale parametro, ovvero di individuare un valore di esso compatibile con i dati osservati, che denoteremo con  $\hat{\theta}$ . Anche in ciò ci sono diversi modi in cui si potrebbe procedere e qui si seguirà l'idea di *stima di massima verosimiglianza*: la nostra stima per il parametro  $\theta$  sarà quel valore che massimizza la probabilità (nel caso discreto, o densità, in

un contesto continuo) di riosservare un campione esattamente concordante, rispetto alle ipotesi fatte, con quello realmente osservato, o, in altri termini, quel valore per il parametro che più si accorda con i dati osservati.

### 1.3.2 Funzione di verosimiglianza

Per rendere tutto ciò agevole è utile descrivere tale probabilità o densità in funzione solo di  $\theta$ , condizionatamente ai dati osservati. Si avrà in questo modo a che fare con una funzione del tipo

$$L(\theta; y^{oss}) = p(\theta; y^{oss})$$

dove per  $p(\theta; y^{oss})$  si deve intendere la funzione di probabilità o densità che descrive il modello, al variare di  $\theta$ , calcolata nel valore  $y^{oss}$  fissato. Tale funzione, detta *funzione di verosimiglianza*, è spesso indicata semplicemente con  $L(\theta)$ , proprio per sottolineare il fatto che è funzione solo del parametro. Ne deriva che la stima di massima verosimiglianza per  $\theta$  sarà semplicemente il valore che massimizza tale funzione, ovvero

$$\hat{\theta} = \arg \max_{\theta} L(\theta).$$

E' interessante notare che, poichè siamo interessati a trovare il massimo della funzione di verosimiglianza, lavorare con una qualsiasi funzione che è trasformazione strettamente monotona crescente di essa darà risultati del tutto analoghi. Per questo usualmente si usa la *funzione di log-verosimiglianza*, definita come

$$l(\theta) = \log L(\theta)$$

con la convenzione  $l(\theta) = -\infty$  se  $L(\theta) = 0$ , da cui si semplificano eventuali costanti additive non dipendenti da  $\theta$ ; in questo modo, solitamente, i calcoli risultano più semplici.

Assumendo che la stima di massima verosimiglianza (S.M.V) esista e sia unica, si pone il problema di come effettivamente calcolarla: essendo un

problema di massimo essa può essere cercata tra le soluzioni del sistema

$$\begin{cases} \frac{\partial l(\theta)}{\partial \theta_1} = 0 \\ \vdots \\ \frac{\partial l(\theta)}{\partial \theta_d} = 0 \end{cases}$$

ovvero tra i punti stazionari della funzione di log-verosimiglianza. Le equazioni che compongono il sistema, chiamate *equazioni di verosimiglianza*, a volte non hanno soluzione esplicita e si dovranno risolvere per via numerica.

### 1.3.3 Condizioni di regolarità

Verranno introdotte ore alcune condizioni di regolarità necessarie per garantire la validità delle proprietà delle procedure di verosimiglianza e, in particolare, dello *stimatore di massima verosimiglianza*, ovvero, seguendo il principio del campionamento ripetuto, dell'oggetto casuale individuato dai valori dalle varie stime che infiniti campioni estratti da  $F_0$  fornirebbero. Ha la stessa forma della S.M.V, in cui però l'osservazione  $y^{oss}$  viene sostituita con la corrispondente variabile casuale  $Y$ , e viene indicato con lo stesso simbolo  $\hat{\theta}$ ; la non ambiguità sarà garantita dal contesto.

Per essere nella situazione di un *problema regolare di stima* devono verificarsi le seguenti condizioni (si veda, ad esempio, Azzalini, 2001, Capitolo 3):

- il modello statistico è identificabile,
- lo spazio parametrico  $\Theta$  è un intervallo aperto di  $\mathbb{R}^d$ , con  $d = \dim(\theta) < \infty$ , indipendente dal parametro  $\theta$ ,
- il supporto delle funzioni di densità/probabilità è lo stesso per tutte quelle appartenenti al modello statistico specificato,
- la log-verosimiglianza è una funzione derivabile almeno tre volte con derivate parziali continue in  $\Theta$ .



### 1.3.4 Quantità notevoli

Ci sono alcune quantità notevoli associate alla funzione di log-verosimiglianza, tra cui:

- lo *score di verosimiglianza*  $l_\theta(\theta)$ , che porta al sistema di equazioni di verosimiglianza le cui radici individuano i punti stazionari tra cui cercare la S.M.V e, più in generale, la forma dello stimatore di massima verosimiglianza;
- l'*informazione osservata di Fisher*, che corrisponde a  $-l_{\theta\theta}(\hat{\theta})$  e si indica con  $j(\hat{\theta})$ , matrice in generale semidefinita positiva, e che rappresenta un indice della rapidità con cui la verosimiglianza cala allontanandosi dalla S.M.V. Il senso di ciò si ha, considerando per semplicità il caso  $\dim(\theta) = 1$ , notando che il comportamento di  $l(\theta)$  in un intorno di  $\hat{\theta}$  è governato proprio da  $j(\hat{\theta})$ :

$$\begin{aligned} l(\theta) &= l(\hat{\theta}) + l_\theta(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{2}l_{\theta\theta}(\hat{\theta})(\theta - \hat{\theta})^2 + \dots = \\ &= l(\hat{\theta}) + \frac{1}{2}l_{\theta\theta}(\hat{\theta})(\theta - \hat{\theta})^2 + \dots = \\ &= l(\hat{\theta}) - \frac{1}{2}j(\hat{\theta})(\theta - \hat{\theta})^2 + \dots \end{aligned}$$

ricordando che  $l_\theta(\hat{\theta}) = 0$ , in quanto  $\hat{\theta}$  è punto stazionario per  $l(\theta)$ ;

- l'*informazione attesa di Fisher*, matrice semidefinita positiva che corrisponde alla varianza dello score di verosimiglianza, inteso ora a sua volta come oggetto casuale in quanto funzione della variabile casuale  $Y$ . Ragionamento analogo sarà, da qui in poi, all'occorrenza fatto anche per le altre quantità, senza più essere specificato. La matrice di informazione attesa si indica con  $i(\theta)$ . Si può mostrare inoltre che

$$i(\theta) = \text{var}_\theta[l_\theta(\theta; Y)] = E_\theta[j(\theta; Y)]$$

### 1.3.5 Proprietà legate alla verosimiglianza

Le proprietà più importanti legate alle funzione di verosimiglianza sono:

- *proprietà di equivarianza*, che dice che, se riscriviamo il parametro  $\theta \in \Theta$  come  $\psi = g(\theta)$ ,  $\psi \in \Psi$ , dove  $g(\cdot)$  è funzione biunivoca da  $\Theta$  in  $\Psi$ , allora si ha che  $\hat{\psi} = g(\hat{\theta})$ , con  $\hat{\psi}$  S.M.V. per  $\psi$ . Vale inoltre la relazione  $i(\psi) = i(\theta)[\psi'(\theta)]^{-2}$ , nel caso  $\dim(\theta) = 1$ , dove si indica con  $\psi(\theta)$  la trasformazione  $g(\theta)$ ; si userà  $\theta(\psi)$  per la sua inversa;
- *consistenza dello stimatore di massima verosimiglianza*, nel senso che, se  $\hat{\theta}_n(Y)$  è lo stimatore di massima verosimiglianza per  $\theta$ , si ha che

$$\lim_{n \rightarrow \infty} Pr(|\hat{\theta}_n(Y) - \theta_0| > \epsilon) = 0$$

per ogni  $\epsilon > 0$ , quando  $\theta_0$  è il vero valore del parametro. Ciò mostra come, al crescere della numerosità campionaria, lo stimatore  $\hat{\theta}_n(Y)$  converga in probabilità a  $\theta_0$ .

La prima di queste proprietà è molto importante in quanto ci garantisce che i risultati ottenuti operando con la verosimiglianza non sono influenzati dal modo in cui si esprime il parametro da studiare, qualità assolutamente ragionevole per la bontà di qualsiasi procedura inferenziale; ciò permette inoltre di riparametrizzare il modello in base alle necessità, per evidenziarne un aspetto piuttosto che un altro, o per semplificare il lavoro analitico o numerico. La seconda proprietà, invece, è la proprietà base richiesta a qualsiasi stimatore e garantisce che i risultati inferenziali a cui esso porta migliorino all'aumentare dell'informazione disponibile.

Le procedure di verosimiglianza seguono inoltre i due seguenti principi, che garantiscono a loro volta una certa stabilità e coerenza per tale metodologia, sottolineando il fatto come, nel primo caso, campioni diversi di uno stesso modello, nel secondo addirittura di modelli diversi, che conducono a funzioni di verosimiglianza proporzionali, portino a risultati equivalenti (Severini, 2000, Capitolo 3):

- *principio debole di verosimiglianza*: due osservazioni  $y_1^{oss}$  e  $y_2^{oss}$  da uno stesso modello  $\{F(\cdot; \theta) : \theta \in \Theta\}$  tali che  $l(\theta; y_1^{oss}) \propto l(\theta; y_2^{oss})$  porteranno alle stesse conclusioni riguardo  $\theta$ ;

- *principio forte di verosimiglianza*: due osservazioni  $y^{oss}$  e  $z^{oss}$  da due diversi modelli  $\{F_Y(\cdot; \theta) : \theta \in \Theta\}$  e  $\{F_Z(\cdot; \theta) : \theta \in \Theta\}$ , in cui il parametro  $\theta$  abbia lo stesso significato, tali per cui  $l(\theta; y^{oss}) \propto \tilde{l}(\theta; z^{oss})$  porteranno alle stesse conclusioni riguardo  $\theta$ .

I metodi frequentisti considerati in questo elaborato rispettano, in genere, il principio debole di verosimiglianza.

### 1.3.6 Distribuzione asintotica

Arrivando a questo punto, dopo aver ottenuto la S.M.V., è naturale chiedersi, consci delle principali caratteristiche della funzione di massima verosimiglianza, se sia possibile avere, accanto al valore della stima puntuale, una stima dell'errore ad essa legato. Per fare ciò bisogna guardare, come già accennato, allo stimatore di massima verosimiglianza  $\hat{\theta}_n(Y)$ , che spesso si indicherà più semplicemente con  $\hat{\theta}(Y)$ , considerando ovvio il suo legame con la numerosità campionaria, e alla sua natura casuale che, secondo i dettami del principio del campionamento ripetuto, avrebbe portato a valori sempre diversi della stima  $\hat{\theta}$ , al variare dei vari campioni che si sarebbero potuti ottenere, attorno al vero valore  $\theta_0$ .

In generale quindi, per avere coscienza dell'errore legato alla stima puntuale della quantità di interesse, bisogna conoscere la distribuzione dello stimatore ad esso associata, che varia da problema a problema. Ottenere tale distribuzione *esatta* è però spesso non semplice. Uno dei maggiori pregi delle procedure di verosimiglianza è che esse portano a risultati di carattere generale, non legati alla specificazione del particolare problema a cui si applicano, e quindi indicano una possibile via da seguire in tutti quei casi in cui la specificazione della esatta distribuzione di  $\hat{\theta}(Y)$  non è possibile o agevole. Per fare questo, però, è necessario essere disposti a perdere la conoscenza esatta di essa, accontentandosi di una *distribuzione asintotica* per  $\hat{\theta}(Y)$ , nel senso che tale distribuzione sarebbe esatta se la numerosità campionaria fos-

se infinita, per  $d = \dim(\theta)$  fissato. Per campioni finiti non si conoscerà la esatta distribuzione di  $\hat{\theta}(Y)$ , ma una sua approssimazione, la cui bontà deriva dalla coscienza del fatto che questa approssimazione è sempre migliore al crescere della numerosità  $n$ , sempre per  $d$  fissato. Seguendo questa strada, all'interno di ipotesi non troppo restrittive (Azzalini, 2001, Capitolo 3), si può mostrare che

$$\hat{\theta}_n \sim N_d(\theta, i(\theta)^{-1}) \quad (1.1)$$

$$l_\theta(\theta_0) \sim N_p(0, i(\theta)) \quad (1.2)$$

Nella pratica spesso  $i(\theta)$  è sostituito con  $i(\hat{\theta})$  o  $j(\hat{\theta})$ , che, essendone stime consistenti, lasciano invariate tali distribuzioni asintotiche grazie al teorema di Slutsky. E' inoltre interessante notare che lo stimatore di massima verosimiglianza è asintoticamente non distorto ed efficiente, in quanto la varianza della sua distribuzione asintotica raggiunge il limite inferiore di Cramer-Rao (L. Pace e A. Salvan, 2001, Capitolo 6).

### 1.3.7 Test statistico

Uno dei modi in cui si può procedere all' inferenza sul parametro di interesse  $\theta$  è quello della *verifica di ipotesi*.

Supponiamo che si sia ottenuta, applicando le procedura di verosimiglianza, la S.M.V.  $\hat{\theta}$  per  $\theta$ . Si può intuire, sempre rifacendosi al principio del campionamento ripetuto, che il vero valore  $\theta_0$  del parametro non sia esattamente  $\hat{\theta}$ , ma che sia un punto dello spazio campionario sperabilmente vicino ad esso; per convincersi di ciò basta pensare al fatto che, se potessimo ricampionare la stessa numerosità campionaria dalla vera  $F_0$  otterremmo, con ogni ragionevolezza, una stima, magari di poco, diversa (nel caso continuo ciò avviene con probabilità 1).

Supponiamo, inoltre, di essere interessati a valutare se  $\theta_0$  fa parte o meno di un qualche insieme  $\Theta_0 \subseteq \Theta$ . Tale ipotesi, che si dice *ipotesi nulla* e si

denota con

$$H_0 : \theta \in \Theta_0 \subseteq \Theta$$

individua un sottomodello  $\mathcal{F}_0$  del nostro modello di partenza  $\mathcal{F}$ , e la domanda che ci si pone è se  $F_0 \in \mathcal{F}_0$  o meno. Si individua, inoltre, la corrispondente *ipotesi alternativa*

$$H_1 : \theta \in \Theta_1 \equiv \Theta \setminus \Theta_0.$$

La verifica di ipotesi è quella procedura statistica che permette, attraverso l'utilizzo di un *test statistico* e fissato un errore di qualche tipo che siamo disposti a commettere, di discriminare univocamente i casi in cui si può accettare da quelli in cui si deve rifiutare l'ipotesi nulla  $H_0$ .

Per fare questo bisogna costruire una statistica test, ovvero una funzione dei soli dati  $t : \mathcal{Y} \rightarrow \mathbb{R}$ , di cui sia nota la distribuzione sotto  $H_0$  (ovvero quando  $\theta \in \Theta_0 \subseteq \Theta$ ; si parla in questo caso di *distribuzione nulla*) e per cui ogni punto di  $t(y)$  porti univocamente ad accettare o rifiutare  $H_0$ . In questo modo si individuano due zone dello spazio campionario, indicate rispettivamente con *regione di accettazione*

$$A = \{y \in \mathcal{Y} : t(y) \text{ porta ad accettare } H_0\}$$

e *regione critica o di rifiuto*

$$R = \{y \in \mathcal{Y} : t(y) \text{ porta a rifiutare } H_0\}$$

dall'ovvio significato (Pace e Salvan, 2001, Capitolo 2).

Le statistiche test basate sulla di verosimiglianza più usate in questo senso sono:

$$W(\theta_0) = 2[l(\hat{\theta}) - l(\theta_0)],$$

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

$$W_u(\theta_0) = l_\theta(\theta_0)^T i(\theta_0)^{-1} l_\theta(\theta_0)$$

e, solo per parametri scalari:

$$\begin{aligned} r(\theta_0) &= \text{sign}(\hat{\theta} - \theta_0)\sqrt{W(\theta_0)} \\ r_e(\theta_0) &= \text{sign}(\hat{\theta} - \theta_0)\sqrt{W_e(\theta_0)} \\ r_u(\theta_0) &= \text{sign}(\hat{\theta} - \theta_0)\sqrt{W_u(\theta_0)}. \end{aligned}$$

Si può mostrare (Pace e Salvan, 2001, Capitolo 6), sfruttando il teorema del limita centrale, lo sviluppo di Taylor e i risultati (1.1) e (1.2), che le prime tre quantità, sotto  $H_0$ , hanno distribuzione asintotica  $\chi_{d-d_0}$ , dove  $d = \dim(\Theta)$  e  $d_0 = \dim(\Theta_0)$  mentre l'ultima, anche detta *test radice con segno del log-rapporto di verosimiglianza*, normale standard.

E' logico chiedersi come distinguere quei valori di  $t$  per i quali accettare o rifiutare  $H_0$ . Considerando le distribuzioni  $\{f_{\theta}^T, \theta \in \Theta_0\}$  della statistica test  $T$  sotto  $H_0$ , si distinguono allora tre casi di ipotesi alternativa nel caso scalare (Pace e Salvan, 2001, Capitolo 2):

- *unilaterale sinistra*:  $T = t(Y)$  ha distribuzioni di probabilità tutte stocasticamente più piccole;
- *unilaterale destra*:  $T$  ha distribuzioni di probabilità tutte stocasticamente più grandi;
- *bilaterale*:  $T$  ha distribuzioni di probabilità tutte stocasticamente più piccole oppure più grandi.

Spesso, nel caso monodimensionale, l'ipotesi nulla è più semplicemente del tipo  $H_0 : \theta = \theta_0 \in \Theta$  e il test  $T$  permette di cogliere la direzione dello scostamento di  $\hat{\theta}$  da  $\theta_0$ . Confrontare  $H_0$ , in questi casi, ad esempio, con un ipotesi alternativa unilaterale sinistra del tipo  $H_1 : \theta < \theta_0$  significa, in questo caso, chiedersi se sia realistico supporre che il vero valore per  $\theta$  sia minore di un qualche scalare  $\theta_0$ . Solitamente, la statistica test  $t(\theta_0; y^{oss})$  è proporzionale rispetto a  $(\hat{\theta} - \theta_0)$ , e quindi si rifiuterà l'ipotesi nulla per valori grandi di essa (oltre una certa soglia). Analogo discorso per un ipotesi

unilaterale destra: si rifiuterà per valori piccoli della statistica test. Quando l'ipotesi è bilaterale, cosa che si indica con  $H_1 : \theta \neq \theta_0$ , si rifiuterà per valori sia grandi che piccoli. Nel caso multidimensionale invece usualmente l'ipotesi nulla è del tipo  $H_0 : \theta = (\theta_1, \dots, \theta_d) = (\theta_{01}, \dots, \theta_{0d})$  e non ha senso in questo caso porre ipotesi alternative direzionali; nei casi dei test  $W$ ,  $W_e$  e  $W_u$ , si rifiuterà  $H_0$  per valori grandi di essi, in quanto essi mostrano forte differenza tra i sottomodelli sotto le due ipotesi.

Esistono due errori legati ad un test statistico. Per *errore del primo tipo* si intende la probabilità di rifiutare  $H_0$  quando è vera, mentre per *errore del secondo tipo* si intende la probabilità di accettare  $H_0$  quando è falsa; solitamente un test è costruito tenendo conto solo dell'errore di primo tipo. Volendo trovare un sottoinsieme di valori di  $t(y)$  per i quali accettare  $H_0$ , si ragiona nel seguente modo. Supponiamo di essere disposti a rifiutare  $H_0$  quando è vera con probabilità  $\alpha$ . Allora, pensando, per semplicità, di avere a che fare con una regione critica unilaterale sinistra, considerando sempre una  $t(\cdot)$  proporzionale a  $(\hat{\theta} - \theta_0)$ , in un contesto monodimensionale, ci si può chiedere quali valori sono contrari ad  $H_0$  tanto quanto, o più, di  $t(y^{oss}) = t^{oss}$ : lo sono tutti quei valori più piccoli di  $t(y^{oss})$ . In questo senso, la probabilità di ottenere valori più contrari ad  $H_0$  rispetto a  $t(y^{oss})$ , se essa è vera, corrisponde alla probabilità associata all'insieme  $[-\infty, t^{oss}]$ , ovvero a  $F_T(t^{oss})$ , funzione di ripartizione di  $T$  calcolata in  $t^{oss}$ . Questo valore, che si indica con  $\alpha^{oss}$  e si dice *p-value* o *livello di significatività osservato*, compreso in  $[0, 1]$ , è una misura del grado di accordo che i dati mostrano rispetto all'ipotesi nulla formulata. Se esso è maggiore di un  $\alpha$  precedentemente fissato si accetterà  $H_0$ , in caso contrario si rifiuterà. In generale:

$$\text{regione critica unilaterale sinistra } \alpha^{oss} = \sup_{\theta \in \Theta_0} Pr_{\theta}(t(Y) \leq t^{oss})$$

$$\text{regione critica unilaterale destra } \alpha^{oss} = \sup_{\theta \in \Theta_0} Pr_{\theta}(t(Y) \geq t^{oss})$$

Per quanto riguarda invece una verifica di ipotesi con regione critica bilaterale, nella quale si rifiuta sia per valori grandi che piccoli del parametro,

$\alpha^{oss}$  sarà della forma

$$\text{regione critica bilaterale } \alpha^{oss} = 2 \sup_{\theta \in \Theta_0} \min(Pr_{\theta}(t(Y) \leq t^{oss}), Pr_{\theta}(T \geq t^{oss}))$$

in quanto i valori più contrari all'ipotesi nulla rispetto a  $t^{oss}$  sono tutti quei valori più piccoli o più grandi di  $t^{oss}$ , a seconda che esso sia situato nella coda sinistra o destra della distribuzione di  $T$ , ma anche tutti quei valori, rispettivamente, più grandi o più piccoli rispetto a  $F^{-1}(1 - Pr(T < t^{oss}))$  o  $F^{-1}(1 - Pr(T > t^{oss}))$ , ovvero al corrispettivo (nel senso di quel valore che individui un insieme del tipo  $[q, +\infty]$  o  $[-\infty, q]$  di pari probabilità di  $[-\infty, t^{oss}]$  o  $[t^{oss}, +\infty]$ ) di  $t^{oss}$  nella coda opposta rispetto a dove è situato.

Se invece si ha a che fare con un parametro multidimensionale si possono usare test il cui legame con il sistema di ipotesi indichi chiaramente cosa si intende per valori più contrari rispetto all'ipotesi nulla di  $t^{oss}$ .

Nel caso di ipotesi del tipo  $H_0 : \theta = \theta_0$  con  $T$  normale standard, per  $\theta$  monodimensionale, come nel caso del test  $r$ , tali espressioni diventano:

$$\text{regione critica unilaterale sinistra } \alpha^{oss} = \Phi(t^{oss})$$

$$\text{regione critica unilaterale destra } \alpha^{oss} = 1 - \Phi(t^{oss})$$

$$\text{regione critica bilaterale } \alpha^{oss} = 2 \min \{ \Phi(t^{oss}), 1 - \Phi(t^{oss}) \}.$$

Nel caso, invece, di ipotesi del tipo  $H_0 : (\theta_1, \dots, \theta_d) = (\theta_{01}, \dots, \theta_{0d})$  con  $T$  chi-quadrato con opportuni gradi di libertà, che esprime discordanza con l'ipotesi nulla per valori grandi, come è il caso dei vari test  $W$ ,  $W_e$  e  $W_u$  presentati, per  $\theta$  multidimensionale, si ha che:

$$\alpha^{oss} = 1 - F_T(t^{oss}).$$

Appare chiaro che, sotto  $H_0$ , essa si rifiuterà con confidenza  $\alpha$ , e si capisce ora inoltre perché sarebbe insensato costruire un test con errore del primo tipo nullo, in quanto ciò corrisponderebbe ad avere un test che accetta sempre  $H_0$ , qualsiasi siano i dati osservati (basti pensare al fatto che, se l' $\alpha$  fissato fosse nullo, non esisterebbe nessun  $t^{oss}$  tale per cui  $\alpha^{oss} < \alpha$ ).



Si noti che le regioni di accettazione e rifiuto possono essere in questo senso riscritte, in relazione ad un  $\alpha$  fissato, come:

$$A_\alpha = \{y \in \mathcal{Y} : \alpha^{oss}(y) > \alpha\}$$

$$R_\alpha = \{y \in \mathcal{Y} : \alpha^{oss}(y) \leq \alpha\}$$

(Pace e Salvan, 2001, Capitolo 2).

Può essere utile, inoltre, sottolineare il fatto che i due tipi di errore menzionati possono essere espressi in relazione alla cosiddetta *funzione di potenza*

$$\pi(\theta_*) = Pr_{\theta_*}(Y \in R_\alpha)$$

che corrisponde alla probabilità di rifiutare  $H_0$  quando il vero valore del parametro è  $\theta_*$ . La funzione che descrive l'errore di secondo è quindi del tipo  $\beta(\theta_*) = 1 - \pi(\theta_*)$  per  $\theta_* \notin \Theta_0$ , mentre l'errore di primo tipo si può scrivere come  $\alpha(\theta_0) = \pi(\theta_0)$  per  $\theta_0 \in \Theta_0$ . In un test ideale, che non commette errori, si ha che

$$\begin{cases} \pi(\theta) = 0, & \text{se } \theta \in \Theta_0 \\ \pi(\theta) = 1, & \text{se } \theta \notin \Theta_0. \end{cases}$$

Come già detto, ciò non avviene mai, se non in casi banali.

Qui si userà la metodologia presentata in questa sezione per fare verifica di ipotesi. Si potrebbe alternativamente confrontare  $t^{oss}$  direttamente con gli opportuni quantili di  $T$  sotto  $H_0$  (a secondo della ipotesi alternativa formulata).

### 1.3.8 Regioni di confidenza

Un'altra tecnica statistica, fortemente legata a quella della verifica di ipotesi appena presentata, che permette di tener conto della stima dell'errore commesso nel calcolare la S.M.V per il vero valore di  $\theta$ , è quella delle *regioni di confidenza*. L'obbiettivo è quello di ottenere un sottoinsieme  $\hat{\Theta}$ , che sarà la regione di confidenza ottenuta, dello spazio parametrico  $\Theta$  tale per cui

$\theta_0 \in \hat{\Theta}$  ( $\theta_0$  vero valore del parametro) con una certa probabilità  $1 - \alpha$ , con  $\alpha \in (0, 1)$ , detta *livello di confidenza* per tale regione, e che corrisponde a

$$Pr_{\theta_0}(\theta_0 \in \hat{\Theta}(Y))$$

ovvero alla *probabilità di copertura* della regione di confidenza (Pace e Salvan, 2001, Capitolo 2). Anche qui, idealmente, si vorrebbe che tale probabilità fosse uguale a uno, ma ciò corrisponderebbe alla soluzione banale  $\hat{\Theta} = \Theta$ .

Per un parametro monodimensionale, ipotizzando che  $\Theta$  sia un intervallo del tipo  $[v_-, v_+]$ , con  $v_- < v_+$ , entrambi appartenenti a  $\mathbb{R}$ , si possono costruire tre diversi tipi di regioni di confidenza, ora anche chiamate *intervalli di confidenza*:

- della forma  $[v_-, \hat{\theta}]$ , che è l'insieme di valori del parametro per i quali si accetterebbe l'ipotesi nulla nel sistema di ipotesi  $H_0 : \theta = \hat{\theta}$  contro  $H_1 : \theta > \hat{\theta}$ ;
- della forma  $[\hat{\theta}, v_+]$ , che è l'insieme di valori del parametro per i quali si accetterebbe l'ipotesi nulla nel sistema di ipotesi  $H_0 : \theta = \hat{\theta}$  contro  $H_1 : \theta < \hat{\theta}$ ;
- della forma  $[c_-, c_+]$ , con  $v_+ \leq c_- \leq \hat{\theta} \leq c_+ \leq v_-$ , che è l'insieme di valori del parametro per i quali si accetterebbe l'ipotesi nulla nel sistema di ipotesi  $H_0 : \theta = \hat{\theta}$  contro  $H_1 : \theta \neq \hat{\theta}$ .

Per un parametro multidimensionale, la regione corrispondente ai valori per i quali si accetta  $H_0 : \theta = (\theta_1, \dots, \theta_d) = (\theta_{01}, \dots, \theta_{0d})$  contro  $H_1 : \bar{H}_0$  e statistica test per la quali si rifiuta per valori grandi, sarà un insieme del tipo

$$\{\theta_0 : T(\theta_0) < F_T^{-1}(1 - \alpha)\}. \quad (1.3)$$

Con simili espressioni si descrivono la maggior parte degli intervalli e delle regioni di confidenza. Alternativamente, si può procedere, più euristicamente, calcolando, per molti valori di  $\theta$ , quali portano ad accettare l'ipotesi nulla corrispondente alla regione che si vuole ottenere. L'idea è sempre quella di

escludere valori del parametro poco verosimili sotto  $H_0$ , di volta in volta nella giusta direzione.

### 1.3.9 Verosimiglianza profilo

Finora si è ragionato sempre pensando che l'interesse inferenziale riguardasse l'intero parametro  $\theta = (\theta_1, \dots, \theta_d)^T$ . Ci sono però dei casi in cui si può essere interessati a fare verifica di ipotesi solo su una parte di tale parametro. In questo senso si può pensare di partizionarlo nel modo  $\theta = (\tau, \rho)$ , dove  $\tau$ , di dimensione  $k$ , rappresenta il *parametro di interesse*, mentre  $\rho$ , di dimensione  $d - k$ , rappresenta il *parametro di disturbo*, ovvero il sottoinsieme di componenti di  $\theta$  che, pur giocando un ruolo nel modello statistico studiato  $\mathcal{F}$ , non sono oggetto di studio della indagine statistica.

Ora, per procedere nella inferenza, è utile riscrivere la funzione di log-verosimiglianza in funzione solo di  $\psi$ , ottenendo, così, la cosiddetta *log-verosimiglianza profilo*

$$l_p(\tau) = l(\hat{\theta}_\tau)$$

dove  $\hat{\theta}_\tau$  rappresenta la S.M.V. per  $\theta$  fissato  $\tau$  e può anche essere scritto come  $(\tau, \hat{\rho}_\tau)$ , con  $\hat{\rho}_\tau$  rappresenta la S.M.V. per  $\rho$  fissato  $\tau$  (Severini, 2000, Capitolo 5). Si può ricavare che

$$W_p(\tau_0) = 2[l_p(\hat{\tau}) - l_p(\tau_0)]$$

ha distribuzione asintotica  $\chi_k^2$  sotto  $H_0 : \tau = \tau_0$ , mentre, nel caso in cui  $k = 1$

$$r_p(\tau_0) = \text{sgn}(\hat{\tau} - \tau_0) \sqrt{W_p(\theta_0)}$$

ha distribuzione asintotica  $N(0, 1)$ , sempre sotto  $H_0 : \tau = \tau_0$ .

### 1.3.10 Modelli a variabili indipendenti

Nel caso in cui il vettore casuale  $Y = (Y_1, \dots, Y_n)$  del modello statistico specificato sia a componente indipendenti, come nel caso del modello

che in questo elaborato sarà studiato la funzione di verosimiglianza per  $Y = (Y_1, \dots, Y_n)$  si può scrivere come prodotto delle singole funzioni di verosimiglianza delle sue componenti

$$L(\theta; y) = \prod_{i=1}^n L_i(\theta; y_i)$$

e, quindi, la log-verosimiglianza come somma delle log-verosimiglianze delle sue componenti. Ciò deriva immediatamente dal fatto che la probabilità che accadano più eventi indipendenti simultaneamente è pari al prodotto delle singole probabilità marginali

## 1.4 Statistiche sufficienti, ancillari e completezza

Come già accennato, una funzione dei dati  $t(Y)$  che non dipende dal parametro  $\theta$  è detta *statistica* e il valore di essa osservato  $t^{oss} = t(y^{oss})$  è detto *valore campionario della statistica* (Azzalini, 2001, Capitolo 1). Alcune di queste statistiche riassumono tutta l'informazione fornita dai dati: nella pratica ciò avviene perché la funzione di verosimiglianza può essere riscritta in funzione di esse, piuttosto che dei dati, nel senso che esiste una funzione  $g$  tale per cui

$$L(\theta; y) \propto g(\theta; t(y)) \quad (1.4)$$

In questi casi si dice che  $t$  è *statistica sufficiente* e si ha che, se due campioni  $\tilde{y}$  e  $\tilde{z}$  sono tali che  $t(\tilde{y}) = t(\tilde{z})$ , allora accade che  $L(\theta; \tilde{y}) \propto L(\theta; \tilde{z})$  e si trarranno conclusioni analoghe dai due campioni. La sufficienza di una statistica è strettamente legata alla specificazione del modello considerato: cambiare tale specificazione può far cessare a tale statistica di essere sufficiente. Dalla (1.4) deriva che  $f(\theta; y)$  può essere scritta, se  $t$  è statistica sufficiente, nel seguente modo (*fattorizzazione di Neyman*):

$$f(\theta; y) = h(y)g(\theta; t(y))$$

per opportune funzioni  $h$  e  $g$ . Si mostra (Azzalini, 2001, Capitolo 2), inoltre, che  $t$  è sufficiente per  $\theta$  se e solo se la distribuzione di  $Y$  condizionata al valore assunto da  $t$  non dipende da  $\theta$ .

E' chiaro che le statistiche sufficienti non sono uniche. Solitamente si vuole lavorare con statistiche che rappresentano la massima sintesi per i dati disponibili; si arriva così alla definizione di *statistica sufficiente minimale*: una statistica si dice essere statistica sufficiente minimale se non è funzione di nessun'altra statistica sufficiente. Ne deriva che, se  $t_1$  e  $t_2$  sono entrambe statistiche sufficienti minimali per un dato modello esse devono essere equivalenti nel senso che

$$t_1(y) = t_1(\tilde{y}) \text{ se e solo se } t_2(y) = t_2(\tilde{y})$$

e che, quindi, tutte le statistiche sufficienti minimali di un modello appartengono a una classe di statistiche equivalenti (Severini, 2000, Capitolo 1). In virtù di tale equivalenza ci si riferirà alla statistica sufficiente minimale come se fosse unica. Tale insieme non è mai vuoto: infatti, una qualsiasi funzione invertibile della partizione, detta *partizione di verosimiglianza*, formata da tutti gli insiemi di punti dello spazio campionario che portano a verosimiglianze equivalenti, è certamente sufficiente e minimale (Azzalini, 2001, Capitolo 2).

Un altro concetto importante è quello di ancillarità: una statistica  $B$  con distribuzione indipendente da  $\theta$  (si dirà *costante in distribuzione*) che è funzione della statistica sufficiente minimale si dice essere una *statistica ancillare*. Una statistica di questo tipo non porta con sé alcuna informazione riguardo al parametro  $\theta$ : si cerca quindi solitamente di considerare la statistica ancillare come una costante fissata, piuttosto che come una variabile casuale, usando la distribuzione di  $Y|A = a$ , piuttosto che quella di  $Y$ , come base dell'inferenza (Severini, 2000, Capitolo 1). In questo senso è importante notare che, nella maggior parte dei casi, nel passaggio tra la statistica sufficiente minimale  $s$  del modello in questione e la S.M.V.  $\hat{\theta} = \hat{\theta}(s)$ , vi è

una perdita di informazione in quanto  $\hat{\theta}$  non è sufficiente minimale. Si può definire allora una *statistica ausiliaria*  $\tilde{a}$  tale per cui la relazione tra  $s$  e  $(\hat{\theta}, \tilde{a})$  è biunivoca;  $\tilde{a}$  rappresenta quindi l'informazione mancante per ricostruire  $s$  a partire da  $\hat{\theta}$ . Se inoltre  $\tilde{a}$  è ancillare, vale la fattorizzazione

$$p_{\hat{\theta}, \tilde{A}}(\hat{\theta}, \tilde{a}; \theta) = p_{\tilde{A}}(\tilde{a})p_{\hat{\theta}|\tilde{A}=\tilde{a}}(\hat{\theta}, \tilde{a}; \theta) \quad (1.5)$$

per la quale  $\hat{\theta}$  è sufficiente minimale condizionatamente al valore assunto da  $\tilde{a}$ . Tale statistica ancillare descrive l'informazione relativa alla precisione della S.M.V. (Pace e Salvan, 1996).

Prima di procedere oltre può essere utile notare che, quando la S.M.V. non è statistica sufficiente per il modello in questione, se la statistica sufficiente minimale  $s$  può essere scritta come  $s = (\hat{\theta}, a)$ , con  $a$  statistica ancillare, la funzione di verosimiglianza (cosiccome le altre quantità di verosimiglianza) possono essere scritte in funzione di  $(\hat{\theta}, a)$  piuttosto che di  $y^{oss}$ , nel senso che

$$l(\theta; y^{oss}) = l(\theta; \hat{\theta}, a)$$

e così via.

Un'ultima definizione che può tornare utile è quella di statistica completa. Una statistica sufficiente  $t$  è *completa* se, per ogni funzione reale  $g$  si ha che

$$E_{\theta}[g(t)] = 0 \iff Pr_{\theta}(g(t) = 0) = 1 \text{ per ogni } \theta$$

Se esiste uno stimatore per un parametro  $\theta$  scalare che è funzione di una statistica sufficiente completa, esso è unico. Inoltre, per il *teorema di Lehmann-Scheffé*, se una statistica  $t$  è sufficiente e completa per un dato parametro  $\theta$ , essa è anche minimale sufficiente (Young e Smith, 2005, Capitolo 6).

## 1.5 Modello per dati dicotomici

Si presenta ora, brevemente, l'usuale modello statistico usato per trattare con dati dicotomici, su cui si basa il modello per la classificazione scorretta

di dati binari, presentato più avanti.

Si ha a che fare, dunque, con un campione di numerosità  $n$ , le cui componenti presentano solo due modalità, che usualmente si codificano con 0 e 1. Spesso accade che esse rappresentino, rispettivamente, l'assenza, o la presenza, di una data caratteristica. Si assume, quindi, che ogni dato sia generato da una variabile  $Y_i \sim Be(\mu_i)$ , dove si intende una variabile distribuita come una bernoulliana di parametro  $\mu_i$ , con  $Y_i \perp Y_j$  per  $i \neq j$ . Le  $n$  variabili casuali da cui proviene il campione sono tutte dello stesso tipo, a meno del parametro di interesse  $\mu_i = \Pr(Y_i = 1)$ , in generale diverso per ognuna di essa. Si ha, quindi, a che fare con dati eteroschedastici e con diversa media, in quanto media e varianza, per la bernoulliana, sono funzione di  $\mu_i$ , nel modo

$$E(Y_i) = \mu_i,$$

$$Var(Y_i) = \mu_i(1 - \mu_i).$$

Si dispone, inoltre, di un certo numero di covariate per ogni dato, che si suppone essere legato con il parametro  $\mu = (\mu_1, \dots, \mu_n)$  di interesse. Tali covariate vengono sintetizzate nella matrice  $X = (\underline{x}_1, \dots, \underline{x}_p)$ .

Seguendo, poi, l'impostazione dei modelli lineari generalizzati (GLM) e con l'obiettivo di modellare il parametro di interesse  $\mu$  a partire dalle covariate  $X$ , si ipotizza che il valore atteso di  $Y_i$  sia una qualche funzione  $g(\cdot)$  lineare dei regressori, tale per cui

$$g(\mu_i) = \underline{x}_i^T \beta = x_{i1}\beta_p + \dots + x_{ip}\beta_p.$$

Con le normali procedure di verosimiglianza si ottiene poi la S.M.V. per  $\beta$ . Bisognerà procedere per via numerica, in quanto l'equazione  $l_\beta(\beta)$  non è, in generale, risolvibile analiticamente, per poi risalire al valore previsto per un generico  $\mu_i$  applicando la relazione  $g^{-1}(x_i^T \beta) = \mu_i$ .

La funzione  $g(\cdot)$ , detta *funzione legame*, può essere una qualsiasi funzione invertibile del tipo  $g : [0, 1] \rightarrow \mathbb{R}$ , come, ad esempio, lo potrebbe essere

l'inversa della funzione di ripartizione di una qualsivoglia variabile casuale, se ne esiste l'inversa. La funzione legame canonica, per questo tipo di modello, è la funzione logit

$$g(\mu_i) = \log \frac{\mu_i}{1 - \mu_i},$$

con sua inversa

$$\mu_i = g^{-1}(x_i^T \beta) = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}.$$

## 1.6 Simboli di Landau

Si danno ora delle definizioni che torneranno utili quando si andrà a confrontare la qualità delle diverse metodologie di test studiate in questo elaborato. In particolare:

- dato  $x_0 \in \mathbb{R}$ , punto di accumulazione per entrambi i domini di due funzioni  $f$  e  $g$ , tali per cui  $\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} g(x) = \infty$ , si dice che  $f$  è un *infinitesimo di ordine superiore* rispetto a  $g$  per  $x \rightarrow x_0$  se si ha che

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 0$$

- dato  $x_0 \in \mathbb{R}$ , punto di accumulazione per entrambi i domini di due funzioni  $f$  e  $g$ , tali per cui  $\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} g(x) = 0$ , (con  $g \neq 0$  definitivamente per  $x \rightarrow x_0$ ), si dice che  $f$  è un *infinitesimo di ordine superiore* rispetto a  $g$  per  $x \rightarrow x_0$  se si ha che

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 0.$$

Inoltre:

- date due funzioni  $f$  e  $g$  e dato  $x_0$  punto di accumulazione del dominio di entrambe le funzioni, si dice che

$$f(x) = o(g(x)) \text{ per } x \rightarrow x_0$$



se si ha che

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = 0$$

- nelle stesse ipotesi, si dice che

$$f(x) = O(g(x)) \text{ per } x \rightarrow x_0$$

se si ha che  $\frac{f(x)}{g(x)}$  è definitivamente limitata per  $x \rightarrow x_0$

I due segni  $o(\cdot)$  e  $O(\cdot)$  (che si leggono, rispettivamente, "o piccolo di ..." e "o grande di ...") sono noti come *simboli di Landau* (M. Bertsch et al., 2007). Si nota fin da subito che è possibile usare tale simbologia per confrontare tra loro le accuratezze di due o più procedure statistiche: se, in generale, si dimostra che una data procedura  $s$  compie un errore  $err_s = O(n^{-\frac{j_s}{2}})$ , mentre un'altra procedura  $v$  compie un errore  $err_v = O(n^{-\frac{j_v}{2}})$  per  $n \rightarrow \infty$ , con  $j_s$  e  $j_v$  numeri reali tali per cui  $j_s > j_v$ , allora si potrà affermare che la procedura  $s$  è migliore di  $v$  in quanto il suo errore tende a 0 più velocemente al crescere della numerosità campionaria ( $err_s$  è un infinitesimo di ordine superiore rispetto a  $err_v$ ).

Questo concetto è puntualizzato dalla nozione di ordine di probabilità (Pace e Salvan, 1996, Capitolo 3): una successione di variabili casuali  $Y_n$  è detta *asintoticamente di ordine  $O(n^\alpha)$  in probabilità*, e si indica con  $O_p(n^\alpha)$ , se per ogni  $\epsilon > 0$  esistono un reale  $A = A_\epsilon > 0$  e un naturale  $n^* = n_\epsilon^*$  tali che per ogni  $n \leq n^*$

$$P(|Y_n/n^\alpha| < A) > 1 - \epsilon.$$

# Capitolo 2

## Classificazione scorretta di dati binari

Viene ora presentato un modello che considera esplicitamente la possibilità della classificazione scorretta di dati binari (McInturff et al., 2003), che sarà poi al centro delle simulazioni svolte nel Capitolo 5.

### 2.1 Il modello

Si suppone di avere a che fare con un fenomeno che può manifestarsi solo in due modalità diverse, codificate con 0 e 1, come accade ad esempio se si vuole fare inferenza sul fatto che una persona sia malata o meno o se un macchinario sia rotto o meno, tenendo conto di altre sue caratteristiche. A differenza, però, di quanto richiamato nel Paragrafo 1.5, si suppone ora di non conoscere esattamente gli esatti valori della risposta dei dati con cui andremo a costruire il modello, ma, al suo posto, il valore di un'altra variabile casuale, sempre a due valori, strettamente legata alla risposta, come ad esempio il valore del test che verifica lo stato di salute/malattia del paziente nello studio. Si avranno così alcuni casi in cui le due variabili non coincidono, in quanto, ad esempio, il test può sbagliare in un senso o nell'altro, ed ignorare questi errori di classificazione, costruendo un modello per dati

binari semplicemente sostituendo alla risposta la variabile ad essa legata, porterebbe a delle stime distorte in modo non irrilevante; si costruirebbe, in sostanza, un modello per predire il valore della variabile test piuttosto che della variabile veramente d'interesse, che è legata alla prima senza però coinciderle. L'obbiettivo è quindi quello di costruire un modello di regressione sulla prima variabile, indicata con  $D$ , a partire dai valori della seconda variabile, indicata con  $Z$ , che tenga conto di quanto detto finora.

Considerando il caso in cui si vuole costruire un modello di regressione sullo stato di presenza/assenza di una data malattia in dei paziente conoscendo delle covariate (ad esempio età, peso, storia clinica, ...) denotate con  $x = (x_1, \dots, x_p)^T$ , dovendo però partire dalla sola conoscenza del valore positivo/negativo del test che verifica tale presenza/assenza, si procede nel seguente modo. Le variabili  $Z$  e  $Y$  saranno, ovviamente, due bernoulliane del tipo

$$\begin{cases} Z = 1, & \text{se il paziente è malato} \\ Z = 0, & \text{se il paziente è sano;} \\ Y = 1, & \text{se il test è positivo,} \\ Y = 0, & \text{se il test è negativo;} \end{cases}$$

con  $\Pr(Z = 1|x) = \pi_{Z|x}$  e  $\Pr(Y = 1|x) = \pi_{Y|x}$ . Definiamo poi le due quantità

$$\eta = Pr(Y = 1|Z = 1), \text{ sensibilità del test}$$

$$\omega = Pr(Y = 0|Z = 0), \text{ specificità del test}$$

che sono due misure, rispettivamente, di quanto il test sia effettivamente legato alla vera presenza o assenza della malattia. Si suppone che sensibilità e specificità, e quindi l'errore di classificazione commesso dal test, siano indipendenti dalle covariate e che siano note.

L'interesse è quindi quello di prevedere  $\pi_{Z|x}$ . Seguendo quanto detto nel Paragrafo 1.5 è conveniente riscrivere  $\pi_{Z|x}$  attraverso una qualsiasi funzione legame: per esempio, nel caso logistico, avremo che  $\pi_{Z|x} = \exp(x^T \beta) / (1 +$

$\exp(x^T \beta)$ ), dove  $\beta = (\beta_1, \dots, \beta_p)^T$  è il vettore dei coefficienti di regressione. Si può ora mostrare, usando la legge della probabilità totale, che un risultato positivo del test per un individuo con covariate  $x$  capita con probabilità

$$\pi_{Y|x} = \Pr(Y = 1|x) = 1 - \omega - (1 - \eta - \omega)\pi_{Z|x}, \quad (2.1)$$

poichè

$$\begin{aligned} \Pr(Y = 1|x) &= \Pr(Y = 1|Z = 1) \Pr(Z = 1|x) + \Pr(Y = 1|Z = 0) \Pr(Z = 0|x) \\ &= \eta\pi_{Z|x} + (1 - \omega)(1 - \pi_{Z|x}) = \eta\pi_{Z|x} + 1 - \omega + \omega\pi_{Z|x} - \pi_{Z|x} \\ &= 1 - \omega - (1 - \eta - \omega)\pi_{Z|x}. \end{aligned}$$

Si osservano poi i valori del test effettuato e delle covariate  $(y_j, x_j)$ ,  $j = 1, \dots, n$  su  $n$  pazienti. Assumendo che  $Y|x_j$  sia una bernulliana di parametro  $\pi_{Y|x_j}$ , descritto dalla (2.1),  $\pi_{Z|x_j} = \Pr(Z = 1|x_j)$  e che le varie  $Y_j$  e  $Z_j$  siano variabili indipendenti, la funzione di verosimiglianza si scrive

$$L(\beta) = \prod_{i=1}^n \pi_{Y|x_i}^{y_i} (1 - \pi_{Y|x_i})^{1-y_i},$$

con, nel caso logistico,

$$\pi_{Y|x_i} = 1 - \omega - (1 - \eta - \omega)\pi_{Z|x_i} = 1 - \omega - (1 - \eta - \omega) \frac{\exp(\beta^T x_i)}{1 + [\exp(\beta^T x_i)]}$$

dove  $\pi_{Y|x_i}$  è quindi funzione di  $\eta$ ,  $\omega$  e  $\beta$ . Nel seguito si considereranno sensibilità e specificità come note; se così non fosse si dovrebbero applicare metodi bayesiani (McInturff et al., 2003). Per quanto riguarda la log-verosimiglianza, si ottiene che

$$l(\beta) = \sum_{i=1}^n y_i \log \pi_{Y|x_i} + \sum_{i=1}^n (1 - y_i) \log(1 - \pi_{Y|x_i}). \quad (2.2)$$

Specificare il problema in questo modo ci permette quindi di costruire un modello di regressione basato sulle risposte del test, a noi perfettamente note, e di ottenere così implicitamente anche il valore previsto per il vero parametro di interesse  $\pi_{Z|x}$ , espresso come funzione degli stessi coefficienti di regressione usati in tale modello che saranno ottenuti a partire dalla (2.2), secondo l'usuale procedura di verosimiglianza.

Si calcola ora lo score di tale funzione di log-verosimiglianza, che tornerà utile per garantire stabilità ed efficienza in alcuni algoritmi di ottimizzazione.

Si ha che

$$\begin{aligned} l_{\beta}(\beta) &= \frac{\partial \beta}{\partial \beta_j} = \sum_{i=1}^n y_i \frac{\frac{\partial \pi_{Y|x_i}}{\partial \beta_j}}{\pi_{Y|x_i}} - \sum_{i=1}^n (1 - y_i) \frac{\frac{\partial \pi_{Y|x_i}}{\partial \beta_j}}{1 - \pi_{Y|x_i}} \\ &= \sum_{i=1}^n y_i \frac{(\omega + \eta - 1) \frac{\partial \pi_{Z|x_i}}{\partial \beta_j}}{1 - \omega - (1 - \eta - \omega) \pi_{Z|x_i}} - \sum_{i=1}^n (1 - y_i) \frac{(\omega + \eta - 1) \frac{\partial \pi_{Z|x_i}}{\partial \beta_j}}{\omega + (1 - \eta - \omega) \pi_{Z|x_i}} \end{aligned}$$

per  $j = 1, \dots, p$ , dove  $\frac{\partial \pi_{Z|x_i}}{\partial \beta_j}$  dipenderà dalla particolare scelta della funzione legame; nel caso di regressione logistica si ha che

$$\begin{aligned} \frac{\partial \pi_{Z|x_i}}{\partial \beta_j} &= \frac{\partial \left( \frac{\exp(\beta^T x)}{1 + \exp(\beta^T x)} \right)}{\partial \beta} = \frac{x \exp(\beta^T x) [1 + \exp(\beta^T x) - \exp(\beta^T x)]}{[1 + \exp(\beta^T x)]^2} = \\ &= \frac{x_j \exp(\beta^T x)}{[1 + \exp(\beta^T x)]^2} \end{aligned}$$

## 2.2 Verifica di ipotesi

Si mostra ora il modo più semplice per fare verifica di ipotesi su questo modello. Come esemplificazione, si vede il caso in cui il parametro  $\dim(\beta) = 3$ , la numerosità campionaria  $n$  è pari a venti e la matrice  $X$ , è formata da un vettore di costanti continue, generate da una normale standard, come secondo, da un vettore dicotomico generato da una bernoulliana di parametro  $\frac{1}{2}$  e come terzo il vettore che funge da intercetta.. La simulazione, inoltre, è svolta fissando  $\beta = (1, -1, 1)$ ,  $\eta = 0.95$  e  $\omega = 0.85$ , in un modello con funzione legame logit. Verranno usate le funzioni messe a disposizione dal pacchetto `likelihoodAsy`. Ottimizzando la funzione di log-verosimiglianza si ottiene che  $\hat{\beta} = (1.595, -1.343, 1.918)$ .

Si fanno ora tre verifiche di ipotesi, che saranno poi riprese anche nel resto delle simulazioni, una per ogni componente di  $\beta$ . Si condurranno le verifiche di ipotesi confrontato ogni volta il parametro di interesse con il suo vero valore: questo tornerà particolarmente utile quando poi si dovrà confrontare la qualità delle diverse metodologie affrontate. Inoltre, le ipotesi alternative fatte, avranno sempre regione critica unilaterale sinistra.

Il test più semplice, previsto dalla teoria della verosimiglianza, per fare ciò con parametri unidimensionali (come in questo caso) è il test  $r_p$  presentato nel Paragrafo 1.3.7. Come già detto, esso ha distribuzione nulla approssimativamente  $N(0, 1)$ . Inoltre si può mostrare (L. Pace e A. Salvan, 1996) che il reale livello di significatività osservato associato all'ipotesi nulla fatta è pari a  $\alpha^{oss} = \Phi(r^{oss}) \{1 + O(n^{-1/2})\}$ .

Per  $H_0 : \beta_1 = 1$  contro  $H_1 : \beta_1 < 1$  si ottiene  $r^{oss} = 1.2$ , per  $H_0 : \beta_2 = -1$  contro  $H_1 : \beta_2 < -1$  si ottiene  $r^{oss} = 0.43$ , mentre per  $H_0 : \beta_3 = 1$  contro  $H_1 : \beta_3 < 1$  si ottiene  $r^{oss} = 1.476$ . I tre valori osservati corrispondono a dei  $p$ -value, rispettivamente, di 0.885, 0.333 e 0.93. Normalmente, si fissa  $\alpha$  a 0.05 o, eventualmente, a 0.1, quindi, in questo caso, si accetterebbero senza dubbio tutte ipotesi nulle.



## Capitolo 3

# Radice con segno modificata del log-rapporto di verosimiglianza

In questo capitolo verrà presentata la statistica  $r_p^*$  e una sua particolare implementazione che, insieme al modello già mostrato ed alle tecniche bootstrap di cui si parlerà in seguito, forma la parte centrale di questo elaborato. Si presenterà, poi, il package `likelihoodAsy`, sviluppato da Bellio e Pierce (2015), che sarà usato nelle simulazioni.

### 3.1 Formula $p^*$

Come già accennato, quando si ha a che fare con un parametro  $\theta = (\tau, \rho)$ , con  $\dim(\tau) = 1$  e  $\rho$  parametro di disturbo di qualsivoglia dimensione e si è interessati a fare inferenza su  $\tau$ , l'usuale test  $r_p(\tau_0) = \text{sign}(\hat{\tau} - \tau_0)\sqrt{W_p(\tau_0)}$  ha distribuzione approssimativa normale standard, con un errore di ordine  $O_p(n^{-\frac{1}{2}})$ , ossia

$$\Pr(r_p(\tau_0) < r^{oss}) = \Phi(r_p^{oss})[1 + O(n^{-\frac{1}{2}})]$$

sotto l'ipotesi nulla  $H_0 : \tau = \tau_0$ . Approssimazioni di questo tipo, dette del *primo ordine*, possono non essere molto accurate, soprattutto in presenza di pochi dati e molti parametri di disturbo. Nel tempo ci si è, quindi, chiesti se fosse possibile giungere a risultati asintotici più precisi, anche detti *di*



*ordine superiore*. Un risultato fondamentale per raggiungere questo scopo è la cosiddetta *formula*  $p^*$

$$p^*(\hat{\theta}; \theta|a) = c(\theta, a) |j(\hat{\theta})|^{\frac{1}{2}} \exp[l(\theta) - l(\hat{\theta})] \quad (3.1)$$

(Barndorff-Nielsen, 1983), dove  $c(\theta, a)$  è una costante di normalizzazione, che esprime una approssimazione molto accurata della distribuzione dello S.M.V., condizionato al valore di una data statistica ancillare  $a$   $f(\hat{\theta}; \theta, a)$ .

Si può mostrare, infatti, che la (3.1) ha un errore di ordine di  $O_p(n^{-1})$  per ogni  $\hat{\theta}$  fissato; mentre, per  $\hat{\theta} = \theta + O_p(n^{-\frac{1}{2}})$ , che è di fatto la situazione a cui si è comunemente interessati, si ha che

$$f(\hat{\theta}; \theta|a) = p^*(\hat{\theta}; \theta|a)[1 + O(n^{-\frac{3}{2}})].$$

Spesso può risultare complesso specificare una statistica ancillare esatta a cui condizionarsi e quindi si può ricorrere a statistiche ancillari approssimative, che portano l'accuratezza della  $p^*$  formula a  $O_p(n^{-1})$  (*secondo ordine*), in luogo di  $O(n^{-\frac{3}{2}})$  (*terzo ordine*) (Severini, 2000, Capitolo 6).

### 3.2 Statistica test $r_p^*$

Utilizzando la formula  $p^*$  si può arrivare a una forma modificata della statistica  $r_p$ , che qui si indicherà con  $r_p^*$ , la cui distribuzione approssimata è normale standard con un errore del terzo ordine. Tale statistica può essere scritta come

$$r_p^*(\tau_0) = r_p(\tau_0) + \frac{1}{r_p(\tau_0)} \log(v(\tau_0)/r_p(\tau_0)).$$

per la quale vale che

$$\Pr[r_p^*(\tau_0) \leq r_p^{*oss}] = \Phi(r_p^{*oss})[1 + O(n^{-\frac{3}{2}})].$$

In generale, al di fuori di casi particolari come può essere, ad esempio, quello delle famiglie esponenziali, la quantità  $v(\tau_0)$  risulta complessa da calcolare, in quanto implica il calcolo di derivate della funzione di verosimiglianza rispetto allo S.M.V.  $\hat{\theta}$ , condizionatamente ad una statistica

ancillare  $a$ . Uno dei modi per ovviare a tale problema può essere quello di condizionarsi ad una statistica approssimativamente ancillare (Severini, 2000, Capitolo 6), o usare un'approssimazione di  $v(\tau_0)$  basata su una riparametrizzazione che renda il parametro di interesse scalare  $\tau$  e il parametro di disturbo  $\rho$  ortogonali (DiCiccio e Martin, 1993), solitamente arrivando ad un errore nell'approssimazione di ordine  $O_p(n^{-1})$  (Severini, 2000, Capitolo 6).

### 3.3 Package likelihoodAsy

Un'altra approssimazione per  $v(\tau_0)$  è quella offerta da Skovgaard (1996), che indicheremo con  $\tilde{v}(\tau_0)$ , che non prevede il condizionamento esplicito ad alcuna statistica ancillare. La sua forma è:

$$\tilde{v}(\tau_0) = [\hat{S}^{-1}\hat{Q}]_1 |j(\hat{\theta})|^{1/2} |i^{-1}(\hat{\theta})| |\hat{S}| |j_{\rho\rho}(\hat{\theta}_{\tau_0})|^{-1/2}$$

con

$$\begin{aligned}\hat{Q} &= \text{cov}_{\theta_0} [l_{\theta}(\theta_0), l(\theta_0) - l(\theta_1)] |_{\theta_0=\hat{\theta}, \theta_1=\hat{\theta}_{\tau_0}} \\ \hat{S} &= \text{cov}_{\theta_0} [l_{\theta}(\theta_0), l_{\theta}(\theta_1)] |_{\theta_0=\hat{\theta}, \theta_1=\hat{\theta}_{\tau_0}}\end{aligned}$$

e dove  $j_{\rho\rho}(\hat{\theta}_{\tau_0})$  è il blocco dell'informazione osservata, valutata in  $\hat{\theta}_{\tau_0}$ , relativo al parametro  $\rho$ . Con questa scrittura  $r_p^*$  ha ancora una accuratezza del secondo ordine.

La cosa più interessante di questa approssimazione per  $v(\tau_0)$  è che non prevede il condizionamento esplicito rispetto ad alcuna statistica ancillare, o il calcolo di derivate rispetto alla S.M.V, ma coinvolge solo valori attesi di quantità di verosimiglianza, pur garantendo un errore dello stesso ordine delle altre approssimazioni. L'unico aspetto difficoltoso può essere rappresentato dal calcolo delle quantità  $\hat{Q}$  e  $\hat{S}$ , che, essendo covarianze, possono

essere riscritte come valori attesi:

$$\begin{aligned}\hat{Q} &= E_{\hat{\theta}} \left\{ l_{\theta}(\hat{\theta}) [l(\hat{\theta}) - l(\hat{\theta}_{\tau_0})] \right\} \\ \hat{S} &= E_{\hat{\theta}} \left\{ l_{\theta}(\hat{\theta}) l_{\theta}(\hat{\theta}_{\tau_0}) \right\}.\end{aligned}$$

In questo senso, può risultare naturale pensare di approssimarli sfruttando i *metodi di integrazione di Montecarlo*, i quali prevedono, se si ha come obiettivo quello di ottenere il valore atteso di una qualche variabile casuale, di approssimare tale valore atteso con la media campionaria tra molte osservazioni generate dalla variabile casuale di interesse. In questo senso, si ha che

$$\begin{aligned}\hat{Q} &\approx \frac{1}{R} \sum_{r=1}^R \left\{ l_{\theta}(\hat{\theta}; y_r^*) [l(\hat{\theta}; y_r^*) - l(\hat{\theta}_{\tau_0}; y_r^*)] \right\} \\ \hat{S} &\approx \frac{1}{R} \sum_{r=1}^R \left\{ l_{\theta}(\hat{\theta}; y_r^*) l_{\theta}(\hat{\theta}_{\tau_0}; y_r^*) \right\}\end{aligned}$$

per  $R$  grande, dove i vari  $y_r^*$  sono campioni generati da  $F_{\hat{\theta}}$ .

Questo è il modo in cui, il pacchetto `likelihoohAsy` (Bellio e Pierce, 2015) calcola la statistica  $r_p^*$  e i relativi  $p$ -value associati al sistema di ipotesi specificato. In questo elaborato è di centrale interesse cercare di capire, tramite simulazioni, se questa metodologia porti effettivamente dei vantaggi in termini di accuratezza rispetto all'utilizzo della statistica  $r_p$ , in particolare modo con il modello per la classificazione scorretta di dati binari già presentato.

### 3.4 Utilizzo di `likelihoohAsy`

Il pacchetto `likelihoohAsy` è essenzialmente composto da due funzioni. La prima funzione di interesse è `rstar()`, che permette di calcolare il  $p$ -value per un qualsivoglia parametro di interesse  $\psi$ , in un sistema di ipotesi del tipo  $H_0 : \psi = \psi_0$  contro  $H_1 : \psi < \psi_0$ . Gli argomenti principali di tale funzione richiedono di specificare essenzialmente due funzioni, una che ritorni il valore della funzione di log-verosimiglianza per un dato valore del parametro e una

che permetta di generare valori pseudo casuali del campione per un dato valore del parametro, di fondamentale importanza per l'implementazione del metodo di integrazione Montecarlo.

Usando tale funzione con 1000 replicazioni Montecarlo, con gli stessi dati usati nel Capitolo 2 e formulando le stesse ipotesi, si ottiene  $r^{*oss} = 1.066$  con  $\alpha^{oss} = 0.857$  per il parametro  $\beta_1$ ,  $r^{*oss} = -0.351$  con  $\alpha^{oss} = 0.363$  per il parametro  $\beta_2$  e  $r^{*oss} = 1.375$  con  $\alpha^{oss} = 0.915$  per il parametro  $\beta_3$ . Con l'usuale statistica  $r$  (che la funzione `rstar()` calcola a sua volta) si era ottenuto  $r^{oss} = 1.2$  con  $\alpha^{oss} = 0.885$  per il parametro  $\beta_1$ ,  $r^{oss} = -0.43$  con  $\alpha^{oss} = 0.333$  per il parametro  $\beta_2$  e  $r^{oss} = 1.476$  con  $\alpha^{oss} = 0.93$  per il parametro  $\beta_3$ . Il comportamento dei due test è, quindi, almeno in questo caso, fortemente simile.

La seconda funzione è `rstar.ci()`, che permette di ottenere intervalli di confidenza per un generico parametro scalare, basati sulla statistica  $r_p^*$ . Gli argomenti di tale funzioni sono, sostanzialmente, gli stessi della precedente. E' interessante mettere in luce come lavora tale funzioni nel caso che il parametro sul quale si vuole costruire l'intervallo di confidenza sia una riparametrizzazione dei parametri originali del modello di interesse. Ad esempio, nel caso del modello per la classificazione scorretta di dati binari, si può volere costruire intervalli di confidenza per  $\psi = \frac{\exp(\beta^T x^*)}{1 + \exp(\beta^T x^*)}$ , per  $x^*$  fissato, che, seguendo l'impostazione data nel Capitolo 2 rappresenta la probabilità che un dato paziente con covariate  $x^*$  sia effettivamente malato. Uno dei modi per ottenere un intervallo di confidenza per  $\psi$ , che è di fatto una funzione da  $\mathbb{R}^3 \rightarrow \mathbb{R}$  di  $\beta = (\beta_1, \beta_2, \beta_3)$ , sarebbe quello di riscrivere la funzione di log-verosimiglianza profilo in funzione del nuovo parametro  $\psi$ , ottenere la S.M.V.  $\hat{\psi}$  e ammettere nell'intervallo di confidenza solo quei valori  $\psi$  tali per cui  $\Phi(r_p^*(\psi))$  è maggiore di un qualche  $\alpha$  fissato, dove  $\Phi$  è la funzione di ripartizione della normale standard. Un modo più veloce, e quello seguito dalla funzione `rstar.ci()`, è però quello di calcolare il valore della statistica  $r_p^*(\psi)$  per una griglia di valori per il parametro  $\psi$ , ottenendo

la log-verosimiglianza di volta in volta tramite una ottimizzazione numerica vincolando  $\psi$  ad ognuno di tali valori, e accettare nell'intervallo di confidenza solo quei valori del parametro per i quali  $r_p^*(\psi)$  sia minore del quantile della normale standard di livello desiderato, dopo una interpolazione non parametrica.

Usando la funzione `rstar.ci()` per il campione descritto in precedenza si ottengono gli intervalli di confidenza in Tabella 3.1 per  $\psi$  basati su  $r^*$  e su  $r$  con covariate fissate  $x^* = (0, 1, 1)$ , che corrispondono a  $\psi = 0.5$ . Il grafico che rappresenta i valori ottenuti per le due statistiche per la griglia di valori per  $\psi$  è mostrato in Figura 3.1.

Tabella 3.1: Intervalli di confidenza per  $\psi$  basati su  $r_p^*$  e  $r_p$ , con  $n = 100$ ,  $\eta = 0.95$ ,  $\omega = 0.85$  e  $x^* = (0, 1, 1)$

Livello di confidenza	$r_p$	$r_p^*$
0.90	[0.465; 0.788]	[0.47; 0.788]
0.95	[0.430; 0.813]	[0.435; 0.813]
0.99	[0.357; 0.857]	[0.361; 0.857]

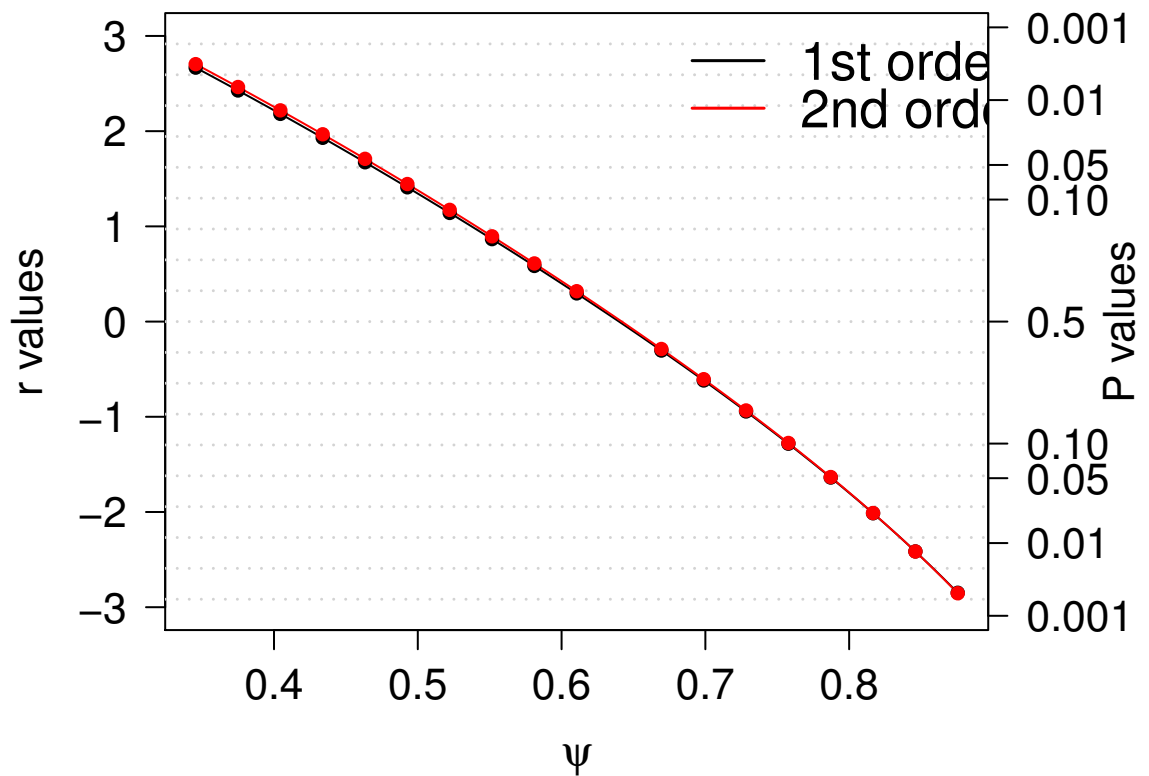


Figura 3.1: Valori di  $r_p$  e  $r_p^*$  per  $\psi$  con  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_1$



# Capitolo 4

## Bootstrap prepivoting di $r$

Dalla fine degli anni '70 è andata ad affermarsi, in statistica, una nuova metodologia per condurre verifiche di ipotesi e costruire intervalli di confidenza che fa largo uso di tecniche computazionali e simulazioni, spesso evitando qualsiasi tipo di ipotesi distributiva riguardo alle variabili interessate dal fenomeno oggetto di studio. Tale metodologia, chiamata *bootstrap*, permette spesso ottimi risultati di accuratezza, in quanto promette errori paragonabili a quelli commessi dalle statistiche di ordine superiore, di cui  $r_p^*$  è un esempio. In questo capitolo si andrà, quindi, a studiare anche il comportamento di due test basati sul bootstrap parametrico della statistica  $r_p$ , per poterli poi confrontare col test  $r_p^*$  implementato nel pacchetto `likelihoodAsy`.

### 4.1 Il bootstrap

Pensando di essere interessati a condurre una verifica di ipotesi o a costruire un intervallo di confidenza di qualche tipo a partire da un campione  $y^{oss}$ , l'idea base del metodo bootstrap è quella di (Young e Smith, 2005, Capitolo 11) :

1. utilizzando il campione osservato, stimare la funzione di ripartizione  $F_0$  della variabile da cui esso è stato generato e ottenere una sua stima  $\hat{F}$ ;



2. attraverso delle simulazioni Monte Carlo da  $\hat{F}$ , ovvero generando campioni bootstrap da  $\hat{F}$  della stessa numerosità di  $y^{oss}$ , ricavare la distribuzione del parametro o della quantità di interesse, la quale sarà poi usata per trarre le conclusioni inferenziali volute.

Quando la stima  $\hat{F}$  è semplicemente la funzione di ripartizione empirica di  $y^{oss}$  si parlerà di *bootstrap non parametrico*, in quanto non viene fatta nessuna assunzione sulle distribuzioni che compongono il modello statistico che descrive la variabile di interesse. Un altro modo di procedere, quando invece si assume di conoscere le distribuzioni che compongono il modello a meno di un parametro  $\theta$ , è quello di stimare tali parametri utilizzando ad esempio il metodo della massima verosimiglianza e usare poi  $F(y, \hat{\theta})$  per generare i campioni bootstrap. Si parlerà in questo caso di *bootstrap parametrico*. Esiste, inoltre, una terza tipologia di bootstrap, quello *semiparametrico*, che si ha quando si formulano ipotesi solo parziali sulla natura di tali distribuzioni, come accade, ad esempio, quando, in presenza di un modello di regressione, si assumono solo le ipotesi di linearità e del secondo ordine.

## 4.2 Bootstrap prepivoting di $r_p$

Si spiega ora come sono costruiti i due test basati sul bootstrap che verranno usati nelle simulazioni. Poichè si è interessati a studiare un particolare modello, quello per la classificazione scorretta di dati binari, si opterà per un bootstrap parametrico. La prima scelta da effettuare è quella di decidere da quale distribuzione generare i campioni bootstrap.

Si può pensare di generare tali campioni bootstrap da

$$Y \sim \text{Bi}(1, \hat{\pi}_{Y|H_0})$$

dove

$$\begin{aligned}\hat{\pi}_{Y|H_0} &= 1 - \omega - (1 - \eta - \omega)\hat{\pi}_{Z|F_0} \\ \hat{\pi}_{Z|H_0} &= \exp(x\hat{\beta}_{H_0}) / (1 + \exp(x\hat{\beta}_{H_0}))\end{aligned}$$

Ovvero dalla distribuzione di  $Y$  stimata sotto  $H_0$ . Per esempio, per  $H_0 : \beta_1 = c$  contro  $H_1 : \beta_1 < c$ , si avrà  $\hat{\beta}_{H_0} = (c, \hat{\beta}_{2c}, \hat{\beta}_{3c})$ , dove  $\hat{\beta}_{2c}$  e  $\hat{\beta}_{3c}$  sono le S.M.V. per  $\beta_2$  e  $\beta_3$  quando  $\beta_1 = c$ . Si può dimostrare che, con questa impostazione, l'errore compiuto dal test bootstrap è di ordine  $O(n^{-\frac{3}{2}})$  (DiCiccio et al., 2001).

Si genera quindi un grande numero  $B$  di campioni, che denoteremo con  $y^b$ , sui quali si andrà a calcolare il valore della statistica  $r_p^b(\hat{\beta}_{H_0})$  che confronta  $H_0 : \beta_1 = c$  con  $H_1 : \beta_1 < c$ . Il valore di  $\alpha^{oss}$  sarà dato da

$$\frac{1}{B} \sum_{b=1}^B I(r_p^b(\hat{\beta}_{H_0}) \leq r_p^{oss}(\hat{\beta}_{H_0}))$$

dove

$$I(r_p^b(\hat{\beta}_{H_0}) \leq r_p^{oss}(\hat{\beta}_{H_0})) = \begin{cases} 1 & \text{se } r_p^b(\hat{\beta}_{H_0}) \leq r_p^{oss}(\hat{\beta}_{H_0}) \\ 0 & \text{altrimenti} \end{cases}$$

con  $r_p^{oss}(\hat{\beta}_{H_0})$  valore di  $r_p$  colcolato con i dati osservati.

Quello che si va a fare è quindi di approssimare il valore di interesse  $\alpha^{oss}$ , che corrisponde alla probabilità di avere un valore per  $r_p$ , sotto  $H_0$ , più contrario (in questo caso minore), di quello osservato  $r_p^{oss}$ , con una stima Montecarlo della probabilità che  $r_p^b$  sia minore di  $r_p^{oss}$ .

Un'altra scelta potrebbe essere quella di usare

$$Y \sim \text{Bi}(1, \hat{\pi}_{Y|x})$$

come distribuzione da cui generali i campioni bootstrap, dove

$$\hat{\pi}_{Y|x} = 1 - \omega - (1 - \eta - \omega)\hat{\pi}_{Z|x}$$

$$\hat{\pi}_{Z|x} = \exp(x\hat{\beta}) / (1 + \exp(x\hat{\beta}))$$

con  $\hat{\beta}$  S.M.V. per  $\beta$ . Il valore di  $\alpha^{oss}$  sarà in questo caso dato da

$$\frac{1}{B} \sum_{b=1}^B I(r_p^b(\hat{\beta}) \leq r_p^{oss}(\hat{\beta}_{H_0}))$$

dove

$$I(r_p^b(\hat{\beta}) \leq r_p^{oss}(\hat{\beta}_{H_0})) = \begin{cases} 1 & \text{se } r_p^b(\hat{\beta}) \leq r_p^{oss}(\hat{\beta}_{H_0}) \\ 0 & \text{altrimenti} \end{cases}$$

con  $r_p^b(\hat{\beta})$  valore di  $r_p$  per i dati originari nella S.M.V. Con questa scelta si ha un errore di ordine  $O(n^{-1})$  (Young e Smith, 2005, Capitolo 11), paragonabile a quello di  $r_p^*$ .

Calcolando gli  $\alpha^{oss}$  con le procedure appena descritte sugli stessi dati del Paragrafo (2.2), con  $B = 5000$  replicazioni, per la verifica di ipotesi sul parametro  $\beta_1$  si ottiene  $\alpha^{oss} = 0.855$  col bootstrap di terzo ordine e  $\alpha^{oss} = 0.804$  per il bootstrap del secondo ordine.

### 4.3 Intervalli di confidenza col bootstrap

Uno dei modi con i quali è possibile costruire intervalli di confidenza attraverso il bootstrap è quello di:

1. calcolare, per una griglia di valori per il parametro  $\psi_1, \dots, \psi_k$  di cui si vuole costruire l'intervallo di confidenza, i corrispondenti  $\alpha^{oss}(\psi_i)$  col test prescelto, e da essi i corrispondenti valori  $r_p(\psi_i) = \Phi^{-1}(\alpha^{oss}(\psi_i))$ ;
2. interpolare non parametricamente i valori così ottenuti, ad esempio con le funzioni fornite dal pacchetto `pspline`;
3. per un intervallo di confidenza di livello 0.05, considerare come estremo inferiore il valore del parametro per il quale  $r^{oss} = z_{0.025}$  e come estremo superiore quello per cui  $r(\psi) = z_{0.975}$ , con  $z_q$  quantile di ordine  $q$  della normale standard.

E' interessante notare che tale procedura è molto più veloce se eseguita utilizzando il bootstrap di secondo ordine, che richiede di generare i campioni bootstrap una sola volta, invece che di dover generare campioni bootstrap differenti per ogni valore che compone la griglia di partenza, come richiede invece il bootstrap di terzo ordine.

Seguendo questo approccio, per il parametro  $\beta_1$ , si ottiene l'intervallo  $[0.572; 2.996]$  col bootstrap di secondo ordine e l'intervallo  $[0.65; 2.977]$  con quello di terzo. Per confronto, con l'usuale statistica  $r$  si ottiene  $[0.696; 3.01]$ ,

mentre con  $r^*$  si ha  $[0.644; 2.946]$ . Il grafico che rappresenta i valori ottenuti per costruire gli intervalli di confidenza con le due procedure basate sul bootstrap per la griglia di valori per  $\psi$  è mostrato in Figura 4.1.

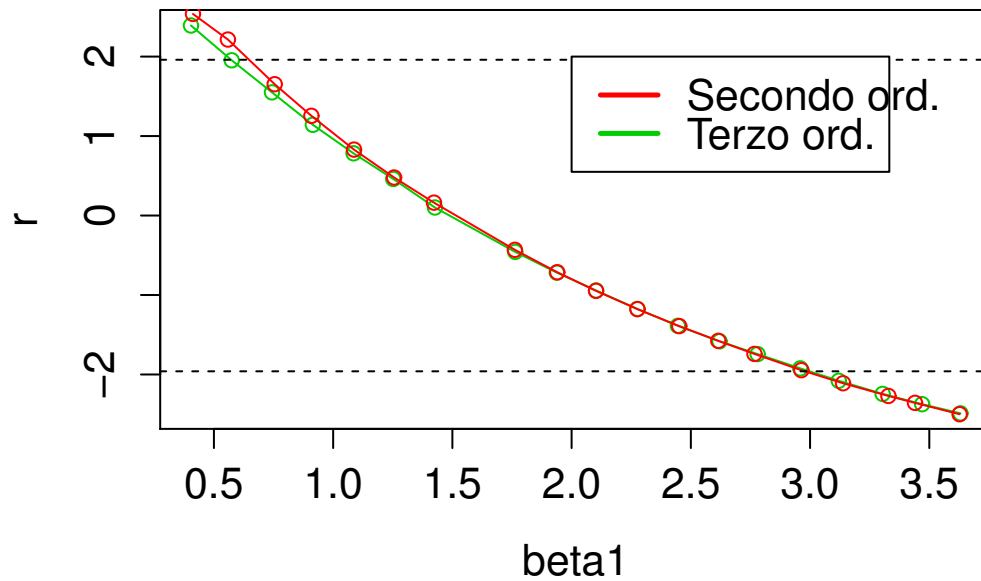


Figura 4.1: Grafico per la costruzione degli intervalli di confidenza con le procedure bootstrap con  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_1$

E' possibile, inoltre, ottenere un'alternativa stima del parametro, come quel valore del parametro per cui  $r(\psi) = 0$ . In questo caso si ha  $\hat{\beta}_1 = 1.481$  col bootstrap del secondo ordine,  $\hat{\beta}_1 = 1.512$  con quello del terzo.



# Capitolo 5

## Studi di simulazione

Gli studi di simulazioni sono molto usati in statistica per valutare le proprietà delle procedure e delle metodologie in essa usate, soprattutto quando sono coinvolte approssimazioni di un qualche tipo, che sarebbe difficoltoso comprendere appieno attraverso uno studio solo analitico. In questo capitolo si effettueranno studi di simulazione per valutare la bontà delle quattro statistiche test già presentate. L'usuale statistica  $r_p$  dovrebbero garantire un errore del primo ordine, il bootstrap di secondo ordine e la statistica  $r_p^*$ , come implementata in `likelihoodAsy`, dovrebbero avere un errore del secondo ordine, mentre il bootstrap del terzo ordine dovrebbe essere il metodo più accurato.

### 5.1 Struttura della simulazione

Il modello considerato è quello già presentato nel Capitolo 2 con  $\beta = (1, -1, 1)$  e matrice delle covariate  $X$  fissata, con primo vettore generato da una normale standard, secondo vettore da una bernoulliana con  $p = 1/2$  e ultimo vettore che funge da intercetta. Si considereranno tre diversi scenari:

- $n = 100$  con  $\eta = 0.95$  e  $\omega = 0.85$
- $n = 250$  con  $\eta = 0.9$  e  $\omega = 0.8$

- $n = 250$  con  $\eta = 0.95$  e  $\omega = 0.85$ .

Per ognuno di essi si genereranno 1000 campioni  $y$  casualmente, su cui si effettueranno, con ognuna delle statistiche test presentate, le tre verifiche di ipotesi, una per ogni componente di  $\beta$ , che contrappongono le relative stime ai veri valori  $(1, -1, 1)$ . Un test di cui si conosce la distribuzione esatta fornirebbe, in questo modo,  $\alpha^{oss}$  distribuiti uniformemente in  $[0, 1]$ . Si valuterà quindi la bontà di ogni statistica test valutando, sia tramite verifica grafica che utilizzando il test di Kolmogorov-Smirnov, quanto la distribuzione degli  $\alpha^{oss}$  da essa forniti per ogni verifica di ipotesi rispetti tale proprietà. Per le metodologie bootstrap si useranno  $B = 5000$  replicazioni. Si calcolerà, inoltre, il livello di copertura degli intervalli di confidenza al 90%, 95% e 99% associati ad ogni test in ciascun scenario. Nei prossimi paragrafi si riporteranno sinteticamente i risultati ottenuti.

## 5.2 Test su $\beta_1$

### 5.2.1 $n = 100$ con $\eta = 0.95$ e $\omega = 0.85$

Tabella 5.1: Risultati del test di Kolmogorov-Smirnov per  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_1$

$r$	$r^*$	Bootstrap secondo ordine	Bootstrap terzo ordine
0	0.807	0.729	0.697

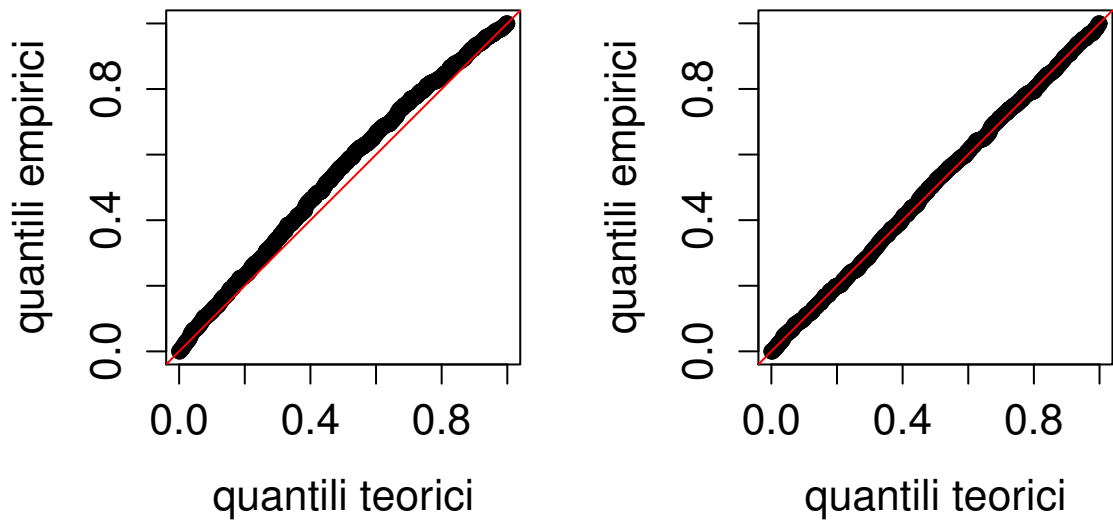


Figura 5.1: qq-plot ottenuto con  $r$  e  $r^*$  per  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_1$

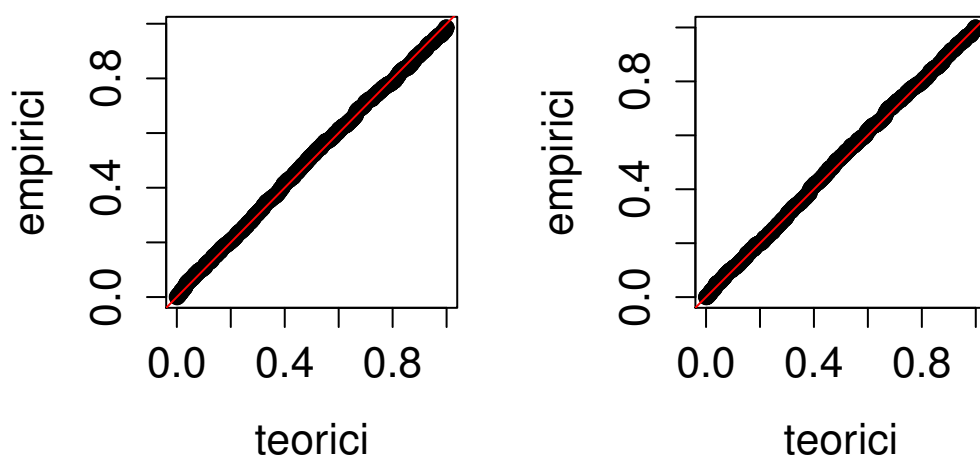


Figura 5.2: qq-plot ottenuto col bootstrap di secondo e terzo ordine per  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_1$



Tabella 5.2:  $\alpha$  percentuali osservati confrontati con quelli teorici con  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_1$

	1	2.5	5	10	50	90	95	97.5	99
$r_p$	1	1.9	3.5	7.4	44	87.3	92.9	96.2	99
$r_p^*$	1.1	2.4	4.2	9.9	48.3	89.8	94.7	98.2	99.3
boot. secondo ord.	0.9	2.2	3.7	9	48.4	91.4	96.5	99.3	100
boot. terzo ord.	1.1	2.3	4	9.1	48.1	89.8	95.3	98.3	99.2

Tabella 5.3: Livelli di copertura con  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_1$

	$r_p$	$r_p^*$	boot. secondo ord.	boot. terzo ord.
90	92.6	90.1	91	90.9
95	96.5	95.8	96.3	96
99	99	98.9	99.1	98.9

### 5.2.2 $n = 250$ con $\eta = 0.9$ e $\omega = 0.8$

Tabella 5.4: Risultati del test di Kolmogorov-Smirnov per  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_1$

$r$	$r^*$	Bootstrap secondo ordine	Bootstrap terzo ordine
0.111	0.927	0.665	0.855

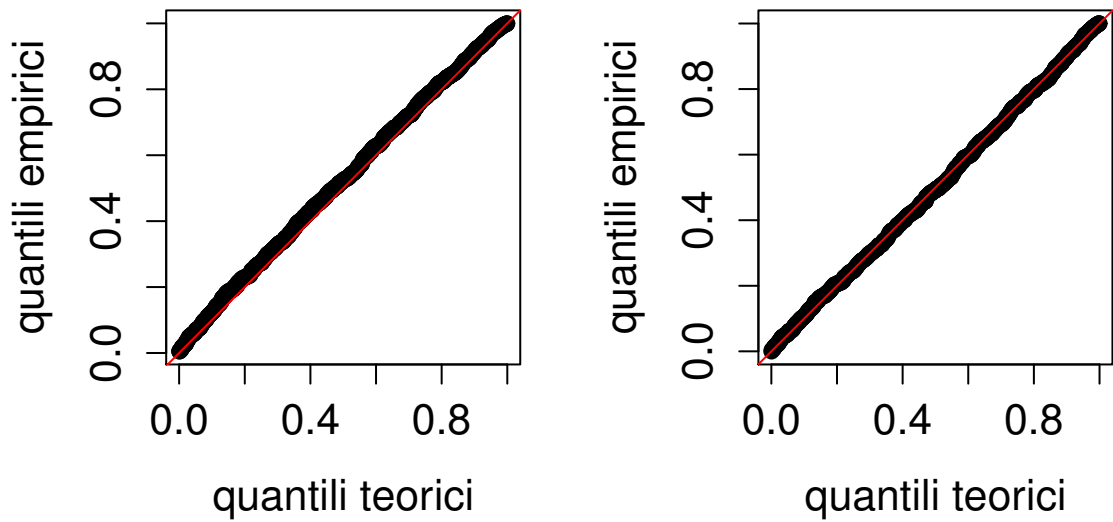


Figura 5.3: qq-plot ottenuto con  $r$  e  $r^*$  per  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_1$

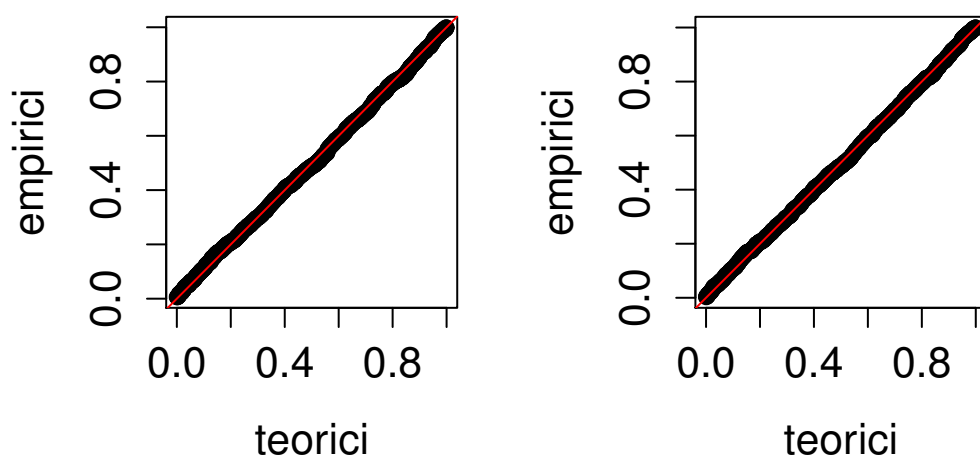


Figura 5.4: qq-plot ottenuto col bootstrap di secondo e terzo ordine per  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_1$

Tabella 5.5:  $\alpha$  percentuali osservati confrontati con quelli teorici per  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_1$

	1	2.5	5	10	50	90	95	97.5	99
$r_p$	0.6	1.7	3.3	8.1	47.3	88.8	93.9	96.3	98.2
$r_p^*$	0.8	2.2	4.5	9.1	51.3	90.1	94.5	96.7	98.4
boot. secondo ord.	0.6	2	4.1	8.8	51.8	90.5	95.2	97.4	98.9
boot. terzo ord.	0.6	1.9	4.3	9.1	51.3	90.4	94.7	96.8	98.8

Tabella 5.6: Livelli di copertura con  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_1$

	$r_p$	$r_p^*$	boot. secondo ord.	boot. terzo ord.
90	91.9	90.9	91.2	90.9
95	96.7	95.5	95.8	95.7
99	99.4	99.2	99.3	99.4

### 5.2.3 $n = 250$ con $\eta = 0.95$ e $\omega = 0.85$

Tabella 5.7: Risultati del test di Kolmogorov-Smirnov per  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_1$

$r$	$r^*$	Bootstrap secondo ordine	Bootstrap terzo ordine
0.207	0.963	0.78	0.902

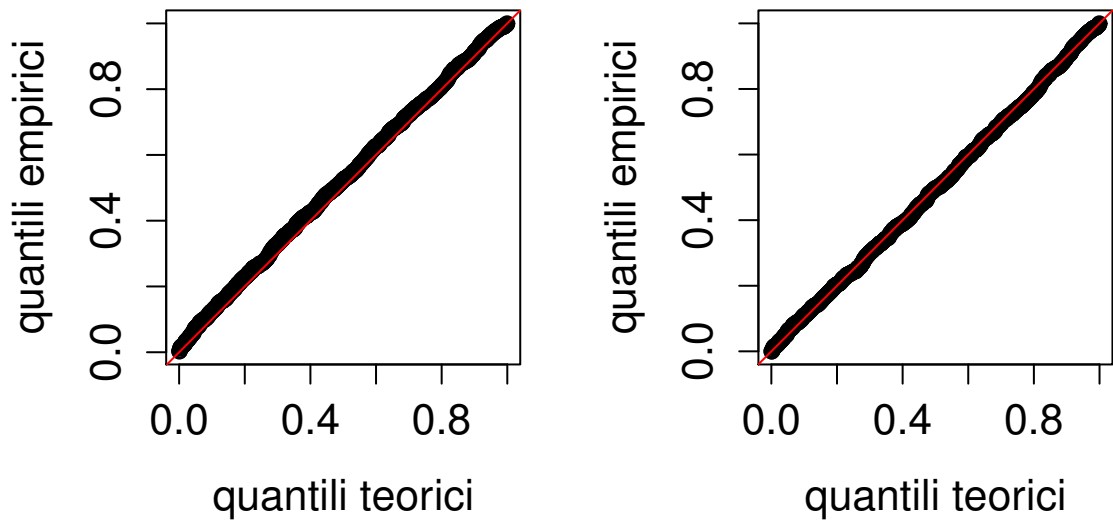


Figura 5.5: qq-plot ottenuto con  $r$  e  $r^*$  per  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_1$

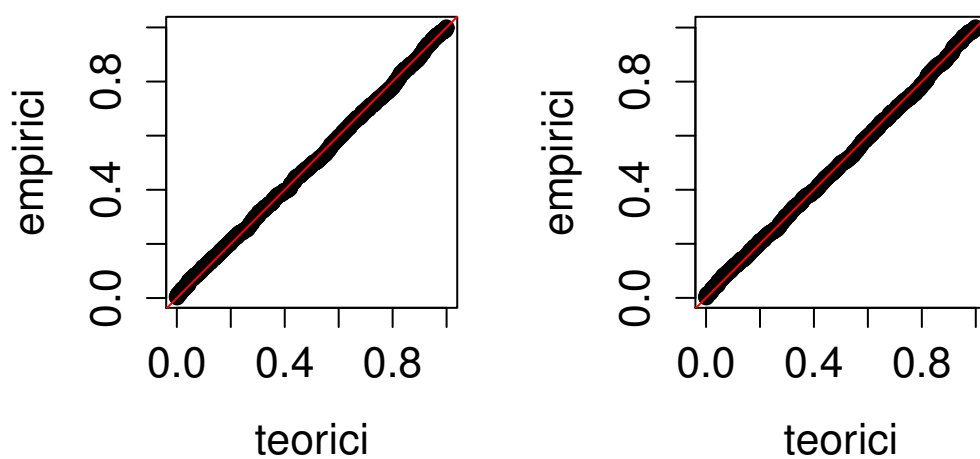


Figura 5.6: qq-plot ottenuto col bootstrap di secondo e terzo ordine per  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_1$

Tabella 5.8:  $\alpha$  percentuali osservati confrontati con quelli teorici per  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_1$

	1	2.5	5	10	50	90	95	97.5	99
$r_p$	0.3	1.7	3.5	7.9	47.5	89	93.6	96.1	98.7
$r_p^*$	0.5	2.1	4.5	9.2	50.6	90.2	94.4	96.9	99.2
boot. secondo ord.	0.3	1.7	4.2	8.9	50.8	90.7	94.6	97.3	99.2
boot. terzo ord.	0.3	1.9	4.3	8.9	50.5	90.3	94.7	96.8	99.2

Tabella 5.9: Livelli di copertura con  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_1$

	$r_p$	$r_p^*$	boot. secondo ord.	boot. terzo ord.
90	92.1	90.8	91.1	91.1
95	96.5	95.5	95.8	95.7
99	99.7	99.5	99.6	99.7

### 5.3 Test su $\beta_2$

#### 5.3.1 $n = 100$ con $\eta = 0.95$ e $\omega = 0.85$

Tabella 5.10: Risultati del test di Kolmogorov-Smirnov per  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_2$

$r$	$r^*$	Bootstrap secondo ordine	Bootstrap terzo ordine
0.241	0.225	0.026	0.209

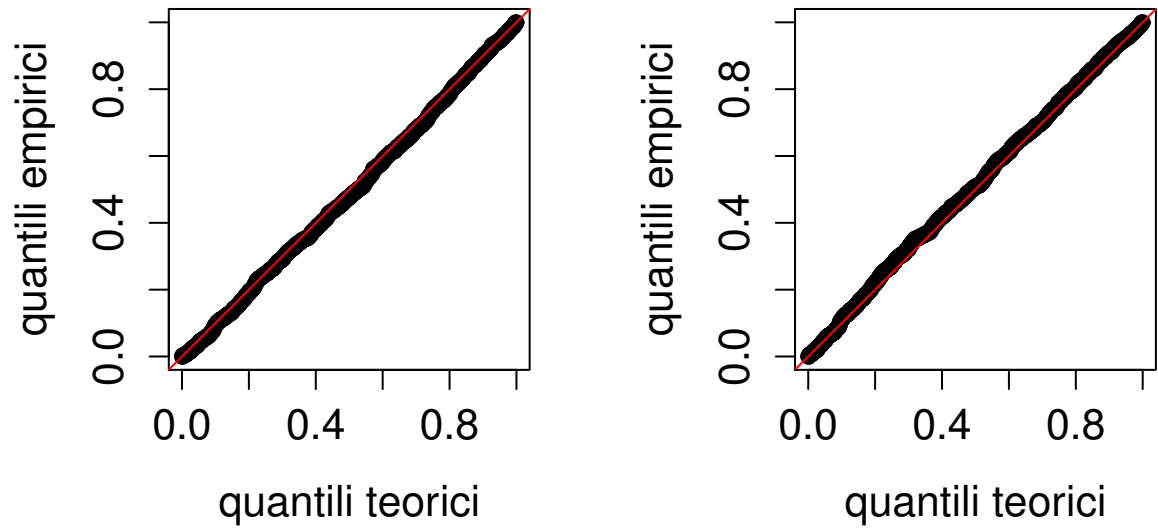


Figura 5.7: qq-plot ottenuto con  $r$  e  $r^*$  per  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_2$

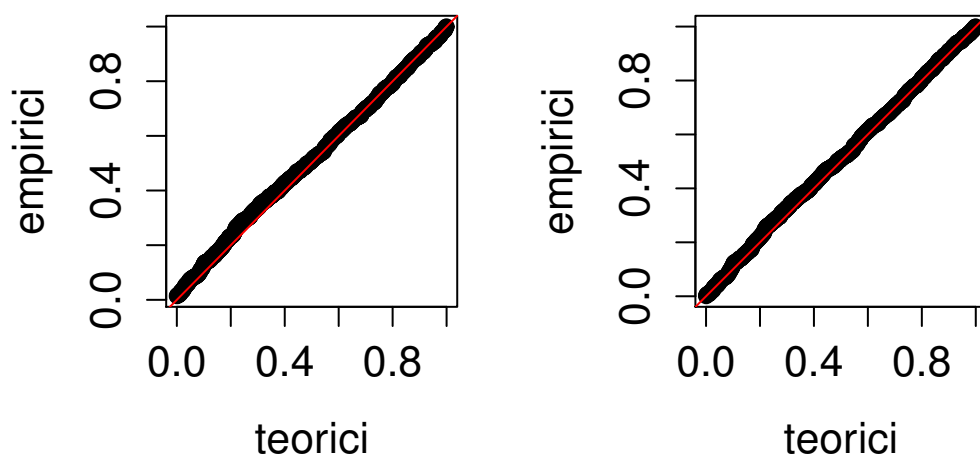


Figura 5.8: qq-plot ottenuto col bootstrap di secondo e terzo ordine per  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_2$

Tabella 5.11:  $\alpha$  percentuali osservati confrontati con quelli teorici per  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_2$

	1	2.5	5	10	50	90	95	97.5	99
$r_p$	1.6	3.2	6.1	10.3	52.6	90.1	95.5	97.8	99.2
$r_p^*$	1.3	2.8	5.2	9.4	49.1	89.1	94.7	97.6	99.1
boot. secondo ord.	0	1.3	3.2	8.3	48.7	90.3	95.9	98	99.4
boot. terzo ord.	1.3	2.7	4.8	9.2	48.6	89.4	94.9	97.6	99.2

Tabella 5.12: Livelli di copertura con  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_2$

	$r_p$	$r_p^*$	boot. secondo ord.	boot. terzo ord.
90	89.7	90.6	91.7	90.8
95	93.9	94.8	96.8	95.2
99	98.4	98.7	100	98.7

### 5.3.2 $n = 250$ con $\eta = 0.9$ e $\omega = 0.8$

Tabella 5.13: Risultati del test di Kolmogorov-Smirnov per  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_2$

$r$	$r^*$	Bootstrap secondo ordine	Bootstrap terzo ordine
0.006	0.136	0.065	0.095

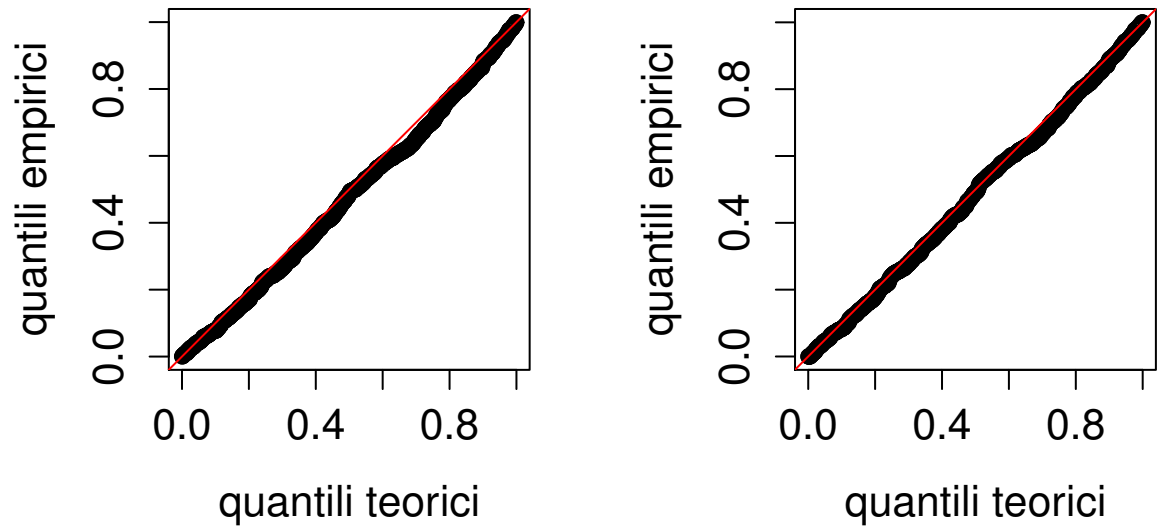


Figura 5.9: qq-plot ottenuto con  $r$  e  $r^*$  per  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_2$

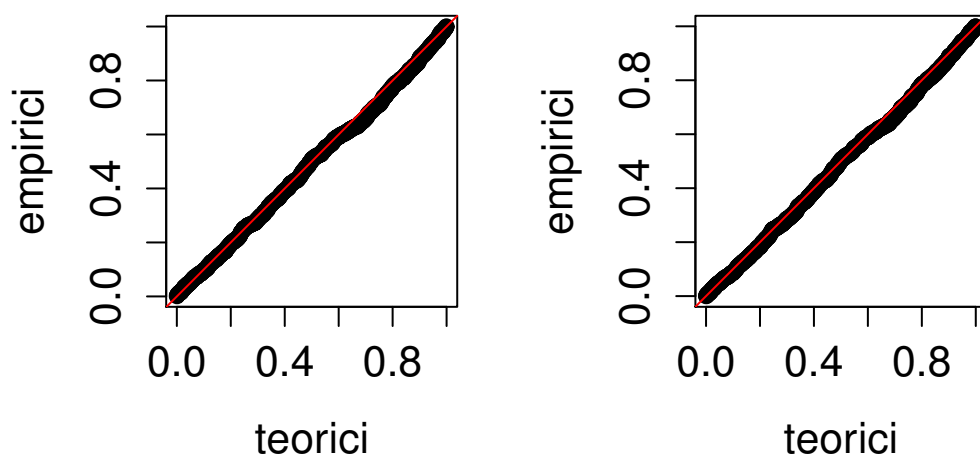


Figura 5.10: qq-plot ottenuto col bootstrap di secondo e terzo ordine per  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_2$



Tabella 5.14:  $\alpha$  percentuali osservati confrontati con quelli teorici per  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_2$

	1	2.5	5	10	50	90	95	97.5	99
$r_p$	1.1	2.5	5.9	11.8	51.7	92	96.4	97.8	99.2
$r_p^*$	1.5	2.8	5.6	11.9	50.7	91.5	95.9	97.6	99.1
boot. secondo ord.	0.6	2.2	4.9	10.5	49.7	91.7	96.1	97.8	99.2
boot. terzo ord.	0.9	2.4	5.1	11.2	49.8	91.7	96	97.8	99.2

Tabella 5.15: Livelli di copertura con  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_2$

	$r_p$	$r_p^*$	boot. secondo ord.	boot. terzo ord.
90	88.2	88.1	89.5	88.8
95	94.1	94.4	95.1	94.9
99	98.9	98.5	99.4	99

### 5.3.3 $n = 250$ con $\eta = 0.95$ e $\omega = 0.85$

Tabella 5.16: Risultati del test di Kolmogorov-Smirnov per  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_2$

$r$	$r^*$	Bootstrap secondo ordine	Bootstrap terzo ordine
0.052	0.385	0.245	0.232

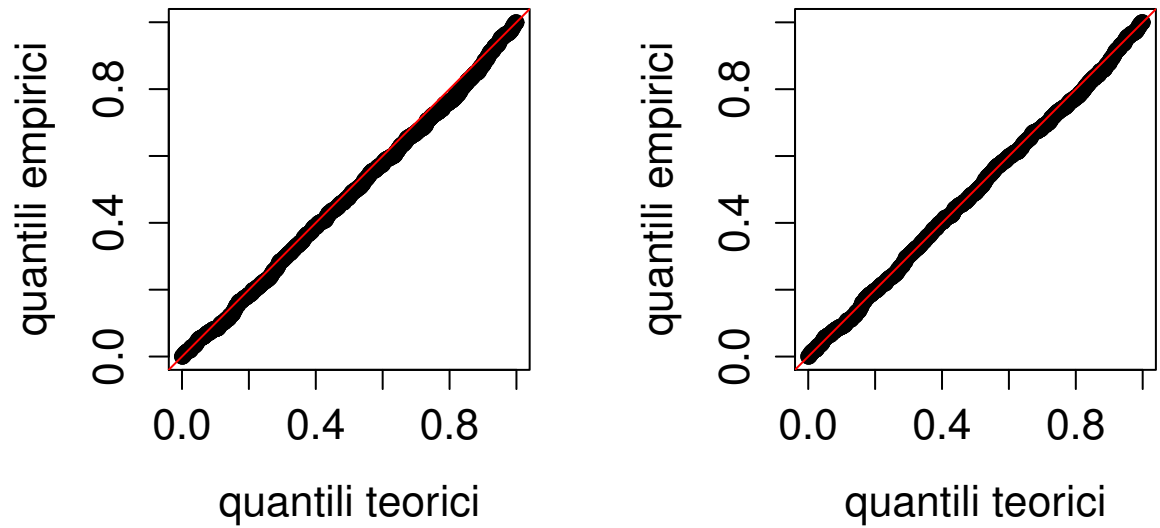


Figura 5.11: qq-plot ottenuto con  $r$  e  $r^*$  per  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_2$

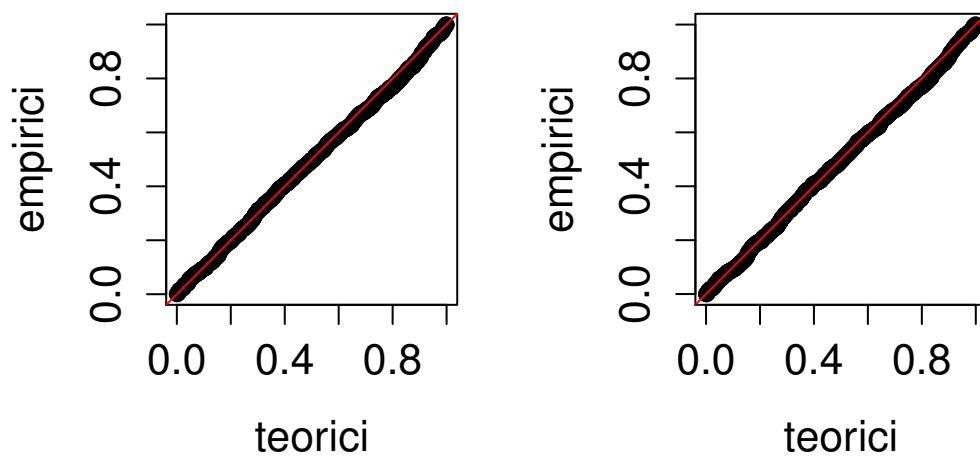


Figura 5.12: qq-plot ottenuto col bootstrap di secondo e terzo ordine per  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_2$

Tabella 5.17:  $\alpha$  percentuali osservati confrontati con quelli teorici per  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_2$

	1	2.5	5	10	50	90	95	97.5	99
$r_p$	0.9	2.7	4.9	11.9	52.3	91.7	95.3	98.4	99.2
$r_p^*$	0.8	2.6	4.6	11.2	50.7	91.2	95	98.1	99
boot. secondo ord.	0.8	2.5	4.5	10.8	50.7	91.7	95.2	98.4	99.2
boot. terzo ord.	0.9	2.7	4.7	11.1	50.5	91.2	95.2	98.1	99.2

Tabella 5.18: Livelli di copertura con  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_2$

	$r_p$	$r_p^*$	boot. secondo ord.	boot. terzo ord.
90	88.1	88.8	89.2	88.9
95	95.1	95.4	95.5	95.3
99	99.1	99.2	99.2	99.1

## 5.4 Test su $\beta_3$

### 5.4.1 $n = 100$ con $\eta = 0.95$ e $\omega = 0.85$

Tabella 5.19: Risultati del test di Kolmogorov-Smirnov per  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_3$

$r$	$r^*$	Bootstrap secondo ordine	Bootstrap terzo ordine
0.067	0.647	0.299	0.739

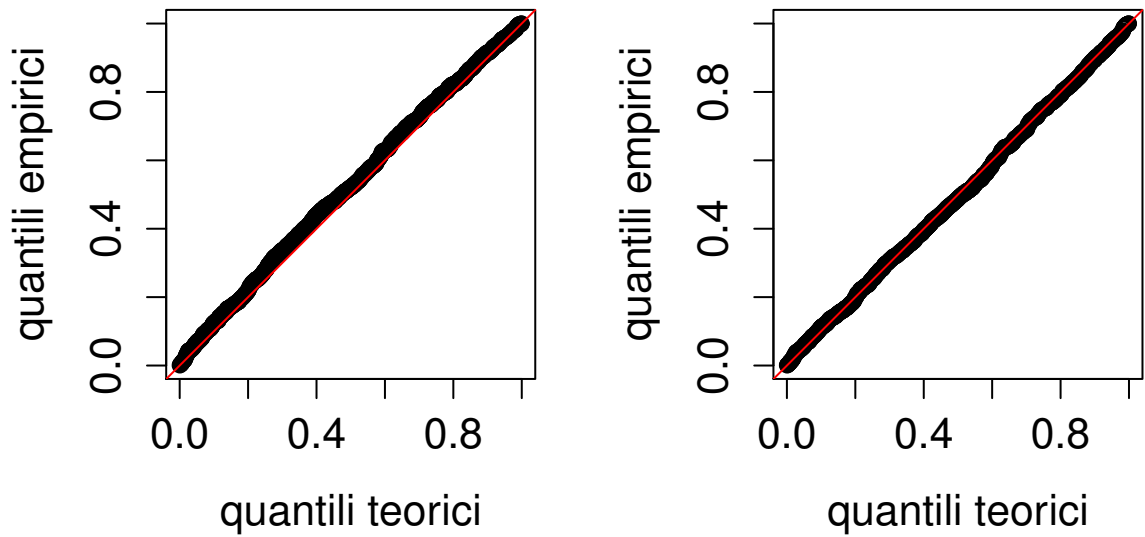


Figura 5.13: qq-plot ottenuto con  $r$  e  $r^*$  per  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_3$

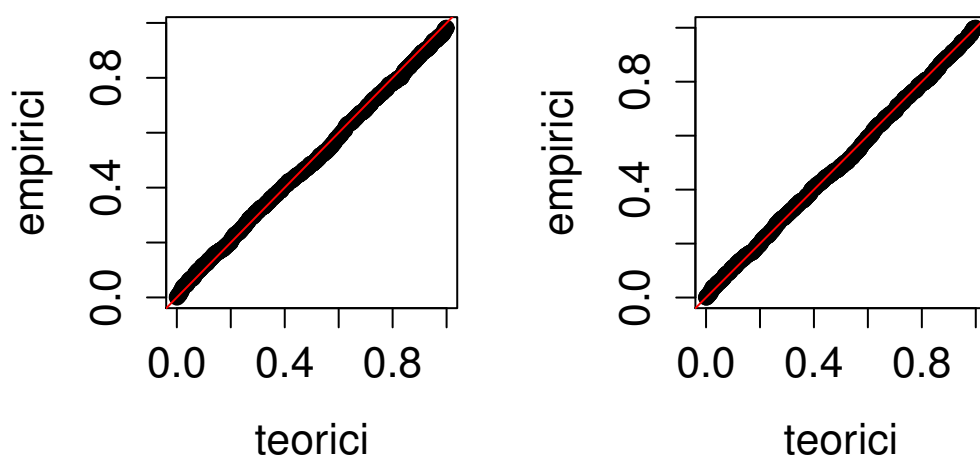


Figura 5.14: qq-plot ottenuto col bootstrap di secondo e terzo ordine per  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_3$

Tabella 5.20:  $\alpha$  percentuali osservati confrontati con quelli teorici per  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_3$

$r_p$	0.8	1.8	3.6	8	47.3	88.1	94.1	97.1	98.5
$r_p^*$	0.9	2	4.3	9	51.5	90.2	95.3	98	98.8
boot. secondo ord.	0.8	1.8	3.8	8.4	51.9	92.1	97.4	99.3	100
boot. terzo ord.	0.9	1.9	4	9.1	51.6	90.8	95.7	98	98.9

Tabella 5.21: Livelli di copertura con  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_3$

	$r_p$	$r_p^*$	boot. secondo ord.	boot. terzo ord.
90	92	91	91.6	90.9
95	96.4	95.7	96.2	96
99	99.2	99.1	99.2	99.1

#### 5.4.2 $n = 250$ con $\eta = 0.9$ e $\omega = 0.8$

Tabella 5.22: Risultati del test di Kolmogorov-Smirnov per  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_3$

$r$	$r^*$	Bootstrap secondo ordine	Bootstrap terzo ordine
0.005	0.086	0.129	0.115

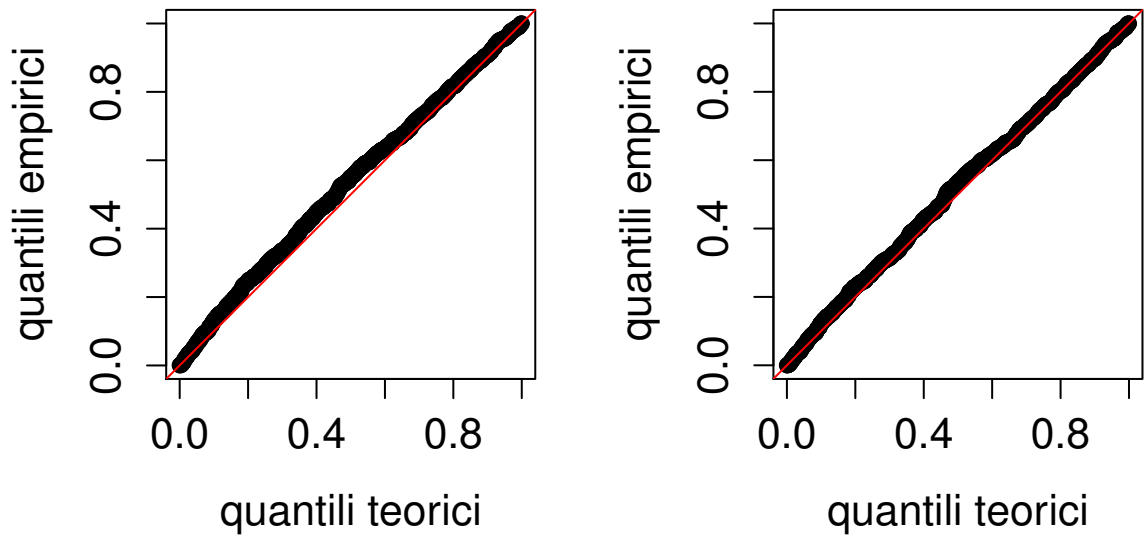


Figura 5.15: qq-plot ottenuto con  $r$  e  $r^*$  per  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_3$

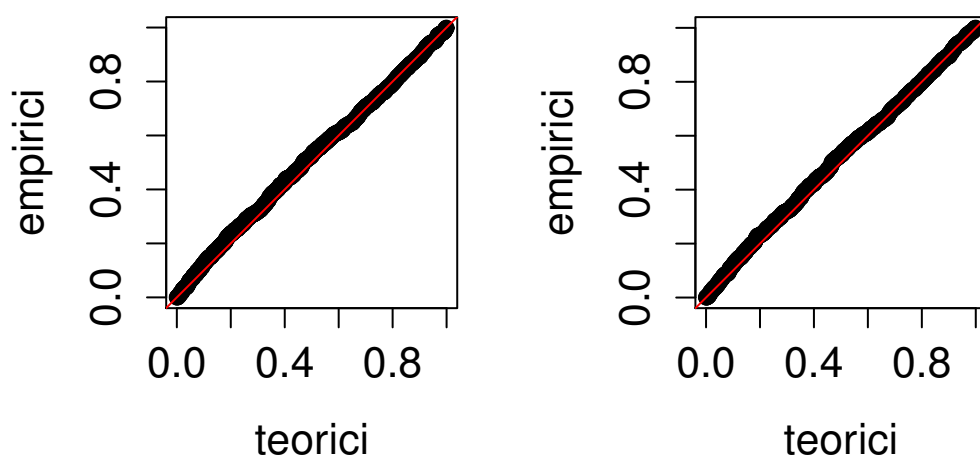


Figura 5.16: qq-plot ottenuto col bootstrap di secondo e terzo ordine per  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_3$

Tabella 5.23:  $\alpha$  percentuali osservati confrontati con quelli teorici per  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_3$

	1	2.5	5	10	50	90	95	97.5	99
$r_p$	1.1	2	4	8.2	45.8	88.9	93.4	96.6	99.1
$r_p^*$	1.2	2.3	4.4	8.7	46.6	90.2	94.7	97.1	99.1
boot. secondo ord.	1.1	2.1	4.1	8.3	46.9	90.5	95.6	97.5	99.4
boot. terzo ord.	1.1	2.2	4.5	8.6	46.6	90.3	94.8	97.2	99.2

Tabella 5.24: Livelli di copertura con  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$  per  $\beta_3$

	$r_p$	$r_p^*$	boot. secondo ord.	boot. terzo ord.
90	91.8	91.3	91.7	91.4
95	96	95.6	95.9	95.5
99	98.9	98.8	98.9	98.9

### 5.4.3 $n = 250$ con $\eta = 0.95$ e $\omega = 0.85$

Tabella 5.25: Risultati del test di Kolmogorov-Smirnov per  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_3$

$r$	$r^*$	Bootstrap secondo ordine	Bootstrap terzo ordine
0.036	0.19	0.266	0.361

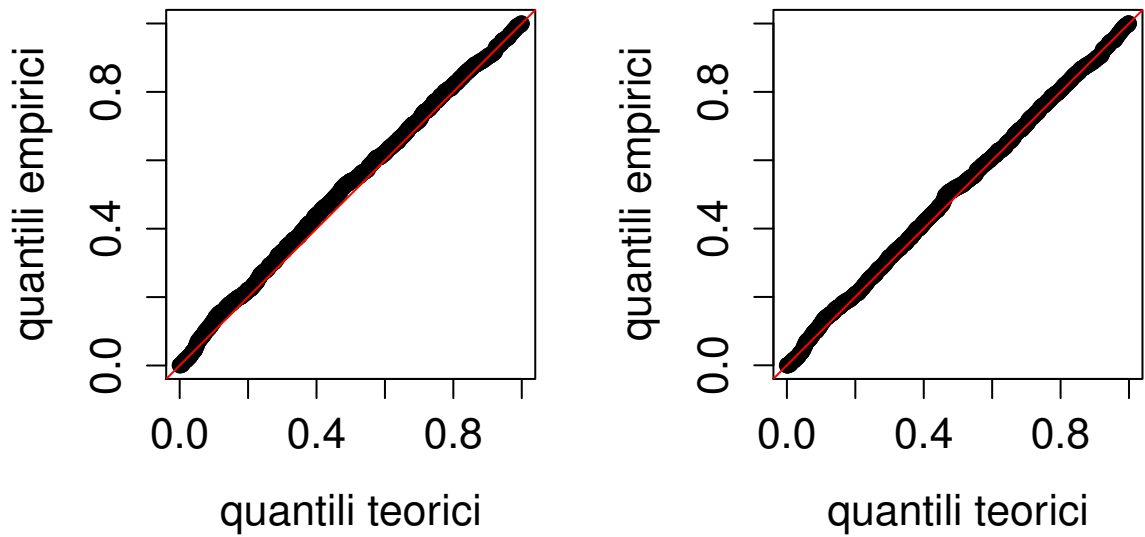


Figura 5.17: qq-plot ottenuto con  $r$  e  $r^*$  per  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_3$

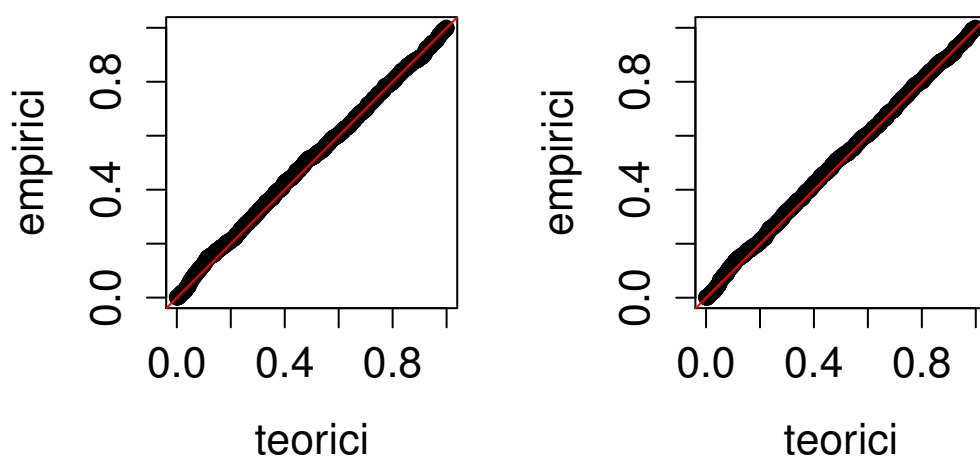


Figura 5.18: qq-plot ottenuto col bootstrap di secondo e terzo ordine per  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_3$



Tabella 5.26:  $\alpha$  percentuali osservati confrontati con quelli teorici per  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_3$

	1	2.5	5	10	50	90	95	97.5	99
r	1.1	2.7	4.6	7.5	46.1	89.6	94.4	97.3	98.6
rs	1.5	3.1	4.9	8.3	46.9	90.7	95.2	97.4	98.6
boot. secondo ord.	1.2	2.9	4.5	8.1	47.6	91.6	95.7	97.6	98.9
boot. terzo ord.	1.2	2.9	4.8	8	47.7	90.7	95.2	97.5	98.6

Tabella 5.27: Livelli di copertura con  $n = 250$ ,  $\eta = 0.95$  e  $\omega = 0.85$  per  $\beta_3$

	$r_p$	$r_p^*$	boot. secondo ord.	boot. terzo ord.
90	92.5	91.7	91.9	92
95	95.4	95.1	95.5	95.2
99	98.9	98.5	98.8	98.8

## 5.5 Commenti

I risultati evidenziati mettono in luce alcuni fatti interessanti. Per quanto riguarda il parametro  $\beta_1$ , associato al vettore di covariate normali, considerando una significatività al cinque per cento, solo gli  $\alpha^{oss}$  forniti dalla statistica  $r$  con  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$ , risultano avere distribuzione non uniforme. In generale, comunque, i  $p$ -value associati al test di Kolmogorov-Smirnov sugli  $\alpha^{oss}$  forniti dalle altre tre statistiche sono del tutto confrontabili e parecchio migliori di quelli ottenuti con  $r$ .

Da quanto emerso, invece, per il parametro  $\beta_2$ , associato al vettore di covariate bernoulliane, sia gli  $\alpha^{oss}$  ottenuti col bootstrap di secondo ordine con  $n = 100$ ,  $\eta = 0.95$  e  $\omega = 0.85$ , che quelli con la statistica  $r$  con  $n = 250$ ,  $\eta = 0.9$  e  $\omega = 0.8$ , risultano non uniformi. Per il resto, anche qui la statistica  $r$  si mostra nettamente peggiore delle altre tre, i due bootstrap sembrano avere un comportamento simile con  $n = 250$  e la statistica  $r^*$  si mostra essere la migliore per ogni configurazione.

Per il parametro  $\beta_3$ , che funge da intercetta, la statistica  $r$  fornisce  $\alpha^{oss}$

non uniformi per entrambi i casi con  $n = 250$  e in generale più distanti dall'uniformità rispetto agli altri tre test. Per il resto la statistica  $r^*$  e il bootstrap di terzo ordine sono apprezzabilmente migliori di quello di secondo ordine con  $n = 100$ , mentre per  $n = 250$  i tre test forniscono risultati simili per  $\eta = 0.9$  e  $\omega = 0.8$  e il bootstrap di terzo ordine risulta lievemente migliore per  $\eta = 0.95$  e  $\omega = 0.85$ .



# Capitolo 6

## Conclusioni

Da quanto emerso il modello presentato da McInturff et al. (2003) risulta essere versatile ed adeguato per trattare il problema della classificazione scorretta di dati binari.

Per quanto riguarda invece i test presentati in relazione a tale modello, il bootstrap di terzo ordine e la statistica  $r_p^*$  sembrano, per quanto riguarda la verifica di ipotesi e secondo le analisi svolte nel Capitolo 5, essere all'incirca equivalenti come bontà e accuratezza. Essi risultano essere i migliori test tra i quattro presentati. Nettamente peggiore degli altri tre risulta essere quello basato sull'usuale statistica  $r_p$ .

Bisogna però sottolineare il fatto che il bootstrap di terzo ordine risulta essere computazionalmente molto più oneroso del test basato sulla statistica  $r_p^*$  e che, quindi, per quanto riguarda la verifica di ipotesi, la scelta migliore potrebbe essere quella di utilizzare quest'ultima che, seppur abbia teoricamente una precisione al secondo ordine, ha fornito risultati sempre soddisfacenti e del tutto confrontabili, se non migliori, di quelli ottenuti con gli altri test.

Può essere interessante notare inoltre che, da quanto visto nelle simulazioni, per tutti e quattro i test i risultati migliori sono arrivati dalle verifiche di ipotesi sui parametri  $\beta_2$  e  $\beta_3$ , associati a una variabile categoriale e all'intercetta, piuttosto che dalla verifica di ipotesi su  $\beta_1$ , associato a una

covariata normale. Ciò potrebbe suggerire che tali test giungano a risultati migliori, in questo modello, se applicati a parametri associati a variabili non continue.

# Appendice A

## Codice R

[Codice R]

```
#capitolo 2
p.range<- function(p,eps=2.22e-15){
  out<-p
  out[p<eps]<-eps
  out[p>(1-eps)]<- (1-eps)
  return(out)
}

log_ver=function(beta, data, eta=eta0, omega=omega0) {
  X <- data$X
  y <- data$y
  arg <- as.vector(X*%beta)
  piz <- plogis(arg)
  piz <- p.range(piz)
  piy <- 1-omega0-(1-omega0-eta0)*piz
  piy <- p.range(piy)
  sum(y*log(piy))+sum((1-y)*log(1-piy))
}
```

```
score <- function(beta,data, eta=eta0, omega=omega0)
{
  X <-data$X
  y <- data$y
  arg <- as.vector(X%%beta)
  piz <- plogis(arg)
  piz <- p.range(piz)
  piy <- 1-omega0-(1-omega0-eta0)*piz
  piy <- p.range(piy)
  (-1+eta0+omega0)*as.vector(t(X)%*(y-piy)/piy/(1-piy)*piz*(1-piz))
}
```

```
library(likelihoodAsy)
```

```
datagen=function(beta,data,eta=eta0,omega=omega0)
{
  out=data
  X <- data$X
  arg <- as.vector(X%%beta)
  piz <- plogis(arg)
  piz <- p.range(piz)
  piy <- 1-omega0-(1-omega0-eta0)*piz
  piy <- p.range(piy)
  out$y <- rbinom(nrow(X),size=1,prob=piy)
  return(out)
}
```

```
set.seed(80)
```

```
n=100
```

```
x1 <- rnorm(n)
```

```

x2 <- rbinom(n,1,0.5)
x3=rep(1,n)
X <- cbind(x1,x2,x3)
beta=c(1,-1,1)
arg=as.vector(X%*%beta)
piz=plogis(arg)
eta0=0.95
omega0=0.85
piy=1-omega0-(1-omega0-eta0)*piz
y=rbinom(n,1,piy)
dat <- list(X=X,y=y)

psifun <- function(theta) theta[1]

obj <- rstar(dat, thetainit=c(1,-1,1), floglik=log_ver,
            fscore=score,psival=1,
            datagen=datagen,fpsi=psifun, ronly=T)

obj$theta.hat
obj$r
pnorm(obj$r)

psifun <- function(theta) theta[2]

obj <- rstar(dat, thetainit=c(1,-1,1), floglik=log_ver,
            fscore=score,psival=-1,
            datagen=datagen,fpsi=psifun, ronly=T)

obj$r
pnorm(obj$r)

```



```
psifun <- function(theta) theta[3]

obj <- rstar(dat, thetainit=c(1,-1,1), floglik=log_ver,
            fscore=score,psival=1,
            datagen=datagen,fpsi=psifun, ronly=T)

obj$r
pnorm(obj$r)

#capitolo 3
set.seed(80)

psifun <- function(theta) theta[1]

obj <- rstar(dat, thetainit=c(1,-1,1), floglik=log_ver,
            fscore=score,psival=1,
            datagen=datagen,fpsi=psifun, R=1000)

obj$rs
pnorm(obj$rs)

psifun <- function(theta) theta[2]

obj <- rstar(dat, thetainit=c(1,-1,1), floglik=log_ver,
            fscore=score,psival=-1,
            datagen=datagen,fpsi=psifun, R=1000)

obj$rs
pnorm(obj$rs)
```

```
psifun <- function(theta) theta[3]

obj <- rstar(dat, thetainit=c(1,-1,1), floglik=log_ver,
            fscore=score,psival=1,
            datagen=datagen,fpsi=psifun, R=1000)

obj$rs
pnorm(obj$rs)

psifun <- function(theta,x0)
{
  arg <- sum(theta*x0)
  plogis(arg)
}

obj2 <- rstar.ci(dat, thetainit=c(1,-1,1),
                floglik=log_ver,fscore=score,
                datagen=datagen,
                fpsi=function(th) psifun(th,x0=c(0,1,1)),
                R=1000)

obj2
plot(obj2)

#capitolo 4
set.seed(80)

psifun <- function(theta) theta[1]

psi0=1
```

```
log_ver_b23=function(b,psi0,data) {
  log_ver(c(psi0,b[1],b[2]),data)
}
```

```
py_change=function(beta,X,eta=eta0,omega=omega0) {
  arg <- as.vector(X%%beta)
  piz=plogis(arg)
  eta=eta0
  omega=omega0
  1-omega-(1-omega-eta)*piz
}
```

```
score_b23=function(beta,data, eta=eta0, omega=omega0) {
  temp=score(beta,data,eta,omega)
  c(temp[2],temp[3])
}
```

```
f_boot_2ordine=function(a=NA) {
  y=rbinom(n,1,piy_boot)
  dat <- list(X=X,y=y)
  mle <- nlminb(obj$theta.hyp,function(th) -log_ver(th,data=dat),
               gradient=function(th) -score(th,data=dat))
  ris=nlminb(objective=function(th) -log_ver_b23(th,psi0=psi0,data=dat),
             start=c(obj$theta.hyp[2],obj$theta.hyp[3])
             ,gradient=function(th) -score_b23(c(psi0,th[1],th[2]),dat)
             )
  xi_hat_0=ris$par
  par=log_ver_b23(xi_hat_0,psi0,dat)
  tot=log_ver(mle$par, dat)
  if(par>tot) {
```

```

mle <- nlminb(c(psi0,xi_hat_0),function(th)
-log_ver(th,data=dat),
            gradient=function(th) -score(th,data=dat))
tot=log_ver(mle$par, dat)
}
W=2*(tot-par)
if(W<0 && W< -abs(tot)/25) cat(tot,par,"2 fuori soglia ")
if(W<0 && W> -abs(tot)/25) {
  cat(tot,log_ver_b23(xi_hat_0,psi0,dat),"2 dentro soglia ")
  W=0
}
sign(mle$par[1]-psi0)*sqrt(W)
}

f_boot_1ordine=function(a=NA) {
  y=rbinom(n,1,piy_boot)
  dat <- list(X=X,y=y)
  mleboot <- nlminb(obj$theta.hat,function(th) -log_ver(th,data=dat),
                  gradient=function(th) -score(th,data=dat))
  ris=nlminb(objective=function(th)
    -log_ver_b23(th,psi0=obj$theta.hat[1],data=dat),
    start=c(mleboot$par[2],mleboot$par[3])
    ,gradient=function(th)
      -score_b23(c(obj$theta.hat[1],th[1],th[2]),dat)
  )
  xi_hat_0=ris$par
  tot=log_ver(mleboot$par, dat)
  par=log_ver_b23(xi_hat_0,obj$theta.hat[1],dat)
  if(par>tot) {
    mleboot <- nlminb(c(obj$theta.hat[1],xi_hat_0),function(th)

```

```

    -log_ver(th,data=dat),
    gradient=function(th) -score(th,data=dat))
  tot=log_ver(mleboot$par, dat)
}
W=2*(tot-par)
if(W<0 && W< -abs(tot)/25) cat(tot,par,"1 fuori soglia ")
if(W<0 && W> -abs(tot)/25) {
  cat(tot,par,"1 dentro soglia ")
  W=0
}
sign(mleboot$par[1]-obj$theta.hat[1])*sqrt(W)
}

obj=rstar(dat, thetainit=beta, floglik=log_ver,
          fscore=score,psival=1,
          datagen=datagen,fpsi=psifun, R=1000,
          psidesc="beta1",trace=F)
obj2 <- rstar.ci(dat, thetainit=mle$par, floglik=log_ver,fscore=score,
                datagen=datagen,fpsi=psifun, R=1000, psidesc="beta1")

par.val <- obj2$psivals

piy_boot=py_change(obj$theta.hat,X)
r_boot_1=sapply(rep(NA, 5000),f_boot_1ordine)

alfa_boot_1=mean(r_boot_1<obj$r)
alfa_boot_1

Jhat <- sqrt(solve(-optimHess(obj$theta.hat,log_ver,data=dat))[1,1])

```

```

c1=obj$theta.hat[1]-2*Jhat
c2=obj$theta.hat[1]+2*Jhat
psi_seq=par.val
alfa_seq=rep(NA,length(par.val))

r_boot=r_boot_1

for(i in 1:length(psi_seq)) {
  print(i)
  r=rstar(dat, thetainit=obj$theta.hat, floglik=log_ver,
          fscore=score,psival=psi_seq[i],
          datagen=datagen,fpsi=psifun, R=1000,
          psidesc="beta1",trace=F,ronly=T)$r
  alfa_seq[i]=mean(r_boot<r)
}

h=cbind(psi_seq, qnorm(p.range(alfa_seq)))

a1=qnorm(0.025)
a2=qnorm(0.975)

while(sum(h[,2]<=a1)==0) {
  c2=c2+Jhat
  psi_seq=alfa_seq=seq(c2-Jhat, c2, length=10)
  for(i in 1:length(psi_seq)) {
    print(i)
    r=rstar(dat, thetainit=obj$theta.hat,
            floglik=log_ver, fscore=score,psival=psi_seq[i],
            datagen=datagen,fpsi=psifun, R=1000,
            psidesc="beta1",trace=F,ronly=T)$r
  }
}

```

```

    alfa_seq[i]=mean(r_boot<r)
  }
  h1=cbind(psi_seq, qnorm(p.range(alfa_seq)))
  h=rbind(h,h1)
}
while(sum(h[,2]>=a2)==0) {
  c1=c1-Jhat
  psi_seq=alfa_seq=seq(c1, c1+Jhat, length=10)
  for(i in 1:length(psi_seq)) {
    print(i)
    psi0=psi_seq[i]
    r=rstar(dat, thetainit=obj$theta.hat, floglik=log_ver,
            fscore=score,psival=psi_seq[i],
            datagen=datagen,fpsi=psifun, R=1000, psidesc="beta1",
            trace=F,ronly=T)$r
    alfa_seq[i]=mean(r_boot<r)
  }
  h1=cbind(psi_seq, qnorm(p.range(alfa_seq)))
  h=rbind(h,h1)
}

library(pspline)
fit1=sm.spline(h[,2], h[,1])
conf.val1=predict(fit1,c(a1,0,a2))
sort(conf.val1)

piy_boot=py_change(obj$theta.hyp,X)
r_boot_2=sapply(rep(NA, 5000),f_boot_2ordine)

alfa_boot_2=mean(r_boot_2<obj$r)

```

```

r.fun2 <- function(data,par0)
{
  mle <- nlminb(rep(0,ncol(data$X)),function(x)
  -log_ver(beta=x,data=data))
  mle0 <- nlminb(rep(0,ncol(data$X)-1),
  function(x) -log_ver(beta=c(par0,x),data=data))
  W <- 2*(mle0$objective-mle$objective)
  if (abs(W)<1e-10) W <- 1e-10
  sign(mle$par[1]-par0)*sqrt(W)
}

sigb1.val <- rep(NA,length(par.val))

library(boot)
library(parallel)
ncpus <- max(1,detectCores()-1)

for (i in 1:length(par.val))
{
  set.seed(123)
  mle0 <- nlminb(rep(0,ncol(dat$X)-1),function(x)
  -log_ver(beta=c(par.val[i],x), data=dat))$par
  mle0 <- c(par.val[i],mle0)
  print(mle0)
  out <- boot(dat, r.fun2, R=4999, sim="parametric",
             ran.gen=function(data,mle) datagen(mle,data),
             mle=mle0,par0=par.val[i],
             parallel="multicore",ncpus=ncpus)
  print(out$t0)
}

```



```
sigb1.val[i] <- mean(out$t <=out$t0,na.rm=TRUE)
}

fit2=sigb1.spline=sm.spline(x=qnorm(sigb1.val),y=par.val)
conf.val2=drop(predict(sigb1.spline,x=qnorm(c(0.975,0.5,0.025))))
sort(conf.val2)

plot(fit1$ysmth,fit1$x,type="o",col=3,ylab="r",xlab="beta1",pty="s")
lines(fit2$ysmth,fit2$x,type="o",col=2)
abline(h=qnorm(0.975),lty=2)
abline(h=-qnorm(0.975),lty=2)
legend(2,2,c("Secondo ord. ","Terzo ord. "),lty=c(1,1),
      lwd=c(2.5,2.5),col=c(2,3))

#capitolo 5
psifun <- function(theta) theta[1]
psi0=1

sim=list(uno=c(100,0.95,0.85),due=c(250,0.9,0.8)
        ,tre=c(250,0.95,0.85))

z=matrix(NA, 4,3000)
j=0
i=1

for(sim_ in sim) {
  set.seed(80)
  n=sim_[1]
  x1 <- rnorm(n)
  x2 <- rbinom(n,1,0.5)
  x3=rep(1,n)
```

```

X <- cbind(x1,x2,x3)
beta=c(1,-1,1)
arg=as.vector(X%*%beta)
piz=plogis(arg)
eta0=sim_[2]
omega0=sim_[3]
piy=1-omega0-(1-omega0-eta0)*piz
for(l in 1:1000) {
  print(i)
  while(is.na(sum(z[,i])))
  {
    j=j+1
    y=rbinom(n,1,piy)
    dat=list(X=X,y=y)
    obj=rstar(dat, thetainit=beta, floglik=log_ver,
              fscore=score,psival=psi0,
              datagen=datagen,fpsi=psifun,
              R=1000,
              psidesc="beta1",trace=F)
    piy_boot=py_change(obj$theta.hyp,X)
    r_boot_2=sapply(rep(NA, 5000),f_boot_2ordine)
    if(sum(is.na(r_boot_2))<25) {
      alfa_boot2=mean(r_boot_2<obj$r,na.rm=T)
    } else alfa_boot2=NA
    piy_boot=py_change(obj$theta.hat,X)
    r_boot_1=sapply(rep(NA, 5000),f_boot_1ordine)
    if(sum(is.na(r_boot_1))<25) {
      alfa_boot1=mean(r_boot_1<obj$r,na.rm=T)
    } else alfa_boot1=NA
    z[,i]=c(alfa_r=pnorm(obj$r),

```

```

        alfa_rstar=pnorm(obj$rs),
        alfa_boot_1=alfa_boot1,
        alfa_boot_2=alfa_boot2)
    }
    i=i+1
  }
}
o=c(0.01,0.025,0.05,0.1,0.5,0.9,0.95,0.975,0.99)
i=c("r","rs","boot. secondo ord.,""boot. terzo ord.")
ii=c(0.90,0.95,0.99)
m=matrix(NA,4,5)
m[1,2:5]=i
m[2:4,1]=ii
m=rbind(m,m,m)
par(pty="s")
test=matrix(NA,4,3)
from=1
to=1000
d=matrix(NA,5,length(o)+1)
d[1,1:length(o)+1]=as.numeric(o)
d[2:5,1]=i
d=rbind(d,d,d)
j=1=2
for(i in 1:3) {
  m[1:(1+2),2]=sapply(ii,function(x) (mean(z[1,from:to]>1-x))*100)
  m[1:(1+2),3]=sapply(ii,function(x) (mean(z[2,from:to]>1-x))*100)
  m[1:(1+2),4]=sapply(ii,function(x) (mean(z[3,from:to]>1-x))*100)
  m[1:(1+2),5]=sapply(ii,function(x) (mean(z[4,from:to]>1-x))*100)
  l=1+4
  d[j,2:10]=sapply(o,function(x) mean(z[1,from:to]<=x))
}

```

```

j=j+1
d[j,2:10]=sapply(o,function(x) mean(z[2,from:to]<=x))
j=j+1
d[j,2:10]=sapply(o,function(x) mean(z[3,from:to]<=x))
j=j+1
d[j,2:10]=sapply(o,function(x) mean(z[4,from:to]<=x))
j=j+2
plot(((1:length(z[1,from:to])-0.5)/
      length(z[1,from:to])),sort(z[1,from:to]),
      xlab="quantili teorici",ylab="quantili empirici")
abline(0,1,col=2)
test[1,i]=ks.test(z[1,from:to],"punif")$p.value
plot(((1:length(z[2,from:to])-0.5)/
      length(z[2,from:to])),sort(z[2,from:to]),
      xlab="quantili teorici",ylab="quantili empirici")
abline(0,1,col=2)
test[2,i]=ks.test(z[2,from:to],"punif")$p.value
plot(((1:length(z[3,from:to])-0.5)/
      length(z[3,from:to])),sort(z[3,from:to]),
      xlab="quantili teorici",ylab="quantili empirici")
abline(0,1,col=2)
test[3,i]=ks.test(z[3,from:to],"punif")$p.value
plot(((1:length(z[4,from:to])-0.5)/
      length(z[4,from:to])),sort(z[4,from:to]),
      xlab="quantili teorici",ylab="quantili empirici")
abline(0,1,col=2)
test[4,i]=ks.test(z[4,from:to],"punif")$p.value
from=from+1000
to=to+1000
}

```

```

psifun <- function(theta) theta[2]

psi0=-1
log_ver_b13=function(b,psi0,data) {
  log_ver(c(b[1],psi0,b[2]),data)
}

score_b13=function(beta,data, eta=eta0, omega=omega0) {
  temp=score(beta,data,eta,omega)
  c(temp[1],temp[3])
}

f_boot_2ordine=function(a=NA) {
  y=rbinom(n,1,piy_boot)
  dat <- list(X=X,y=y)
  mle <- nlminb(obj$theta.hyp,function(th) -log_ver(th,data=dat),
               gradient=function(th) -score(th,data=dat))
  ris=nlminb(objective=function(th)
    -log_ver_b13(th,psi0=psi0,data=dat),
    start=c(obj$theta.hyp[1],obj$theta.hyp[3])
    ,gradient=function(th)
      -score_b13(c(th[1],psi0,th[2]),dat)
  )
  xi_hat_0=ris$par
  par=log_ver_b13(xi_hat_0,psi0,dat)
  tot=log_ver(mle$par, dat)
  if(par>tot) {
    mle <- nlminb(c(xi_hat_0[1],psi0,xi_hat_0[2]),function(th)
      -log_ver(th,data=dat),

```

```

        gradient=function(th) -score(th,data=dat))
    tot=log_ver(mle$par, dat)
}
W=2*(tot-par)
if(W<0 && W< -abs(tot)/25) cat(tot,par,"2 fuori soglia ")
if(W<0 && W> -abs(tot)/25) {
    cat(tot,par,"2 dentro soglia ")
    W=0
}
sign(mle$par[2]-psi0)*sqrt(W)
}

f_boot_1ordine=function(a=NA) {
    y=rbinom(n,1,piy_boot)
    dat <- list(X=X,y=y)
    mleboot <- nlminb(obj$theta.hat,function(th)
        -log_ver(th,data=dat),
        gradient=function(th) -score(th,data=dat))
    ris=nlminb(objective=function(th)
        -log_ver_b13(th,psi0=obj$theta.hat[2],data=dat),
        start=c(mleboot$par[1],mleboot$par[3])
        ,gradient=function(th)
            -score_b13(c(th[1],obj$theta.hat[2],th[2]),dat)
    )
    xi_hat_0=ris$par
    tot=log_ver(mleboot$par, dat)
    par=log_ver_b13(xi_hat_0,obj$theta.hat[2],dat)
    if(par>tot) {
        mleboot <- nlminb(c(xi_hat_0[1],obj$theta.hat[2],xi_hat_0[2]),
            function(th) -log_ver(th,data=dat),

```

```
      gradient=function(th) -score(th,data=dat))
    tot=log_ver(mleboot$par, dat)
  }
  W=2*(tot-par)
  if(W<0 && W< -abs(tot)/25) cat(tot,par,"1 fuori soglia ")
  if(W<0 && W> -abs(tot)/25) {
    cat(tot,par,"1 dentro soglia ")
    W=0
  }
  sign(mleboot$par[2]-obj$theta.hat[2])*sqrt(W)
}

z=matrix(NA, 4,3000)
j=0
i=1

for(sim_ in sim) {
  set.seed(80)
  n=sim_[1]
  x1 <- rnorm(n)
  x2 <- rbinom(n,1,0.5)
  x3=rep(1,n)
  X <- cbind(x1,x2,x3)
  beta=c(1,-1,1)
  arg=as.vector(X%*%beta)
  piz=plogis(arg)
  eta0=sim_[2]
  omega0=sim_[3]
  piy=1-omega0-(1-omega0-eta0)*piz
  for(l in 1:1000) {
```

```

print(i)
while(is.na(sum(z[,i])))
{
  j=j+1
  y=rbinom(n,1,piy)
  dat=list(X=X,y=y)
  obj=rstar(dat, thetainit=beta, floglik=log_ver,
            fscore=score,psival=psi0,
            datagen=datagen,fpsi=psifun, R=1000,
            psidesc="beta2",trace=F)
  piy_boot=py_change(obj$theta.hyp,X)
  r_boot_2=sapply(rep(NA, 5000),f_boot_2ordine)
  if(sum(is.na(r_boot_2))<25) {
    alfa_boot2=mean(r_boot_2<obj$r,na.rm=T)
  } else alfa_boot2=NA
  piy_boot=py_change(obj$theta.hat,X)
  r_boot_1=sapply(rep(NA, 5000),f_boot_1ordine)
  if(sum(is.na(r_boot_1))<25) {
    alfa_boot1=mean(r_boot_1<obj$r,na.rm=T)
  } else alfa_boot1=NA
  z[,i]=c(alfa_r=pnorm(obj$r),
           alfa_rstar=pnorm(obj$rs),
           alfa_boot_1=alfa_boot1,
           alfa_boot_2=alfa_boot2)
}
i=i+1
}
}

```

```
o=c(0.01,0.025,0.05,0.1,0.5,0.9,0.95,0.975,0.99)
```



```

i=c("r","rs","boot. secondo ord.,""boot. terzo ord.")
ii=c(0.90,0.95,0.99)
m=matrix(NA,4,5)
m[1,2:5]=i
m[2:4,1]=ii
m=rbind(m,m,m)
par(pty="s")
test=matrix(NA,4,3)
from=1
to=1000
d=matrix(NA,5,length(o)+1)
d[1,1:length(o)+1]=as.numeric(o)
d[2:5,1]=i
d=rbind(d,d,d)
j=1=2
for(i in 1:3) {
  m[1:(1+2),2]=sapply(ii,function(x) (mean(z[1,from:to]>1-x))*100)
  m[1:(1+2),3]=sapply(ii,function(x) (mean(z[2,from:to]>1-x))*100)
  m[1:(1+2),4]=sapply(ii,function(x) (mean(z[3,from:to]>1-x))*100)
  m[1:(1+2),5]=sapply(ii,function(x) (mean(z[4,from:to]>1-x))*100)
  l=1+4
  d[j,2:10]=sapply(o,function(x) mean(z[1,from:to]<=x))
  j=j+1
  d[j,2:10]=sapply(o,function(x) mean(z[2,from:to]<=x))
  j=j+1
  d[j,2:10]=sapply(o,function(x) mean(z[3,from:to]<=x))
  j=j+1
  d[j,2:10]=sapply(o,function(x) mean(z[4,from:to]<=x))
  j=j+2
  plot(((1:length(z[1,from:to]))-0.5)/

```

```

        length(z[1,from:to])),sort(z[1,from:to]),
        xlab="quantili teorici",ylab="quantili empirici")
abline(0,1,col=2)
test[1,i]=ks.test(z[1,from:to],"punif")$p.value
plot(((1:length(z[2,from:to])-0.5)/
        length(z[2,from:to])),sort(z[2,from:to]),
        xlab="quantili teorici",ylab="quantili empirici")
abline(0,1,col=2)
test[2,i]=ks.test(z[2,from:to],"punif")$p.value
plot(((1:length(z[3,from:to])-0.5)/
        length(z[3,from:to])),sort(z[3,from:to]),
        xlab="quantili teorici",ylab="quantili empirici")
abline(0,1,col=2)
test[3,i]=ks.test(z[3,from:to],"punif")$p.value
plot(((1:length(z[4,from:to])-0.5)/
        length(z[4,from:to])),sort(z[4,from:to]),
        xlab="quantili teorici",ylab="quantili empirici")
abline(0,1,col=2)
test[4,i]=ks.test(z[4,from:to],"punif")$p.value
from=from+1000
to=to+1000
}

psifun <- function(theta) theta[3]

psi0=1
log_ver_b12=function(b,psi0,data) {
  log_ver(c(b[1],b[2],psi0),data)
}

```

```

score_b12=function(beta,data, eta=eta0, omega=omega0) {
  temp=score(beta,data,eta,omega)
  c(temp[1],temp[2])
}

f_boot_2ordine=function(a=NA) {
  y=rbinom(n,1,piy_boot)
  dat <- list(X=X,y=y)
  mle <- nlminb(obj$theta.hyp,function(th)
    -log_ver(th,data=dat),
    gradient=function(th) -score(th,data=dat))
  ris=nlminb(objective=function(th)
    -log_ver_b12(th,psi0=psi0,data=dat),
    start=c(obj$theta.hyp[1],obj$theta.hyp[2])
    ,gradient=function(th)
      -score_b12(c(th[1],th[2],psi0),dat)
  )
  xi_hat_0=ris$par
  par=log_ver_b12(xi_hat_0,psi0,dat)
  tot=log_ver(mle$par, dat)
  if(par>tot) {
    mle <- nlminb(c(xi_hat_0,psi0),function(th)
      -log_ver(th,data=dat),
      gradient=function(th) -score(th,data=dat))
    tot=log_ver(mle$par, dat)
  }
  W=2*(tot-par)
  if(W<0 && W< -abs(tot)/25) cat(tot,par,"2 fuori soglia ")
  if(W<0 && W> -abs(tot)/25) {
    cat(tot,par,"2 dentro soglia ")
  }
}

```

```

    W=0
  }
  sign(mle$par[3]-psi0)*sqrt(W)
}

f_boot_1ordine=function(a=NA) {
  y=rbinom(n,1,piy_boot)
  dat <- list(X=X,y=y)
  mleboot <- nlminb(obj$theta.hat,function(th)
    -log_ver(th,data=dat),
    gradient=function(th) -score(th,data=dat))
  ris=nlminb(objective=function(th)
    -log_ver_b12(th,psi0=obj$theta.hat[3],data=dat),
    start=c(mleboot$par[1],mleboot$par[2])
    ,gradient=function(th)
      -score_b12(c(th[1],th[2],obj$theta.hat[3]),dat)
  )
  xi_hat_0=ris$par
  tot=log_ver(mleboot$par, dat)
  par=log_ver_b12(xi_hat_0,obj$theta.hat[3],dat)
  if(par>tot) {
    mleboot <- nlminb(c(xi_hat_0,obj$theta.hat[3]),
      function(th) -log_ver(th,data=dat),
      gradient=function(th) -score(th,data=dat))
    tot=log_ver(mleboot$par, dat)
  }
  W=2*(tot-par)
  if(W<0 && W< -abs(tot)/25) cat(tot,par,"1 fuori soglia ")
  if(W<0 && W> -abs(tot)/25) {
    cat(tot,par,"1 dentro soglia ")
  }
}

```

```
      W=0
    }
    sign(mleboot$par[3]-obj$theta.hat[3])*sqrt(W)
  }

z=matrix(NA, 4,3000)
j=0
i=1

for(sim_ in sim) {
  set.seed(80)
  n=sim_[1]
  x1 <- rnorm(n)
  x2 <- rbinom(n,1,0.5)
  x3=rep(1,n)
  X <- cbind(x1,x2,x3)
  beta=c(1,-1,1)
  arg=as.vector(X%%beta)
  piz=plogis(arg)
  eta0=sim_[2]
  omega0=sim_[3]
  piy=1-omega0-(1-omega0-eta0)*piz
  for(l in 1:1000) {
    print(i)
    while(is.na(sum(z[,i])))
    {
      j=j+1
      y=rbinom(n,1,piy)
      dat=list(X=X,y=y)
      obj=rstar(dat, thetainit=beta, floglik=log_ver,
```

```

        fscore=score,psival=psi0,
        datagen=datagen,fpsi=psifun, R=1000,
        psidesc="beta3",trace=F)
    piy_boot=py_change(obj$theta.hyp,X)
    r_boot_2=sapply(rep(NA, 5000),f_boot_2ordine)
    if(sum(is.na(r_boot_2))<25) {
        alfa_boot2=mean(r_boot_2<obj$r,na.rm=T)
    } else alfa_boot2=NA
    piy_boot=py_change(obj$theta.hat,X)
    r_boot_1=sapply(rep(NA, 5000),f_boot_1ordine)
    if(sum(is.na(r_boot_1))<25) {
        alfa_boot1=mean(r_boot_1<obj$r,na.rm=T)
    } else alfa_boot1=NA
    z[,i]=c(alfa_r=pnorm(obj$r),
            alfa_rstar=pnorm(obj$rs),
            alfa_boot_1=alfa_boot1,
            alfa_boot_2=alfa_boot2)
    }
    i=i+1
}
}

o=c(0.01,0.025,0.05,0.1,0.5,0.9,0.95,0.975,0.99)
i=c("r","rs","boot. secondo ord.,""boot. terzo ord.")
ii=c(0.90,0.95,0.99)
m=matrix(NA,4,5)
m[1,2:5]=i
m[2:4,1]=ii
m=rbind(m,m,m)
par(pty="s")

```

```

test=matrix(NA,4,3)
from=1
to=1000
d=matrix(NA,5,length(o)+1)
d[1,1:length(o)+1]=as.numeric(o)
d[2:5,1]=i
d=rbind(d,d,d)
j=1=2
for(i in 1:3) {
  m[1:(1+2),2]=sapply(ii,function(x) (mean(z[1,from:to]>1-x))*100)
  m[1:(1+2),3]=sapply(ii,function(x) (mean(z[2,from:to]>1-x))*100)
  m[1:(1+2),4]=sapply(ii,function(x) (mean(z[3,from:to]>1-x))*100)
  m[1:(1+2),5]=sapply(ii,function(x) (mean(z[4,from:to]>1-x))*100)
  l=1+4
  d[j,2:10]=sapply(o,function(x) mean(z[1,from:to]<=x))
  j=j+1
  d[j,2:10]=sapply(o,function(x) mean(z[2,from:to]<=x))
  j=j+1
  d[j,2:10]=sapply(o,function(x) mean(z[3,from:to]<=x))
  j=j+1
  d[j,2:10]=sapply(o,function(x) mean(z[4,from:to]<=x))
  j=j+2
  plot(((1:length(z[1,from:to])-0.5)/
        length(z[1,from:to])),sort(z[1,from:to]),
        xlab="quantili teorici",ylab="quantili empirici")
  abline(0,1,col=2)
  test[1,i]=ks.test(z[1,from:to],"punif")$p.value
  plot(((1:length(z[2,from:to])-0.5)/
        length(z[2,from:to])),sort(z[2,from:to]),
        xlab="quantili teorici",ylab="quantili empirici")

```

```
abline(0,1,col=2)
test[2,i]=ks.test(z[2,from:to],"punif")$p.value
plot(((1:length(z[3,from:to])-0.5)/
      length(z[3,from:to])),sort(z[3,from:to]),
      xlab="quantili teorici",ylab="quantili empirici")
abline(0,1,col=2)
test[3,i]=ks.test(z[3,from:to],"punif")$p.value
plot(((1:length(z[4,from:to])-0.5)/
      length(z[4,from:to])),sort(z[4,from:to]),
      xlab="quantili teorici",ylab="quantili empirici")
abline(0,1,col=2)
test[4,i]=ks.test(z[4,from:to],"punif")$p.value
from=from+1000
to=to+1000
}
```





# Bibliografia

- [1] A. Azzalini. *Inferenza Statistica. Un'Introduzione Basata sul Concetto di Verosimiglianza*. Springer Verlag, 2 edition, 2001.
- [2] Barndorff-Nielsen. On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, (70 (2): 343-365), 1983.
- [3] Ruggero Bellio and Donald Pierce. *likelihoodAsy: Functions for Likelihood Asymptotics*, 2015. R package version 0.40.
- [4] R. Bellio D. A. Pierce. Modern likelihood-frequentist inference. *Sottoposto per pubblicazione*, 2015.
- [5] A. Salvan L. Pace. *Teoria della Statistica. Metodi, modelli, approssimazioni asintotiche*. CEDAM, 1996.
- [6] A. Salvan L. Pace. *Introduzione alla Statistica II. Inferenza, verosimiglianza, modelli*. CEDAM, 2001.
- [7] L. Giacomelli M. Bertsch, R. Dal Passo. *Analisi Matematica*. McGraw-Hill, 2007.
- [8] I. A. Gardner D. Cowling P. McInturff, W. O. Johnson. Modelling risk when binary outcomes are subject to error. *Statistics in Medicine*, (23, 1095-1109), 2004.
- [9] T. A. Severini. *Likelihood Methods in Statistics*. Oxford University Press, 2000.

- [10] M.A. Martin T. J. DiCiccio. Simple modifications for signed roots of likelihood ratio statistics. *J.R. Statist. Soc.*, (55, 305-316), 1993.
- [11] S. E Stern T. J. DiCiccio, M. A. Martin. Simple and accurate one-sided inference from signed roots of likelihood ratios. *Can. J. Stat.*, (29, 67-76), 2001.