

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE
CORSO DI LAUREA MAGISTRALE IN
SCIENZE STATISTICHE



Metodi di verosimiglianza di ordine
superiore per modelli Stress-Strength

Relatore:

Prof. Nicola Sartori

Dipartimento di Scienze Statistiche

Laureando:

Vito Fanelli

Matricola N. 1058877

A.A 2014-2015

Alla mia famiglia

Indice

Introduzione	1
1 I modelli Stress-Strength	3
1.1 Che cosa sono	3
1.2 Assunzioni e formulazioni matematiche	4
1.3 Un po' di storia	6
1.4 Problemi	6
2 L' inferenza di verosimiglianza	9
2.1 Introduzione	9
2.2 Il modello statistico parametrico	10
2.3 Verosimiglianza statistica e quantità collegate	12
2.3.1 La funzione di verosimiglianza	12
2.3.2 Statistiche sufficienti	13
2.3.3 Stima di massima verosimiglianza	14
2.3.4 Statistiche ancillari	16
2.4 Invarianza rispetto a riparametrizzazioni	16
2.5 Parametri di disturbo e verosimiglianza profilo	17
2.6 Problemi e procedure inferenziali basate sulla verosimiglianza .	20
2.6.1 Verifica di ipotesi	20
2.6.2 Stima intervallare	24
2.7 Approssimazioni di ordine superiore	27
2.7.1 Integrazione Monte Carlo	30

3	Esempi di modelli Stress-Strength	35
3.1	Distribuzione Esponenziale	35
3.2	Distribuzione Normale	36
3.2.1	Normali indipendenti con varianze uguali	36
3.2.2	Normali indipendenti con varianze diverse	36
3.2.3	Normali dipendenti con varianze diverse	37
3.3	Distribuzione Weibull	38
3.4	Distribuzioni normali multivariate	38
3.5	Variabili X e Y dipendenti da covariate	39
3.5.1	Caso normale	39
4	Studi di simulazione	41
4.1	Introduzione	41
4.2	Esponenziali	42
4.3	Normali indipendenti con varianze uguali	46
4.4	Normali indipendenti con varianze diverse	50
4.5	Normali dipendenti	54
4.6	Weibull	58
5	Un caso reale: confronto tra PET/CT e PET/RM su pazienti pediatrici	63
5.1	PET/CT e PET/RM	63
5.2	Il campione	65
5.3	L'analisi	66
	Conclusioni	69
	Bibliografia	71
	A Codice R	75
	Ringraziamenti	91

Introduzione

L'importanza dei metodi legati alla verosimiglianza e basati sulle approssimazioni di ordine superiore (al primo) è cresciuta sempre di più nella letteratura recente e nelle applicazioni pratiche. Molte procedure inferenziali di verosimiglianza si basano su approssimazioni, dette del primo ordine, con le quali, attraverso l'applicazione del teorema del limite centrale, si afferma che la statistica di interesse si distribuisce approssimativamente come una normale, con media e varianza stimabili dai dati. Naturalmente tale approssimazione è soggetta ad un certo ordine di errore. Nell'ultimo quarto di secolo si è sempre più sviluppata la teoria asintotica di ordine superiore che, applicata all'inferenza tramite verosimiglianza, consente di effettuare analisi più precise e ottenere risultati più accurati, riducendo il grado di errore delle approssimazioni. Le approssimazioni di ordine superiore riguardano la funzione di verosimiglianza e le quantità ad essa collegate, quali le statistiche test o i p-value. In particolare, in questa tesi verrà discussa e utilizzata la teoria asintotica di ordine superiore nell'ambito dei *modelli stress-strength*, allo scopo di effettuare una migliore inferenza sul parametro di interesse $\psi = P(X < Y)$, dove X e Y sono due variabili casuali indipendenti. Tale parametro rappresenta, solitamente, una misura di affidabilità, ad esempio in applicazioni ingegneristiche, o una semplice misura di differenza tra due distribuzioni di probabilità. Nel primo caso, si pensi ad un componente meccanico o elettronico, all'interno di un sistema più complesso, che viene sottoposto ad un certo stress (X). Lo scopo dei modelli in esame è stimare la probabilità che la forza del componente (Y) sia maggiore dello stress a cui è sottoposto, in modo tale che esso potrà continuare a funzionare; viceversa il componente si romperà. Nel secondo caso, si immagini di osservare due

distribuzioni di probabilità originate da dati provenienti da due diversi trattamenti terapeutici, che per comodità indicheremo con X e Y . Lo scopo è quello di capire quale dei due trattamenti è più efficace, ad esempio in termini di tempo. Allora ψ rappresenterà una misura di differenza tra la distribuzione di X e la distribuzione di Y , e tanto maggiore sarà tale differenza tanto più efficace sarà il trattamento X rispetto al trattamento Y . Nel nostro caso l'attenzione sarà rivolta alla statistica test del test rapporto di verosimiglianza per verifiche di ipotesi unidirezionali e l'obiettivo di questa tesi è mostrare come la teoria asintotica di ordine superiore consenta di avere un'inferenza più accurata rispetto alla teoria classica.

Nel primo capitolo verrà fornita una descrizione formale dei modelli stress-strength, con una panoramica sulle varie estensioni e sui problemi che ne possono scaturire. Nel secondo capitolo si parlerà dell'inferenza tramite verosimiglianza, fornendo i principali concetti ad essa legati utili per il proseguo della lettura, e delle approssimazioni di ordine superiore. Inoltre sarà data una breve descrizione dei metodi di *Integrazione Monte Carlo*, la cui utilità sarà chiara più avanti. Nel capitolo 3 verranno esaminati i modelli stress-strength per le distribuzioni più comuni quali esponenziale, normale e Weibull, con vari sottocasi. Nel capitolo 4 saranno presentati i risultati delle simulazioni condotte per valutare la differenza tra la teoria asintotica di ordine superiore e la teoria del primo ordine, per l'inferenza su ψ . In particolare, verrà valutata la precisione degli intervalli di confidenza, di tipo Wald e quelli basati su r_p e r_p^* , in termini di copertura empirica, simulando quello che è uno dei principi cardine dell'inferenza statistica ovvero il *principio del campionamento ripetuto*. Nel quinto ed ultimo capitolo sarà esaminata un'applicazione su dati reali dei modelli stress-strength utilizzando le procedure viste in precedenza e confrontandole.

Il software utilizzato per le simulazioni e le analisi del caso pratico è R 3.1.0, liberamente scaricabile all'indirizzo <http://www.r-project.org> alla sezione *Download*. In particolare è stata utilizzata la libreria `LikelihoodAsy`, che fa uso dei metodi di integrazione Monte Carlo per calcolare r_p^* , versione modificata di r_p , specificando solamente la funzione di log-verosimiglianza del modello, una funzione per generare i dati e la funzione ψ .

Capitolo 1

I modelli Stress-Strength

1.1 Che cosa sono

I modelli Stress-Strength sono modelli statistici utilizzati per stimare la probabilità che una certa distribuzione X sia minore di una certa distribuzione Y . Una variabile Y è stocasticamente maggiore di una variabile X se la funzione di ripartizione di Y non è mai maggiore di quella di X .

In altri termini, $F_Y(t) \leq F_X(t)$ per ogni t . Da qui ne consegue che

$$P(X < Y) \geq \frac{1}{2}$$

poichè,

$$P(X < Y) = \int_{-\infty}^{+\infty} F_X(t) dF_Y(t) \geq \int_{-\infty}^{+\infty} F_Y(t) dF_Y(t) = \frac{1}{2}.$$

Si può pensare che la variabile X rappresenti uno "stress" a cui un particolare componente, all'interno di un sistema più complesso, avente "forza" Y è sottoposto. Se lo stress eccede la forza il componente si rompe, viceversa continua a funzionare. Tali valutazioni dunque riguardano l'affidabilità dei componenti, ad esempio in ambito ingegneristico. Da ciò deriva il nome dei modelli considerati e la definizione di affidabilità come probabilità di non fallimento:

$$R = P(X < Y). \quad (1.1)$$

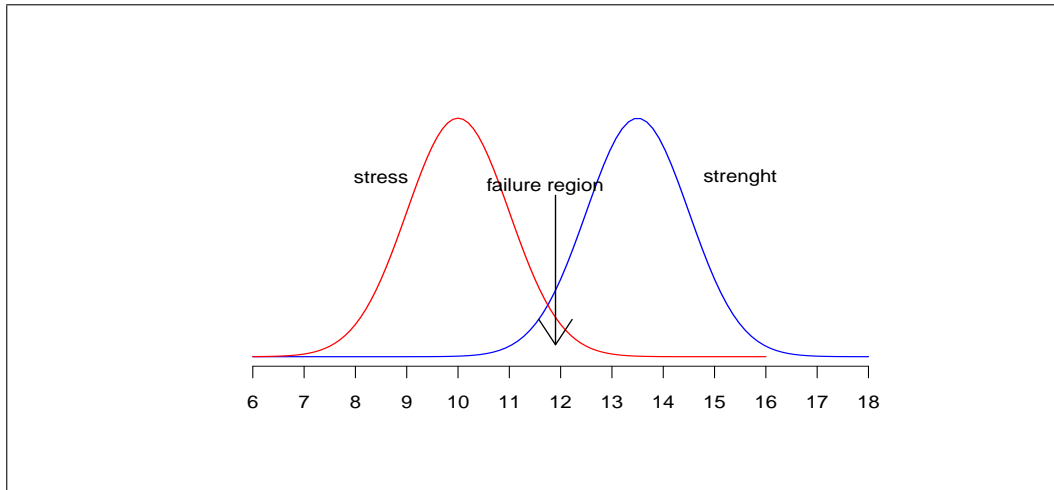


Figura 1.1: Distribuzioni di stress e forza, e regione di fallimento

In realtà questi modelli possono essere usati anche in altri contesti, quali la biostatistica, come misura di differenza tra due distribuzioni: una relativa ai soggetti sottoposti ad un certo trattamento (X) e l'altra relativa ai soggetti non trattati o sottoposti ad un altro trattamento (Y). Chiaramente più grande sarà il valore della (1.1) maggiore sarà la differenza tra la distribuzione X e la distribuzione Y e minore sarà l'area della regione di fallimento, indicata nella Figura 1.1.

1.2 Assunzioni e formulazioni matematiche

L'assunzione principale che vale per un modello stress-strength è che X e Y siano due distribuzioni indipendenti e appartenenti alla stessa famiglia (Normale, Esponenziale, Weibull,...). Proprio la possibilità di stimare la (1.1) sotto diverse ipotesi distributive, non solo quella di normalità, rende preferibile l'utilizzo di questi modelli rispetto al classico test *t di Student*, che al contrario richiede che l'assunzione di normalità sia verificata.

È anche possibile estendere tali modelli a casi più realistici, ad esempio casi in

cui è ipotizzabile una relazione di dipendenza tra X e Y , o in cui X e Y sono distribuzioni multivariate, o ancora quando X e Y dipendono da covariate.

La quantità (1.1) non è l'unica quantità di interesse. Questa può rappresentare l'affidabilità di un singolo componente all'interno di un sistema di k componenti. In tal caso, se il sistema è in *parallelo* esso continuerà a funzionare finché almeno uno dei k componenti funzionerà; se invece il sistema è in *serie* allora funzionerà solo se tutti i componenti che lo costituiscono funzioneranno. L'affidabilità di un sistema in parallelo o in serie può essere espressa, rispettivamente, come:

$$P(X < \max(Y_1, \dots, Y_k)), \quad (1.2)$$

$$P(X < \min(Y_1, \dots, Y_k)). \quad (1.3)$$

Un'altra situazione abbastanza realistica prevede che la forza Y non sia solo maggiore di uno stress X ma sia anche minore di un ulteriore stress Z . Si può pensare ad esempio alla pressione sanguigna, che ha due limiti (sistolica e diastolica), oppure ad alcuni componenti che lavorano solo a temperature comprese in un certo range. In questo caso la quantità oggetto di studio potrebbe essere

$$P(X < Y < Z). \quad (1.4)$$

Infine si può immaginare che X e Y siano vettori indipendenti di variabili casuali e che la quantità di interesse sia la probabilità che una combinazione lineare di X sia minore di una combinazione lineare di Y . Un esempio può essere un ponte, la cui forza è rappresentata da supporti speciali o altri componenti e che è sottoposto a diversi tipi di stress, ovvero a diversi pesi (automobili, bus, camion, moto,...). Allora l'affidabilità della struttura sarà

$$P(A^T X < B^T Y), \quad (1.5)$$

dove A e B sono vettori contenenti i coefficienti delle combinazioni lineari di X e di Y .

1.3 Un po' di storia

I modelli stress-strength sono nati intorno alla metà del XX secolo in un contesto non parametrico. I primi ad utilizzarli sono stati Wilcoxon (1945), Mann e Whitney (1947) allo scopo di confrontare due variabili casuali X e Y che descrivevano i risultati derivanti da due trattamenti. La loro analisi si poggiava su una statistica basata sui ranghi delle osservazioni di X e Y nel campione congiunto. Essi, inoltre, furono i primi ad evidenziare la relazione:

$$F_X = F_Y \Rightarrow P(X < Y) = \frac{1}{2}.$$

I metodi non parametrici erano robusti nel senso che non richiedevano assunzioni sulla distribuzione di X e di Y ma risultavano inefficienti nelle applicazioni pratiche.

Il primo tentativo di studiare la (1.1) da un punto di vista parametrico avvenne negli anni '60 con Owen (1964), il quale costruì gli intervalli di confidenza per $P(X < Y)$ con X e Y indipendenti e normalmente distribuite. Alla fine degli anni '70 la stima della (1.1) era disponibile per la maggior parte delle distribuzioni, quali esponenziale, normale, Pareto, sotto l'assunzione di indipendenza tra X e Y . Successivamente intorno agli anni '80 gli sforzi furono indirizzati verso lo studio di situazioni più realistiche che prevedevano una dipendenza tra X e Y e con X e Y distribuzioni bivariate. Sempre in quegli anni Bilikam (1985) accennò ai *time-dependent model* per X e Y , ossia a modelli che dipendevano dal tempo. Alla fine degli anni '80 i modelli stress-strength furono estesi ai dati categoriali e a X e Y che dipendevano da variabili esplicative. Negli anni '90 e 2000 la stima della (1.1) fu estesa a nuove distribuzioni come gaussiana inversa, normale asimmetrica, esponenziale multivariata, gamma generalizzata.

1.4 Problemi

I modelli stress-strength presentano alcuni inconvenienti, alcuni legati ad aspetti concettuali ed altri riguardanti questioni tecniche. In ogni caso sono

state trovate delle soluzioni per superare queste difficoltà. Uno dei problemi concettuali riguarda l'assunzione di indipendenza tra X e Y , di cui si è parlato prima. Secondo alcuni tale assunzione non è realistica e concettualmente sbagliata poiché la forza è necessariamente condizionata allo stress, in quanto la sua realizzazione fisica si manifesta solo quando lo stress viene applicato. Tuttavia si è già accennato della possibile estensione dei modelli ai casi in cui X e Y sono dipendenti. Inoltre è possibile che stress e forza cambino nel tempo, ovvero siano caratterizzati da parametri dipendenti dal tempo. Gli studi in questo senso sono ancora rari e tale problema costituisce una base di partenza per ricerche future.

Un altro problema è di natura inferenziale. I modelli in esame sono caratterizzati da un parametro di interesse, ψ , e da uno o più parametri di disturbo. Per molti dispositivi è richiesta un'affidabilità quasi pari ad 1 per poter essere funzionali, per cui è necessaria una stima molto accurata di ψ , ottenibile solo con una numerosità campionaria sufficientemente elevata. Da notare che spesso in letteratura il parametro di interesse è indicato con R , da *Reliability* ovvero affidabilità. L'inferenza tramite verosimiglianza basata sulle approssimazioni del primo ordine mostra delle difficoltà nei casi in cui la numerosità campionaria è bassa ed i parametri di disturbo sono molti, in quanto si cade nel classico problema statistico della difficoltà di stima in presenza di molti parametri e pochi dati. A tal proposito si dimostra - e sarà l'obiettivo di questa tesi - che l'inferenza tramite verosimiglianza basata sulle approssimazioni di ordine superiore fornisce delle stime (puntuali ed intervallari) più accurate e risultati più affidabili specialmente nei casi sopracitati.

Capitolo 2

L' inferenza di verosimiglianza

2.1 Introduzione

Nel capitolo precedente si è accennato all' inferenza tramite verosimiglianza. Uno degli scopi principali che la statistica si propone di raggiungere è quello di spiegare un fenomeno osservato su una popolazione. Talvolta però, a causa di diversi motivi, difficilmente il fenomeno che vogliamo spiegare è osservabile sull'intera popolazione di riferimento. Per questo motivo è necessario misurare il fenomeno su un campione, cioè su una parte della popolazione, per poi estendere i risultati ad essa. L'operazione mediante la quale le informazioni rilevate sul campione vengono generalizzate è detta *inferenza statistica*. In particolare in questa tesi, ci serviremo delle procedure inferenziali basate sulla verosimiglianza, introdotta da Fisher (1922), quali stime puntuali, intervalli di confidenza e verifiche di ipotesi.

Per capire meglio cosa si intende per inferenza di verosimiglianza e perchè ci aiuta nello studio ed applicazione dei modelli stress-strength di seguito verranno illustrate alcune definizioni, proprietà e concetti legati alla verosimiglianza. Per prima cosa verrà data una definizione di modello statistico parametrico; a seguire introdurremo il concetto di verosimiglianza, di alcune quantità ad essa collegate e proprietà di cui ci serviremo, come l' invarianza rispetto a riparametrizzazioni. Successivamente parleremo di inferenza statistica in presenza di parametri di disturbo - che ha un ruolo centrale

in questa tesi - e di verosimiglianza profilo. Infine analizzeremo i problemi tipici dell' inferenza come le stime e le verifiche di ipotesi, dando una breve rassegna dei test utilizzati. Tutti questi temi rientrano nella teoria classica della verosimiglianza, ovvero di quella basata sulle approssimazioni del primo ordine.

Nell' ultima parte del capitolo verranno introdotte delle modifiche ad alcune quantità connesse alla verosimiglianza viste precedentemente, basate sulle cosiddette approssimazioni di ordine superiore - che giocano un ruolo cruciale nell' applicazione dei modelli stress-strength, in quanto permettono di ottenere stime più accurate, specialmente in caso di piccoli campioni. Al termine verrà spiegato brevemente il funzionamento della libreria `LikelihoodAsy` utilizzata dal software *R* per le successive analisi inferenziali e i metodi di integrazione *Monte Carlo*, di cui si serve la libreria sopracitata.

2.2 Il modello statistico parametrico

Si è detto che l' obiettivo dell' inferenza statistica è spiegare un fenomeno misurato da una variabile casuale Y a partire da un campione di osservazioni empiriche $y = (y_1, \dots, y_n)$. In altri termini lo scopo è quello di descrivere la variabilità all' interno dei dati osservati attraverso un modello, ovvero individuando la distribuzione teorica (ignota) di Y , sulla base delle sue realizzazioni. Per fare ciò occorre innanzitutto *specificare il modello statistico*, se possibile correttamente. A tal proposito definiamo il **modello statistico** \mathcal{F} come una collezione di distribuzioni di probabilità, compatibili con Y , e che contiene tutte le possibili forme ipotizzate per il **modello probabilistico** $p^0(y)$, vera distribuzione di probabilità di Y (Azzalini, 2001):

$$\mathcal{F} = \{p(y; \theta), \theta \in \Theta \subseteq \mathbb{R}^p\}.$$

Se $p^0(y) \in \mathcal{F}$ allora il modello statistico è **correttamente specificato**.

In altri termini \mathcal{F} è un insieme di funzioni di densità di probabilità $p(y; \theta)$, opportunamente scelte, con Θ spazio parametrico p -dimensionale all' interno del quale θ varia. Spesso non ci vengono fornite indicazioni precise su come

effettuare una corretta specificazione del modello e questo rappresenta uno dei problemi che emergono all'interno di un'analisi di tipo inferenziale, ovvero l'individuazione del modello statistico adatto a Y , osservata tramite il vettore di realizzazioni campionarie y . Un modo per restringere la classe \mathcal{F} è basarsi su considerazioni preliminari riguardo il fenomeno oggetto di studio oppure utilizzare vari strumenti per testare la bontà di adattamento del modello stimato, ma queste tematiche sono fuori dagli scopi di questo elaborato. A questo punto, supponendo che il modello statistico sia correttamente specificato, si può procedere alla stima di θ , così da individuare un unico modello probabilistico all'interno di \mathcal{F} , $p(y; \hat{\theta})$, con $\hat{\theta}$ che si spera essere il più vicino possibile a θ_0 , **vero parametro** corrispondente alla vera distribuzione di Y , ossia $p^0(y) = p(y; \theta_0)$. Una buona soluzione è assegnare a θ la sua **stima di massima verosimiglianza** (SMV). La teoria asintotica di verosimiglianza, infatti, garantisce nei problemi regolari di stima che lo stimatore $\hat{\theta}$ sia consistente, ovvero converge in probabilità al vero valore del parametro. Poiché l'operazione di stima di un parametro ignoto mediante un singolo valore è destinata a fornire un risultato sbagliato con pratica certezza allora conviene cercare un insieme di valori plausibili per θ , ovvero cercare un **intervallo di confidenza**, oppure testare la plausibilità di un certo valore attraverso una **verifica di ipotesi**. Simili procedure fanno uso di *quantità pivotali*, ovvero da funzioni del tipo $q(y; \theta)$ che dipendono sia dai dati sia dal parametro ma che hanno distribuzione nota, indipendente dal parametro stesso, e *test statistici* spesso basati su approssimazioni asintotiche **del primo ordine**. Tali approssimazioni in certe situazioni risultano poco accurate e possono fornire risultati imprecisi; per ovviare a questo problema si ricorre alle approssimazioni asintotiche **di ordine superiore**, che vedremo in seguito.

2.3 Verosimiglianza statistica e quantità collegate

2.3.1 La funzione di verosimiglianza

Una volta osservato il campione $y = (y_1, \dots, y_n)$, la funzione di densità $p(y; \theta)$ dipende solo dal parametro θ ; questa ci dà la probabilità di osservare, prima della realizzazione di Y , ciò che poi è stato osservato. Si può quindi definire **funzione di verosimiglianza** di θ , relativamente a y osservato, una funzione $L : \Theta \rightarrow \mathbb{R}$, dove (Azzalini, 2001)

$$L(\theta) = L(\theta; y) = c(y)p(y; \theta) \quad (2.1)$$

con $c(y)$ costante positiva che non dipende da θ e che sarà trascurabile ai fini dell'inferenza sul parametro. Si noti che la scrittura $L(\theta; y)$ serve a sottolineare la dipendenza di $L(\theta)$ da y .

In caso di campionamento casuale semplice le osservazioni y_1, \dots, y_n sono indipendenti ed identicamente distribuite (i.i.d.) e quindi la funzione di verosimiglianza è esprimibile come

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

dove $p_{\psi}(y_i; \theta)$ è la funzione di densità di probabilità di Y .

Se $L(\theta_1) > L(\theta_2)$ allora diremo che θ_1 è più plausibile di θ_2 , relativamente ai dati osservati, poichè sotto θ_1 è maggiore la probabilità di osservare y (prima della sua realizzazione). La preferenza attribuita a θ_1 è determinata dal rapporto $L(\theta_1)/L(\theta_2)$. Da notare che nel rapporto le costanti che non dipendono da θ si semplificano e allora le due funzioni di verosimiglianza, che differiscono per una costante moltiplicativa, si dicono *equivalenti*.

Per ragioni di utilità pratica si preferisce utilizzare il logaritmo della funzione di verosimiglianza, ovvero la **funzione di log-verosimiglianza**

$$l(\theta) = l(\theta; y) = \log L(\theta; y) = \log c(y) + \log p(y; \theta),$$

con $l(\theta) = -\infty$ quando $L(\theta) = 0$. Come in precedenza, se le osservazioni sono i.i.d. la funzione di log-verosimiglianza è pari a

$$l(\theta) = \sum_{i=1}^n \log p_{\psi}(y_i; \theta).$$

È infine utile accennare alla definizione di **problema regolare di stima**. Siamo in tale situazione quando valgono le seguenti condizioni (Azzalini, 2001, Paragrafo 3.2.3):

1. il modello statistico è identificabile, ovvero esiste una corrispondenza biunivoca tra Θ e gli elementi di \mathcal{F} ;
2. lo spazio parametrico Θ è un sottoinsieme aperto dello spazio euclideo \mathbb{R}^p , con p finito e indipendente da n ;
3. le funzioni di densità specificate in \mathcal{F} hanno tutte lo stesso supporto, che non dipende quindi da θ ;
4. la log-verosimiglianza è una funzione derivabile almeno tre volte con derivate parziali continue in Θ , e con la possibilità di scambiare, per la funzione $p(y; \theta)$, due volte il segno di integrale con quello di derivata rispetto a θ .

2.3.2 Statistiche sufficienti

Una funzione $T : \mathcal{Y} \rightarrow \mathbb{R}^p$, tale che $T(y)$ non dipenda da θ si dice **statistica** e il valore $t = T(y)$ è detto **valore campionario** della statistica. L'indipendenza di t da θ garantisce l'effettiva computabilità della statistica. Quindi una statistica non è altro che una trasformazione dei dati o in certi casi è rappresentata dai dati stessi.

Una caratteristica desiderabile sarebbe che la statistica riassume l'informazione su θ contenuta nei dati senza disperderne una parte. Una statistica si dice **sufficiente** per θ se assume lo stesso valore in due punti dello spazio campionario solo se tali punti hanno verosimiglianze equivalenti, ovvero se per ogni $y, z \in \mathcal{Y}$

$$T(y) = T(z) \implies L(\theta; y) \propto L(\theta; z)$$

per ogni $\theta \in \Theta$. Pertanto vale la seguente fattorizzazione

$$p_Y(y; \theta) = p_T(t; \theta) p_{Y|T=t}(y; t),$$

per tutti gli y tali che $t(y) = t$. Quindi se T è statistica sufficiente, due osservazioni y_1, y_2 tali che $t(y_1) = t(y_2)$ devono condurre alle stesse conclusioni inferenziali per θ . Per qualunque modello statistico, y è sempre statistica sufficiente.

Una statistica sufficiente si dice **minimale** per θ se è sufficiente e se assume valori distinti solamente in punti dello spazio campionario a cui corrispondono verosimiglianze non equivalenti, cioè se per ogni $y, z \in \mathcal{Y}$

$$T(y) = T(z) \iff L(\theta; y) \propto L(\theta; z)$$

per ogni $\theta \in \Theta$. In altri termini t è statistica sufficiente minimale se è funzione di ogni altra statistica sufficiente, ed è la massima sintesi dei dati senza perdita di informazione sul parametro che ci permette di fare inferenza su di esso.

2.3.3 Stima di massima verosimiglianza

Si è detto che l'obiettivo dell'inferenza di verosimiglianza è trovare un valore di θ più vicino possibile a θ_0 all'interno di Θ , o trovare un insieme di valori plausibili per θ , le cosiddette *regioni di confidenza*, o in alternativa testare la plausibilità di un certo valore tramite verifica di ipotesi. Il metodo utilizzato per il primo di questi scopi è quello della **massima verosimiglianza** che fornisce una stima del parametro a partire dalla funzione di verosimiglianza relativa ai dati osservati e al modello statistico scelto. Per una funzione di verosimiglianza $L(\theta)$ con $\theta \in \Theta$, si dice *stima di massima verosimiglianza* (SMV) di θ un valore $\hat{\theta}$ che massimizza la funzione $L(\theta)$ nello spazio Θ , ovvero

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

La SMV massimizza la probabilità di osservare y prima della sua realizzazione. Non è detto che la SMV esista o che sia unica, e talvolta non è possibile ottenere una sua espressione analitica e quindi occorre pervenire ad essa in modo numerico. Se $T(y)$ è una statistica sufficiente per θ allora $\hat{\theta}$ è funzione di $T(y)$. In pratica è preferibile ed equivalente massimizzare la funzione di log-verosimiglianza $l(\theta)$.

Nel contesto di un problema regolare di stima, descritto in precedenza, si possono calcolare alcune quantità (Pace e Salvani, 2001):

- la **funzione score**, ovvero il vettore delle derivate parziali prime di $l(\theta)$, indicata con $l_\theta(\theta) = l_\theta(\theta; y) = \frac{\partial l(\theta)}{\partial \theta} = (l_{\theta_1}(\theta), \dots, l_{\theta_p}(\theta))^T$ con $l_{\theta_r}(\theta) = \frac{\partial l(\theta)}{\partial \theta_r}$. Uguagliando a zero queste derivate (unica nel caso θ sia unidimensionale), $l_\theta(\theta) = 0$, si ottiene la cosiddetta **equazione (o sistema di equazioni) di verosimiglianza**, la cui risoluzione rispetto a θ fornisce la SMV;
- la **matrice di informazione osservata**, indicata con $j(\theta) = -l_{\theta\theta}(\theta) = -\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} = [-l_{\theta_r \theta_s}]$ dove $l_{\theta_r \theta_s} = \frac{\partial^2 l(\theta)}{\partial \theta_r \partial \theta_s}$ è l'elemento di posizione (r, s) , contenente l'opposto delle derivate parziali seconde di $l(\theta)$;
- la **matrice di informazione attesa**, indicata con $i(\theta) = \mathbb{E}_\theta[j(\theta)]$, ovvero il valore atteso della matrice di informazione osservata.

L'informazione osservata di Fisher calcolata in $\hat{\theta}$ è una matrice definita non negativa che può essere vista come un indice della rapidità di caduta della log-verosimiglianza man mano che ci si allontana da $\hat{\theta}$, ovvero della misura di preferibilità di cui $\hat{\theta}$ gode rispetto ad altri valori di θ . Si dimostra che la distribuzione asintotica dello SMV per n sufficientemente elevato è

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

dove p indica la dimensione del parametro. In tale espressione $i(\theta)$ può essere sostituita da $i(\hat{\theta})$, $j(\theta)$ o $j(\hat{\theta})$. Nel seguito si assumerà che la stima di massima verosimiglianza esista e sia unica.

2.3.4 Statistiche ancillari

Un'altra classe di statistiche a cui faremo riferimento in seguito è quella delle statistiche ancillari. Una statistica **ancillare** rispetto a $t = T(y)$, sufficiente minimale, è una statistica a tale per cui la sua distribuzione non dipende da θ ed esiste una relazione biunivoca tra $(\hat{\theta}, a)$ e t . Se a è ancillare allora la sua distribuzione è costante e vale la seguente fattorizzazione

$$p_{\hat{\theta}, A}(\hat{\theta}, a; \theta) = p_A(a)p_{\hat{\theta}|A=a}(\hat{\theta}; a, \theta),$$

dove $\hat{\theta}$ è quindi statistica sufficiente minimale nel modello condizionato ad a osservato. Non è detto che esista un'unica statistica ancillare esatta o che la sua individuazione sia semplice; in tal caso è possibile ricorrere a statistiche ancillari asintoticamente costanti in distribuzione.

2.4 Invarianza rispetto a riparametrizzazioni

Nel modello statistico \mathcal{F} , esiste una relazione biunivoca tra i suoi elementi, $p(y; \theta)$, e il parametro θ . Scegliere una parametrizzazione significa determinare un sistema di coordinate geometriche all'interno del quale gli elementi di \mathcal{F} possano sempre essere individuati univocamente. In alcuni casi è utile riparametrizzare il modello, ad esempio per evitare vincoli sul parametro rendendo lo spazio parametrico illimitato oppure per dare maggiore rilevanza ad un particolare aspetto all'interno della distribuzione dei dati e del modello. Naturalmente è lecito aspettarsi che le conclusioni inferenziali non cambino, cioè che non dipendano dalla parametrizzazione.

Il nuovo parametro su cui fare inferenza dev'essere una funzione biunivoca e regolare del parametro originario, $\omega = \omega(\theta)$, con $\omega : \Theta \rightarrow \Omega$ e $\Theta, \Omega \in \mathbb{R}^p$. Si parla quindi di **invarianza rispetto alla riparametrizzazione** quando le conclusioni inferenziali su ψ sono le stesse di quelle su θ e possono essere espresse in funzione di $\theta = \theta(\omega)$.

Tale proprietà, che per quanto detto risulta molto interessante, è tipica della SMV (è detta anche *proprietà di equivarianza*), in contesti di stima puntuale, e in generale di gran parte delle procedure inferenziali basate sulla

verosimiglianza. Ne deriva quindi che se $\hat{\theta}$ è la stima di massima verosimiglianza di θ allora $\hat{\omega} = \omega(\hat{\theta})$ sarà la SMV di ω e $L_{\Omega}(\omega(\hat{\theta})) = L_{\Theta}(\hat{\theta})$, dove $L_{\Omega}(\omega)$ è la funzione di verosimiglianza riscritta nella nuova parametrizzazione e $L_{\Theta}(\theta)$ è la funzione di verosimiglianza scitta nella parametrizzazione originale. Ovviamente le conclusioni restano invariate anche considerando la funzione di log-verosimiglianza.

2.5 Parametri di disturbo e verosimiglianza profilo

Spesso si può avere a che fare con dei parametri detti di *disturbo*, ossia dei parametri che rappresentano aspetti del fenomeno oggetto di studio ma che non sono di diretto interesse inferenziale. Abbiamo visto che si può rappresentare la realtà studiata attraverso un modello statistico \mathcal{F} , cioè una collezione di modelli probabilistici del tipo $p(y; \theta)$, dove y sono i dati osservati e θ è parametro fissato e che di fatto determina y . Il nostro interesse quindi è su θ . La sua dimensione p dipende dal grado di complessità del fenomeno che, a sua volta, dipende dal numero di aspetti di cui il modello dovrà tenere conto per descrivere in modo adeguato i dati osservati. Tuttavia però spesso l'analisi statistica non mira ad indagare la realtà in tutta la sua complessità ma solo particolari aspetti di essa.

A tal proposito si consideri $\theta = (\psi, \lambda)$, dove ψ è il **parametro di interesse**, di dimensione $1 \leq k < p$ chiamato così appunto perchè rappresenta l'aspetto a cui siamo interessati e su cui fare inferenza, e λ è il **parametro di disturbo** di dimensione $p - k$ che rappresenta un altro aspetto della realtà studiata ma che non è nostro interesse indagare. A questo punto le procedure inferenziali utilizzate sono ancora legate alla verosimiglianza e alla sua funzione $L(\theta) = L(\psi, \lambda)$. Poichè l'inferenza però riguarda solo il parametro ψ sarebbe utile riscrivere la verosimiglianza in funzione del solo parametro di interesse. Per questo, per un fissato valore di ψ , troveremo la stima di massima verosimiglianza di $\lambda(\hat{\lambda}_{\psi})$ attraverso le equazioni di verosimiglianza viste in precedenza e scriveremo $L(\psi, \hat{\lambda}_{\psi}) = L_p(\psi)$ dove $L_p(\psi)$ sarà la **fun-**

zione di verosimiglianza profilo per ψ e ci servirà per fare inferenza su di esso. Una volta trovata la SMV di ψ si perverrà conseguentemente alla SMV di θ , ovvero $\hat{\theta} = (\hat{\psi}, \hat{\lambda}) = (\hat{\psi}, \hat{\lambda}_{\hat{\psi}})$. La stessa cosa vale per la funzione di log-verosimiglianza profilo che indicheremo con $l_p(\psi)$. La verosimiglianza profilo non è una vera e propria verosimiglianza ma gode di alcune importanti proprietà che la rendono simile ad una verosimiglianza propria (Pace e Salvan, 1996):

1. la stima di massima verosimiglianza profilo coincide con la stima di massima verosimiglianza di ψ basata su $L(\psi, \lambda)$, ovvero

$$\sup_{\psi} L_p(\psi) = L_p(\hat{\psi});$$

2. l'informazione osservata profilo è definita come

$$j_p(\psi) = -\frac{\partial^2 l_p(\psi)}{\partial \psi \partial \psi^T} = -\frac{\partial^2 l(\psi, \hat{\lambda}_{\psi})}{\partial \psi \partial \psi^T}.$$

Inoltre la matrice inversa di $j_p(\psi)$ coincide col blocco (ψ, ψ) dell'inversa della matrice di informazione osservata totale. Si può infatti notare che

$$\frac{\partial l_p(\psi)}{\partial \psi} = l_{\psi}(\psi, \hat{\lambda}_{\psi}) + l_{\lambda}(\psi, \hat{\lambda}_{\psi}) \frac{\partial \hat{\lambda}_{\psi}}{\partial \psi};$$

tuttavia il secondo addendo del membro di destra è uguale a zero poichè $l_{\lambda}(\psi, \hat{\lambda}_{\psi}) = 0$. Risulta quindi

$$\frac{\partial l_p(\psi)}{\partial \psi} = l_{\psi}(\psi, \hat{\lambda}_{\psi}).$$

La matrice delle derivate seconde è data da

$$\frac{\partial^2 l_p(\psi)}{\partial \psi \partial \psi^T} = l_{\psi\psi}(\psi, \hat{\lambda}_{\psi}) + l_{\psi\lambda}(\psi, \hat{\lambda}_{\psi}) \frac{\partial \hat{\lambda}_{\psi}}{\partial \psi}.$$

Poichè $l_{\lambda}(\psi, \hat{\lambda}_{\psi}) = 0$, anche la sua derivata parziale calcolata rispetto a

ψ sarà nulla. Quindi

$$l_{\lambda\psi}(\psi, \hat{\lambda}_\psi) + l_{\lambda\lambda}(\psi, \hat{\lambda}_\psi) \frac{\partial \hat{\lambda}_\psi}{\partial \psi} = 0,$$

da cui

$$\frac{\partial \hat{\lambda}_\psi}{\partial \psi} = -(l_{\lambda\lambda}(\psi, \hat{\lambda}_\psi))^{-1} l_{\lambda\psi}(\psi, \hat{\lambda}_\psi);$$

di conseguenza

$$j_p(\psi) = -(l_{\psi\psi} - l_{\psi\lambda}(l_{\lambda\lambda})^{-1}l_{\lambda\psi}),$$

dove tutte le derivate indicate sono valutate in $(\psi, \hat{\lambda}_\psi)$. Si può pertanto concludere che

$$[j_p(\psi)]^{-1} = j^{\psi\psi}(\psi, \hat{\lambda}_\psi),$$

dove $j^{\psi\psi}(\psi, \hat{\lambda}_\psi)$ è il blocco (ψ, ψ) dell'inversa della matrice di informazione osservata complessiva calcolata in $(\psi, \hat{\lambda}_\psi)$;

3. il test log-rapporto di verosimiglianza profilo è

$$W_p(\psi) = 2\{l_p(\hat{\psi}) - l_p(\psi)\} = 2\{l(\hat{\psi}, \hat{\lambda}) - l(\psi, \hat{\lambda}_\psi)\}$$

e coincide con il test log-rapporto di verosimiglianza basato su $l(\psi, \lambda)$ utilizzato per la verifica di ipotesi su ψ con λ ignoto.

Anche nel caso di presenza di parametri di disturbo la riparametrizzazione non deve alterare le conclusioni inferenziali sul parametro di interesse. A tale scopo è necessario far riferimento alle sole riparametrizzazioni che non alterano l'interesse, ovvero quelle del tipo $\omega = \omega(\theta) = \omega(\psi, \lambda)$, con $\omega = (\rho, \xi)$ tali che $\rho = \rho(\psi)$ e $\xi = \xi(\psi, \lambda)$ con $\rho(\cdot)$ funzione biunivoca, ψ parametro di interesse e λ componente di disturbo.

2.6 Problemi e procedure inferenziali basate sulla verosimiglianza

Come accennato in precedenza, per stimare θ possiamo optare per diverse soluzioni: stima puntuale (SMV), stima intervallare o verifica di ipotesi. Vediamo ora nel dettaglio gli ultimi due metodi citati.

2.6.1 Verifica di ipotesi

In una verifica di ipotesi vogliamo testare la plausibilità di un certo valore, θ_0 , per il parametro ignoto θ . Una verifica di ipotesi è caratterizzata da tre elementi:

1. il **sistema di ipotesi**, formulato prima di osservare i dati, relativo al parametro θ , costituito da un' **ipotesi nulla**, del tipo $H_0 : \theta = \theta_0$, e da un' **ipotesi alternativa**, $H_1 : \theta \neq \theta_0$ (bilaterale in questo caso);
2. una **statistica test**, ovvero una statistica T a valori in \mathbb{R} che bipartisce lo spazio campionario \mathcal{Y} in due parti, e la **distribuzione nulla**, cioè la distribuzione di probabilità di T sotto H_0 . Questa distribuzione può essere nota oppure ricavata tramite opportune simulazioni;
3. i **valori critici** della distribuzione nulla, ossia i quantili di tale distribuzione che, ad un certo **livello di significatività**, α , individuano le **regioni di accettazione e di rifiuto** dell'ipotesi nulla, in cui lo spazio campionario \mathcal{Y} si divide tramite T .

Se il valore campionario $T(y^{oss})$ appartiene alla regione di accettazione allora diremo che c'è abbastanza evidenza empirica a favore dell'ipotesi nulla, che accetteremo, altrimenti la rifiuteremo. Rifiutare l'ipotesi nulla non significa automaticamente accettare l'ipotesi alternativa ma solo che probabilmente quest'ultima è più verosimile e bisognerebbe fare altre verifiche in tal senso; ecco perchè si preferisce dire "rifiutare l'ipotesi nulla" piuttosto che "accettare l'ipotesi alternativa". La formulazione delle due ipotesi dipende dalla natura del problema e del modello; solitamente il valore che testiamo (ipotesi nulla)

è quello su cui abbiamo qualche informazione in più o che, per qualche motivo, pensiamo sia essere quello "giusto", mentre sull'ipotesi alternativa non sappiamo molto.

Un altro strumento per decidere se accettare o rifiutare l'ipotesi nulla, una volta ricavato il valore osservato t^{oss} della statistica test sulla base dei dati a disposizione, è il **livello di significatività osservato** (p -value). Immaginando di voler testare un'ipotesi del tipo $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_1$, a seconda della statistica test utilizzata potremo avere una regione di rifiuto dell'ipotesi nulla *unilaterale destra*, *unilaterale sinistra* oppure *bilaterale*. Nel primo caso valori elevati di $t(y)$ sono sintomo di poca evidenza nei dati a supporto dell'ipotesi nulla, e il p -value è dato da (Pace e Salvan, 2001)

$$\alpha^{oss} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta[T \geq t^{oss}], \quad (2.2)$$

dove t^{oss} è il valore osservato della statistica test relativamente a y^{oss} , cioè $t^{oss} = t(y^{oss})$, mentre T è una variabile casuale. In altri termini il p -value non è altro che la probabilità di osservare una statistica test maggiore o uguale a quella osservata, sotto l'ipotesi nulla. Possiamo ancora definire il p -value come l'area, a destra del valore t^{oss} , sottesa alla curva che rappresenta la distribuzione di probabilità di T , sotto H_0 (distribuzione nulla). Analogamente se la regione critica è unilaterale sinistra allora rifiuteremo H_0 per valori piccoli di t^{oss} , e il p -value sarà

$$\alpha^{oss} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta[T \leq t^{oss}], \quad (2.3)$$

ovvero l'area a sinistra della statistica test osservata, sottesa alla curva della distribuzione di T . Nel caso in cui la regione di rifiuto è bilaterale il p -value sarà

$$\alpha^{oss} = 2 \sup_{\theta \in \Theta_0} \min(\mathbb{P}_\theta[T \leq t^{oss}], \mathbb{P}_\theta[T \geq t^{oss}]). \quad (2.4)$$

Confrontando il p -value con un livello di significatività α fissato, potremo decidere se accettare o meno l'ipotesi nulla; in particolare accetteremo se $\alpha^{oss} > \alpha$ e rifiuteremo se $\alpha^{oss} \leq \alpha$.

Per testare un'ipotesi nulla semplice del tipo $H_0 : \theta = \theta_0$ esistono tre test legati alla verosimiglianza. Quello più utilizzato è sicuramente il **test rapporto di verosimiglianza** (o TRV), che consiste nel confrontare la verosimiglianza calcolata in θ_0 con quella calcolata nella sua stima di massima verosimiglianza, $\hat{\theta}$. Si ottiene così la statistica

$$\frac{L(\hat{\theta})}{L(\theta_0)} \quad (2.5)$$

che ha una regione di rifiuto unilaterale destra. Utilizzando, come ormai è consuetudine, la log-verosimiglianza, e un sistema di ipotesi che prevede un'ipotesi alternativa bilaterale ($H_1 : \theta \neq \theta_0$), otteniamo la statistica **log-rapporto di verosimiglianza**

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

la quale è una trasformazione monotona della (2.5). Il test $W(\theta_0)$ si distribuisce asintoticamente come un chi-quadrato con p gradi di libertà, dove p è la dimensione del parametro, e rappresenta una quantità approssimativamente pivotale.

Gli altri due test utilizzati sono il **test score** (o test di Rao)

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

e il **test di Wald**,

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

entrambi con distribuzione nulla asintotica chi-quadrato con p gradi di libertà.

Quando $p = 1$ e l'ipotesi alternativa è unilaterale (destra o sinistra) esistono le versioni

$$\begin{aligned} r(\theta_0) &= \text{sign}(\hat{\theta} - \theta_0) \sqrt{W(\theta_0)}, \\ r_u(\theta_0) &= l_\theta(\theta_0) \sqrt{i(\theta_0)}, \end{aligned} \quad (2.6)$$

$$r_e(\theta_0) = (\hat{\theta} - \theta_0)\sqrt{i(\theta_0)}.$$

e tutti e tre hanno distribuzione asintotica nulla $\mathcal{N}(0, 1)$.

Anche nel caso di presenza di parametri di disturbo, grazie alle proprietà della verosimiglianza profilo viste in precedenza, si possono ricavare statistiche test per il parametro di interesse ψ asintoticamente equivalenti a quelle costruite per il parametro θ . Se si vuole verificare l'ipotesi $H_0 : \psi = \psi_0$, il test log-rapporto di verosimiglianza profilo diventa

$$2\{l(\hat{\psi}, \hat{\lambda}) - l(\psi_0, \hat{\lambda}_{\psi_0})\}$$

che coincide con la statistica log-verosimiglianza profilo

$$W_p(\psi_0) = 2 \{l_p(\hat{\psi}) - l_p(\psi_0)\}$$

la cui distribuzione nulla asintotica è χ_k^2 , dove k è la dimensione di ψ , parametro di interesse. Analogamente si possono calcolare le statistiche di Rao e di Wald:

$$W_{p_u}(\psi_0) = l_\psi(\psi_0, \hat{\lambda}_{\psi_0})^T i^{\psi\psi}(\psi_0, \hat{\lambda}_{\psi_0}) l_\psi(\psi_0, \hat{\lambda}_{\psi_0}),$$

$$W_{p_e}(\psi_0) = (\hat{\psi} - \psi_0)^T [i^{\psi\psi}(\psi_0, \hat{\lambda}_{\psi_0})]^{-1} (\hat{\psi} - \psi_0),$$

entrambi con distribuzione asintotica nulla chi-quadrato con k gradi di libertà, e dove per $i^{\psi\psi}$ si intende il blocco relativo a ψ di $i(\theta)^{-1}$.

Nei casi unidirezionali le statistiche corrispondenti sono:

$$r_p(\psi_0) = \text{sign}(\hat{\psi} - \psi_0)\sqrt{W_p(\psi_0)},$$

$$r_{p_u}(\psi_0) = l_\psi(\psi_0, \hat{\lambda}_{\psi_0})\sqrt{i^{\psi\psi}(\psi_0, \hat{\lambda}_{\psi_0})},$$

$$r_{p_e}(\psi_0) = (\hat{\psi} - \psi_0)[i^{\psi\psi}(\psi_0, \hat{\lambda}_{\psi_0})]^{-1/2},$$

tutti e tre con distribuzione asintotica nulla normale standard.

Test di Rao, test di Wald e TRV sono asintoticamente equivalenti, ma in generale si preferisce l'ultimo per la proprietà di invarianza alla riparamet-

trizzazione di cui gode e per il suo diretto collegamento alla verosimiglianza. Anche il test di Rao è invariante e in più non richiede il calcolo di $\hat{\theta}$, seppure questo sia in generale un vantaggio trascurabile in fase di verifica, ma presenta frequenti instabilità numeriche. Il test di Wald invece, per quanto semplice, non gode della proprietà di invarianza rispetto alle riparametrizzazioni. Le stesse considerazioni valgono per le versioni unidirezionali dei test e per le versioni profilo. In quest'ultimo caso si fa riferimento all'invarianza rispetto a trasformazioni che non alterano l'interesse.

2.6.2 Stima intervallare

Come detto in precedenza il tentativo di individuare il vero valore del parametro θ in modo puntuale fallirà quasi certamente. È più opportuno cercare un insieme di valori plausibili per θ , in modo da avere più possibilità di individuare il suo vero valore. L'insieme di questi valori costituisce una cosiddetta **regione di confidenza** per il vero valore del parametro; nel caso in cui θ fosse unidimensionale allora parleremo di **intervallo di confidenza**.

La stima intervallare e la verifica di ipotesi, che rappresentano due procedure apparentemente indipendenti l'una dall'altra, in realtà presentano una forte connessione tra di loro. Risolvendo rispetto a θ l'equazione che collega la quantità pivotale $q(y; \theta)$ ai percentili della distribuzione nulla di q , scelti in modo opportuno rispetto ad un **livello di confidenza** $1 - \alpha$, si ricava facilmente l'intervallo di confidenza per θ

$$\{y : \theta \in C(y)\}, \quad (2.7)$$

dove $C(y)$ è un sottoinsieme di Θ e tale che

$$\mathbb{P}_\theta[\theta \in C(Y)] = 1 - \alpha,$$

dove $1 - \alpha$ esprime il livello di confidenza nominale della regione, cioè la probabilità che l'insieme casuale dei possibili valori di θ , $C(Y)$, contenga il vero parametro. Da notare che questa definizione fa riferimento all'insieme casuale $C(Y)$, quindi prima della realizzazione di Y , e non allo specifico

intervallo (c_1, c_2) ottenuto dopo che abbiamo osservato y . Si può dire quindi che l'intervallo (c_1, c_2) deriva da un procedimento che prevede la costruzione di intervalli che nell' $(1 - \alpha)\%$ dei casi contengono il vero valore di θ , ma come si capisce tale affermazione non è più vera per l'intervallo specifico, poichè questo o contiene il vero parametro oppure no. Per capire meglio questo si pensi al *principio del campionamento ripetuto*: se prendessimo N campioni, ripetutamente campionati in modo casuale da una popolazione di parametro fissato θ , allora, in media, l' $(1 - \alpha)\%$ degli N intervalli così costruiti conteranno il parametro θ . Di fatto però il nostro campione è unico e lo osserviamo una sola volta e quindi abbiamo confidenza $(1 - \alpha)\%$ che (c_1, c_2) contenga l'ignoto θ .

I principali intervalli di confidenza basati sulla verosimiglianza sono quelli alla Wald e quelli basati sul TRV. I primi si basano sulla normalità asintotica dello stimatore $\hat{\theta}$ e sono del tipo

$$\hat{\theta} \mp z_{1-\alpha/2} \sqrt{j(\hat{\theta})^{-1}}, \quad (2.8)$$

dove $z_{1-\alpha/2}$ è il quantile di una normale standard di ordine $1 - \alpha/2$. Questi intervalli sono di semplice costruzione e simmetrici, anche quando la verosimiglianza è asimmetrica ma presentano alcuni inconvenienti: potrebbero escludere valori di θ che hanno una verosimiglianza maggiore dei valori inclusi e non sono invarianti rispetto alle riparametrizzazioni, che abbiamo visto essere una proprietà abbastanza desiderabile.

Gli intervalli di confidenza basati sul TRV includono quei valori di θ che hanno verosimiglianza maggiore di una certa soglia, legata al livello di confidenza $1 - \alpha$ desiderato. Fanno uso della statistica $W(\theta)$, vista in precedenza, e sono del tipo (Azzalini, 2001)

$$\{\theta \in \Theta : W(\theta) \leq \chi_{p,1-\alpha}^2\}, \quad (2.9)$$

dove $\chi_{p,1-\alpha}^2$ è il quantile di ordine $1 - \alpha$ della distribuzione chi-quadrato con p gradi di libertà. Esplicitando la (2.9), si può scrivere

$$\{\theta \in \Theta : l(\theta) \geq l(\hat{\theta}) - \frac{1}{2}\chi_{p,1-\alpha}^2\}.$$

Questi intervalli sono in generale asimmetrici e tutti i valori di θ inclusi nell'intervallo hanno verosimiglianza maggiore dei valori esclusi; inoltre presentano tipicamente una migliore copertura empirica (più vicina al valore nominale) e cosa più importante sono invarianti rispetto alle riparametrizzazioni.

Gli intervalli alla Wald sono ottenuti approssimando la log-verosimiglianza con una funzione quadratica. I due tipi di intervalli visti sono tanto più simili quanto maggiore è l'accuratezza di tale approssimazione, ovvero quanto più quadratica è la funzione di log-verosimiglianza nell'intervallo dei valori di interesse per il parametro, con n sufficientemente elevato. Inoltre va ricordato che tali intervalli sono di livello nominale approssimato (e non esatto) poichè si basano su quantità asintoticamente pivotali, ovvero con distribuzione nota approssimativamente.

Quanto esposto sulla stima intervallare per il parametro θ può essere esteso al caso in cui siano presenti parametri di disturbo. Quello che interessa è individuare una regione di confidenza per il parametro di interesse ψ . Questa volta utilizzeremo la statistica log-rapporto di verosimiglianza profilo per cercare gli intervalli di confidenza basati sul TRV di livello $1 - \alpha$

$$\{\psi \in \Theta_\psi : W_p(\psi) \leq \chi_{k,1-\alpha}^2\},$$

che nel caso in cui $k = 1$ diventa

$$\{\psi \in \Theta_\psi : |r_p(\psi)| \leq z_{1-\alpha/2}\},$$

dove $z_{1-\alpha/2}$ è il quantile di ordine $1 - \alpha/2$ della $\mathcal{N}(0, 1)$.

In generale, il livello di significatività osservato è una quantità pivotale che sotto H_0 ha distribuzione uniforme $\mathcal{U}(0, 1)$, sia se il test è unilaterale sia se è bilaterale. Ciò vuol dire che un insieme di valori $\tilde{\theta}$ che non sono stati rifiutati da una classe di test $t_{\tilde{\theta}}$, ad un livello di significatività α , per le ipotesi semplici $H_0 : \theta = \tilde{\theta}$, ovvero i valori

$$\{\theta \in \Theta : \alpha_\theta^{oss}(y) \geq \alpha\},$$

definiscono una regione di confidenza di livello $1 - \alpha$ per il parametro ignoto. Quindi dati i p-value di un test, è possibile risalire ad una regione di confidenza di livello $1 - \alpha$ per θ e viceversa. In questo senso esiste una dualità tra verifiche di ipotesi e stime intervallari.

2.7 Approssimazioni di ordine superiore

Le procedure inferenziali viste fin qui si basano su approssimazioni asintotiche, dette del primo ordine. Si è visto inoltre che può risultare utile riparametrizzare il modello, per focalizzare l'attenzione su un particolare parametro (aspetto) di interesse. Immaginiamo di volere fare inferenza, in particolare una verifica di ipotesi, sul parametro di interesse ψ , testando l'ipotesi $H_0 : \psi = \psi_0$, con ψ di dimensione 1 e λ parametro di disturbo. Se si considera, ad esempio, un'ipotesi alternativa unidirezionale allora possiamo utilizzare la statistica test *log-rapporto di verosimiglianza segnata*, ovvero

$$r(\psi_0) = \text{sign}(\hat{\psi} - \psi_0) \sqrt{W_p(\psi_0)}. \quad (2.10)$$

La (2.10) si distribuisce asintoticamente come una normale standard con un errore pari a $O(n^{-1/2})$. In particolare

$$Pr_{\psi_0}[r(\psi_0) < r(\psi_0)^{oss}] = \Phi(r(\psi_0)^{oss}) + O(n^{-1/2}),$$

dove $\Phi(\cdot)$ è la funzione di ripartizione di una $\mathcal{N}(0, 1)$ e $r(\psi_0)^{oss}$ è il valore osservato di $r(\psi_0)$. Tale ordine di errore implica approssimazioni potenzialmente poco accurate, soprattutto con n piccolo e una dimensione di λ elevata, e quindi risultati imprecisi.

Da un'approssimazione base di ordine superiore della densità di probabilità dello stimatore di massima verosimiglianza deriva una modificazione di $r(\psi_0)$ (Brazzale et al., 2007)

$$r^*(\psi_0) = r(\psi_0) + \frac{1}{r(\psi_0)} \log \frac{q(\psi_0)}{r(\psi_0)}, \quad (2.11)$$

che ha sempre distribuzione asintotica normale standard ma con un errore di ordine $O(n^{-3/2})$. In particolare,

$$Pr_{\psi_0}[r^*(\psi_0) < r^*(\psi_0)^{oss}] = \Phi(r^*(\psi_0)^{oss}) + O(n^{-3/2}).$$

Tale approssimazione risulta più accurata in quanto l'errore di approssimazione tende a zero più velocemente al crescere di n .

Il calcolo di r^* implica il calcolo della quantità $q(\cdot)$. In generale il suo calcolo è complesso e ci sono varie derivazioni. Se la log-verosimiglianza appartiene alla *famiglia esponenziale lineare*, ovvero è del tipo

$$l(\psi, \lambda) = \psi u + \lambda^T v - c(\psi, \lambda),$$

dove (ψ, λ) è il parametro canonico e (u, v) sono le componenti della statistica sufficiente minimale, allora possiamo scrivere

$$q(\psi) = t(\psi) \rho(\psi, \hat{\psi}),$$

con

$$t(\psi) = j_p^{1/2}(\hat{\psi}) (\hat{\psi} - \psi)$$

e

$$\rho(\psi, \hat{\psi}) = \left\{ \frac{|j_{\lambda\lambda}(\hat{\theta})|}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|} \right\}^{1/2},$$

dove $j_{\lambda\lambda}(\cdot)$ è il blocco (λ, λ) della matrice di informazione osservata, calcolata in $\hat{\theta}$ al numeratore e in $\hat{\theta}_\psi$ al denominatore. Il termine r^* può essere scritto come

$$r^*(\psi) = r(\psi) + INF(\psi) + NP(\psi),$$

dove il secondo e terzo addendo del membro di destra rappresentano, rispetti-

vamente, un aggiustamento per migliorare l'approssimazione normale e un aggiustamento relativo alla presenza di parametri di disturbo.

In generale, se $\phi(\theta)$ è il parametro canonico di un modello esponenziale tangente al modello nei dati osservati e $\psi = \psi(\theta)$, si ha

$$q(\psi) = \frac{|\phi(\hat{\theta}) - \phi(\hat{\theta}_\psi) \quad \phi_\lambda(\hat{\theta}_\psi)|}{|\phi_\theta(\hat{\theta})|} \left\{ \frac{|j_{\theta\theta}(\hat{\theta})|}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|} \right\}^{1/2}, \quad (2.12)$$

dove $\phi_\theta = \frac{\partial\phi(\theta)}{\partial\theta^T}$ e $\phi_\lambda = \frac{\partial\phi(\theta)}{\partial\lambda^T}$. La matrice $p \times p$ che appare nel numeratore del primo termine consiste in un vettore colonna contenente le derivate sullo spazio campionario e da una matrice $p \times (p - 1)$ di derivate miste.

Infatti, si ha

$$\phi(\theta)^T = \sum_{i=1}^n \frac{\partial l(\theta; y)}{\partial y_i} V_i,$$

con $y = y^0$ dati osservati e V_1, \dots, V_n sono vettori $1 \times p$ che dipendono dai dati osservati e che rappresentano dei vettori tangenti ad una statistica ancillare. In effetti $\phi(\theta)^T$ richiede il calcolo delle derivate rispetto ai dati. Per evitare ciò possiamo esprimere la statistica sufficiente y come funzione biunivoca di $(\hat{\theta}, a)$, dove a è una statistica ancillare che fa in modo che la relazione tra y e θ sia biunivoca quando hanno dimensioni diverse. Allora si può scrivere

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

$$\frac{\partial l(\theta; \hat{\theta}, a)}{\partial \hat{\theta}} = l_{;\hat{\theta}}(\theta; \hat{\theta}, a)$$

e

$$l_{\theta;\hat{\theta}}(\theta; \hat{\theta}, a) = \frac{\partial^2 l(\theta; \hat{\theta}, a)}{\partial \theta \partial \hat{\theta}^T}.$$

Quindi la (2.12) diventa

$$q(\psi) = \frac{|l_{;\hat{\theta}}(\hat{\theta}) - l_{;\hat{\theta}}(\hat{\theta}_\psi) \quad l_{\lambda;\hat{\theta}}(\hat{\theta}_\psi)|}{|l_{\theta;\hat{\theta}}(\hat{\theta})|} \left\{ \frac{|j_{\theta\theta}(\hat{\theta})|}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|} \right\}^{1/2}. \quad (2.13)$$

La (2.13) è la forma di q più utilizzata. Skovgaard (1996) fa notare che il numeratore del primo termine di q può essere espresso come

$$|l_{\theta;\hat{\theta}}(\hat{\theta}_\psi)|[\{l_{;\hat{\theta}}(\hat{\theta}) - l_{;\hat{\theta}}(\hat{\theta}_\psi)\}^T \{l_{\theta;\hat{\theta}}(\hat{\theta}_\psi)\}^{-1}]_1, \quad (2.14)$$

dove il secondo fattore è il primo componente del vettore tra parentesi quadre (indicato con 1 alla fine). Inoltre Skovgaard ha sviluppato delle approssimazioni basate sui momenti delle quantità coinvolte nella (2.14):

$$l_{;\hat{\theta}}(\hat{\theta}) - l_{;\hat{\theta}}(\hat{\theta}_\psi) \doteq i(\hat{\theta})^{-1}j(\hat{\theta})Q(\hat{\theta}, \hat{\theta}_\psi),$$

$$l_{\theta;\hat{\theta}}(\hat{\theta}_\psi)^T \doteq i(\hat{\theta})^{-1}j(\hat{\theta})S(\hat{\theta}, \hat{\theta}_\psi),$$

dove

$$Q(\theta_1, \theta_2) = cov_{\theta_1}\{l_\theta(\theta_1), l(\theta_1) - l(\theta_2)\},$$

$$S(\theta_1, \theta_2) = cov_{\theta_1}\{l_\theta(\theta_1), l_\theta(\theta_2)\}.$$

Otteniamo quindi

$$q(\psi) \doteq [\hat{S}^{-1}\hat{Q}]_1 |j(\hat{\theta})|^{1/2} |i^{-1}(\hat{\theta})| |\hat{S}| |j_{\lambda\lambda}(\hat{\theta}_\psi)|^{-1/2}, \quad (2.15)$$

con $\hat{Q} = Q(\hat{\theta}, \hat{\theta}_\psi)$ e $\hat{S} = S(\hat{\theta}, \hat{\theta}_\psi)$.

L'approssimazione di Skovgaard evita il calcolo delle derivate rispetto allo spazio campionario e il condizionamento alla statistica ancillare, in quanto questa può essere di difficile individuazione, non essendo sempre unica. Il prezzo da pagare, però, è il calcolo di alcuni valori attesi non standard, quali \hat{Q} e \hat{S} , che tuttavia possono essere facilmente approssimati.

2.7.1 Integrazione Monte Carlo

Le covarianze \hat{Q} e \hat{S} nella (2.15) utili al calcolo di q all'interno di r^* possono essere calcolate numericamente, e come abbiamo visto evitano il

ricorso ad una statistica ancillare. Possiamo scrivere

$$Q(\theta_1, \theta_2) = E_{\theta_1}[l_{\theta}(\theta_1) (l(\theta_1) - l(\theta_2))] = \int l_{\theta}(\theta_1)[l(\theta_1) - l(\theta_2)] p(y; \theta_1) dy \quad (2.16)$$

e

$$S(\theta_1, \theta_2) = E_{\theta_1}[l_{\theta}(\theta_1) l_{\theta}(\theta_2)] = \int l_{\theta}(\theta_1) l_{\theta}(\theta_2) p(y; \theta_1) dy. \quad (2.17)$$

La soluzione numerica dei due integrali mostrati sopra può essere ottenuta sfruttando i **metodi di integrazione Monte Carlo**, quindi via simulazione. L'idea di questi metodi è simulare tanti campioni da una densità e calcolare l'integrale utilizzando la media empirica della funzione che ci interessa integrare, calcolata sui campioni generati. In altri termini, supponendo di voler ottenere la soluzione numerica dell'integrale

$$I = \int m(y)g(y)dy,$$

dove $g(\cdot)$ è una funzione di densità, possiamo simulare R campioni *i.i.d.* Y_1, \dots, Y_R da g e utilizzare come approssimazione di I la media empirica

$$\hat{I} = R^{-1} \sum_{r=1}^R m(Y_r).$$

Poichè utilizziamo la media di campioni *i.i.d.*, per la *Legge forte dei grandi numeri* \hat{I} converge **quasi certamente** a I . Inoltre

- $E(\hat{I}) = I$
- $Var(\hat{I}) = R^{-1} \{ \int m^2(y)g(y)dy - I^2 \},$

dove la varianza dello stimatore può essere calcolata attraverso i valori simulati.

La (2.16) e la (2.17), utilizzate nella (2.15), si possono calcolare, rispettivamente, come

$$Q(\theta_1, \theta_2) = R^{-1} \sum_{r=1}^R l_{\theta}(\theta_1; y_r) [l(\theta_1; y_r) - l(\theta_2; y_r)]$$

e

$$S(\theta_1, \theta_2) = R^{-1} \sum_{r=1}^R l_{\theta}(\theta_1; y_r) l_{\theta}(\theta_2; y_r)$$

con Y_r generati da $p(y; \theta_1)$ per $r = 1, \dots, R$, dove $\theta_1 = \hat{\theta}$ e $\theta_2 = \hat{\theta}_{\psi}$.

Va sottolineato il fatto che, in ogni simulazione, $\hat{\theta}$ e $\hat{\theta}_{\psi}$ sono sempre le stesse stime di massima verosimiglianza, ovvero quelle ottenute con il campione originale. Quindi le covarianze empiriche vengono calcolate utilizzando i campioni simulati ma con le stesse SMV ottenute a partire da y osservato. Questo rappresenta una sostanziale differenza col metodo *bootstrap* applicato a $r_p(\psi)$ che invece calcolerebbe ad ogni simulazione la verosimiglianza calcolata sui nuovi dati (quelli simulati), e ottenendo di conseguenza $\hat{\theta}$ e $\hat{\theta}_{\psi}$ sempre diversi.

La libreria `LikelihoodAsy` (Bellio e Pierce, 2015) contiene funzioni utili al calcolo di r_p e r_p^* , e dei relativi p-value ed intervalli di confidenza. Il tutto avviene semplicemente indicando la funzione di log-verosimiglianza del modello, che dipende dai dati in esame, una funzione che genera dati a seconda della distribuzione ipotizzata, e la funzione ψ . In particolare questa libreria implementa l'integrazione Monte Carlo per il calcolo della (2.16) e (2.17), con le quali ottenere la (2.15) necessaria al calcolo di r_p^* , versione modificata di r_p . Infatti, sfruttiamo i dati generati tramite la apposita funzione per calcolare la log-verosimiglianza in quei punti, necessaria alla stima di Q e S . Possiamo inoltre decidere il numero di simulazioni di dati da effettuare, ma diciamo già che opteremo per 1000 simulazioni, numero sufficiente per una buona precisione. Infine, a seconda della funzione ψ , verranno calcolate le stime di massima verosimiglianza di θ_1 , θ_2 e ψ con i relativi standard error, i valori di r_p e r_p^* con i corrispondenti p-value a seconda dei valori di ψ testati, gli intervalli di confidenza al 90, 95, 99 percento per ψ basati sulle approssimazioni del primo ordine e di ordine superiore, e le quantità $INF(\psi)$ e $NP(\psi)$. Il pacchetto di Bellio e Pierce e, in generale, il metodo di

2.7 Approssimazioni di ordine superiore

integrazione Monte Carlo ci consentono di ottenere agevolmente l'espressione di q introdotta da Skovgaard per il calcolo di r_p^* , a cui sarebbe complicato pervenire a causa di calcoli lunghi e non banali.

Capitolo 3

Esempi di modelli Stress-Strength

3.1 Distribuzione Esponenziale

Supponiamo che X e Y siano due variabili casuali indipendenti e con distribuzione esponenziale di parametro α e di parametro β , ossia $X \sim Exp(\alpha)$ e $Y \sim Exp(\beta)$, rispettivamente. Supponiamo inoltre che $x = (x_1, \dots, x_{n_1})$ sia un campione casuale di n_1 osservazioni estratte da X e $y = (y_1, \dots, y_{n_2})$ sia un campione casuale di n_2 osservazioni estratte da Y . Sotto queste assunzioni, la probabilità $P(X < Y)$ può essere scritta come (Cortese e Ventura, 2012)

$$\psi = \psi(\theta) = \frac{E(X)}{E(X) + E(Y)} = \frac{\alpha}{\alpha + \beta}, \quad (3.1)$$

con $\theta = (\alpha, \beta)$.

Per procedere con l'inferenza su ψ può essere utile riparametrizzare la funzione di log-verosimiglianza $l(\theta)$ con $\theta = (\psi, \lambda)$, dove ψ , parametro di interesse, è mostrato nella (3.1) e $\lambda = \alpha + \beta$ è il parametro di disturbo. Sfruttando la proprietà di invarianza rispetto alle riparametrizzazioni, descritta nel capitolo precedente, possiamo pervenire alle stime di massima verosimiglianza di ψ e λ ; in particolare $\hat{\psi} = \hat{\alpha}/(\hat{\alpha} + \hat{\beta})$ e $\hat{\lambda} = \hat{\alpha} + \hat{\beta}$, dove le SMV di α e β sono date da $\hat{\alpha} = n_1/\sum x_i$ e $\hat{\beta} = n_2/\sum y_i$, rispettivamente.

3.2 Distribuzione Normale

In questa sezione verranno esaminati tre casi: X e Y variabili casuali indipendenti e distribuite normalmente con varianze uguali; X e Y variabili casuali indipendenti e distribuite normalmente con varianze diverse; X e Y variabili casuali dipendenti con varianze diverse. In tutti e tre i casi ipotizzeremo che tutti i parametri siano ignoti.

3.2.1 Normali indipendenti con varianze uguali

Assumiamo che X e Y siano due variabili casuali indipendenti e normalmente distribuite con media μ_1 , μ_2 , rispettivamente, e stessa varianza σ^2 , ossia $X \sim N(\mu_1, \sigma^2)$ e $Y \sim N(\mu_2, \sigma^2)$. Siano $x = (x_1, \dots, x_{n_1})$ un campione casuale di n_1 osservazioni provenienti da X e $y = (y_1, \dots, y_{n_2})$ un campione casuale di n_2 osservazioni provenienti da Y . Allora si ha che la probabilità $P(X < Y)$ è data da

$$\psi = \psi(\theta) = \Phi\left(\frac{\mu_2 - \mu_1}{\sigma\sqrt{2}}\right), \quad (3.2)$$

dove $\Phi(\cdot)$ è la funzione di ripartizione di una normale standard e $\theta = (\mu_1, \mu_2, \sigma^2)$.

La SMV della (3.2) è data da (Cortese e Ventura, 2012)

$$\hat{\psi} = \psi(\hat{\theta}) = \Phi\left(\frac{\hat{\mu}_2 - \hat{\mu}_1}{\hat{\sigma}\sqrt{2}}\right),$$

dove $\hat{\mu}_1 = n_1^{-1} \sum x_i$, $\hat{\mu}_2 = n_2^{-1} \sum y_i$ e $\hat{\sigma}^2 = (1/n)(\sum(x_i - \hat{\mu}_1)^2 + \sum(y_i - \hat{\mu}_2)^2)$ sono le SMV di μ_1 , μ_2 e σ^2 , rispettivamente, con $n = n_1 + n_2$. Anche in questo caso possiamo riparametrizzare il modello con $\theta = (\psi, \lambda)$, dove ψ è dato dalla (3.2) e $\lambda = (\lambda_1, \lambda_2) = (\mu_1/\sigma\sqrt{2}, \sigma\sqrt{2})$.

3.2.2 Normali indipendenti con varianze diverse

Un caso più realistico del precedente prevede che X e Y abbiano varianze diverse, ovvero $X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$, pur mantenendo l'assunzione di indipendenza tra X e Y . In questo caso si ha (Cortese e Ventura, 2012)

$$\psi = \psi(\theta) = \Phi\left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right), \quad (3.3)$$

dove $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$.

La SMV di ψ è

$$\hat{\psi} = \psi(\hat{\theta}) = \Phi\left(\frac{\hat{\mu}_2 - \hat{\mu}_1}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}}\right),$$

dove $\hat{\sigma}_1^2 = n_1^{-1} \sum (x_i - \hat{\mu}_1)^2$ e $\hat{\sigma}_2^2 = n_2^{-1} \sum (y_i - \hat{\mu}_2)^2$ sono le SMV di σ_1^2 e σ_2^2 , rispettivamente, mentre $\hat{\mu}_1$ e $\hat{\mu}_2$ sono le stime del paragrafo precedente. In questo caso la riparametrizzazione utilizzata è $\theta = (\psi, \lambda)$ con ψ dato dalla (3.3) e $\lambda = (\lambda_1, \lambda_2, \lambda_3) = (\mu_1, \sigma_1^2, \sigma_2^2)$.

3.2.3 Normali dipendenti con varianze diverse

Si assuma ora che X e Y siano due variabili casuali dipendenti con $X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$. In alternativa possiamo considerare un vettore casuale $Z = (X, Y)$ con valore atteso $\mu = (\mu_1, \mu_2)$ e matrice di varianze e covarianze Σ . Quindi si può scrivere $\psi = P(X < Y) = P(A^T Z > 0)$ con $A = (-1, 1)$. Supponiamo sempre che $x = (x_1, \dots, x_{n_1})$ sia un campione casuale di n_1 osservazioni estratte da X e $y = (y_1, \dots, y_{n_2})$ sia un campione casuale di n_2 osservazioni estratte da Y , con $n_1 = n_2$. La probabilità $P(X < Y)$ è data da (Kotz et al., 2003)

$$\psi = \psi(\theta) = \Phi\left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 - 2\sigma_{12} + \sigma_2^2}}\right), \quad (3.4)$$

dove σ_1^2 , σ_2^2 e σ_{12} sono le varianze di X e di Y e la covarianza tra X e Y , rispettivamente. La SMV della (3.4) è data da

$$\hat{\psi} = \psi(\hat{\theta}) = \Phi\left(\frac{\hat{\mu}_2 - \hat{\mu}_1}{\sqrt{\hat{\sigma}_1^2 - 2\hat{\sigma}_{12} + \hat{\sigma}_2^2}}\right),$$

dove $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1^2$ e $\hat{\sigma}_2^2$ sono le SMV di μ_1 , μ_2 , σ_1^2 e σ_2^2 già viste nel caso precedente, e $\hat{\sigma}_{12} = n^{-1} \sum (x_i - \hat{\mu}_1)(y_i - \hat{\mu}_2)$ è la SMV della covarianza tra X e Y .

3.3 Distribuzione Weibull

Consideriamo il caso in cui X e Y siano due variabili casuali indipendenti e con distribuzione Weibull, con stesso parametro di forma α , e parametri di scala ignoti σ_1, σ_2 , rispettivamente. Siano $x = (x_1, \dots, x_{n_1})$ e $y = (y_1, \dots, y_{n_2})$ due campioni casuali di n_1 ed n_2 osservazioni estratte da X e da Y , rispettivamente. Allora $P(X < Y)$ è (Kotz et al., 2003)

$$\psi = \psi(\theta) = \frac{\sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \quad (3.5)$$

con $\theta = (\alpha, \sigma_1, \sigma_2)$.

Riparametrizziamo il modello in $\theta = (\psi, \lambda)$, dove il parametro di interesse ψ è dato dalla (3.5) e $\lambda = (\sigma_1^\alpha \sigma_2^\alpha) / (\sigma_1^\alpha + \sigma_2^\alpha)$ è il parametro di disturbo. La stima di massima verosimiglianza della (3.5) è data da

$$\hat{\psi} = \psi(\hat{\theta}) = \frac{\hat{\sigma}_2^\alpha}{\hat{\sigma}_1^\alpha + \hat{\sigma}_2^\alpha},$$

dove α lo ipotizziamo noto mentre $\hat{\sigma}_1^\alpha = n_1^{-1} \sum x_i^\alpha$ e $\hat{\sigma}_2^\alpha = n_2^{-1} \sum y_j^\alpha$.

3.4 Distribuzioni normali multivariate

Una naturale estensione dei casi visti fin ora è quella alle distribuzioni multivariate. Un classico esempio è quello del ponte: si pensi ad un ponte, il quale è soggetto a diverse fonti di stress, ad esempio i pesi delle varie tipologie di veicoli (automobili, camion, moto), e possiede una propria forza, rappresentata da piloni di sostegno, tiranti, ecc. Si pensi dunque a due vettori casuali indipendenti, rispettivamente di k_1 e k_2 componenti, ossia $X = (X^{(1)}, \dots, X^{(k_1)})$ e $Y = (Y^{(1)}, \dots, Y^{(k_2)})$. Si è interessati a studiare $P(A^T X < B^T Y)$, ovvero la probabilità che un'opportuna combinazione lineare delle componenti di X , con vettore di coefficienti A , sia minore di un'opportuna combinazione lineare di Y , con vettore dei coefficienti B . In alternativa si potrebbe essere interessati a studiare la probabilità che la forza ecceda lo stress per una particolare costante C fissata, in altri termini $P(A^T X + B^T Y + C > 0)$.

Si nota facilmente che se $k_1 = k_2 = 1, C = 0, A = (-1)$ e $B = 1$ il problema si riduce alla stima di $P(X < Y)$ con X e Y variabili casuali indipendenti. Se invece $k_1 = 2, A = (-1, 1)$ e $B = 0$ allora siamo nel caso in cui X e Y sono variabili casuali dipendenti.

Supponiamo ora che X e Y siano due variabili casuali multivariate e indipendenti tra loro, con vettore delle medie μ_1 e μ_2 e matrici di varianze e covarianze Σ_1 e Σ_2 , rispettivamente. La probabilità $P(X < Y)$ è (Kotz et al., 2003)

$$\psi = \psi(\theta) = \Phi\left(\frac{c}{\sqrt{a^2 + b^2}}\right), \quad (3.6)$$

dove $a = \sqrt{A^T \Sigma_1 A}$, $b = \sqrt{B^T \Sigma_2 B}$ e $c = A^T \mu_1 + B^T \mu_2 + C$.

La SMV della (3.6), nel caso in cui tutti i parametri siano ignoti e quindi da stimare, è

$$\hat{\psi} = \psi(\hat{\theta}) = \Phi\left(\frac{A^T \hat{\mu}_1 + B^T \hat{\mu}_2 + C}{\sqrt{n_1^{-1} A^T S_X A + n_2^{-1} B^T S_Y B}}\right),$$

dove $\hat{\mu}_1 = n_1^{-1} \sum X_i$, $\hat{\mu}_2 = n_2^{-1} \sum Y_i$ per ogni componente dei vettori X, Y , e $S_X = \sum (X_i - \hat{\mu}_1)(X_i - \hat{\mu}_1)^T$ e $S_Y = \sum (Y_i - \hat{\mu}_2)(Y_i - \hat{\mu}_2)^T$.

3.5 Variabili X e Y dipendenti da covariate

Nei casi reali, soprattutto nelle moderne applicazioni mediche, si può ipotizzare che forza e stress siano influenzati da altri fattori, ovvero da variabili esplicative, il cui ruolo nell'analisi diventa fondamentale. Supponiamo che X dipenda da k_1 variabili esplicative contenute nel vettore z_1 e Y dipenda da k_2 variabili esplicative contenute nel vettore z_2 , ossia $X = \beta_1^T z_1 + \epsilon_1$, $Y = \beta_2^T z_2 + \epsilon_2$, dove β_1 e β_2 sono i vettori dei coefficienti di regressione, e ϵ_1 ed ϵ_2 sono variabili casuali con una specifica distribuzione.

3.5.1 Caso normale

Supponiamo che ϵ_1 e ϵ_2 siano due variabili casuali indipendenti e normalmente distribuite, con medie zero e varianze σ_1^2 e σ_2^2 , rispettivamente. Suppo-

niamo inoltre che siano disponibili i vettori delle osservazioni $x = (x_1, \dots, x_{n_1})$ e $y = (y_1, \dots, y_{n_2})$. Allora $P(X < Y)$ è (Guttman et al., 1988)

$$\psi = \psi(\theta) = \Phi\left(\frac{\beta_2^T z_2 - \beta_1^T z_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right). \quad (3.7)$$

Nel caso in cui X e Y hanno stessa varianza, allora la SMV della (3.7) è data da

$$\hat{\psi} = \psi(\hat{\theta}) = \Phi\left(\frac{\hat{\beta}_2^T z_2 - \hat{\beta}_1^T z_1}{S\sqrt{2}}\right),$$

dove $\hat{\beta}_1 = (Z_1^T Z_1)^{-1} Z_1^T x$, $\hat{\beta}_2 = (Z_2^T Z_2)^{-1} Z_2^T y$, e $(N - k)S^2 = (n_1 - k_1)S_1^2 + (n_2 - k_2)S_2^2$, con $N = n_1 + n_2$ e $k = k_1 + k_2$, dove $(n_1 - k_1)S_1^2 = X^T X - \beta_1^T Z_1^T X$ e $(n_2 - k_2)S_2^2 = Y^T Y - \beta_2^T Z_2^T Y$.

Se invece X e Y hanno varianze diverse, allora la SMV della (3.7) diventa

$$\hat{\psi} = \psi(\hat{\theta}) = \Phi\left(\frac{\hat{\beta}_2^T z_2 - \hat{\beta}_1^T z_1}{\sqrt{S_1^2 + S_2^2}}\right).$$

Capitolo 4

Studi di simulazione

4.1 Introduzione

La teoria vista fino ad ora è servita per fare chiarezza sui modelli oggetto di questa tesi. Dopo aver visto cosa si intende per modelli stress-strength e dato alcuni concetti riguardanti l'inferenza tramite verosimiglianza, passiamo allo studio del loro comportamento in termini di stima, nei casi in cui si ha una bassa numerosità campionaria, applicando sia la teoria asintotica del primo ordine sia quella di ordine superiore. A tale scopo in questo capitolo viene presentato uno studio di simulazione che ci permetterà di investigare sull'accuratezza di stima dei modelli stress-strength; in particolare saranno valutate le coperture empiriche degli intervalli di confidenza alla Wald, di quelli basati sulla statistica r_p e sulla sua modifica r_p^* . L'obiettivo è verificare se tali valori sono vicini ai valori nominali (0.90, 0.95, 0.99) e quanto accurata è l'approssimazione alla normale delle statistiche sopracitate. I valori di ψ testati sono 0.5, 0.8 e 0.95 per numerosità campionarie pari a 5, 10 e 30. I casi trattati sono modelli stress-strength con variabili casuali esponenziali, normali indipendenti con varianze uguali, normali indipendenti con varianze diverse, normali dipendenti con varianze diverse e Weibull con stesso parametro di scala. Ogni simulazione prevede 2000 ripetizioni, mentre per il calcolo delle covarianze necessarie al calcolo di r_p^* , ottenute tramite integrazione Monte Carlo, sono state utilizzate 1000 ripetizioni. Tra parentesi vengono indicati

gli errori di copertura, sinistro e destro.

4.2 Esponenziali

Nel primo studio di simulazione si vuole stimare $P(X < Y)$, dove $X \sim Exp(\alpha)$ e $Y \sim Exp(\beta)$. La Tabella 4.1 mostra i risultati delle simulazioni effettuate considerando il vero valore di ψ pari a 0.5. L'intervallo alla Wald è quello che fornisce coperture meno soddisfacenti con bassa numerosità campionaria, anche se già con $n_1 = n_2 = 10$ i risultati migliorano. L'intervallo basato su r_p^* presenta coperture empiriche più vicine ai livelli nominali rispetto all'intervallo basato su r_p con $n_1 = n_2 = 5$ e pertanto risulta preferibile mentre con numerosità maggiori i due intervalli si equivalgono. In generale gli errori di copertura sono abbastanza simmetrici.

Tabella 4.1: Livelli di copertura empirica con $\psi = 0.5$, $\alpha = 1$, $\beta = 1$

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.873 (0.064, 0.0625)	0.928 (0.0355, 0.0365)	0.981 (0.011, 0.008)
	r_p	0.881 (0.0565, 0.062)	0.936 (0.0315, 0.0325)	0.984 (0.065, 0.0095)
	r_p^*	0.89 (0.052, 0.0575)	0.941 (0.029, 0.0295)	0.985 (0.006, 0.0095)
(10, 10)	w_p	0.894 (0.05, 0.056)	0.95 (0.022, 0.0275)	0.985 (0.007, 0.0075)
	r_p	0.899 (0.0535, 0.0475)	0.954 (0.025, 0.021)	0.988 (0.006, 0.006)
	r_p^*	0.902 (0.052, 0.0455)	0.956 (0.023, 0.021)	0.989 (0.0065, 0.0065)
(30, 30)	w_p	0.902 (0.0553, 0.0426)	0.946 (0.0236, 0.02)	0.992 (0.0043, 0.003)
	r_p	0.906 (0.0406, 0.053)	0.956 (0.0196, 0.0236)	0.993 (0.003, 0.004)
	r_p^*	0.907 (0.0406, 0.0523)	0.955 (0.0206, 0.0236)	0.993 (0.004, 0.0043)

Le Tabelle 4.2 e 4.3 mostrano i risultati relativi alle simulazioni condotte fissando il vero valore di ψ pari a 0.8 e 0.95. Anche in questi casi gli intervalli alla Wald si dimostrano poco precisi, con livelli di copertura empirica al di sotto dei valori nominali, con basse numerosità. Gli intervalli basati su r_p e r_p^* invece sono leggermente più accurati in termini di copertura, seppure i loro livelli empirici restano un pò al di sotto di quelli teorici, e in generale sembrano essere equivalenti. Gli errori di copertura non sempre risultano simmetrici.

Tabella 4.2: Livelli di copertura empirica con $\psi = 0.8$, $\alpha = 1$, $\beta = 0.25$

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.887 (0.0576, 0.0553)	0.944 (0.0313, 0.0243)	0.985 (0.0086, 0.006)
	r_p	0.898 (0.049, 0.0526)	0.95 (0.022, 0.0273)	0.988 (0.0053, 0.006)
	r_p^*	0.902 (0.0483, 0.049)	0.95 (0.022, 0.0273)	0.989 (0.0043, 0.0063)
(10, 10)	w_p	0.88 (0.0623, 0.0573)	0.935 (0.0366, 0.0276)	0.985 (0.009, 0.006)
	r_p	0.885 (0.0553, 0.059)	0.939 (0.026, 0.346)	0.987 (0.0046, 0.0076)
	r_p^*	0.888 (0.0523, 0.0593)	0.939 (0.0256, 0.0356)	0.988 (0.005, 0.0086)
(30, 30)	w_p	0.889 (0.0576, 0.053)	0.945 (0.0283, 0.0263)	0.983 (0.0076, 0.0093)
	r_p	0.891 (0.052, 0.0563)	0.945 (0.0263, 0.0283)	0.984 (0.0086, 0.007)
	r_p^*	0.898 (0.053, 0.0586)	0.942 (0.0273, 0.03)	0.981 (0.010, 0.0083)

Tabella 4.3: Livelli di copertura empirica con $\psi = 0.95$, $\alpha = 1$, $\beta = 0.05$

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.88 (0.068, 0.0513)	0.939 (0.0323, 0.0286)	0.983 (0.0083, 0.0083)
	r_p	0.891 (0.0467, 0.0614)	0.949 (0.023, 0.027)	0.989 (0.0043, 0.0063)
	r_p^*	0.895 (0.0474, 0.0574)	0.949 (0.0233, 0.0267)	0.986 (0.0066, 0.007)
(10, 10)	w_p	0.892 (0.0653, 0.0423)	0.939 (0.038, 0.0223)	0.985 (0.009, 0.0043)
	r_p	0.895 (0.4101, 0.06302)	0.944 (0.02, 0.0353)	0.989 (0.0026, 0.0073)
	r_p^*	0.891 (0.044, 0.0646)	0.95 (0.023, 0.0363)	0.989 (0.0053, 0.0083)
(30, 30)	w_p	0.882 (0.082, 0.0353)	0.94 (0.043, 0.017)	0.984 (0.0106, 0.005)
	r_p	0.885 (0.0346, 0.08)	0.941 (0.016, 0.0426)	0.987 (0.004, 0.009)
	r_p^*	0.888 (0.0343, 0.07)	0.942 (0.015, 0.0413)	0.988 (0.004, 0.008)

Nelle Figure 4.1 e 4.2 vengono mostrati i Normal QQ-Plot relativi ai valori di r_p e r_p^* ottenuti tramite simulazione nei casi in cui i veri valori di ψ sono 0.5 e 0.8 e le numerosità campionarie pari a 5. In generale i punti sembrano disporsi bene sulla diagonale, con piccole deviazioni nelle code destre per quanto riguarda r_p^* , ma comunque l'approssimazione alla normale è buona.

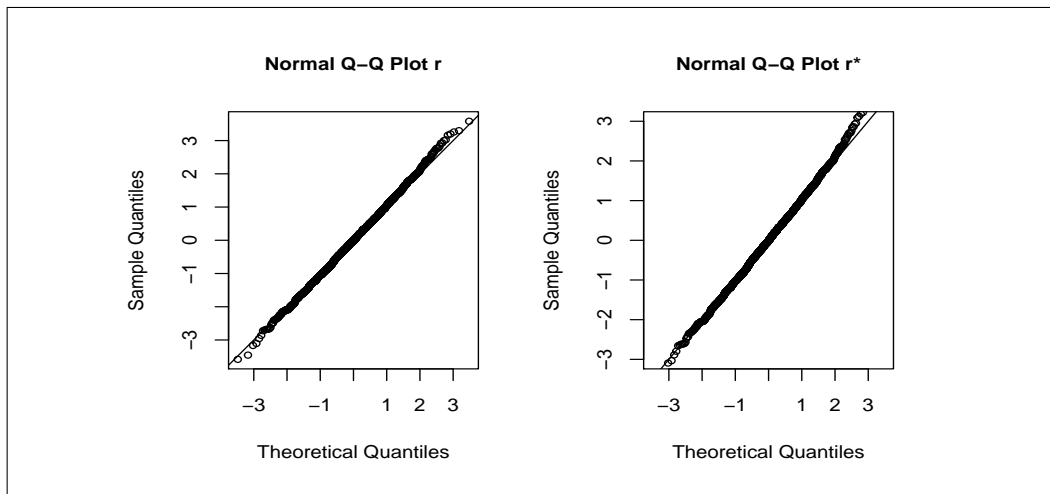


Figura 4.1: Normal qq-plot dei valori di r_p e r_p^* ottenuti tramite simulazione con $n_1 = n_2 = 5$ e $\psi = 0.5$.

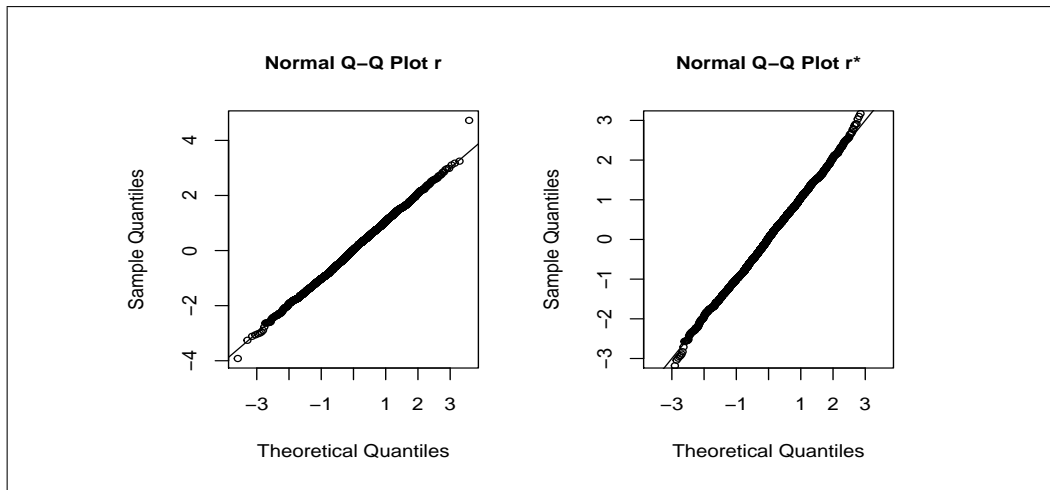


Figura 4.2: Normal qq-plot dei valori di r_p e r_p^* ottenuti tramite simulazione con $n_1 = n_2 = 5$ e $\psi = 0.8$.

4.3 Normali indipendenti con varianze uguali

Il secondo studio di simulazione prevede che $X \sim N(\mu_1, \sigma^2)$ e $Y \sim N(\mu_2, \sigma^2)$, ovvero che le due variabili casuali X e Y abbiano distribuzione normale con medie diverse e varianze uguali. La Tabella 4.4 mostra i risultati delle simulazioni effettuate considerando il vero valore di ψ uguale a 0.5. Per ogni livello di confidenza l'intervallo alla Wald fornisce livelli di copertura empirica lontani dai valori nominali. Gli intervalli basati su r_p e r_p^* forniscono risultati più soddisfacenti, anche se il secondo risulta preferibile in quanto i suoi livelli di copertura sono in linea con i valori nominali, anche con piccoli campioni, mentre quelli del primo sono buoni per $n_1 = n_2 = 30$ ma un po' inferiori per basse numerosità.

Tabella 4.4: Livelli di copertura empirica con $\psi = 0.5$, $\mu_1 = 5$, $\mu_2 = 5$, $\sigma = 1$.

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.835 (0.0930, 0.0719)	0.876 (0.073, 0.051)	0.958 (0.0241, 0.0179)
	r_p	0.877 (0.067, 0.057)	0.929 (0.0341, 0.0372)	0.971 (0.0151, 0.0138)
	r_p^*	0.895 (0.0505, 0.0524)	0.949 (0.0255, 0.0255)	0.989 (0.0050, 0.0059)
(10, 10)	w_p	0.848 (0.05, 0.103)	0.898 (0.0447, 0.0563)	0.957 (0.0231, 0.01999)
	r_p	0.876 (0.0649, 0.0586)	0.933 (0.0315, 0.0352)	0.985 (0.0060, 0.0080)
	r_p^*	0.898 (0.0524, 0.0524)	0.949 (0.0267, 0.0267)	0.989 (0.0066, 0.0060)
(30, 30)	w_p	0.88 (0.035, 0.095)	0.936 (0.039, 0.025)	0.981 (0.0090, 0.01)
	r_p	0.889 (0.0608, 0.05)	0.947 (0.0308, 0.0216)	0.985 (0.0083, 0.0066)
	r_p^*	0.898 (0.0575, 0.0516)	0.952 (0.0261, 0.0271)	0.989 (0.0059, 0.0551)

Nelle Tabelle 4.5 e 4.6 vengono mostrati i risultati delle simulazioni

4.3 Normali indipendenti con varianze uguali

effettuate con il vero valore di ψ fissato a 0.8 e 0.95. Anche in questi casi l'intervallo alla Wald risulta il peggiore in termini di copertura empirica. Gli intervalli basati su r_p e r_p^* si comportano meglio, con una leggera preferenza per il secondo, soprattutto quando la numerosità campionaria è bassa. Nel caso $\psi = 0.95$ l'intervallo basato su r_p presenta valori di copertura empirica inferiori a quelli nominali, in particolare con numerosità pari a 5 e 10, mentre l'intervallo basato su r_p^* ha valori più vicini a quelli teorici. In generale gli errori di copertura di w_p e r_p risultano un po' asimmetrici, soprattutto con basse numerosità, mentre quelli di r_p^* risultano più o meno simmetrici.

Tabella 4.5: Livelli di copertura empirica con $\psi = 0.8$, $\mu_1 = 5$, $\mu_2 = 6.55$, $\sigma = 1.3$.

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.847 (0.0731, 0.0801)	0.91 (0.0438, 0.0461)	0.942 (0.042, 0.0275)
	r_p	0.886 (0.0571, 0.0568)	0.938 (0.0325, 0.0308)	0.984 (0.0085, 0.0075)
	r_p^*	0.89 (0.0557, 0.0553)	0.944 (0.0284, 0.0286)	0.989 (0.0055, 0.0055)
(10, 10)	w_p	0.855 (0.0461, 0.0689)	0.913 (0.0328, 0.0543)	0.951 (0.015, 0.0349)
	r_p	0.885 (0.0623, 0.052)	0.943 (0.038, 0.0186)	0.982 (0.0015, 0.0066)
	r_p^*	0.888 (0.0533, 0.058)	0.949 (0.02548, 0.02552)	0.987 (0.0062, 0.0058)
(30, 30)	w_p	0.864 (0.0587, 0.0783)	0.911 (0.0551, 0.0339)	0.965 (0.007, 0.028)
	r_p	0.891 (0.0391, 0.0691)	0.939 (0.0233, 0.0375)	0.993 (0.0015, 0.0048)
	r_p^*	0.903 (0.0476, 0.0495)	0.95 (0.0255, 0.0255)	0.993 (0.005, 0.0041)

Nelle Figure 4.3 e 4.4 vengono mostrati i Normal QQ-Plot relativi ai valori di r_p e r_p^* ottenuti tramite simulazione nei casi in cui $\psi = 0.5$ e $\psi = 0.95$ con numerosità campionaria pari a 10. In entrambi i casi i punti si distribuiscono

Capitolo 4. Studi di simulazione

Tabella 4.6: Livelli di copertura empirica con $\psi = 0.95$, $\mu_1 = 5$, $\mu_2 = 8.5$, $\sigma = 1.5$.

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.803 (0.0783, 0.1197)	0.865 (0.0478, 0.0873)	0.901 (0.0475, 0.0505)
	r_p	0.834 (0.0326, 0.1327)	0.896 (0.0213, 0.082)	0.973 (0.0036, 0.0226)
	r_p^*	0.89 (0.0543, 0.0550)	0.945 (0.0270, 0.0277)	0.99 (0.0043, 0.0047)
(10, 10)	w_p	0.848 (0.0864, 0.0656)	0.902 (0.0388, 0.0593)	0.95 (0.048, 0.002)
	r_p	0.867 (0.0306, 0.102)	0.929 (0.014, 0.0563)	0.982 (0.0023, 0.0156)
	r_p^*	0.898 (0.0496, 0.0516)	0.948 (0.0237, 0.0283)	0.988 (0.0057, 0.0063)
(30, 30)	w_p	0.88 (0.0505, 0.0694)	0.937 (0.0218, 0.0412)	0.971 (0.0138, 0.0153)
	r_p	0.892 (0.034, 0.074)	0.947 (0.0176, 0.0353)	0.989 (0.0033, 0.0073)
	r_p^*	0.899 (0.0476, 0.0514)	0.949 (0.0243, 0.0263)	0.991 (0.0046, 0.0043)

bene lungo la diagonale, seppur con qualche lieve differenza nelle code, segno che l'approssimazione normale di r_p e r_p^* è adeguata.

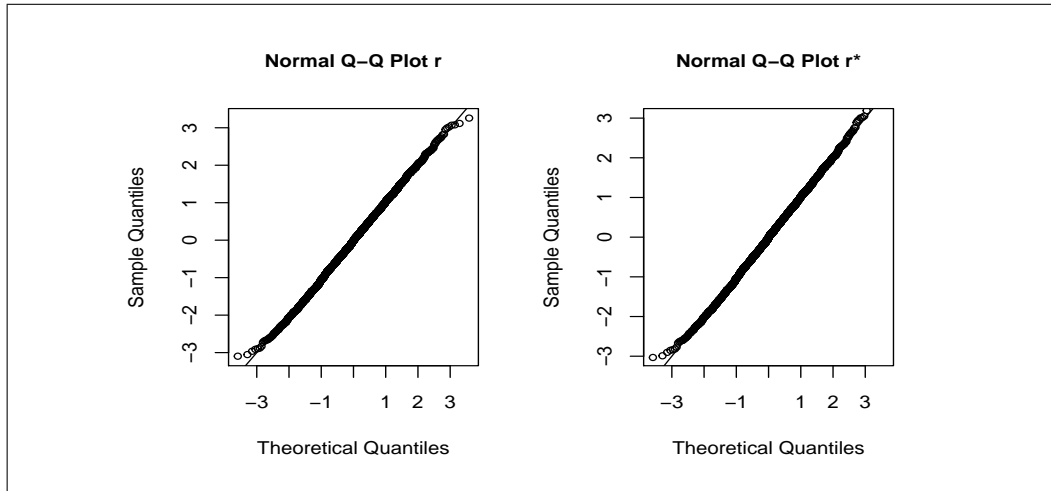


Figura 4.3: Normal qq-plot dei valori di r_p e r_p^* ottenuti tramite simulazione con $n_1 = n_2 = 10$ e $\psi = 0.5$.

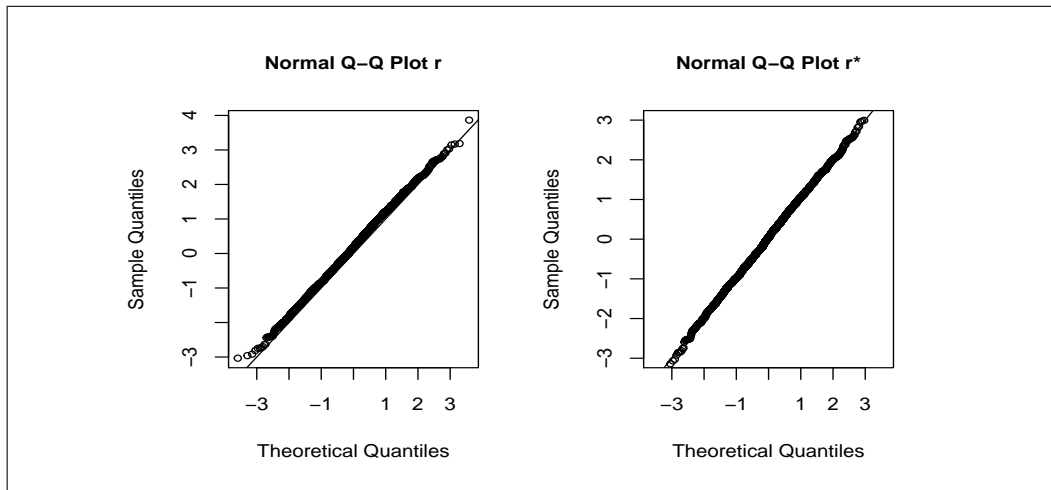


Figura 4.4: Normal qq-plot dei valori di r_p e r_p^* ottenuti tramite simulazione con $n_1 = n_2 = 10$ e $\psi = 0.95$.

4.4 Normali indipendenti con varianze diverse

Il terzo studio di simulazione prevede una situazione più realistica, in quanto X e Y sono due variabili casuali indipendenti e normalmente distribuite con medie e varianze diverse, ovvero $X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$. La Tabella 4.7 mostra i risultati delle simulazioni effettuate con il vero valore di ψ pari a 0.5. In generale l'intervallo di confidenza alla Wald fornisce dei livelli di copertura empirica lontani dai valori nominali, in particolare per basse numerosità campionarie. L'intervallo basato su r_p fornisce livelli empirici vicini a quelli nominali solo con numerosità pari a 30 mentre con piccoli campioni risulta impreciso. L'intervallo basato su r_p^* invece fornisce livelli empirici di copertura in linea con i valori teorici già con basse numerosità e pertanto risulta preferibile.

Tabella 4.7: Livelli di copertura empirica con $\psi = 0.5$, $\mu_1 = 5$, $\mu_2 = 5$, $\sigma_1 = 0.5$, $\sigma_2 = 1.5$.

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.808 (0.1142, 0.0778)	0.859 (0.1009, 0.0402)	0.902 (0.0635, 0.0345)
	r_p	0.865 (0.0528, 0.0822)	0.921 (0.0245, 0.0546)	0.964 (0.021, 0.015)
	r_p^*	0.891 (0.0515, 0.0575)	0.953 (0.024, 0.023)	0.989 (0.0057, 0.0053)
(10, 10)	w_p	0.85 (0.09, 0.06)	0.902 (0.0725, 0.0255)	0.961 (0.0013, 0.0377)
	r_p	0.876 (0.0642, 0.0597)	0.939 (0.0241, 0.0395)	0.979 (0.012, 0.0091)
	r_p^*	0.893 (0.0536, 0.0534)	0.957 (0.0205, 0.0225)	0.995 (0.0028, 0.0022)
(30, 30)	w_p	0.884 (0.0523, 0.0637)	0.939 (0.0278, 0.03320)	0.977 (0.0018, 0.0050)
	r_p	0.903 (0.0653, 0.0317)	0.95 (0.03, 0.02)	0.987 (0.0076, 0.0054)
	r_p^*	0.91 (0.048, 0.043)	0.955 (0.0225, 0.0225)	0.989 (0.0051, 0.0059)

4.4 Normali indipendenti con varianze diverse

La Tabella 4.8 mostra i risultati delle simulazioni condotte con il vero valore di ψ pari a 0.8. L'intervallo alla Wald è quello che in generale fornisce i risultati peggiori, con livelli di copertura empirica lontani dai valori nominali. L'intervallo basato su r_p è più soddisfacente in termini di copertura seppure nei casi con piccoli campioni ($n_1 = n_2 = 5$ e $n_1 = n_2 = 10$) i livelli non sono ottimali. Invece l'intervallo basato su r_p^* fornisce, già con bassa numerosità campionaria, livelli empirici in linea con quelli teorici, e pertanto risulta preferibile.

Tabella 4.8: Livelli di copertura empirica con $\psi = 0.8$, $\mu_1 = 5$, $\mu_2 = 6.33$, $\sigma_1 = 0.5$, $\sigma_2 = 1.5$.

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.821 (0.1152, 0.0638)	0.905 (0.0235, 0.0751)	0.944 (0.0147, 0.0413)
	r_p	0.881 (0.0549, 0.0641)	0.931 (0.045, 0.024)	0.966 (0.0213, 0.0127)
	r_p^*	0.887 (0.0552, 0.0578)	0.949 (0.0262, 0.0248)	0.97 (0.020, 0.010)
(10, 10)	w_p	0.864 (0.0897, 0.0463)	0.926 (0.0548, 0.0192)	0.96 (0.0035, 0.0365)
	r_p	0.893 (0.0328, 0.0742)	0.934 (0.0466, 0.0195)	0.979 (0.0018, 0.0193)
	r_p^*	0.903 (0.0505, 0.0465)	0.967 (0.017, 0.016)	0.993 (0.0035, 0.0035)
(30, 30)	w_p	0.863 (0.0152, 0.1218)	0.915 (0.0541, 0.031)	0.97 (0.0236, 0.0064)
	r_p	0.89 (0.0588, 0.0512)	0.945 (0.0305, 0.0245)	0.987 (0.0067, 0.0063)
	r_p^*	0.90 (0.0518, 0.0482)	0.951 (0.025, 0.025)	0.987 (0.006, 0.007)

La Tabella 4.9 mostra i risultati delle simulazioni con vero valore di ψ pari a 0.95. Anche in questo caso l'intervallo alla Wald è il peggiore in termini di copertura con livelli empirici nettamente al di sotto di quelli nominali. In realtà si nota una certa difficoltà di tutti e tre i metodi di

mantenere i livelli di copertura empirica in linea con quelli teorici quando la numerosità campionaria è bassa e probabilmente tale difficoltà può essere data dal fatto che il vero valore di ψ è molto estremo, vicino a 1, per cui occorre una numerosità campionaria più elevata per ottenere risultati più accurati. Comunque già con $n_1 = n_2 = 30$ la situazione migliora e l'intervallo basato su r_p^* risulta ancora una volta preferibile agli altri due. In generale gli errori di copertura risultano simmetrici per r_p^* e piuttosto asimmetrici per w_p e r_p .

Tabella 4.9: Livelli di copertura empirica con $\psi = 0.95$, $\mu_1 = 5$, $\mu_2 = 5$, $\sigma_1 = 0.5$, $\sigma_2 = 1.5$.

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.741 (0.1025, 0.1565)	0.889 (0.0421, 0.069)	0.953 (0.0128, 0.0342)
	r_p	0.874 (0.1026, 0.0234)	0.932 (0.0228, 0.0452)	0.966 (0.020, 0.014)
	r_p^*	0.881 (0.0611, 0.0579)	0.939 (0.020, 0.040)	0.979 (0.0112, 0.0098)
(10, 10)	w_p	0.854 (0.0237, 0.1223)	0.909 (0.0244, 0.0666)	0.962 (0.0267, 0.0113)
	r_p	0.875 (0.0586, 0.0664)	0.934 (0.034, 0.032)	0.969 (0.011, 0.020)
	r_p^*	0.889 (0.0542, 0.0568)	0.94 (0.024, 0.036)	0.979 (0.0117, 0.0093)
(30, 30)	w_p	0.891 (0.0788, 0.0302)	0.924 (0.0129, 0.0631)	0.963 (0.0255, 0.0115)
	r_p	0.909 (0.056, 0.0350)	0.944 (0.0237, 0.0323)	0.982 (0.008, 0.011)
	r_p^*	0.908 (0.045, 0.0045)	0.955 (0.0232, 0.00218)	0.991 (0.0046, 0.0044)

Nelle Figure 4.6 e 4.7 vengono mostrati i Normal QQ-Plot dei valori di r_p e di r_p^* ottenuti tramite le simulazioni; in particolare in Figura 4.6 i valori rappresentati sono quelli risultanti dalle simulazioni effettuate con $n_1 = n_2 = 30$ e vero valore di ψ pari a 0.5, mentre nella Figura 4.7 $n_1 = n_2 = 30$ e $\psi = 0.95$. I punti si dispongono bene sulla diagonale per cui possiamo essere

4.4 Normali indipendenti con varianze diverse

soddisfatti dell'approssimazione normale di r_p e di r_p^* , seppure i punti del primo deviano leggermente nella coda sinistra.

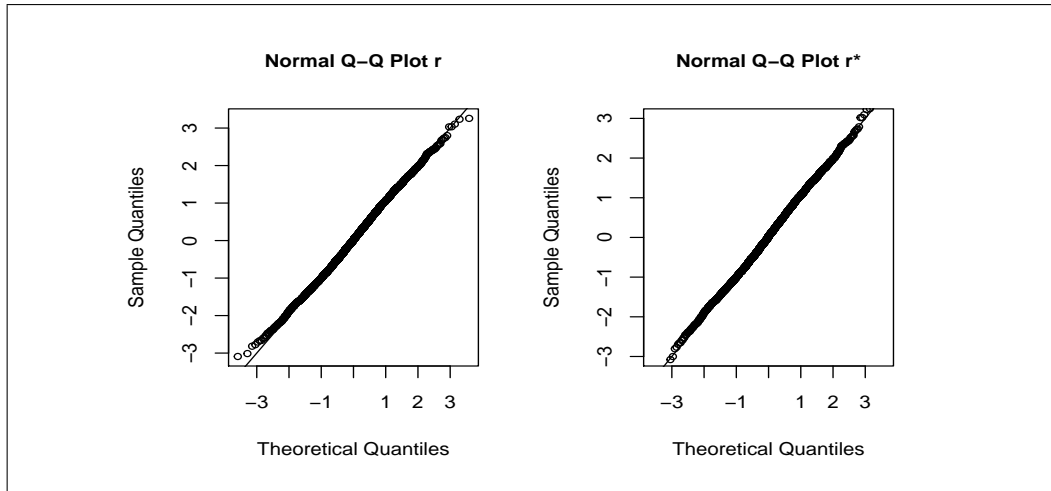


Figura 4.5: Normal qq-plot dei valori di r_p e r_p^* ottenuti tramite simulazione con $n_1 = n_2 = 30$ e $\psi = 0.5$.

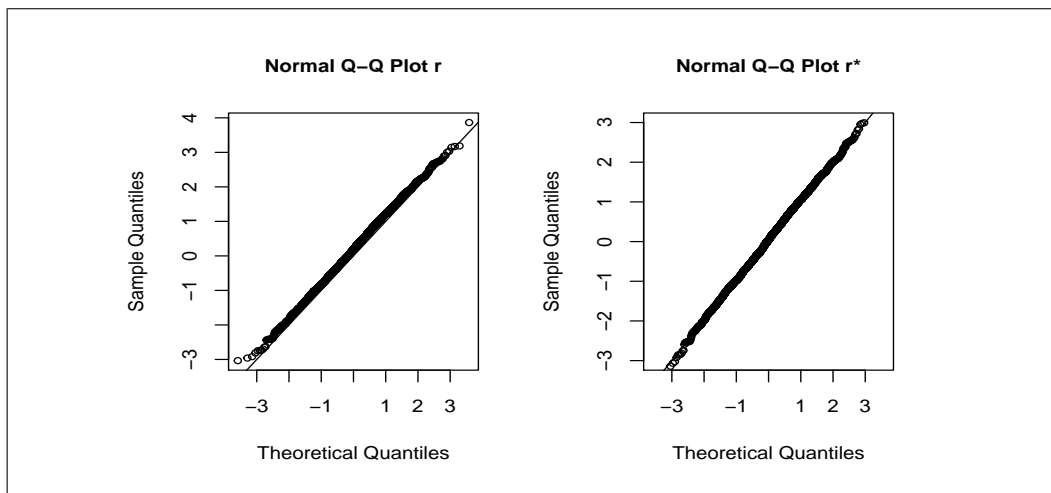


Figura 4.6: Normal qq-plot dei valori di r_p e r_p^* ottenuti tramite simulazione con $n_1 = n_2 = 30$ e $\psi = 0.95$.

4.5 Normali dipendenti

In questo studio di simulazione esaminiamo il caso in cui X e Y siano due variabili casuali normali, rispettivamente di medie μ_1, μ_2 e varianze σ_1^2, σ_2^2 , dipendenti tra loro. La Tabella 4.10 mostra i risultati delle simulazioni condotte considerando il vero valore di ψ pari a 0.5 e con coefficiente di correlazione pari a 0.8. Si nota che in generale l'intervallo alla Wald è quello meno preciso in termini di copertura con livelli empirici lontani da quelli teorici, in particolare con basse numerosità campionarie. L'intervallo basato su r_p^* è invece quello che fornisce risultati migliori anche con piccoli campioni, infatti i suoi livelli di copertura empirica sono in linea con quelli nominali già con $n_1 = n_2 = 5$ e quindi risulta preferibile anche all'intervallo basato su r_p , che fornisce risultati apprezzabili solo con una numerosità maggiore. Le Tabelle 4.11 e 4.12 mostrano i risultati delle simulazioni dove i veri valori di ψ sono 0.8 e 0.95. Anche in questi casi l'intervallo alla Wald fornisce livelli di copertura empirica molto distanti da quelli nominali, soprattutto con basse numerosità campionarie. Nel caso con $\psi = 0.8$ l'intervallo basato su r_p^* si dimostra abbastanza preciso, con livelli empirici in linea con quelli teorici già con piccoli campioni e in generale risulta preferibile agli altri due; con $\psi = 0.95$ mostra una minore accuratezza per $n_1 = n_2 = 5$, seppure rimane il più preciso, mentre all'aumentare della numerosità campionaria i livelli empirici di copertura rispecchiano quelli teorici. L'intervallo basato su r_p risulta apprezzabile con numerosità maggiori o uguali a 10, ed è comunque preferibile all'intervallo alla Wald con numerosità pari a 5. Si nota come per valori estremi di ψ tutti e tre i tipi di intervallo siano poco precisi in caso di piccoli campioni. In generale, gli errori di copertura di w_p e di r_p risultano asimmetrici, mentre quelli di r_p^* sono abbastanza simmetrici.

Tabella 4.10: Livelli di copertura empirica con $\psi = 0.5$, $\mu_1 = 5$, $\mu_2 = 5$, $\sigma_1 = 0.5$, $\sigma_2 = 1.5$, $\rho = 0.8$.

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.832 (0.1025, 0.0655)	0.886 (0.0456, 0.0685)	0.953 (0.0124, 0.0346)
	r_p	0.889 (0.0781, 0.0329)	0.937 (0.0415, 0.0215)	0.976 (0.0053, 0.0188)
	r_p^*	0.897 (0.0409, 0.0621)	0.949 (0.0214, 0.0269)	0.981 (0.01, 0.009)
(10, 10)	w_p	0.847 (0.0637, 0.0893)	0.901 (0.0358, 0.0632)	0.959 (0.0166, 0.0244)
	r_p	0.883 (0.0754, 0.0415)	0.937 (0.0478, 0.0152)	0.987 (0.0087, 0.0044)
	r_p^*	0.906 (0.05, 0.044)	0.958 (0.027, 0.016)	0.995 (0.0022, 0.0028)
(30, 30)	w_p	0.875 (0.0783, 0.0466)	0.925 (0.0206, 0.0544)	0.975 (0.0128, 0.0122)
	r_p	0.895 (0.0742, 0.0308)	0.945 (0.025, 0.029)	0.986 (0.0068, 0.0071)
	r_p^*	0.901 (0.04, 0.06)	0.955 (0.022, 0.023)	0.99 (0.0042, 0.0058)

1

Dai Normal Q-Q Plot dei valori di r_p e r_p^* , mostrati nelle Figure 4.8 e 4.9, si nota che i punti di punti relativi a r_p in un paio di tratti non si dispongono perfettamente sulla diagonale, mentre i punti relativi a r_p^* hanno una migliore disposizione. Si potrebbe concludere che, in questo caso, l'approssimazione alla normale di r_p^* è migliore di quella di r_p .

Tabella 4.11: Livelli di copertura empirica con $\psi = 0.8$, $\mu_1 = 5$, $\mu_2 = 6.3$, $\sigma_1 = 0.4$, $\sigma_2 = 1.8$, $\rho = 0.8$.

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.834 (0.1023, 0.0637)	0.866 (0.0896, 0.0444)	0.903 (0.0145, 0.0825)
	r_p	0.887 (0.0632, 0.0498)	0.942 (0.0258, 0.0323)	0.976 (0.0123, 0.0118)
	r_p^*	0.897 (0.0529, 0.0501)	0.951 (0.026, 0.023)	0.993 (0.006, 0.008)
(10, 10)	w_p	0.859 (0.0271, 0.114)	0.903 (0.0866, 0.0104)	0.959 (0.0125, 0.0285)
	r_p	0.89 (0.0758, 0.0342)	0.941 (0.03, 0.03)	0.986 (0.0086, 0.0054)
	r_p^*	0.901 (0.048, 0.052)	0.955 (0.023, 0.022)	0.989 (0.006, 0.005)
(30, 30)	w_p	0.86 (0.0923, 0.0477)	0.908 (0.0561, 0.036)	0.958 (0.0122, 0.03)
	r_p	0.895 (0.0369, 0.0681)	0.947 (0.0344, 0.0187)	0.989 (0.0050, 0.060)
	r_p^*	0.905 (0.048, 0.047)	0.957 (0.022, 0.023)	0.992 (0.005, 0.005)

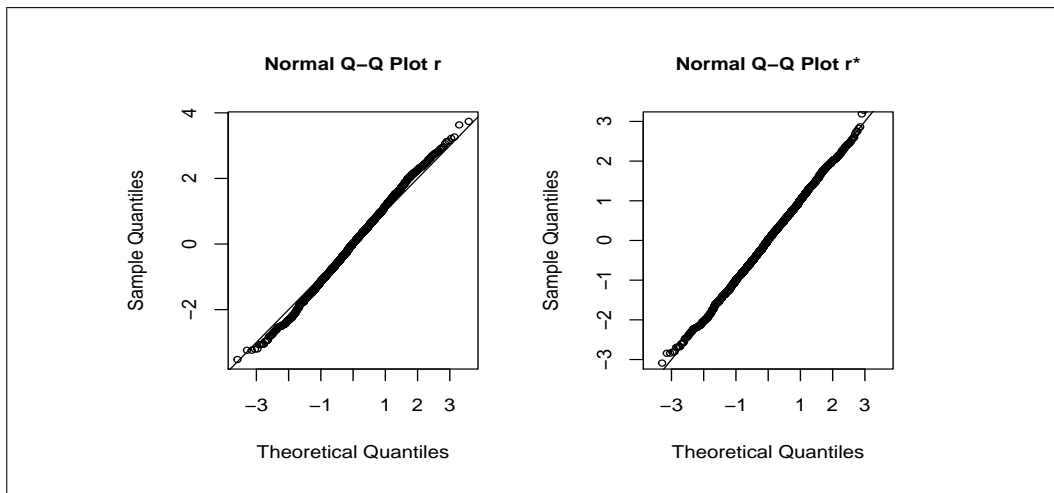


Figura 4.7: Normal qq-plot dei valori di r_p e r_p^* ottenuti tramite simulazione con $n_1 = n_2 = 5$ e $\psi = 0.8$.

Tabella 4.12: Livelli di copertura empirica con $\psi = 0.95$, $\mu_1 = 5$, $\mu_2 = 7.5$, $\sigma_1 = 0.4$, $\sigma_2 = 1.8$, $\rho = 0.8$.

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.811 (0.125, 0.064)	0.875 (0.0235, 0.1015)	0.912 (0.0611, 0.0179)
	r_p	0.854 (0.0422, 0.1037)	0.891 (0.0856, 0.0235)	0.933 (0.0213, 0.0457)
	r_p^*	0.877 (0.0722, 0.0508)	0.921 (0.07, 0.009)	0.963 (0.02, 0.016)
(10, 10)	w_p	0.857 (0.1088, 0.0342)	0.90 (0.07, 0.03)	0.943 (0.0176, 0.0395)
	r_p	0.881 (0.0748, 0.0442)	0.921 (0.056, 0.023)	0.976 (0.0023, 0.0217)
	r_p^*	0.889 (0.054, 0.057)	0.942 (0.03, 0.03)	0.987 (0.0035, 0.0095)
(30, 30)	w_p	0.87 (0.0184, 0.1117)	0.934 (0.0452, 0.0208)	0.968 (0.0287, 0.0033)
	r_p	0.895 (0.0631, 0.0419)	0.941 (0.025, 0.0345)	0.98 (0.008, 0.012)
	r_p^*	0.905 (0.048, 0.047)	0.953 (0.025, 0.022)	0.989 (0.0055, 0.0055)

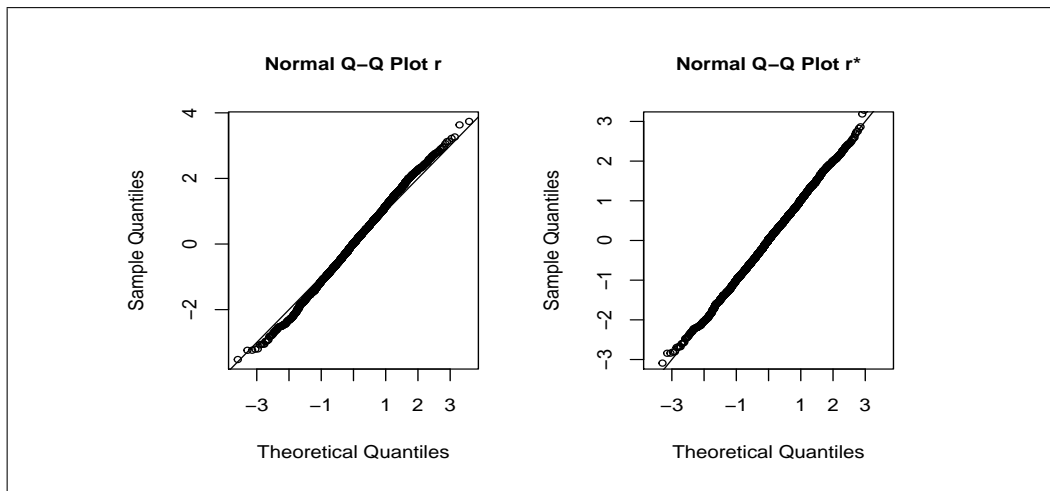


Figura 4.8: Normal qq-plot dei valori di r_p e r_p^* ottenuti tramite simulazione con $n_1 = n_2 = 5$ e $\psi = 0.95$.

4.6 Weibull

Nell'ultimo studio simulativo analizzeremo il caso in cui X e Y siano due variabili casuali indipendenti di tipo Weibull con stesso parametro di forma α e parametri di scala σ_1 e σ_2 , rispettivamente. La Tabella 4.13 mostra i risultati delle simulazioni condotte fissando il vero valore di ψ pari a 0.5.

Tabella 4.13: Livelli di copertura empirica con $\psi = 0.5$, $\alpha = 1$, $\sigma_1 = 0.5$, $\sigma_2 = 0.5$.

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.827 (0.0762, 0.0978)	0.898 (0.0456, 0.0514)	0.964 (0.0166, 0.0284)
	r_p	0.829 (0.0813, 0.0503)	0.901 (0.0473, 0.0163)	0.968 (0.014, 0.0207)
	r_p^*	0.88 (0.059, 0.061)	0.928 (0.0399, 0.0321)	0.971 (0.017, 0.02)
(10, 10)	w_p	0.862 (0.0863, 0.0537)	0.93 (0.053, 0.0316)	0.971 (0.0051, 0.0239)
	r_p	0.873 (0.0654, 0.0909)	0.93 (0.055, 0.015)	0.982 (0.0058, 0.0124)
	r_p^*	0.897 (0.0501, 0.0499)	0.943 (0.029, 0.0271)	0.985 (0.0072, 0.0081)
(30, 30)	w_p	0.885 (0.0616, 0.0523)	0.937 (0.0343, 0.0276)	0.986 (0.008, 0.005)
	r_p	0.895 (0.0505, 0.0549)	0.946 (0.03, 0.0238)	0.988 (0.0035, 0.0029)
	r_p^*	0.901 (0.0495, 0.0485)	0.95 (0.025, 0.025)	0.993 (0.003, 0.004)

Come si vede, in generale l'intervallo alla Wald fornisce i risultati peggiori, con livelli di copertura empirica più bassi di quelli nominali, in particolare con piccoli campioni. Anche l'intervallo basato su r_p non risulta molto preciso. La situazione migliore aumentando la numerosità campionaria. Ancora una volta l'intervallo basato su r_p^* risulta quello preferibile, almeno in termini di copertura, in particolare con bassa numerosità campionaria, seppure anche

i suoi livelli sono un po' al di sotto dei valori teorici. All' aumentare della numerosità campionaria i livelli empirici sono in linea con quelli nominali.

Le Tabelle 4.14 e 4.15 mostrano i risultati delle simulazioni dove il vero valore di ψ è stato fissato a 0.8 e 0.95. In entrambi i casi emergono le difficoltà dell'intervallo alla Wald, che presenta livelli di copertura empirica al di sotto dei valori nominali. Invece l'intervallo basato su r_p^* presenta livelli empirici in linea con quelli teorici già con numerosità pari a 5, e risulta quindi quello preferibile. L'intervallo basato su r_p presenta discreti risultati con numerosità maggiore o uguale a 10 ma pecca di precisione con $n_1 = n_2 = 5$. In generale gli errori di copertura di w_p e r_p risultano asimmetrici, mentre quelli di r_p^* sono abbastanza simmetrici.

Tabella 4.14: Livelli di copertura empirica con $\psi = 0.8$, $\alpha = 1$, $\sigma_1 = 0.5$, $\sigma_2 = 2$.

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.83	0.904	0.968
		(0.1156, 0.0543)	(0.0613, 0.0343)	(0.026, 0.006)
	r_p	0.841	0.909	0.974
(10, 10)	w_p	(0.029, 0.129)	(0.0081, 0.0822)	(0.0000, 0.026)
		r_p^*	0.906	0.964
	r_p^*	(0.0451, 0.0487)	(0.016, 0.02)	(0.0038, 0.0042)
(30, 30)	w_p	0.867	0.936	0.977
		(0.0836, 0.0493)	(0.0373, 0.0263)	(0.0217, 0.0013)
	r_p	0.869	0.937	0.988
(30, 30)	w_p	(0.0434, 0.0868)	(0.0213, 0.0417)	(0.0026, 0.0083)
		r_p^*	0.903	0.953
	r_p^*	(0.049, 0.048)	(0.0257, 0.0203)	(0.0038, 0.0042)
(30, 30)	w_p	0.881	0.936	0.985
		(0.071, 0.0476)	(0.0393, 0.0246)	(0.0093, 0.0056)
	r_p	0.882	0.935	0.985
(30, 30)	w_p	(0.0466, 0.0716)	(0.0243, 0.0406)	(0.0056, 0.01)
		r_p^*	0.886	0.94
	r_p^*	(0.0571, 0.0563)	(0.0296, 0.0303)	(0.0048, 0.0052)

Tabella 4.15: Livelli di copertura empirica con $\psi = 0.95$, $\alpha = 2$, $\sigma_1 = 0.85$, $\sigma_2 = 3.8$.

(n_1, n_2)	Metodo	90%	95%	99%
(5, 5)	w_p	0.818 (0.0236, 0.1576)	0.889 (0.0123, 0.098)	0.964 (0.003, 0.0323)
	r_p	0.85 (0.123, 0.0273)	0.926 (0.0586, 0.015)	0.988 (0.0083, 0.0036)
	r_p^*	0.895 (0.0513, 0.0533)	0.945 (0.0263, 0.0283)	0.987 (0.0053, 0.007)
(10, 10)	w_p	0.86 (0.026, 0.1133)	0.919 (0.0126, 0.068)	0.979 (0.004, 0.017)
	r_p	0.869 (0.102, 0.0283)	0.932 (0.0526, 0.0153)	0.989 (0.0063, 0.004)
	r_p^*	0.897 (0.0504, 0.0526)	0.947 (0.024, 0.0286)	0.99 (0.005, 0.005)
(30, 30)	w_p	0.883 (0.028, 0.0883)	0.941 (0.012, 0.0466)	0.982 (0.0033, 0.014)
	r_p	0.887 (0.0826, 0.0303)	0.946 (0.041, 0.013)	0.986 (0.01, 0.0033)
	r_p^*	0.894 (0.0546, 0.0514)	0.95 (0.024, 0.026)	0.993 (0.0034, 0.0036)

Le Figure 4.10 e 4.11 mostrano i Normal QQ-Plot dei valori r_p e r_p^* ottenuti dalle simulazioni; in particolare facciamo riferimento al caso di piccoli campioni, ovvero con $n_1 = n_2 = 10$, e dove ψ è pari a 0.5 e 0.8. Si nota come i punti relativi a r_p^* di sovrappongono meglio alla diagonale rispetto ai punti di r_p nelle code. Inoltre nel secondo caso si nota un leggero spostamento dei punti dalla diagonale, probabilmente dovuto a problemi di stima. Comunque l'approssimazione normale sembra buona per entrambi.

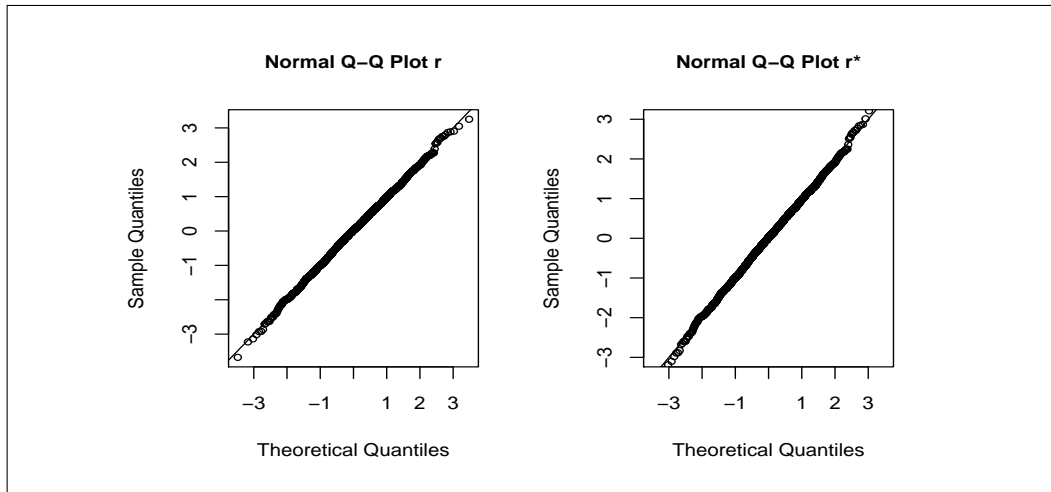


Figura 4.9: Normal qq-plot dei valori di r_p e r_p^* ottenuti tramite simulazione con $n_1 = n_2 = 10$ e $\psi = 0.5$.

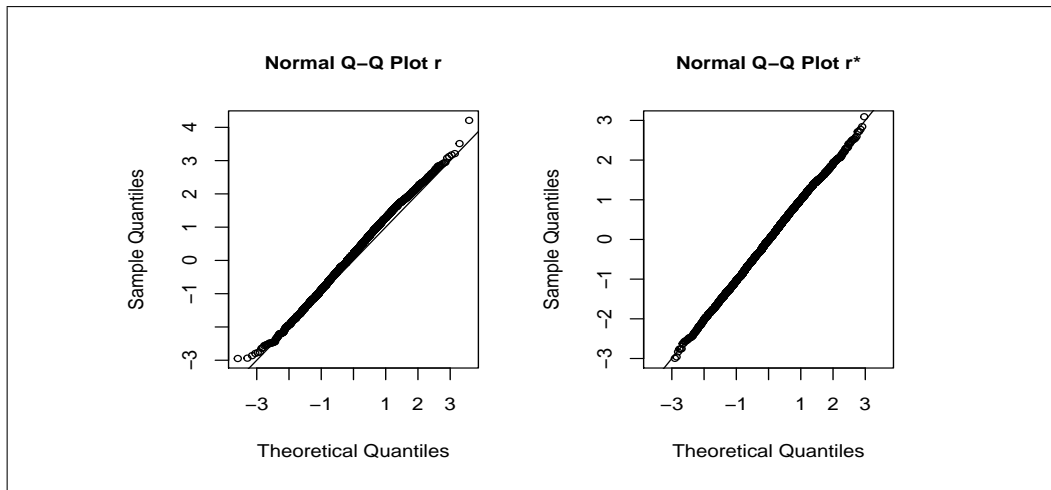


Figura 4.10: Normal qq-plot dei valori di r_p e r_p^* ottenuti tramite simulazione con $n_1 = n_2 = 10$ e $\psi = 0.8$.

Capitolo 5

Un caso reale: confronto tra PET/CT e PET/RM su pazienti pediatrici

5.1 PET/CT e PET/RM

In quest'ultimo capitolo utilizzeremo un modello stress-strength su dei dati reali. In particolare, i dati riguardano misurazioni effettuate con due tecniche di medicina nucleare su pazienti in età pediatrica con neoplasie solide extracraniche, ovvero tumori di diverse forme. Queste tecniche sono la PET/CT (Positron Emission Tomography - Computed Tomography) e la PET/RM (Positron Emission Tomography - Risonanza Magnetica), le quali non sono altro che macchine di diagnostica per immagini che integrano alla Tomografia ad Emissione di Positroni, nel primo caso, un tomografo TAC Spirale, e nel secondo caso, un tomografo a risonanza magnetica. Tali macchine vengono utilizzate in ambito oncologico, neurologico e cardiologico per l'individuazione di neoplasie o lesioni. La PET/CT è la tecnica più utilizzata da molti anni, soprattutto in campo oncologico, e consente di definire in termini di metabolismo cellulare la natura di una lesione. In oncologia con la PET CT è pertanto possibile differenziare le lesioni tumorali da lesioni di altra natura, nonché, nello studio delle lesioni tumorali accertate, analizzare

Capitolo 5. Un caso reale: confronto tra PET/CT e PET/RM su pazienti pediatrici

l'evoluzione della malattia anche dopo l'asportazione del tumore primitivo, e monitorare l'efficacia di una terapia antitumorale. L'esame comporta per il paziente la somministrazione in vena di isotopi emittenti positroni e per il resto si svolge come una TAC. Il tracciante di gran lunga più utilizzato è il 18-fluorodesossiglucosio, sigla 18-FDG. Mediante i radioisotopi è quindi possibile esplorare il metabolismo cellulare che spesso nelle cellule tumorali si differenzia da quello del tessuto normale.

La PET/RM, invece, è una tecnica abbastanza recente, i cui aspetti sono ancora in evoluzione e tutti da studiare ma con buone prospettive future. Essa permette l'esecuzione simultanea di indagini PET e di Risonanza Magnetica: la RM produce immagini del corpo dettagliate al millimetro mentre la PET indica al medico esaminatore l'attività di quelle immagini, ossia l'attività metabolica cioè il consumo di energia di quella cellula di quel dato distretto corporeo esaminato. Un'attività metabolica che oltrepassa un determinato, ma non standard, parametro è un'attività patologica, verosimilmente di natura oncologica. Le immagini RM, oltre a non comportare la radioesposizione del paziente, forniscono un migliore contrasto nella visualizzazione dei tessuti molli, rispetto a quelle CT. Tuttavia la RM presenta alcuni svantaggi in certi distretti corporei, quali ad esempio l'apparato scheletrico e quello polmonare, per difficoltà legate al riconoscimento del tessuto osseo e dell'aria. Studi recenti (Pfluger et al., 2012) hanno dimostrato che le due tecniche sono equivalenti dal punto di vista qualitativo, ossia individuano correttamente le stesse patologie, mentre dal punto di vista quantitativo la PET/RM sottostima la PET/CT nella misurazione del SUV (Standardized Uptake Value).

Il SUV misura la concentrazione di FDG in una lesione, normalizzata per la dose effettivamente erogata ed il peso del paziente. Per misurare il SUV si posiziona un volume di interesse 3D (VOI) sulla lesione o sul tumore e si calcola il rapporto tra la radioattività misurata all'interno della VOI e l'attività di dose iniettata divisa per il peso del paziente. In particolare viene utilizzato il SUV MEAN, ovvero la media dei SUV all'interno di una VOI.

5.2 Il campione

I dati utilizzati per questa applicazione provengono da un campione di 26 pazienti di età compresa tra i 4 e i 13 anni, affetti da: linfoma, sarcoma, leucemia, fibrosarcomatosi desmoide e adenocarcinoma del rinofaringe. Sui pazienti sono state effettuate diagnosi, a debita distanza temporale, con entrambe le tecniche discusse in precedenza. Le osservazioni sono tra loro indipendenti, mentre, da indagini preliminari, possiamo ipotizzare che le misurazioni (SUV) ottenute sui pazienti siano realizzazioni campionarie di una variabile casuale normale bivariata $Z = (X, Y)$ con vettore delle medie $\mu = (\mu_1, \mu_2)$, varianze σ_1^2, σ_2^2 , rispettivamente, e coefficiente di correlazione ρ . I due campioni inoltre risultano essere appaiati, poichè, come detto in precedenza, le due misurazioni sono state eseguite sugli stessi pazienti. Il nostro scopo è capire quanto diverse sono le distribuzioni del SUV delle due tecniche, al fine di trarre delle conclusioni di tipo quantitativo. Nelle Figure 5.1 e 5.2 vengono mostrati gli istogrammi dei SUV e le funzioni di densità marginali ipotizzate, relative a PET/RM e PET/CT, con medie e varianze stimate tramite metodo dei momenti.

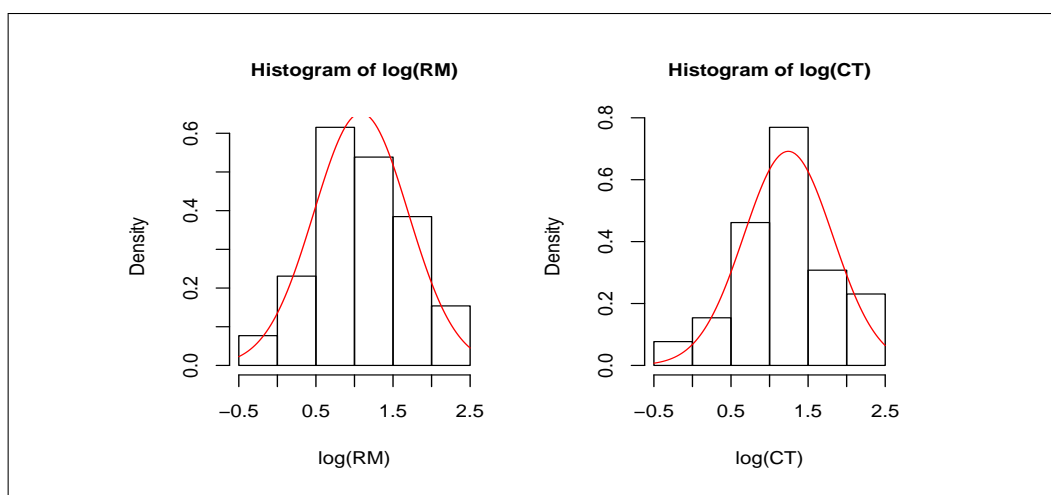


Figura 5.1: Istogrammi di SUV RM e SUV CT in scala logaritmica

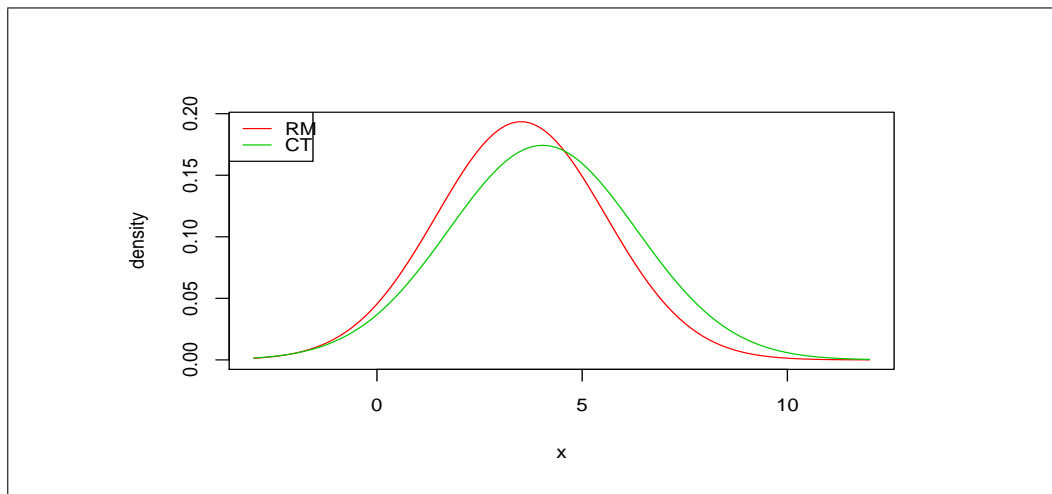


Figura 5.2: Funzioni di densità ipotizzate per SUV RM e SUV CT.

5.3 L'analisi

Utilizziamo un modello stress-strength per stimare la probabilità che il SUV della PET/RM, che indicheremo con X , sia minore del SUV della PET/CT, che indicheremo con Y , mentre indichiamo tale probabilità con ψ . Come accennato prima, ipotizziamo che i due campioni siano normalmente distribuiti e dipendenti. Tramite le funzioni presenti nella libreria `LikelihoodAsy` testiamo l'ipotesi nulla $\psi = 0.5$, ovvero di uguaglianza delle due distribuzioni, contro l'alternativa $\psi > 0.5$. La SMV di ψ è pari a 0.7983 con un errore standard di 0.065; le SMV di μ_1 , μ_2 , σ_1 e σ_2 sono 3.5111, 4.0380, 2.0213 e 2.2441, rispettivamente. La SMV di ρ è pari a 0.945. Successivamente vengono calcolate le tre statistiche test studiate nei capitoli precedenti: w_p è pari a 4.655 con un p-value di 1.619904e-06, r_p è uguale a 3.7112 con p-value pari a 0.00010314 e r_p^* è pari a 3.6058 con un p-value di 0.0001556. Tutte le statistiche test calcolate portano al rifiuto dell'ipotesi nulla di uguaglianza delle due distribuzioni, ad un livello di significatività del 5%. Infine vengono calcolati gli intervalli di confidenza al 95%, del primo ordine e di ordine superiore, per ψ , che sono (0.6525, 0.9006) e (0.6435, 0.8960), rispettivamente. Come prevedibile, entrambi gli intervalli escludono il valore testato 0.5 e fanno propendere per l'ipotesi alternativa. La Figura 5.3 è una rappresentazione grafica degli intervalli di confidenza per ψ .

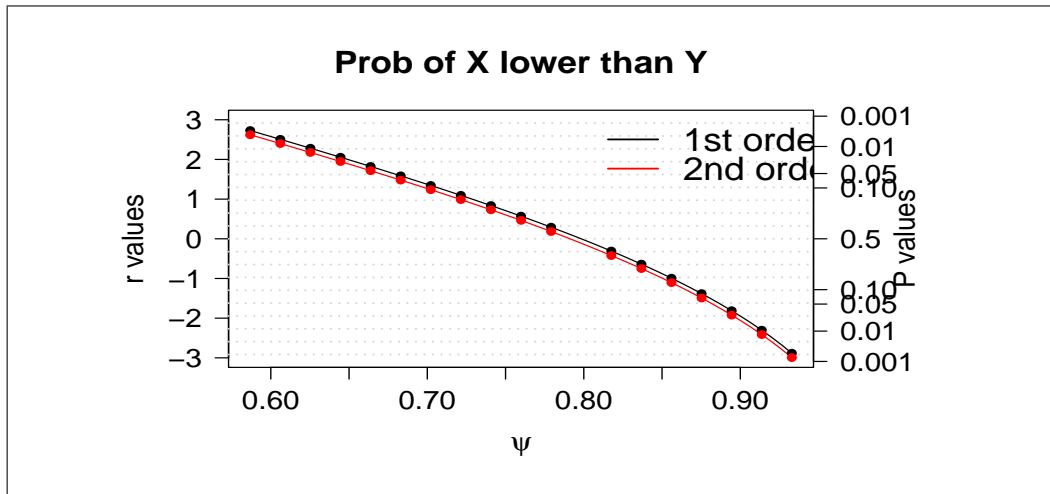


Figura 5.3: Intervalli di confidenza per ψ

In conclusione, quindi, si può affermare che, con ogni probabilità, la distribuzione dei SUV relativi alla PET/RM è minore di quella relativa alla PET/CT, a conferma dei risultati discussi nella letteratura, di cui si rimandano al lettore eventuali approfondimenti. Nonostante la differenza tra i due metodi risulti statisticamente significativa, da un punto di vista qualitativo le due tecniche vengono considerate equivalenti. Sulla base dei dati attuali si può ipotizzare che la PET/RM, probabilmente, non sostituirà la PET/TC in tutte le sue applicazioni ma sarà adoperata come integrazione in presenza di determinate indicazioni e come modalità a sé in altre, anche in età pediatrica.

Conclusioni

Nel presente elaborato sono stati discussi i modelli Stress-Strength, utilizzati per stimare la probabilità che una distribuzione sia minore di un'altra. In particolare, in molti campi, si potrebbe voler stimare la probabilità che lo stress a cui un oggetto è sottoposto sia minore della sua forza; viceversa l'oggetto in questione si romperebbe. Si sono viste varie formulazioni matematiche di questi modelli, a seconda dei vari utilizzi e obiettivi, e alcuni problemi che li caratterizzano, come ad esempio la dipendenza tra stress e forza - spesso ignorata - o le difficoltà di stima in presenza di piccoli campioni e/o parametri di disturbo. Successivamente sono stati illustrati alcuni concetti chiave riguardanti l'inferenza statistica tramite verosimiglianza, utili per la comprensione dei successivi temi. Dopo aver passato in rassegna, nel Capitolo 3, le formulazioni teoriche dei modelli in esame per diverse distribuzioni, si è passati a studiare il comportamento dei modelli in termini di precisione di stima. A tal proposito, abbiamo effettuato delle simulazioni utilizzando per la costruzione degli intervalli di confidenza per ψ , parametro di interesse, sia l'inferenza di verosimiglianza di primo ordine sia la teoria asintotica di ordine superiore. Quest'ultima è di aiuto soprattutto in presenza di basse numerosità campionarie; infatti gli intervalli di confidenza calcolati in questo modo presentano livelli di copertura empirica vicini ai valori nominali, più di quanto facciano gli intervalli basati sulle approssimazioni del primo ordine. Inoltre abbiamo notato anche una maggiore simmetria degli errori di copertura. Gli intervalli alla Wald, invece, si sono dimostrati essere, in generale, i peggiori, con livelli empirici di copertura inferiori, a volte di molto, rispetto ai valori teorici. La teoria asintotica di ordine superiore afferma che la statistica test r_p^* , che consiste in una modifica della classica statistica test r_p , ha un ordine

di errore nell'approssimazione alla normale che tende a zero più velocemente; grazie a ciò otteniamo stime più accurate, soprattutto con piccoli campioni e parametri di disturbo. Un altro aspetto emerso dagli studi di simulazione, e che conferma l'importanza della teoria di ordine superiore, è che, in certi casi, con basse numerosità campionarie e valori di ψ estremi la stima risulta difficoltosa o instabile e spesso invece è proprio in questi casi che necessitiamo di una maggiore accuratezza, in quanto molti dispositivi hanno bisogno di un'alta affidabilità per poter essere funzionali, quasi vicino ad 1, ed una stima attendibile con i metodi classici è ottenibile solo con grandi campioni. Molto utile è stata la libreria `LikelihoodAsy` che, sfruttando l'integrazione Monte Carlo nel calcolo delle covarianze, ci ha permesso un calcolo agevole della quantità q , necessaria alla computabilità di r_p^* . Infine nel Capitolo 5 abbiamo applicato un modello Stress-Strength ad un caso pratico, evidenziando come tali modelli, non sono utili solo in campo ingegneristico, ma anche in ambito medico, utilizzandoli come misura di differenza tra due distribuzioni. Inoltre va sottolineato come la teoria asintotica di ordine superiore risulti molto importante, in quanto in molti contesti pratici spesso non si dispone di una numerosità campionaria elevata, per cui è necessario un metodo che abbia la migliore accuratezza possibile al fine di ottenere delle stime affidabili. Nell'esempio studiato nell'ultimo capitolo, si ricade nel caso di campioni normali dipendenti. In tal caso si può notare che l'intervallo di confidenza al 95% basato su r_p è leggermente più corto di quello basato su r_p^* allo stesso livello di fiducia e quindi risulterebbe preferibile ma le simulazioni hanno mostrato una migliore accuratezza del secondo, seppure la differenza sia lieve avendo una buona numerosità campionaria. In conclusione si può affermare che i metodi di verosimiglianza di ordine superiore nei modelli Stress-Strength risultano preferibili a quelli del primo ordine in presenza di piccoli campioni e/o molti parametri di disturbo, mentre con buone numerosità campionarie - ossia con un numero di osservazioni maggiori o uguali a 30 - i due metodi sembrano equivalersi.

Bibliografia

- [1] Azzalini A. (2001). *Inferenza statistica. Una presentazione basata sul concetto di verosimiglianza*. Springer-Verlag Italia, Milano.
- [2] Bellio R.; Pierce D.A. (2015). Practical likelihood asymptotics with the LikelihoodAsy package. (Sottoposto per pubblicazione)
- [3] Bilikam J.E. (1985). Some stochastic stress-strength processes. *IEEE Trans. Reliab.*, **34**, 269-274.
- [4] Brazzale A.R.; Davison A.C.; Reid N. (2007). *Applied Asymptotics Case Studies in Small-Sample Statistics*. Cambridge University Press.
- [5] Cortese G.; Ventura L. (2012). Accurate higher-order likelihood inference on $P(Y < X)$. *Computational statistics*, **28**, 1035-1059.
- [6] Fisher R.A. (1922). On the mathematical foundation of theoretical statistics. *Phil. Trans. Roy. Soc.*, **A**, 309-368.
- [7] Guttman I.; Johnson R.A.; Bhattacharyya G.K.; Reiser B. (1988). Confidence limits for stress-strength models with explanatory variables. *Technometrics*, **30**, 161-168.
- [8] Kotz S.; Lumelskii Y.; Pensky M. (2003). *The Stress-Strength model and its generalizations. Theory and applications*. World Scientific, Singapore.
- [9] Mann H.B.; Whitney D.R. (1947). On a test whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.*, **18**, 50-60.

Bibliografia

- [10] Owen D.B.; Craswell K.J.; Hanson D.L. (1964). Nonparametric upper confidence bounds for $P(Y < X)$ and confidence limits for $P(Y < X)$ when X and Y are normal. *J. Amer. Statist. Assoc.*, **59**, 906-924.
- [11] Pace L.; Salvan A. (1996). *Teoria della statistica: Metodi, Modelli, Approssimazioni Asintotiche*. Cedam, Padova.
- [12] Pace L.; Salvan A. (2001). *Introduzione alla statistica. II Inferenza, verosimiglianza, modelli*. Cedam, Padova.
- [13] Pfluger T.; Melzer H.; Mueller W.; Coppenrath E.; Bartenstein P.; Albert M.; Schmidt I. (2012). Diagnostic value of combined ^{18}F -FDG PET/MRI for staging and restaging in paediatric oncology. *EUR J Nucl Med Mol Imaging.*, **39**, 1745-1755.
- [14] Skovgaard I.M. (1996). A general large deviation approximation to one parameter tests. *Bernoulli*, **2**, 145-165.
- [15] Wilcoxon F. (1945). Individual comparisons by ranking methods. *Biometrical Bull.*, **1**, 80-83.

Elenco delle tabelle

4.1	Livelli di copertura empirica con $\psi = 0.5, \alpha = 1, \beta = 1$	42
4.2	Livelli di copertura empirica con $\psi = 0.8, \alpha = 1, \beta = 0.25$. . .	43
4.3	Livelli di copertura empirica con $\psi = 0.95, \alpha = 1, \beta = 0.05$. .	44
4.4	Livelli di copertura empirica con $\psi = 0.5, \mu_1 = 5, \mu_2 = 5, \sigma = 1.$	46
4.5	Livelli di copertura empirica con $\psi = 0.8, \mu_1 = 5, \mu_2 = 6.55,$ $\sigma = 1.3.$	47
4.6	Livelli di copertura empirica con $\psi = 0.95, \mu_1 = 5, \mu_2 = 8.5,$ $\sigma = 1.5.$	48
4.7	Livelli di copertura empirica con $\psi = 0.5, \mu_1 = 5, \mu_2 = 5,$ $\sigma_1 = 0.5, \sigma_2 = 1.5.$	50
4.8	Livelli di copertura empirica con $\psi = 0.8, \mu_1 = 5, \mu_2 = 6.33,$ $\sigma_1 = 0.5, \sigma_2 = 1.5.$	51
4.9	Livelli di copertura empirica con $\psi = 0.95, \mu_1 = 5, \mu_2 = 5,$ $\sigma_1 = 0.5, \sigma_2 = 1.5.$	52
4.10	Livelli di copertura empirica con $\psi = 0.5, \mu_1 = 5, \mu_2 = 5,$ $\sigma_1 = 0.5, \sigma_2 = 1.5, \rho = 0.8.$	55
4.11	Livelli di copertura empirica con $\psi = 0.8, \mu_1 = 5, \mu_2 = 6.3,$ $\sigma_1 = 0.4, \sigma_2 = 1.8, \rho = 0.8.$	56
4.12	Livelli di copertura empirica con $\psi = 0.95, \mu_1 = 5, \mu_2 = 7.5,$ $\sigma_1 = 0.4, \sigma_2 = 1.8, \rho = 0.8.$	57
4.13	Livelli di copertura empirica con $\psi = 0.5, \alpha = 1, \sigma_1 = 0.5,$ $\sigma_2 = 0.5.$	58
4.14	Livelli di copertura empirica con $\psi = 0.8, \alpha = 1, \sigma_1 = 0.5,$ $\sigma_2 = 2.$	59

Elenco delle tabelle

4.15 Livelli di copertura empirica con $\psi = 0.95$, $\alpha = 2$, $\sigma_1 = 0.85$,
 $\sigma_2 = 3.8$ 60

Appendice A

Codice R

Di seguito verranno riportati il codice R e le funzioni della libreria `LikelihoodAsy` utilizzate per gli studi di simulazione e l'analisi dei dati reali, presentati in questa tesi.

```
#### Funzione di verosimiglianza caso esponenziale
llik.exp=function(theta,data)
{
  alpha=log(theta[1])
  beta=log(theta[2])
  sum(dexp(data$x,alpha,log=T))+sum(dexp(data$y,beta,log=T))
}

#### Funzione generazione campioni esponenziali

datagen.exp=function(theta,data)
{
  out=data
  out$x=rexp(length(data$x),theta[1])
  out$y=rexp(length(data$y),theta[2])
  out
}

#### Funzione calcolo psi caso esponenziale

logit.psi.fun.exp<-function(theta,data){
  psi=theta[1]/(theta[1]+theta[2])
  log(psi/(1-psi))
}

#### Funzione per la stima di theta, psi e per il calcolo di w, r, rstar
```

Appendice A. Codice R

```
obj<- rstar(dati, thetainit=c(1,1), floglik=llik.exp, psival=logit(0.5),
           fsim=datagen.exp, fpsi=logit.psi.fun.exp, R=1000)

#### Funzione per il calcolo degli intervalli di confidenza di primo ordine
#### e di ordine superiore

obj2<- rstar.ci(dati, thetainit=c(1,1), floglik=llik.exp,
              fsim=datagen.exp, fpsi=logit.psi.fun.exp, R=1000)

#### Funzione di verosimiglianza caso normali indipendenti con varianze uguali

llik.normivu=function(theta, data)
{
  m1=theta[1]
  m2=theta[2]
  sigma=log(theta[3])
  sum(dnorm(data$x, m1, sigma, log=T))+sum(dnorm(data$y, m2, sigma, log=T))
}

#### Funzione generazione campioni normali indipendenti con varianze uguali

datagen.normivu=function(theta, data)
{
  out=data
  out$x=rnorm(length(data$x), theta[1], theta[3])
  out$y=rnorm(length(data$y), theta[2], theta[3])
  out
}

#### Funzione calcolo psi caso normali indipendenti varianze uguali

logit.psi.fun.normivu<-function(theta){
  q<-(theta[2]-theta[1])/theta[3]*sqrt(2)
  psi=pnorm(q)
  log(psi/(1-psi))
}

#### Funzione per la stima di theta, psi e per il calcolo di w, r, rstar

obj<- rstar(dati, thetainit=c(1,1,1), floglik=llik.normivu, psival=logit(0.5),
           fsim=datagen.normivu, fpsi=logit.psi.fun.normivu, R=1000)

#### Funzione per il calcolo degli intervalli di confidenza di primo ordine
#### e di ordine superiore

obj2<- rstar.ci(dati, thetainit=c(1,1,1), floglik=llik.normivu,
              fsim=datagen.normivu, fpsi=logit.psi.fun.normivu, R=1000)
```

Funzione di verosimiglianza caso normali indipendenti con varianze diverse

```
llik.normivd=function(theta , data)
{
  m1=theta [1]
  m2=theta [2]
  sigma1=log (theta [3])
  sigma2=log (theta [4])
  sum(dnorm(data$x ,m1 ,sigma1 , log=T))+sum(dnorm (data$y ,m2 ,sigma2 , log=T))
}
```

Funzione generazione campioni normali indipendenti con varianze diverse

```
datagen.normivd=function(theta , data)
{
  out=data
  out$x=rnorm(length (data$x) ,theta [1] ,theta [3])
  out$y=rnorm(length (data$y) ,theta [2] ,theta [4])
  out
}
```

Funzione calcolo psi caso normali indipendenti varianze diverse

```
logit.psi.fun.normivd<-function(theta)
{
  q<-(theta [2] -theta [1]) / (sqrt (theta [3]^2 +theta [4]^2))
  psi=pnorm(q)
  log (psi / (1 -psi))
}
```

Funzione per la stima di theta , psi e per il calcolo di w , r , rstar

```
obj<- rstar(dati , thetainit=c(1,1,1,1) , floglik=llik.normivd , psival=logit (0.5) ,
  fsim=datagen.normivd , fpsi=logit.psi.fun.normivd , R=1000)
```

Funzione per il calcolo degli intervalli di confidenza di primo ordine
e di ordine superiore

```
obj2<- rstar.ci(dati , thetainit=c(1,1,1,1) , floglik=llik.normivd ,
  fsim=datagen.normivd , fpsi=logit.psi.fun.normivd , R=1000)
```

Funzione di verosimiglianza caso normali dipendenti

```
llik.dipnorm<-function(theta , data){
  f<-(1/(2*pi*log (theta [3]) * log (theta [4]) * sqrt (1 -theta [5]^2)))*exp((-1/(2*(1 -theta [5]^2))
  (((data$x -theta [1]) / log (theta [3]))^2 +((data$y -theta [2]) / log (theta [4]))^2 -2*theta [5]
  ((data$x -theta [1]) / log (theta [3]))*((data$y -theta [2]) / log (theta [4]))))
  sum(log (f))
}
```

Appendice A. Codice R

```
}

#### Funzione generazione campioni normali dipendenti

datagen.bvnorm<-function(theta, data)
{
  out=data
  out$x<-rnorm(length(data$x), theta[1], theta[3])
  out$y<-rnorm(length(data$y), theta[2]+theta[5]*theta[4]/theta[3]*
  (data$x-theta[1]), theta[4]*sqrt(1-theta[5]^2))
  out
}

#### Funzione calcolo psi caso normali dipendenti

logit.psi.fun.dipnorm<-function(theta)
{
  q<-(theta[2]-theta[1])/sqrt(theta[3]^2+theta[4]^2-2*theta[5]*
  theta[3]*theta[4])
  log(psi/(1-psi))
}

#### Funzione per la stima di theta, psi e per il calcolo di w, r, rstar

obj<- rstar(dati, thetainit=c(1,1,1,1,1), floglik=llik.dipnorm,
  psival=logit(0.5), fsim=datagen.bvnorm,
  fpsi=logit.psi.fun.dipnorm, R=1000)

#### Funzione per il calcolo degli intervalli di confidenza di primo ordine
#### e di ordine superiore

obj2<- rstar.ci(dati, thetainit=c(1,1,1,1,1), floglik=llik.dipnorm,
  fsim=datagen.bvnorm, fpsi=logit.psi.fun.dipnorm, R=1000)

#### Funzione verosimiglianza caso weibull

llik.weib=function(theta, data)
{
  a=log(theta[1]) # forma
  b1=log(theta[2]) # scala x
  b2=log(theta[3]) # scala y
  sum(dweibull(data$x, a, b1, log=T))+sum(dweibull(data$y, a, b2, log=T))
}

#### Funzione generazione campioni weibull

datagen.weib=function(theta, data)
{
  out=data
```

```

    out$x=rweibull(length(data$x),theta[1],theta[2])
    out$y=rweibull(length(data$y),theta[1],theta[3])
    out
  }

#### Funzione calcolo psi caso weibull

logit.psi.fun.weib=function(theta){
  psi=theta[3]^theta[1]/(theta[2]^theta[1]+
  theta[3]^theta[1])
  log(psi/(1-psi))
}

#### Funzione per la stima di theta, psi e per il calcolo di w, r, rstar

obj<-rstar(dati,thetainit=c(1,1,1),floglik=llik.weib,psival=logit(0.5),
  fsim=datagen.weib,fpsi=logit.psi.fun.weib,R=1000)

#### Funzione per il calcolo degli intervalli di confidenza di primo ordine
#### e di ordine superiore

obj2<-rstar.ci(dati,thetainit=c(1,1,1),floglik=llik.weib,
  fsim=datagen.weib,fpsi=logit.psi.fun.weib, R=1000)

#### Funzione inversa logit

logit.inv=function(eta) exp(eta)/(1+exp(eta))

#### SIMULAZIONI : esponenziale

a<- # parametri
b<-
teta<-c(a,b)
R<-

smv<-NULL # statistiche
se<-NULL
rclass<-NULL
rstar<-NULL

nx<- # numerosita
ny<-
N<-

mat.wald<-matrix(rep(0,6*N),N,6) # intervalli per wald

for(i in 1:N){
  print(i)
  x<-rexp(nx,teta[1],teta[2])

```

Appendice A. Codice R

```
y<-rexp(ny, teta [1], teta [3])
dati=list(x=x,y=y)
obj<-try(rstar(dati, thetainit=c(1,1), floglik=llik.exp, psival=logit(R),
fsim=datagen.exp, fpsi=logit.psi.fun.exp, R=1000))
if(is.character(obj)==T) next
rclass[i]<-obj$r
rstar[i]<-as.vector(obj$rs)
smv[i]<-obj$psi.hat
se[i]<-as.vector(obj$se.psi.hat)
mat.wald[i,1]<-smv[i]-1.64*se[i]
mat.wald[i,2]<-smv[i]+1.64*se[i]
mat.wald[i,3]<-smv[i]-1.96*se[i]
mat.wald[i,4]<-smv[i]+1.96*se[i]
mat.wald[i,5]<-smv[i]-2.57*se[i]
mat.wald[i,6]<-smv[i]+2.57*se[i]
}

# livelli di copertura empirica
mat.wald<-mat.wald[apply(is.na(mat.wald), 1, sum)==0,]
N<-nrow(mat.wald)
a<-sum((logit(R)>mat.wald[,1])&(logit(R)<mat.wald[,2])/N
# 0.90
b<-sum((logit(R)>mat.wald[,3])&(logit(R)<mat.wald[,4])/N
# 0.95
c<-sum((logit(R)>mat.wald[,5])&(logit(R)<mat.wald[,6])/N
# 0.99

err.a<-c(sum(logit(R)<mat.wald[,1])/N, sum(logit(R)>mat.wald[,2])/N)
err.b<-c(sum(logit(R)<mat.wald[,3])/N, sum(logit(R)>mat.wald[,4])/N)
err.c<-c(sum(logit(R)<mat.wald[,5])/N, sum(logit(R)>mat.wald[,6])/N

matrice<-matrix(c(rclass, rstar), N, 2)
matrice<-matrice[apply(is.na(matrice), 1, sum)==0,]
rclass<-matrice[,1]
rstar<-matrice[,2]
N<-nrow(matrice)

d<-sum((rclass>qnorm(0.05))&(rclass<qnorm(0.95)))/N
# 0.90
e<-sum((rclass>qnorm(0.025))&(rclass<qnorm(0.975)))/N
# 0.95
f<-sum((rclass>qnorm(0.005))&(rclass<qnorm(0.995)))/N
# 0.99

err.d<-c(sum(rclass<qnorm(0.05))/N, sum(rclass>qnorm(0.95))/N)
err.e<-c(sum(rclass<qnorm(0.025))/N, sum(rclass>qnorm(0.975))/N)
err.f<-c(sum(rclass<qnorm(0.005))/N, sum(rclass>qnorm(0.995))/N)
```

```

g<-sum((rstar>qnorm(0.05))&(rstar<qnorm(0.95)))/N
# 0.90
h<-sum((rstar>qnorm(0.025))&(rstar<qnorm(0.975)))/N
# 0.95
i<-sum((rstar>qnorm(0.005))&(rstar<qnorm(0.995)))/N
# 0.99

err.g<-c(sum(rstar<qnorm(0.05))/N,sum(rstar>qnorm(0.95))/N)
err.h<-c(sum(rstar<qnorm(0.025))/N,sum(rstar>qnorm(0.975))/N)
err.i<-c(sum(rstar<qnorm(0.005))/N,sum(rstar>qnorm(0.995))/N)

mat.cop<-matrix(c(a,b,c,d,e,f,g,h,i),3,3,byrow=T)
colnames(mat.cop)<-c("0.90","0.95","0.99")
rownames(mat.cop)<-c("wald","r","r*")

# qqplot

par(mfrow=c(1,2))
qqnorm(rclass,main="Normal Q-Q Plot r")
abline(0,1)
qqnorm(rstar,main="Normal Q-Q Plot r*",ylim=c(-3,3))
abline(0,1)

### SIMULAZIONI : weibull

a<- # parametri
b1<-
b2<-
teta<-c(a,b1,b2)
R<-

smv<-NULL # statistiche
se<-NULL
rclass<-NULL
rstar<-NULL

nx<- # numerosita
ny<-
N<-

mat.wald<-matrix(rep(0,6*N),N,6) # intervalli

for(i in 1:N){
  print(i)
  x<-rweibull(nx,teta[1],teta[2])
  y<-rweibull(ny,teta[1],teta[3])
  dati=list(x=x,y=y)
  obj<-try(rstar(dati,thetainit=c(1,1,1),floglik=llik.weib,psival=logit(R),
  fsim=datagen.weib,fpsi=logit.psi.fun.weib,R=1000,psidesc="Prob of X lower than Y"))
}

```

Appendice A. Codice R

```
if (is.character(obj)==T) next
rclass[i]<-obj$r
rstar[i]<-as.vector(obj$rs)
smv[i]<-obj$psi.hat
se[i]<-as.vector(obj$se.psi.hat)
mat.wald[i,1]<-smv[i]-1.64*se[i]
mat.wald[i,2]<-smv[i]+1.64*se[i]
mat.wald[i,3]<-smv[i]-1.96*se[i]
mat.wald[i,4]<-smv[i]+1.96*se[i]
mat.wald[i,5]<-smv[i]-2.57*se[i]
mat.wald[i,6]<-smv[i]+2.57*se[i]
}

# livelli di copertura empirica
mat.wald<-mat.wald[apply(is.na(mat.wald),1,sum)==0,]
N<-nrow(mat.wald)
a<-sum((logit(R)>mat.wald[,1])&(logit(R)<mat.wald[,2]))/N
# 0.90
b<-sum((logit(R)>mat.wald[,3])&(logit(R)<mat.wald[,4]))/N
# 0.95
c<-sum((logit(R)>mat.wald[,5])&(logit(R)<mat.wald[,6]))/N
# 0.99

err.a<-c(sum(logit(R)<mat.wald[,1])/N,sum(logit(R)>mat.wald[,2])/N)
err.b<-c(sum(logit(R)<mat.wald[,3])/N,sum(logit(R)>mat.wald[,4])/N)
err.c<-c(sum(logit(R)<mat.wald[,5])/N,sum(logit(R)>mat.wald[,6])/N)

matrice<-matrix(c(rclass,rstar),N,2)
matrice<-matrice[apply(is.na(matrice),1,sum)==0,]
rclass<-matrice[,1]
rstar<-matrice[,2]
N<-nrow(matrice)

d<-sum((rclass>qnorm(0.05))&(rclass<qnorm(0.95)))/N
# 0.90
e<-sum((rclass>qnorm(0.025))&(rclass<qnorm(0.975)))/N
# 0.95
f<-sum((rclass>qnorm(0.005))&(rclass<qnorm(0.995)))/N
# 0.99

err.d<-c(sum(rclass<qnorm(0.05))/N,sum(rclass>qnorm(0.95))/N)
err.e<-c(sum(rclass<qnorm(0.025))/N,sum(rclass>qnorm(0.975))/N)
err.f<-c(sum(rclass<qnorm(0.005))/N,sum(rclass>qnorm(0.995))/N)

g<-sum((rstar>qnorm(0.05))&(rstar<qnorm(0.95)))/N
# 0.90
h<-sum((rstar>qnorm(0.025))&(rstar<qnorm(0.975)))/N
# 0.95
```

```

i<-sum((rstar>qnorm(0.005))&(rstar<qnorm(0.995)))/N
# 0.99

err.g<-c(sum(rstar<qnorm(0.05))/N,sum(rstar>qnorm(0.95))/N)
err.h<-c(sum(rstar<qnorm(0.025))/N,sum(rstar>qnorm(0.975))/N)
err.i<-c(sum(rstar<qnorm(0.005))/N,sum(rstar>qnorm(0.995))/N)

mat.cop<-matrix(c(a,b,c,d,e,f,g,h,i),3,3,byrow=T)
colnames(mat.cop)<-c("0.90","0.95","0.99")
rownames(mat.cop)<-c("wald","r","r*")

# qqplot

par(mfrow=c(1,2))
qqnorm(rclass,main="Normal Q-Q Plot r")
abline(0,1)
qqnorm(rstar,main="Normal Q-Q Plot r*",ylim=c(-3,3))
abline(0,1)

#### SIMULAZIONI : normali indipendenti varianze uguali

mu1<-                                # parametri
mu2<-
sigma<-
teta<-c(mu1,mu2,sigma)
R<-

smv<-NULL                             # statistiche
se<-NULL
rclass<-NULL
rstar<-NULL

nx<-                                    # numerosita
ny<-
N<-

mat.wald<-matrix(rep(0,6*N),N,6)      # intervalli

for(i in 1:N){
  print(i)
  x<-rnorm(nx,teta[1],teta[3])
  y<-rnorm(ny,teta[2],teta[3])
  dati=list(x=x,y=y)
  obj<-try(rstar(dati,thetainit=c(1,1,1),floglik=llik.normivu,psival=logit(R),
  fsim=datagen.normivu,fpsi=logit.psi.fun.normivu,R=1000,
  psidesc="Prob of X lower than Y"))
  if(is.character(obj)==T) next
  rclass[i]<-obj$r
  rstar[i]<-as.vector(obj$rs)
}

```

Appendice A. Codice R

```
smv[i]<-obj$psi.hat
se[i]<-as.vector(obj$se.psi.hat)
mat.wald[i,1]<-smv[i]-1.64*se[i]
mat.wald[i,2]<-smv[i]+1.64*se[i]
mat.wald[i,3]<-smv[i]-1.96*se[i]
mat.wald[i,4]<-smv[i]+1.96*se[i]
mat.wald[i,5]<-smv[i]-2.57*se[i]
mat.wald[i,6]<-smv[i]+2.57*se[i]
}

# livelli di copertura empirica
mat.wald<-mat.wald[apply(is.na(mat.wald),1,sum)==0,]
N<-nrow(mat.wald)
a<-sum((logit(R)>mat.wald[,1])&(logit(R)<mat.wald[,2]))/N
# 0.90
b<-sum((logit(R)>mat.wald[,3])&(logit(R)<mat.wald[,4]))/N
# 0.95
c<-sum((logit(R)>mat.wald[,5])&(logit(R)<mat.wald[,6]))/N
# 0.99

err.a<-c(sum(logit(R)<mat.wald[,1])/N,sum(logit(R)>mat.wald[,2])/N)
err.b<-c(sum(logit(R)<mat.wald[,3])/N,sum(logit(R)>mat.wald[,4])/N)
err.c<-c(sum(logit(R)<mat.wald[,5])/N,sum(logit(R)>mat.wald[,6])/N)

matrice<-matrix(c(rclass,rstar),N,2)
matrice<-matrice[apply(is.na(matrice),1,sum)==0,]
rclass<-matrice[,1]
rstar<-matrice[,2]
N<-nrow(matrice)

d<-sum((rclass>qnorm(0.05))&(rclass<qnorm(0.95)))/N
# 0.90
e<-sum((rclass>qnorm(0.025))&(rclass<qnorm(0.975)))/N
# 0.95
f<-sum((rclass>qnorm(0.005))&(rclass<qnorm(0.995)))/N
# 0.99

err.d<-c(sum(rclass<qnorm(0.05))/N,sum(rclass>qnorm(0.95))/N)
err.e<-c(sum(rclass<qnorm(0.025))/N,sum(rclass>qnorm(0.975))/N)
err.f<-c(sum(rclass<qnorm(0.005))/N,sum(rclass>qnorm(0.995))/N)

g<-sum((rstar>qnorm(0.05))&(rstar<qnorm(0.95)))/N
# 0.90
h<-sum((rstar>qnorm(0.025))&(rstar<qnorm(0.975)))/N
# 0.95
i<-sum((rstar>qnorm(0.005))&(rstar<qnorm(0.995)))/N
# 0.99

err.g<-c(sum(rstar<qnorm(0.05))/N,sum(rstar>qnorm(0.95))/N)
```

```

err.h<-c(sum(rstar<qnorm(0.025))/N,sum(rstar>qnorm(0.975))/N)
err.i<-c(sum(rstar<qnorm(0.005))/N,sum(rstar>qnorm(0.995))/N)

mat.cop<-matrix(c(a,b,c,d,e,f,g,h,i),3,3,byrow=T)
colnames(mat.cop)<-c("0.90","0.95","0.99")
rownames(mat.cop)<-c("wald","r","r*")

# qqplot

par(mfrow=c(1,2))
qqnorm(rclass,main="Normal Q-Q Plot r")
abline(0,1)
qqnorm(rstar,main="Normal Q-Q Plot r*",ylim=c(-3,3))
abline(0,1)

### SIMULAZIONI : normali indipendenti varianze diverse

mu1<-                                # parametri
mu2<-
sigma1<-
sigma2<-
teta<-c(mu1,mu2,sigma1,sigma2)
R<-

smv<-NULL                             # statistiche
se<-NULL
rclass<-NULL
rstar<-NULL

nx<-                                    # numerosita
ny<-
N<-

mat.wald<-matrix(rep(0,6*N),N,6)       # intervalli

for(i in 1:N){
  print(i)
  x<-rnorm(nx,teta[1],teta[3])
  y<-rnorm(ny,teta[2],teta[4])
  dati=list(x=x,y=y)
  obj<-try(rstar(dati,thetainit=c(1,1,1,1),floglik=llik.normivd,psival=logit(R),
  fsim=datagen.normivd,fpsi=logit.psi.fun.normivd,R=1000,
  psidesc="Prob of X lower than Y"))
  if(is.character(obj)==T) next
  rclass[i]<-obj$r
  rstar[i]<-as.vector(obj$rs)
  smv[i]<-obj$psi.hat
  se[i]<-as.vector(obj$se.psi.hat)
  mat.wald[i,1]<-smv[i]-1.64*se[i]
}

```

Appendice A. Codice R

```
mat.wald[i,2]<-smv[i]+1.64*se[i]
mat.wald[i,3]<-smv[i]-1.96*se[i]
mat.wald[i,4]<-smv[i]+1.96*se[i]
mat.wald[i,5]<-smv[i]-2.57*se[i]
mat.wald[i,6]<-smv[i]+2.57*se[i]
}

# livelli di copertura empirica
mat.wald<-mat.wald[apply(is.na(mat.wald),1,sum)==0,]
N<-nrow(mat.wald)
a<-sum((logit(R)>mat.wald[,1])&(logit(R)<mat.wald[,2]))/N
# 0.90
b<-sum((logit(R)>mat.wald[,3])&(logit(R)<mat.wald[,4]))/N
# 0.95
c<-sum((logit(R)>mat.wald[,5])&(logit(R)<mat.wald[,6]))/N
# 0.99

err.a<-c(sum(logit(R)<mat.wald[,1])/N,sum(logit(R)>mat.wald[,2])/N)
err.b<-c(sum(logit(R)<mat.wald[,3])/N,sum(logit(R)>mat.wald[,4])/N)
err.c<-c(sum(logit(R)<mat.wald[,5])/N,sum(logit(R)>mat.wald[,6])/N)

matrice<-matrix(c(rclass,rstar),N,2)
matrice<-matrice[apply(is.na(matrice),1,sum)==0,]
rclass<-matrice[,1]
rstar<-matrice[,2]
N<-nrow(matrice)

d<-sum((rclass>qnorm(0.05))&(rclass<qnorm(0.95)))/N
# 0.90
e<-sum((rclass>qnorm(0.025))&(rclass<qnorm(0.975)))/N
# 0.95
f<-sum((rclass>qnorm(0.005))&(rclass<qnorm(0.995)))/N
# 0.99

err.d<-c(sum(rclass<qnorm(0.05))/N,sum(rclass>qnorm(0.95))/N)
err.e<-c(sum(rclass<qnorm(0.025))/N,sum(rclass>qnorm(0.975))/N)
err.f<-c(sum(rclass<qnorm(0.005))/N,sum(rclass>qnorm(0.995))/N)

g<-sum((rstar>qnorm(0.05))&(rstar<qnorm(0.95)))/N
# 0.90
h<-sum((rstar>qnorm(0.025))&(rstar<qnorm(0.975)))/N
# 0.95
i<-sum((rstar>qnorm(0.005))&(rstar<qnorm(0.995)))/N
# 0.99

err.g<-c(sum(rstar<qnorm(0.05))/N,sum(rstar>qnorm(0.95))/N)
err.h<-c(sum(rstar<qnorm(0.025))/N,sum(rstar>qnorm(0.975))/N)
err.i<-c(sum(rstar<qnorm(0.005))/N,sum(rstar>qnorm(0.995))/N)
```

```

mat.cop<-matrix(c(a,b,c,d,e,f,g,h,i),3,3,byrow=T)
colnames(mat.cop)<-c("0.90","0.95","0.99")
rownames(mat.cop)<-c("wald","r","r*")

# qqplot

par(mfrow=c(1,2))
qqnorm(rclass,main="Normal Q-Q Plot r")
abline(0,1)
qqnorm(rstar,main="Normal Q-Q Plot r*",ylim=c(-3,3))
abline(0,1)

### SIMULAZIONI : normali dipendenti

mu1<-                                # parametri
mu2<-
sigma1<-
sigma2<-
ro<-0.8
teta<-c(mu1,mu2,sigma1,sigma2,ro)
R<-

smv<-NULL                             # statistiche
se<-NULL
rclass<-NULL
rstar<-NULL

nx<-                                   # numerosita
ny<-
N<-

mat.wald<-matrix(rep(0,6*N),N,6)      # intervalli

for(i in 1:N){
  print(i)
  x<-rnorm(nx,teta[1],teta[3])
  y<-rnorm(ny,teta[2]+teta[5]*teta[4]/teta[3]*(x-teta[1]),teta[4]*sqrt(1-teta[5]^2))
  dati=list(x=x,y=y)
  obj<-try(rstar(dati,thetainit=c(1,1,1,1,0.3),floglik=llik.dipnorm,psival=logit(R),
  fsim=datagen.bvnorm,fpsi=logit.psi.fun.dipnorm,R=1000,
  psidesc="Prob of X lower than Y"))
  if(is.character(obj)==T) next
  rclass[i]<-obj$r
  rstar[i]<-as.vector(obj$rs)
  smv[i]<-obj$psi.hat
  se[i]<-as.vector(obj$se.psi.hat)
  mat.wald[i,1]<-smv[i]-1.64*se[i]
  mat.wald[i,2]<-smv[i]+1.64*se[i]
  mat.wald[i,3]<-smv[i]-1.96*se[i]
}

```

Appendice A. Codice R

```
    mat.wald[i,4]<-smv[i]+1.96*se[i]
    mat.wald[i,5]<-smv[i]-2.57*se[i]
    mat.wald[i,6]<-smv[i]+2.57*se[i]
  }

# livelli di copertura empirica
mat.wald<-mat.wald[apply(is.na(mat.wald),1,sum)==0,]
N<-nrow(mat.wald)
a<-sum((logit(R)>mat.wald[,1])&(logit(R)<mat.wald[,2]))/N
# 0.90
b<-sum((logit(R)>mat.wald[,3])&(logit(R)<mat.wald[,4]))/N
# 0.95
c<-sum((logit(R)>mat.wald[,5])&(logit(R)<mat.wald[,6]))/N
# 0.99

err.a<-c(sum(logit(R)<mat.wald[,1])/N,sum(logit(R)>mat.wald[,2])/N)
err.b<-c(sum(logit(R)<mat.wald[,3])/N,sum(logit(R)>mat.wald[,4])/N)
err.c<-c(sum(logit(R)<mat.wald[,5])/N,sum(logit(R)>mat.wald[,6])/N)

matrice<-matrix(c(rclass,rstar),N,2)
matrice<-matrice[apply(is.na(matrice),1,sum)==0,]
rclass<-matrice[,1]
rstar<-matrice[,2]
N<-nrow(matrice)

d<-sum((rclass>qnorm(0.05))&(rclass<qnorm(0.95)))/N
# 0.90
e<-sum((rclass>qnorm(0.025))&(rclass<qnorm(0.975)))/N
# 0.95
f<-sum((rclass>qnorm(0.005))&(rclass<qnorm(0.995)))/N
# 0.99

err.d<-c(sum(rclass<qnorm(0.05))/N,sum(rclass>qnorm(0.95))/N)
err.e<-c(sum(rclass<qnorm(0.025))/N,sum(rclass>qnorm(0.975))/N)
err.f<-c(sum(rclass<qnorm(0.005))/N,sum(rclass>qnorm(0.995))/N)

g<-sum((rstar>qnorm(0.05))&(rstar<qnorm(0.95)))/N
# 0.90
h<-sum((rstar>qnorm(0.025))&(rstar<qnorm(0.975)))/N
# 0.95
i<-sum((rstar>qnorm(0.005))&(rstar<qnorm(0.995)))/N
# 0.99

err.g<-c(sum(rstar<qnorm(0.05))/N,sum(rstar>qnorm(0.95))/N)
err.h<-c(sum(rstar<qnorm(0.025))/N,sum(rstar>qnorm(0.975))/N)
err.i<-c(sum(rstar<qnorm(0.005))/N,sum(rstar>qnorm(0.995))/N)

mat.cop<-matrix(c(a,b,c,d,e,f,g,h,i),3,3,byrow=T)
colnames(mat.cop)<-c("0.90","0.95","0.99")
```

```

rownames(mat.cop)<-c("wald","r","r*")

# qqplot

par(mfrow=c(1,2))
qqnorm(rclass ,main="Normal Q-Q Plot r ")
abline(0,1)
qqnorm(rstar ,main="Normal Q-Q Plot r *",ylim=c(-3,3))
abline(0,1)

#### APPLICAZIONE

# carica i dati

pet<-read.csv("paul.csv",h=T,sep=";",dec=",")
rm<-pet[,5]
ct<-pet[,6]

# grafici

par(mfrow=c(1,2))
hist(log(rm),prob=T,cex.main=0.7,cex.axis=0.7,cex.lab=0.7,
main="Histogram of log(RM)",xlab="log(RM)")
curve(dnorm(x,mean(log(rm)),sd(log(rm))),add=T,col=2)
hist(log(ct),prob=T,cex.main=0.7,cex.axis=0.7,cex.lab=0.7,
main="Histogram of log(CT)",xlab="log(CT)")
curve(dnorm(x,mean(log(ct)),sd(log(ct))),add=T,col=2)

par(mfrow=c(1,1))
curve(dnorm(x,mean(rm),sd(rm)),-3,12,col=2,cex.main=0.7,
cex.axis=0.7,cex.lab=0.7,ylab="density")
curve(dnorm(x,mean(ct),sd(ct)),add=T,col=3,cex.main=0.7,
cex.axis=0.7,cex.lab=0.7)
legend("topleft",col=c(2,3),legend=c("RM","CT"),lty=1,cex=0.7)

dati<-list(x=rm,y=ct)

obj<-rstar(dati,thetainit=c(1,1,1,1,0.5),floglik=llik.dipnorm,
psival=0.5,fsim=datagen.bvnorm,
fpsi=psi.fun.dipnorm,R=1000)
summary(obj)

obj2<-rstar.ci(dati,thetainit=c(1,1,1,1,0.5),floglik=llik.dipnorm,
fsim=datagen.bvnorm,fpsi=psi.fun.dipnorm,R=1000)
summary(obj2)
plot(obj2)

```

Tesi scritta con L^AT_EX 2_ε

Ringraziamenti

Vorrei ringraziare tutte le persone che mi hanno aiutato nella stesura di questa tesi e coloro che hanno reso speciali questi tre anni di studio a Padova. In particolare ringrazio, prima di tutti, il mio relatore, prof. Nicola Sartori, sempre disponibile e puntuale, e senza il quale questa tesi non esisterebbe.

Un enorme grazie va sicuramente dato alle persone più importanti della mia vita, che mi hanno permesso di arrivare sino a qui con grandi sacrifici e senza mai farmi mancare nulla, ossia i miei genitori. Ringrazio anche mio fratello Enzo e mia cognata Jenny per il supporto, e le mie piccole nipoti Giada e Simona che hanno sempre accolto con gioia i miei ritorni a casa.

Un ringraziamento speciale va ad una persona speciale, la mia fidanzata Maria, che sopporta le mie continue prese in giro, che mi è stata vicina in ogni mia scelta e con la quale ho passato, passo e spero passerò tantissimi momenti allegri e felici.

Ringrazio enormemente le persone con cui ho condiviso vita universitaria e non, ovvero i miei coinquilini, ex ed attuali: Cesare per le continue sconfitte a Fifa, Silvia per aver mantenuto le redini della casa evitando il degrado, e Giuseppe (Fenomeno) per le risate e l'istigazione all'illegalità.

Ringrazio le tante persone che ho conosciuto in questi anni: i "vecchi" che mi hanno subito fatto sentire a casa, ossia Nicola (Nik) per gli autoinviti a casa mia, Riccardo (Rik) per le partite a subbuteo che puntualmente perdeva, Antonio (Zigro) per i continui "vi raggiungo dopo..", Maddalena, persona su cui poter sempre contare nonostante sia salentina, Andrea (Andre) per gli esami fatti assieme e i bei momenti in montagna. Ringrazio inoltre Elisabetta (Eli), Alice, Fabio, Francesco (Contin), Francesco (Gatti), Ema, Tiziana, Beppe, Michele (Michi), Paolo, Annamaria, Alessia e Teresa. Ringrazio anche

i "nuovi" arrivati: Max (Prof) per la sua saggezza nel campo della statistica e femminile, Max (Ausiella) per le domande incomprensibili, Annalisa, Tamara, Luca (Bissa), Alex e Stefano.

Infine, un sincero ringraziamento va agli amici di "casa" che ci sono sempre, nonostante le distanze ci allontanino: Lorenzo, Davide, Giovanni, Nicola, Enzo, Daniele e Gianni.

