

Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Magistrale in  
Scienze Statistiche



**Eliminazione dei parametri di disturbo: inferenza sul  
rapporto di coefficienti di regressione nel modello  
lineare**

Relatore Prof. Laura Ventura  
Dipartimento di Scienze Statistiche

Laureando: Andrea Calzavara  
Matricola N. 1036158

Anno Accademico 2012/2013



# Indice

<b>Introduzione</b> .....	<b>1</b>
---------------------------	----------

## Capitolo 1

<b>Eliminazione dei parametri di disturbo</b> .....	<b>4</b>
---	----------

1.1 Inferenza statistica .....	4
--------------------------------	---

1.2 Paradigma fisheriano .....	6
--------------------------------	---

1.2.1 Principio d'invarianza rispetto alla parametrizzazione .....	10
--	----

1.3 Paradigma bayesiano .....	11
-------------------------------	----

1.4 Eliminazione dei parametri di disturbo .....	13
--	----

1.4.1 Verosimiglianza condizionata .....	14
--	----

1.4.2 Verosimiglianza marginale .....	16
---------------------------------------	----

1.4.3 Verosimiglianza profilo .....	18
-------------------------------------	----

1.4.4 Verosimiglianza condizionata approssimata .....	20
---	----

1.4.5 Verosimiglianza profilo modificata .....	23
--	----

1.4.6 Verosimiglianza integrata .....	26
---------------------------------------	----

1.4.7 Alcuni esempi di confronto .....	31
--	----

1.5 Conclusioni .....	38
-----------------------	----

## Capitolo 2

<b>Metodi ibridi per l'eliminazione dei parametri di disturbo</b> .....	<b>40</b>
---	-----------

2.1 Distribuzione a posteriori marginale .....	41
--	----

2.2 Uso di pseudo-verosimiglianze alternative a $L_I(\psi)$ .....	42
---	----

2.3 Matching priors .....	43
---------------------------	----

2.3.1	Parametrizzazione ortogonale .....	43
2.3.1	Parametrizzazione non ortogonale .....	44
2.3.2	Interpretazione frequentista delle <i>matching priors</i> .....	50
2.4	Approssimazione della funzione di ripartizione di $\pi(\psi y)$ .....	53
2.4.1	Simulazione attraverso la <i>tail area approximation</i> .....	54
2.5	Conclusioni.....	57
<b>Capitolo 3</b>		
<b>Inferenza sul rapporto di coefficienti in un modello di regressione normale .....</b>		<b>59</b>
3.1	Il modello.....	60
3.2	Parametrizzazione ortogonale.....	62
3.3	Metodologia classica .....	63
3.3.1	Verosimiglianza profilo.....	63
3.3.2	Verosimiglianza condizionata approssimata .....	64
3.3.3	Verosimiglianza integrata.....	65
3.4	Metodologia ibrida .....	67
3.5	Intervalli di confidenza.....	68
3.6	Esempi .....	72
3.7	Studi di simulazione .....	84
3.7.1	Primo studio di simulazione. ....	85
3.7.2	Secondo studio di simulazione. ....	89
3.7.3	Terzo studio di simulazione.....	93
3.8	Conclusioni.....	102
<b>Conclusioni.....</b>		<b>104</b>
<b>Bibliografia.....</b>		<b>106</b>

## Introduzione

In diversi ambiti applicativi è frequente dover utilizzare, modelli complessi. La complessità, ad esempio, può essere data dal numero di parametri che vengono utilizzati per spiegare un determinato fenomeno.

Si pensi, ad esempio, al modello di regressione, dove è sovente trovare un numero di regressori elevato e dove si è spesso interessati ad analizzare un solo effetto o una sua trasformazione.

Per risolvere questo problema, si possono applicare i metodi classici dell'inferenza statistica in ambito frequentista o bayesiano (Pace e Salvan, 1996 e Severini, 2000). I principali strumenti sono le funzioni di verosimiglianza condizionata, marginale, profilo, condizionata approssimata, profilo modificata in ambito frequentista e la funzione di verosimiglianza integrata per l'approccio bayesiano.

Recenti risultati, in una visione più moderna nell'ambito dell'eliminazione dei parametri di disturbo, hanno sviluppato una nuova metodologia inferenziale ibrida, che unisce elementi della visione frequentista con elementi dell'area bayesiana. Il risultato è una funzione del solo parametro d'interesse che può essere interpretato sia come una versione penalizzata di alcune funzioni di pseudo-verosimiglianza, sia come una pseudo-distribuzione a posteriori marginale (Ventura *et al.*, 2009, 2012 e 2013). Seppur non ortodossa, questa tecnica inferenziale ha diversi pregi. All'interno del contesto bayesiano non richiede di dover elicitarne una distribuzione a priori per tutto il parametro, che può essere oneroso sia nell'approccio oggettivo che soggettivo, né di calcolare integrali multidimensionali. Nell'approccio frequentista permette di ottenere una versione penalizzata delle funzioni di pseudo-verosimiglianza marginale, condizionata, condizionata approssimata e profilo modificata.

L'obiettivo principale della tesi è quello di indagare e applicare il metodo ibrido per l'eliminazione dei parametri di disturbo all'inferenza sul rapporto di coefficienti di

regressione nel modello lineare. Un problema, infatti, tale per cui l'utilizzo della metodologia classica può portare a risultati poco utili (Ghosh *et al* 2006).

La tesi è articolata in tre capitoli.

Il primo capitolo fornisce una panoramica sui principali concetti delle metodologie frequentista e bayesiana. In particolare, nella prima parte si illustrano le nozioni di base dell'inferenza frequentista e bayesiana, che vengono poi utilizzate nella tesi. Nella seconda parte del capitolo si presentano le tecniche classiche per l'eliminazione dei parametri di disturbo, ossia le funzioni di verosimiglianza condizionata, marginale, profilo, condizionata approssimata, profilo modificata per l'approccio frequentista e la funzione di verosimiglianza integrata in ambito bayesiano.

Nel secondo capitolo si presentano alcuni recenti sviluppi per l'eliminazione dei parametri di disturbo, introducendo ad una nuova tecnica ibrida. Tale metodologia ha il pregio di fondere i due paradigmi inferenziali e di poter essere utilizzata in entrambi gli ambiti. Inoltre, viene mostrata un'approssimazione della funzione di ripartizione della distribuzione a posteriori per il parametro d'interesse a partire dalla quale si può costruire un intervallo di credibilità bilaterale. Infine, viene proposto un nuovo schema di simulazione (HOTA) basato sulla *tail area approximation* (Brazzale *et al.*, 2007 e Ruli *et al.*, 2013).

Il terzo capitolo presenta l'inferenza sul rapporto di coefficienti di regressione nel modello lineare, mostrando che l'applicazione delle tecniche moderne porta ad avere risultati migliori rispetto ai metodi di eliminazione dei parametri di disturbo classici. In particolare, confrontando le due metodologie, classica e moderna, al caso di studio, si dimostra che l'utilizzo di alcune funzioni di pseudo-verosimiglianza classiche possono portare a conclusioni inutilizzabili. Questi risultati sono confermati da alcuni studi di simulazione effettuati sulla distorsione e l'errore quadratico medio delle stime e sull'ampiezza e la copertura degli intervalli di confidenza ottenuti dalla statistica  $F$  e dalle statistiche log-rapporto di pseudo-verosimiglianza. Dai risultati degli studi di simulazione emerge che a partire dalla funzione di verosimiglianza profilo modificata penalizzata si ottengono mediamente stime meno variabili e un'ampiezza media degli intervalli più piccola. Queste conclusioni confermano alcuni risultati teorici descritti da Ventura e Racugno (2012), dove si dimostra che l'errore quadratico medio della stima di massima verosimiglianza profilo modificata penalizzata e l'ampiezza media

dell'intervallo ottenuto dalla statistica log-rapporto di verosimiglianza profilo modificata penalizzata sono inferiori in media alle stesse quantità derivate dalla funzione di log-verosimiglianza profilo modificata.

Inoltre, gli studi mostrano che le quantità analizzate sono particolarmente sensibili alla numerosità campionaria e al valore assunto dal parametro d'interesse. Infatti, si mostra che all'aumentare di  $n$  o al ridursi di  $\psi$  tutte le stime di massima pseudo-verosimiglianza analizzate tendono a mostrare un errore quadratico medio minore e le ampiezze medie degli intervalli di confidenza considerati evidenziano un andamento decrescente.

Per le analisi statistiche presentate nella tesi è stato utilizzato il programma statistico R. In questa tesi è stata utilizzata la versione R.15.2 per Windows.

# Capitolo 1

## Eliminazione dei parametri di disturbo

In questo capitolo sono sinteticamente richiamati i principali metodi inferenziali per l'eliminazione dei parametri di disturbo. Tali procedure sono accompagnate da alcuni esempi e saranno utilizzate nel Capitolo 3.

In primo luogo vengono analizzati i due principali paradigmi inferenziali, quello fisheriano e quello bayesiano, e le più importanti quantità che li caratterizzano. Per tali argomenti, alcuni riferimenti sono, ad esempio, Azzalini (2008 cap. 3), Pace e Salvan (1996, capp. 2 e 3) e Severini (2000, cap. 3).

In secondo luogo sono presentati alcuni metodi per l'eliminazione dei parametri di disturbo basati su opportune funzioni di pseudo-verosimiglianza, quali la verosimiglianza condizionata, marginale, profilo, condizionata approssimata, profilo modificata e integrata. Ogni funzione di pseudo-verosimiglianza viene illustrata con un esempio. Infine, sono proposti degli esempi di confronto tra i diversi metodi. Un'ampia rassegna su queste tematiche si può trovare, ad esempio, in Pace e Salvan (1996, cap. 4) e Severini (2000, capp.8 e 9).

### 1.1 Inferenza statistica

Si consideri un insieme di dati osservati  $y = (y_1, \dots, y_n)$ , a componenti indipendenti ed identicamente distribuite (i.i.d.) da una variabile casuale  $Y$ . Obiettivo dell'analisi statistica è la ricostruzione della densità di probabilità  $p^0(y)$  della variabile casuale  $Y$ , in modo da ottenere una descrizione sintetica del fenomeno oggetto di studio, che ne permetta l'interpretazione e la previsione.

Delimitando le forme possibili per  $p^0(y)$ , si delinea una famiglia di distribuzioni di probabilità compatibili con i dati osservati  $y$ . La famiglia  $\mathcal{F}$ , così formata, è detta **modello statistico**. Se la vera funzione di densità  $p^0(y)$  che ha generato  $y$  appartiene a tale famiglia  $\mathcal{F}$ , allora si dice che il modello statistico è **correttamente specificato**.

Il modello statistico  $\mathcal{F}$  può avere una estensione più o meno ampia a seconda delle informazioni disponibili. Si possono delineare tre diversi livelli specificazione: parametrica, semiparametrica e non parametrica. In questa tesi si considerano esclusivamente i modelli parametrici.

Un **modello statistico parametrico** (m.s.p.) è una famiglia  $\mathcal{F} = \{p(y; \theta); \theta \in \Theta \subseteq \mathbb{R}^p, y \in \mathcal{Y}\}$ , dove  $\Theta$  costituisce lo spazio parametrico e  $\mathcal{Y}$  lo spazio campionario. Il **parametro**  $\theta$  rappresenta una qualche quantità che caratterizza sia l'esperimento casuale che produce i dati, sia la popolazione da cui i dati sono tratti.

Se in un m.s.p. c'è corrispondenza biunivoca tra  $\Theta$  e  $\mathcal{F}$ , allora si dice che il parametro  $\theta$  è **identificabile**. Inoltre, un m.s.p. è **correttamente specificato** se  $p^0(y) = p(y; \theta_0)$  per un valore  $\theta_0 \in \Theta$ , detto **vero valore del parametro**.

Vi sono diversi paradigmi dell'inferenza statistica. In questa tesi ci si limita a richiamare il paradigma fisheriano e quello bayesiano. In particolare:

- il **paradigma fisheriano** si basa sul principio del campionamento ripetuto e assume grande rilievo il concetto di verosimiglianza. Il campionamento ripetuto considera il campione osservato come uno dei possibili campioni che si sarebbero potuti ottenere ripetendo un gran numero di volte, nelle stesse condizioni, l'operazione di campionamento. La verosimiglianza racchiude tutte le informazioni che un generico campione può fornire;
- il **paradigma bayesiano** si fonda sull'assegnazione di una distribuzione di probabilità a priori sugli elementi di  $\mathcal{F}$ . Tale distribuzione può essere scelta direttamente dal ricercatore in base alle informazioni in possesso (paradigma bayesiano personalistico), oppure può riflettere una procedura oggettiva (paradigma bayesiano non personalistico).

## 1.2 Paradigma fisheriano

Il concetto di verosimiglianza è centrale per l'inferenza statistica parametrica nella visione fisheriana.

Dato un m.s.p., una volta osservato il campione casuale semplice  $y = (y_1, \dots, y_n)$ , la funzione di densità si riduce ad una legge che dipende unicamente dal parametro  $\theta = (\theta_1, \dots, \theta_p)$ . Si definisce **funzione di verosimiglianza** di  $\theta$  basata sui dati  $y$ , la funzione da  $\Theta \rightarrow \mathbb{R}^+$  data da

$$L(\theta) = L(\theta; y) = c(y) \prod_{i=1}^n p(y_i; \theta), \quad (1.1)$$

dove  $c(y)$  è una costante di proporzionalità arbitraria che non dipende da  $\theta$ . Se due funzioni di verosimiglianza differiscono esclusivamente per la costante moltiplicativa, allora si definiscono **equivalenti**. Infatti, per il **principio debole di verosimiglianza**, con riferimento ad un dato m.s.p., se due punti  $y$  e  $z$ , appartenenti al medesimo spazio campionario, sono caratterizzati dal fatto che  $L(\theta; y) \propto L(\theta; z)$ , allora devono condurre alle medesime conclusioni inferenziali. Il fatto che i due modelli siano proporzionali indica che possono differire unicamente dalla costante  $c(\cdot)$ , che perciò può essere trascurata nella funzione di verosimiglianza.

Spesso è conveniente considerare la trasformata logaritmica di  $L(\theta)$ , ottenendo la **funzione di log-verosimiglianza**

$$l(\theta) = \log L(\theta) = \sum_{i=1}^n \log p(y_i; \theta), \quad (1.2)$$

dove per convenzione  $l(\theta) = -\infty$  se  $L(\theta) = 0$ . Si definiscono equivalenti due funzioni di log-verosimiglianza che differiscono tra loro per una costante additiva.

Assumendo che la funzione di log-verosimiglianza sia una funzione di  $\theta$  sufficientemente regolare, ossia che ammetta derivate parziali fino agli ordini richiesti, si possono definire le seguenti quantità di verosimiglianza:

- **Funzione punteggio:** è il vettore delle derivate parziali prime di  $l(\theta)$  rispetto alle componenti di  $\theta$ , ed è indicato con  $l_* = l_*(\theta) = (l_1, \dots, l_p)^T$  con generico elemento  $l_s = \partial l(\theta) / \partial \theta_s$ , per  $s = 1, \dots, p$ ;

- **Informazione osservata:** è la matrice di dimensione  $p \times p$  delle derivate parziali seconde di  $l(\theta)$  cambiate di segno, ed è indicata con  $j = j(\theta) = [-l_{sr}]$ , dove  $l_{sr} = \partial^2 l(\theta) / (\partial \theta_s \partial \theta_r)$ , per  $r, s = 1, \dots, p$ . In forma matriciale

$$j(\theta) = \left[ -\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \right] = [j_{sr}]; \quad (1.3)$$

- **Informazione attesa (di Fisher):** è la matrice di dimensioni  $p \times p$  data dal valore atteso dell'informazione osservata, ossia

$$i = i(\theta) = \mathbb{E}_\theta [j(\theta)] = \mathbb{E}_\theta \left[ -\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \right] = [i_{sr}]. \quad (1.4)$$

Si dice che si è in presenza di un problema regolare di stima se sono verificate le seguenti condizioni (per maggiori approfondimenti si veda Severini, 2000, § 3.4, oppure Azzalini, 2008, § 3.2.3):

- 1) il modello ammette verosimiglianza regolare, ossia  $\Theta$  è aperto e  $l(\theta)$  è differenziabile almeno tre volte rispetto a  $\theta$ , con derivate parziali continue;
- 2) è soddisfatta l'identificabilità, ossia le funzioni di probabilità definite in corrispondenza di due differenti valori di  $\theta$  sono distinte. Ossia, se  $\theta' \neq \theta \in \Theta$ , allora  $\mathbb{P}_\theta \left[ \frac{p_Y(Y; \theta')}{p_Y(Y; \theta)} = 1 \right] < 1$ ;
- 3) il modello è correttamente specificato, ossia  $\theta_0 \in \Theta$ ;
- 4) lo spazio campionario  $\mathcal{Y}$  non dipende da  $\theta$ ;
- 5) la dimensione di  $\Theta$  non dipende dalla numerosità campionaria;
- 6) è possibile scambiare le operazioni di derivazione e di integrazione.

Sotto queste assunzioni, si possono ricavare alcuni risultati di notevole importanza per l'inferenza:

- la **funzione punteggio ha valore atteso nullo**, ossia

$$\mathbb{E}_\theta [l_*(\theta)] = 0, \quad \forall \theta \in \Theta.$$

Questa proprietà stabilisce che l'equazione di verosimiglianza  $l_*(\theta) = 0$  è un'equazione di stima non distorta;

- vale l'**identità dell'informazione**, ossia

$$\mathbb{E}_\theta [j(\theta)] = \mathbb{E}_\theta [l_*(\theta) l_*(\theta)^T], \quad \forall \theta \in \Theta.$$

Più precisamente, la matrice dell'informazione attesa è pari alla matrice di covarianza della funzione punteggio e come tale è una matrice definita non negativa.

- vale la **diseguaglianza di Wald**, ossia

$$\mathbb{E}_\theta[l(\theta)] > \mathbb{E}_\theta[l(\theta')], \quad \theta' \neq \theta.$$

I primi due risultati sono noti in letteratura anche come le prime due identità di Bartlett. Il valore di  $\theta$  che massimizza  $L(\theta)$ , o in modo equivalente  $l(\theta)$ , è detto **stima di massima verosimiglianza** (s.m.v.) di  $\theta$ , ossia

$$\hat{\theta} = \operatorname{argsup}_{\theta \in \Theta} L(\theta).$$

In un modello con verosimiglianza regolare,  $\hat{\theta}$  va cercato tra le soluzioni dell'equazione di verosimiglianza  $l_*(\theta) = 0$ .

Sempre sotto le condizioni di regolarità, si possono ottenere alcuni risultati asintotici rilevanti per  $\hat{\theta}$ :

- la s.m.v.  $\hat{\theta}$  converge in probabilità al vero valore del parametro, ossia

$$\hat{\theta} \xrightarrow{P} \theta;$$

- la distribuzione asintotica dello stimatore di massima verosimiglianza è normale, ossia

$$\hat{\theta} \xrightarrow{d} N_p(\theta, i(\theta)^{-1}),$$

dove al posto di  $i(\theta)$  si può sostituire  $i(\hat{\theta})$ ,  $j(\theta)$  o  $j(\hat{\theta})$ , ovvero le diverse quantità di verosimiglianza valutate in  $\hat{\theta}$ ;

- se  $\theta = (\psi, \lambda)$ , dove  $\psi$  e  $\lambda$  rappresentano due componenti di  $\theta$  di dimensione rispettivamente  $k$  e  $n - k$ , la distribuzione asintotica per la sola componente  $\hat{\psi}$  è

$$\hat{\psi} \xrightarrow{d} N_k(\psi, i^{\psi\psi}(\psi, \lambda)),$$

dove  $i^{\psi\psi}(\psi, \lambda) = [i(\psi, \lambda)^{-1}]_{(\psi, \psi)}$  definisce il blocco  $(\psi, \psi)$  dell'inversa della matrice di informazione attesa. Un analogo risultato vale per  $\hat{\lambda}$ .

Tali risultati hanno delle implicazioni asintotiche importanti per  $\hat{\theta}$ , ossia:

- $\hat{\theta}$  è non distorto:  $\lim_{n \rightarrow +\infty} \mathbb{E}_\theta[\hat{\theta}] = \theta$ ;
- $\hat{\theta}$  è consistente.

Per approfondimenti si veda Pace e Salvani (1996, § 9).

Attraverso la funzione di verosimiglianza  $L(\theta)$  è possibile:

1. basare la verifica di un'ipotesi della forma  $H_0: \theta \in \Theta_0 \subset \Theta$  sul confronto fra il massimo valore di  $L(\theta)$  per  $\theta \in \Theta_0$ ,  $L(\hat{\theta}_0)$ , ed il massimo valore di  $L(\theta)$  per  $\theta \in \Theta$ , ossia  $L(\hat{\theta})$ ;
2. ottenere, come stima intervallare, una regione di confidenza  $\hat{\Theta}(y)$  per  $\theta$ .

Per la verifica dell'ipotesi nulla semplice  $H_0: \theta = \theta_0$  contro l'alternativa  $H_1: \theta \neq \theta_0$  è possibile utilizzare la statistica **log-rapporto di verosimiglianza**

$$W(\theta_0) = 2\{l(\hat{\theta}) - l(\theta_0)\} \quad (1.5)$$

È possibile dimostrare che (Pace e Salvan, 1996, § 3.5)  $W(\theta_0)$  ha distribuzione nulla approssimata chi-quadrato con  $p$  gradi di libertà.

Si possono, anche, costruire regioni di confidenza di livello nominale approssimato  $1 - \alpha$ , della forma

$$\hat{\Theta}(y) = \{\theta \in \Theta: W(\theta) \leq \chi_{p;1-\alpha}^2\}, \quad (1.6)$$

con  $\chi_{p;1-\alpha}^2$  quantile- $(1 - \alpha)$  della distribuzione  $\chi_p^2$ . La regione  $\hat{\Theta}(y)$  può anche essere riscritta come

$$\hat{\Theta}(y) = \left\{ \theta \in \Theta: l(\theta) \geq l(\hat{\theta}) - \frac{1}{2} \chi_{p;1-\alpha}^2 \right\}.$$

Un'altra statistica su cui basare test e regioni di confidenza per  $\theta$ , con distribuzione asintotica nulla  $\chi_p^2$ , è la statistica **test di Wald**, definita come

$$W_e(\theta_0) = (\hat{\theta} - \theta_0)^T i(\theta_0) (\hat{\theta} - \theta_0). \quad (1.7)$$

La statistica  $W_e(\theta_0)$  misura la distanza fra la stima  $\hat{\theta}$  ed il valore ipotizzato  $\theta_0$ , e si basa sulla distribuzione asintotica dello stimatore di massima verosimiglianza.

In modo analogo, si può costruire una regione di confidenza di livello nominale approssimato  $1 - \alpha$ , come

$$\hat{\Theta}(y) = \{\theta \in \Theta: W_e(\theta) \leq \chi_{p;1-\alpha}^2\}. \quad (1.8)$$

Nonostante l'equivalenza asintotica fra le due statistiche  $W(\theta_0)$  e  $W_e(\theta_0)$ , solo la statistica  $W(\theta)$  è invariante rispetto a riparametrizzazioni (cfr. § 1.2.1) e non pone, nelle regioni di confidenza, valori per il parametro  $\theta$  inammissibili.

Se il parametro  $\theta$  è scalare ( $p = 1$ ), si possono definire le versioni unilaterali di  $W(\theta_0)$  e  $W_e(\theta_0)$  (Pace e Salvan, 2001, §6.6), ossia date da

$$r = r(\theta_0) = \text{sgn}(\hat{\theta} - \theta_0)\sqrt{W(\theta_0)} \quad (1.9)$$

e

$$r_e = r_e(\theta_0) = (\hat{\theta} - \theta_0)\sqrt{i(\theta_0)}, \quad (1.10)$$

che hanno entrambe distribuzione asintotica nulla  $N(0,1)$ .

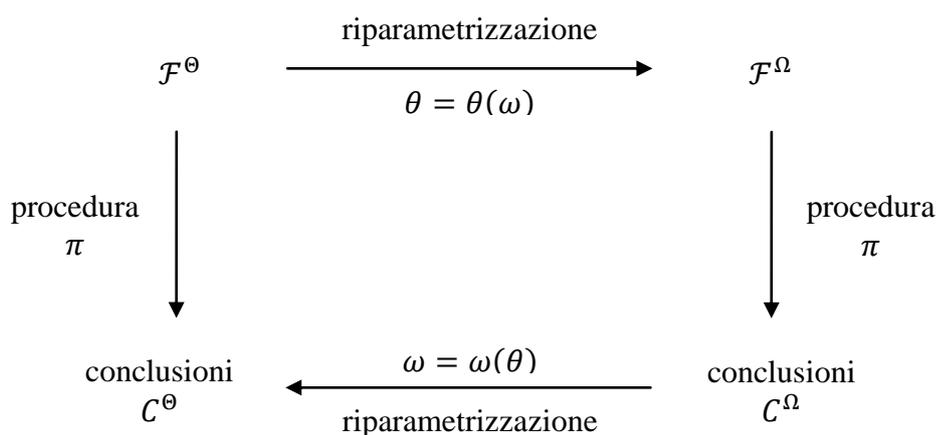
È possibile costruire intervalli di confidenza di livello nominale approssimato  $(1 - \alpha)$  per il parametro scalare  $\theta$  basati su  $r(\theta)$ , della forma

$$\hat{\Theta}(y) = \{\theta \in \Theta: -z_{1-\alpha/2} \leq r(\theta) \leq z_{1-\alpha/2}\}, \quad (1.11)$$

con  $z_{1-\alpha/2}$  quantile- $(1 - \alpha/2)$  di una  $N(0,1)$ . Per approfondimenti si vedano Pace e Salvan (1996, § 3.5 e 2001, §3.6 e 6.8).

### 1.2.1 Principio d'invarianza rispetto alla parametrizzazione

La parametrizzazione di un m.s.p.  $\mathcal{F}$  è arbitraria. Infatti, fissare una parametrizzazione equivale a scegliere un sistema di coordinate nello spazio parametrico. Se  $\theta$  costituisce una parametrizzazione di  $\mathcal{F}$ , qualsiasi trasformazione biunivoca e continua,  $\omega = \omega(\theta)$ , può fungere da sistema di coordinate alternativo. È allora importante valutare se una data procedura statistica risulta invariante rispetto alla parametrizzazione usata, ossia se le conclusioni inferenziali a cui si perviene sono invarianti alla parametrizzazione. È infatti una proprietà desiderabile che riparametrizzando il modello di partenza, ed applicando la procedura statistica, si arrivi al medesimo risultato che si sarebbe ottenuto prima applicando la procedura, e successivamente trasferendo la conclusione inferenziale sulla nuova parametrizzazione. Se una procedura statistica è invariante rispetto alla parametrizzazione vale:



Esempi di procedure invarianti sono la funzione di log-verosimiglianza e la stima di verosimiglianza; al contrario i metodi basati sulla normalità asintotica dello stimatore di massima verosimiglianza in genere non risultano invarianti alla parametrizzazione.

### 1.3 Paradigma bayesiano

Nell'ottica classica fisheriana dell'inferenza il parametro  $\theta$  rappresenta una quantità fissata ma ignota, che indicizza le diverse distribuzioni di probabilità. Il fine principale del paradigma classico è quello di individuare quale è la vera distribuzione che ha generato i dati, ossia individuare, per i modelli statistici parametrici, il vero valore  $\theta_0$ .

Nel paradigma bayesiano, il parametro  $\theta$  non rimane vincolato ad essere un elemento fisso, ma assume i connotati di una variabile aleatoria, la cui distribuzione iniziale dipende dalle informazioni in possesso dal ricercatore. La **distribuzione a priori** su  $\theta$  rappresenta la sintesi delle informazioni che si hanno sul problema specifico.

Nella visione bayesiana dell'inferenza, le informazioni espresse dalla distribuzione a priori si combinano con la densità delle osservazioni, mediante il teorema di Bayes. In questo modo viene fornita la distribuzione del parametro  $\theta$  condizionata ai dati, detta **distribuzione a posteriori** del parametro.

Sia  $\pi(\theta)$  una distribuzione a priori sul parametro  $\theta$  e sia  $L(\theta)$  la funzione di verosimiglianza. Il teorema di Bayes fornisce la **distribuzione a posteriori** per  $\theta$  che, nel caso in cui  $\theta$  sia una variabile aleatoria assolutamente continua, risulta data da

$$\pi(\theta|y) = \frac{\pi(\theta)L(\theta)}{\int_{\Theta} \pi(\theta)L(\theta)d\theta}. \quad (1.12)$$

Si noti che  $\int_{\Theta} \pi(\theta)L(\theta)d\theta$  è una costante di normalizzazione che non dipende da  $\theta$  e che definisce la distribuzione marginale di  $y$ .

La distribuzione a posteriori  $\pi(\theta|y)$  rappresenta la distribuzione del parametro  $\theta$  condizionata al risultato dell'esperimento. In essa è racchiusa tutta l'informazione su  $\theta$  e su di essa ci si basa per produrre indicatori sintetici su  $\theta$ . Alcuni esempi possono essere il valore atteso a posteriori  $\mathbb{E}_{\theta|Y}[\theta|y]$ , la mediana a posteriori  $Mediana(\theta|y)$ , oppure la moda a posteriori  $Moda(\theta|y)$ , che è pari a  $argsup_{\theta \in \Theta} \pi(\theta|y)$ .

È evidente che le informazioni a priori hanno un ruolo importante nell'inferenza su  $\theta$ . Due ricercatori, di fronte al medesimo esperimento, potrebbero scegliere diverse distribuzioni a priori per il parametro, arrivando perciò a risultati diversi. La scelta di una distribuzione iniziale ha anche ripercussioni di tipo computazionale; ad esempio, per il calcolo dei momenti della distribuzione oppure della costante di normalizzazione. È indubbio che questa componente risulti essere il punto cruciale di tutta l'impostazione bayesiana.

Per ovviare ai problemi di soggettività, sono state proposte delle procedure per la costruzione di distribuzioni a priori non informative. Tali distribuzioni possono essere anche improprie, ossia la densità  $\pi(\theta)$  può non avere integrale finito su  $\Theta$ . L'importante è verificare che la distribuzione a posteriori risulti una distribuzione propria, qualunque sia il risultato campionario osservato. Nel seguito, si fornisce un breve cenno alle distribuzioni non informative uniformi e di Jeffreys per il caso di  $\theta$  scalare; per maggiori approfondimenti si veda Liseo (2008) oppure Berger *et al.* (1999).

Una prima e semplice soluzione per ottenere distribuzioni non informative è assegnare a  $\theta$  una distribuzione a priori uniforme su  $\Theta$ , ossia

$$\pi^U(\theta) \propto 1.$$

Questa scelta, però, risulta criticabile. Infatti:

1. risulta essere una legge di probabilità impropria se lo spazio parametrico  $\Theta$  non è limitato;
2. non è invariante rispetto a riparametrizzazioni.

Se il primo punto, in genere, può non risultare un vero problema, la mancanza di invarianza alla parametrizzazione è il punto debole di tale scelta.

La mancanza di invarianza della legge uniforme rispetto a trasformazioni biunivoche di  $\theta$  ha indotto a formulare criteri diversi per la costruzione di distribuzioni a priori non

informative. Una soluzione è la distribuzione non informativa di Jeffreys (Jeffreys, 1946), definita come

$$\pi^J(\theta) \propto \sqrt{|i(\theta)|}, \quad \theta \in \Theta. \quad (1.13)$$

La principale proprietà di questa distribuzione consiste nell'invarianza rispetto a riparametrizzazioni. Infatti, data una trasformazione biunivoca di  $\theta$ , tale che  $i(\omega)$  esista e sia definita positiva, la distribuzione non informativa di Jeffreys per la nuova parametrizzazione  $\omega = \omega(\theta)$  risulta (Liseo, 2008)

$$\pi_\omega^J(\omega) \propto \pi_\theta^J(\theta(\omega)) \left| \frac{\partial \omega}{\partial \theta} \right|.$$

Una rassegna su distribuzioni non informative è disponibile in Liseo (2008).

#### **1.4 Eliminazione dei parametri di disturbo**

In molte situazioni, come ad esempio in un modello di regressione, il parametro  $\theta$  è un vettore  $p$ -dimensionale. Tuttavia, gli aspetti d'interesse sono esprimibili tramite un parametro  $k$ -dimensionale  $\psi$ , detto parametro d'interesse. Il parametro  $\lambda$ , tale che  $\theta = (\psi, \lambda)$ , costituisce il parametro di disturbo. In tal caso, il m.s.p. è indicizzato come  $\mathcal{F} = \{p(y; \psi, \lambda): \psi \in \Psi \subseteq \mathbb{R}^k, \lambda \in \Lambda \subseteq \mathbb{R}^{p-k}, y \in \mathcal{Y}\}$ .

Per l'inferenza solo su  $\psi$ , una possibile soluzione consiste nel costruire una funzione, che dipende dal solo parametro di interesse e dai dati, in cui la componente di disturbo viene eliminata. In letteratura sono state proposte diverse soluzioni, sia in ambito frequentista che bayesiano. Si vedano Pace e Salvan (1996, cap. 4) e Severini, (2000, cap. 8) per rassegne e approfondimenti.

Le funzioni che dipendono da  $\psi$  solamente e dai dati sono dette **funzioni di pseudo-verosimiglianza**. Più precisamente, una funzione di pseudo-verosimiglianza è una funzione che dipende dal solo parametro d'interesse  $\psi$  e che si comporta, sotto uno o più aspetti, come una verosimiglianza propria.

Si possono distinguere due diverse situazioni:

- la pseudo-verosimiglianza è ottenuta a partire da un modello statistico definito come riduzione del modello  $\mathcal{F}$  originario. La pseudo-verosimiglianza così ottenuta è una verosimiglianza in senso proprio;

- non è definito un modello ridotto da cui la pseudo-verosimiglianza possa pensarsi dedotta in via diretta, ovvero la funzione non è dedotta da una fattorizzazione del modello originale. Tutte le proprietà della pseudo-verosimiglianza vanno indagate caso per caso.

Nel seguito si presentano, sinteticamente, le principali pseudo-verosimiglianze trattate in letteratura: condizionata, marginale, profilo, condizionata approssimata, profilo modificata ed integrata. Solo l'ultima rientra nel paradigma bayesiano, mentre le altre sono soluzioni proposte in ambito frequentista. Per approfondimenti si rinvia a Pace e Salvan (1996, cap. 4) e Severini (2000, cap. 8 e 9).

### 1.4.1 Verosimiglianza condizionata

Un possibile approccio per derivare una pseudo-verosimiglianza per  $\psi$  consiste nell'eliminare il parametro  $\lambda$  mediante il condizionamento ad una statistica.

Si supponga che per il modello  $p(y; \theta)$  si possano individuare le statistiche sufficienti minimali  $(t, s)$  tali che, per il teorema di fattorizzazione, la densità possa essere scomposta in

$$p(y; \theta) = p_{(T,S)}(t, s; \theta) p_{Y|(T,S)=(t,s)}(y|t, s),$$

per cui la verosimiglianza risulta essere proporzionale al solo fattore  $p_{(T,S)}(t, s; \theta)$ . Ora si supponga che la densità congiunta di  $(T, S)$  possa essere fattorizzata come

$$p_{(T,S)}(t, s; \theta) = p_{T|S=s}(t|s; \psi) p_S(s; \psi, \lambda).$$

La statistica  $s$  risulta sufficiente per il modello quando  $\psi$  è fissato. Purché la componente corrispondente a  $p_S(\cdot)$  sia trascurabile, la funzione di verosimiglianza corrispondente a  $p_{T|S=s}(t|s; \psi)$  non dipende da  $\lambda$ , ma solo da  $\psi$ . La funzione di verosimiglianza corrispondente risulta una verosimiglianza propria poiché derivata da una funzione di densità.

Le funzioni di verosimiglianza e di log-verosimiglianza condizionata risultano, rispettivamente, date da

$$L_C(\psi) \propto p_{(T|S)}(t|s; \psi) \tag{1.14}$$

e

$$l_C(\psi) = \log L_C(\psi). \quad (1.15)$$

Per costruire  $L_C(\psi)$  non si tiene in considerazione il fattore  $p_S(s; \psi, \lambda)$ , che dipende anche da  $\psi$ . È importante quindi che la statistica  $s$  non contenga informazione utile su  $\psi$ , in presenza di  $\lambda$ , ossia che risulti ancillare per  $\psi$ . La proprietà di ancillarità in questo contesto, caratterizzato da partizioni d'interesse e di disturbo, risulta un concetto diverso dal caso generale. Una statistica  $s$  è detta  $S$ -ancillare per  $\psi$ , in presenza del parametro di disturbo  $\lambda$ , se la famiglia delle funzioni di densità  $\{p_S(s; \psi, \lambda); \lambda \in \Lambda\}$  non cambia al variare di  $\psi$ . Una statistica  $s$ , perciò, viene definita  $S$ -ancillare se e solo se esiste una riparametrizzazione del modello che non altera l'interesse (*interest respecting*)  $(\psi, \phi)$  tale che la distribuzione di  $s$  dipenda solo da  $\phi$ .

Per approfondimenti si veda Pace e Salvani (1996, § 4.2.3-4.2.4) e Severini (2000, § 8.2).

### Esempio 1.1

Siano  $t$  e  $v$  due osservazioni indipendenti da due variabili casuali con distribuzione di Poisson di medie, rispettivamente,  $\mu$  e  $\eta$ . La densità congiunta risulta

$$p_{(T,V)}(t, v; \mu, \eta) = p_T(t; \mu)p_S(s; \eta) = \exp\{t \log \mu + v \log \eta - (\mu + \eta)\} (t! v!)^{-1}.$$

Sia  $\psi = \log \frac{\mu}{\eta}$  il parametro d'interesse e sia  $\lambda = \log \eta$  il parametro di disturbo. La coppia  $(t, v)$  risulta essere una statistica sufficiente minimale per  $\theta = (\psi, \lambda)$ , così come  $(t, t + v) = (t, s)$ . La funzione di verosimiglianza per  $(\psi, \lambda)$  è

$$L(\psi, \lambda) \propto \exp\{t\psi + s\lambda - e^\lambda(e^\psi + 1)\}.$$

Sfruttando la proprietà di additività della distribuzione di Poisson, si può dimostrare che la statistica  $S = T + V$  si distribuisce come una Poisson di media  $e^\lambda(e^\psi + 1)$ . Si può quindi verificare che

$$p_{T|S=s}(t|s; \psi, \lambda) = \frac{e^{\psi t + \lambda s - e^\lambda(e^\psi + 1)} s!}{e^{\lambda s - e^\lambda(e^\psi + 1)} (e^\psi + 1)^s t! (s - t)!} = \binom{s}{t} \left( \frac{e^\psi}{e^\psi + 1} \right)^t \left( \frac{1}{e^\psi + 1} \right)^{s-t},$$

ossia, la distribuzione condizionata di  $T|S = s$  dipende solo dal parametro d'interesse  $\psi$ . Perciò la verosimiglianza condizionata risulta essere proporzionale alla distribuzione di  $T$  dato  $S = s$ , ossia

$$L_C(\psi) \propto \exp\{t\psi - s \log(e^\psi + 1)\}.$$

Si noti che la statistica  $s$  risulta essere  $S$ -ancillare, in quanto si può operare una riparametrizzazione *interest respecting*  $(\psi, \phi)$ , con  $\phi = e^\lambda(e^\psi + 1)$ , tale che la distribuzione di  $S$  dipende solo da  $\phi$ .

■

In generale, la verosimiglianza condizionata è facile da reperire nel caso in cui il modello appartenga alla famiglia esponenziale.

Si consideri il caso di una famiglia esponenziale regolare con parametrizzazione canonica e osservazioni naturali  $v(y)$ , della forma  $v(y) = (t(y), s(y)) = (t, s)$ , con  $t$  e  $s$  di dimensione, rispettivamente,  $k$  e  $p - k$ . Si consideri la corrispondente partizione  $\theta = (\psi, \lambda)$ . La densità di un generico elemento della famiglia è

$$p(y; \psi, \lambda) = h(y) \exp\{\psi^T t(y) + \lambda^T s(y) - K(\psi, \lambda)\},$$

dove  $h(y) \geq 0$  è una funzione solo di  $y$  e  $K(\psi, \lambda)$  è una funzione solo di  $(\psi, \lambda)$  che non dipende da  $y$ .

Per una famiglia esponenziale con elementi di questo tipo, si può dimostrare che (Pace e Salvan, 1996, § 5.4.2) la distribuzione marginale di  $S$  appartiene alla famiglia esponenziale naturale (di ordine  $p - k$ ), ossia

$$p_S(s; \psi, \lambda) = h_S(s) \exp\{\lambda^T s - K_S(\lambda)\},$$

e che la distribuzione condizionata di  $T|S = s$  appartiene ancora alla famiglia esponenziale (di ordine  $k$ ) ed è indipendente da  $\lambda$ , ossia

$$p_{T|S=s}(t; \psi) = h_T(t) \exp\{\psi^T t - K_T(\psi)\},$$

dove  $h_S(s) \geq 0$  è una funzione solo di  $s$ ,  $h_T(t) \geq 0$  è una funzione solo di  $t$ ,  $K_S(\lambda)$  è una funzione di  $\lambda$  mentre  $K_T(\psi)$  è una funzione solo di  $\psi$ .

È possibile quindi ricavare la funzione di verosimiglianza condizionata, data da

$$L_C(\psi) \propto \exp\{\psi^T t - K_T(\psi)\}.$$

La forma della funzione  $K_T(\psi)$  non è sempre desumibile in modo immediato. Per approfondimenti si veda Pace e Salvan (1996, cap. 5).

#### 1.4.2 Verosimiglianza marginale

Un altro approccio per la costruzione di una pseudo-verosimiglianza per  $\psi$  è derivare la distribuzione marginale di una statistica ancillare, che dipenda solamente dal parametro d'interesse. Anche in questo caso il risultato è una verosimiglianza propria, in quanto derivata direttamente da una distribuzione di probabilità.

Si supponga che esista una statistica  $T$  tale che la densità di  $Y$  possa essere scritta come

$$p(y; \psi, \lambda) = p_T(t; \psi) p_{Y|T}(y|t; \psi, \lambda).$$

Nel caso in cui il fattore di verosimiglianza  $p_{Y|T}(\cdot)$  sia trascurabile, l'inferenza su  $\psi$  può essere basata esclusivamente sulla distribuzione marginale di  $T$ , che non dipende dal parametro  $\lambda$ .

La funzione di verosimiglianza marginale per  $\psi$  basata su  $t$  risulta

$$L_M(\psi) \propto p_T(t; \psi). \quad (1.16)$$

La funzione di log-verosimiglianza marginale è

$$l_M(\psi) = \log L_M(\psi). \quad (1.17)$$

### Esempio 2.1

Siano  $Y_1, Y_2, \dots, Y_n$  delle variabili casuali i.i.d. normali con media  $\lambda$  e varianza  $\psi$ , con  $\psi > 0$ . La funzione di verosimiglianza per  $(\psi, \lambda)$ , a meno di costanti moltiplicative, è pari a

$$L(\psi, \lambda) = \psi^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\psi} \sum_{i=1}^n (y_i - \lambda)^2 \right\}.$$

Il modello ammette la statistica sufficiente minimale  $s = (\sum y_i/n, \sum (y_i - \bar{y})^2/n) = (\bar{y}, \hat{\sigma}^2)$ . L'inferenza su  $\psi$  può essere basata sulla legge marginale di  $T = \sum (Y_i - \bar{Y})^2/\psi$ , che non dipende dal parametro di disturbo  $\lambda$ . Infatti, è noto che  $T \sim \chi_{n-1}^2$  e, quindi, per l'inferenza su  $\psi$  può essere usata la funzione di verosimiglianza marginale

$$L_M(\psi) = \psi^{-\frac{n-1}{2}} \exp \left\{ -\frac{1}{2\psi} \sum_{i=1}^n (y_i - \bar{y})^2 \right\}. \quad \blacksquare$$

In generale, la funzione di verosimiglianza marginale è facilmente reperibile solo nei modelli di gruppo. Per approfondimenti si veda Pace e Salvan (1996, cap. 6)

### 1.4.3 Verosimiglianza profilo

Nelle situazioni in cui le funzioni di verosimiglianza marginale e condizionata per il parametro d'interesse  $\psi$  sono calcolabili, esse risultano validi strumenti inferenziali, in quanto si comportano come funzioni di verosimiglianza propria. Purtroppo, queste funzioni sono calcolabili solo in poche situazioni speciali (famiglie esponenziali o di gruppo). Per approfondimenti si veda, ad esempio, Severini (2000, § 8.3).

Un approccio più generale, molto usato nella pratica, consiste nel sostituire la componente di disturbo con una sua stima consistente. Questa procedura è alla base della verosimiglianza profilo.

Si definisce verosimiglianza profilo per  $\psi$  la funzione

$$L_P(\psi) = L(\psi, \hat{\lambda}_\psi), \quad (1.18)$$

dove  $\hat{\lambda}_\psi$  è la s.m.v. di  $\lambda$  con  $\psi$  fissato. Si indica inoltre con  $l_P(\psi) = \log L_P(\psi)$  la funzione di log-verosimiglianza profilo.

Poiché  $L_P(\psi)$  non è dedotta da una funzione di densità, essa non è una verosimiglianza in senso proprio. La verosimiglianza profilo, tuttavia, possiede una serie di proprietà che l'assimilano ad una verosimiglianza propria. In particolare:

- 1) la s.m.v. profilo di  $\psi$  coincide con la s.m.v.  $\hat{\psi}$  basata su  $L(\theta)$ ;
- 2) nei problemi regolari,  $\hat{\psi}$  è la soluzione di  $\partial l_P(\psi)/\partial \psi = 0$ ;
- 3) l'informazione osservata profilo è data da

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

È possibile dimostrare che l'inversa dell'informazione osservata profilo è uguale al blocco  $(\psi, \psi)$  dell'inversa della matrice d'informazione osservata complessiva, valutato in  $(\psi, \hat{\lambda}_\psi)$ , ovvero

$$[j_P(\psi)]^{-1} = \left[ j(\psi, \hat{\lambda}_\psi)^{-1} \right]_{(\psi, \psi)} = j^{\psi\psi}(\psi, \hat{\lambda}_\psi);$$

4.  $L_P(\psi)$  è invariante rispetto a riparametrizzazione;
5. il log-rapporto di verosimiglianza profilo  $W_P(\psi) = 2\{l_P(\hat{\psi}) - l_P(\psi)\}$  coincide con  $W(\theta)$  calcolato utilizzando  $l(\psi, \lambda)$ , per verificare ipotesi su  $\psi$  con  $\lambda$  ignoto, e risulta data da

$$W_p(\psi) = 2\{l_p(\hat{\psi}) - l_p(\psi)\} = 2\{l(\hat{\psi}, \hat{\lambda}) - l(\psi, \hat{\lambda}_\psi)\}, \quad (1.19)$$

con distribuzione asintotica nulla  $\chi_k^2$ ;

6. una regione di confidenza approssimata per  $\psi$  di livello nominale  $(1 - \alpha)$  può essere ottenuta mediante il log-rapporto di verosimiglianza profilo, come

$$\hat{\Psi}(y) = \{\psi \in \Psi: W_p(\psi) \leq \chi_{k;1-\alpha}^2\}. \quad (1.20)$$

Il punto di debolezza della  $l_p(\psi)$  è che non è soddisfatta la prima identità di Bartlett, ossia

$$\mathbb{E}_\theta[\partial l_p(\psi)/\partial \psi] \neq 0.$$

Costruire test e regioni di confidenza con  $l_p(\psi)$  coincide a comportarsi come se  $\lambda$  fosse noto e pari a  $\hat{\lambda}_\psi$ . La  $l_p(\psi)$  non tiene in considerazione dell'incertezza del parametro di disturbo  $\lambda$ , poiché esso viene sostituito da una sua stima. Ciò può essere non appropriato se i dati sono carenti di informazioni su  $\lambda$ , cosa che avviene se la dimensione del parametro di disturbo è elevata rispetto alla numerosità campionaria. Per maggiori approfondimenti si veda Pace e Salvani (1996, § 4.6).

### Esempio 3.1

Siano  $Y_1, Y_2, \dots, Y_n$  variabili casuali i.i.d. normali con media  $\psi$  e varianza  $\lambda$ . La funzione di log-verosimiglianza per  $(\psi, \lambda)$  è data da

$$l(\psi, \lambda) \propto -\frac{n}{2} \log \lambda - \frac{1}{2\lambda} \sum_{i=1}^n (y_i - \psi)^2.$$

Il vettore punteggio ha componenti

$$l_\psi(\psi, \lambda) = \frac{n}{\lambda} (\bar{y} - \psi)$$

e

$$l_\lambda(\psi, \lambda) = -\frac{n}{2\lambda} + \frac{1}{2\lambda^2} \sum_{i=1}^n (y_i - \psi)^2.$$

Risolvendo rispetto a  $\lambda$  l'equazione di verosimiglianza parziale  $l_\lambda(\psi, \lambda) = 0$ , si ottiene la s.m.v. per  $\lambda$  con  $\psi$  fissato, data da

$$\hat{\lambda}_\psi = \frac{1}{n} \sum_{i=1}^n (y_i - \psi)^2.$$

Quindi la log-verosimiglianza profilo per  $\psi$  è

$$l_P(\psi) = l(\psi, \hat{\lambda}_\psi) = -\frac{n}{2} \log \hat{\lambda}_\psi - \frac{1}{2\hat{\lambda}_\psi} \sum_{i=1}^n (y_i - \psi)^2 = -\frac{n}{2} \log \hat{\lambda}_\psi - \frac{n}{2}.$$

Per calcolare la log-verosimiglianza profilo per la varianza  $\lambda$ , invece, si può osservare che la s.m.v. per la media  $\psi$  con  $\lambda$  fissato coincide con la s.m.v.  $\hat{\psi}$  e risulta essere pari alla media campionaria  $\bar{y}$ . Allora

$$l_P(\lambda) = l(\hat{\psi}_\lambda, \lambda) = -\frac{n}{2} \log \lambda - \frac{1}{2\lambda} \sum_{i=1}^n (y_i - \bar{y})^2 = -\frac{n}{2} \log \lambda - \frac{n\hat{\lambda}}{2\lambda}.$$

Il valore di  $\lambda$  che massimizza  $l_P(\lambda)$  coincide con la s.m.v., ovvero la varianza campionaria

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2.$$

■

#### 1.4.4 Verosimiglianza condizionata approssimata

Per compensare i limiti della verosimiglianza profilo, in particolare la distorsione della funzione punteggio profilo, sono state pubblicate varie versioni modificate della verosimiglianza profilo. I due esempi più importanti sono: la **verosimiglianza condizionata approssimata** (Cox e Reid, 1987) e la **verosimiglianza profilo modificata** (cfr. § 1.3.5).

La prima modificazione della funzione di verosimiglianza profilo può essere realizzata sfruttando le proprietà derivanti da una parametrizzazione ortogonale.

Sia dato un m.s.p.  $\mathcal{F}$ , con  $\theta = (\psi, \lambda)$ , e per semplicità si assuma  $p = 2$  e  $k = 1$ . I parametri  $\psi$  e  $\lambda$  sono detti **ortogonali** se (Pace e Salvan, 1996, § 4.7)

$$i_{\lambda\psi} = \mathbb{E}_\theta \left[ -\frac{\partial^2 l(\psi, \lambda)}{\partial \lambda \partial \psi} \right] = 0. \quad (1.21)$$

Tale definizione può essere immediatamente estesa al caso in cui  $p > 2$  e  $k = 1$ , ma una definizione di ortogonalità globale per  $k > 1$  non è possibile da ottenere in generale.

Se le due componenti di  $\theta$  sono ortogonali, i parametri godono di due importanti proprietà:

- 1)  $\hat{\psi}$  e  $\hat{\lambda}$  sono asintoticamente indipendenti;
- 2) la varianza di  $\hat{\psi}$  può essere valutata come se il parametro di disturbo fosse noto.

Infatti,

$$i^{\psi\psi} = [i(\psi, \lambda)^{-1}]_{\psi\psi} = [i_{\psi\psi}(\psi, \lambda)]^{-1}$$

e, in modo equivalente vale per la matrice di varianza-covarianza di  $\lambda$ ,

$$i^{\lambda\lambda} = [i(\psi, \lambda)^{-1}]_{\lambda\lambda} = [i_{\lambda\lambda}(\psi, \lambda)]^{-1}.$$

Si può allora pensare di basare l'inferenza su  $\psi$  sul modello condizionato a  $\hat{\lambda}$ , che costituisce una pseudo-verosimiglianza più accurata rispetto alla profilo. In particolare, si dimostra che (si veda Pace e Salvan, 1996, § 4.7) la log-verosimiglianza condizionata a  $\hat{\lambda}$  è approssimativamente pari a

$$\log p_{Y|\hat{\lambda}}(y|\hat{\lambda}; \psi, \lambda) \doteq l(\psi, \lambda) - \frac{1}{2} \log |j_{\lambda\lambda}(\psi, \lambda)| + \frac{1}{2} (\hat{\lambda} - \lambda)^T j_{\lambda\lambda}(\psi, \lambda) (\hat{\lambda} - \lambda).$$

È evidente che tale funzione dipende ancora da  $\lambda$ . Per ovviare a questo inconveniente risulta naturale sostituire il valore di  $\lambda$  con la s.m.v. del modello con  $\psi$  fissato, ovvero  $\hat{\lambda}_\psi$ . Così facendo il contributo dell'ultimo addendo di destra diventa asintoticamente trascurabile, pervenendo così alla log-verosimiglianza condizionata approssimata (Cox e Reid, 1987)

$$l_{CA}(\psi) = l_P(\psi) - \frac{1}{2} \log |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|. \quad (1.22)$$

Questa funzione di log-verosimiglianza risulta essere una modifica della profilo, dove si è aggiunto un termine di penalità, per tenere conto del comportamento del parametro di disturbo al variare di  $\psi$ .

La log-verosimiglianza condizionata approssimata presenta però l'inconveniente di richiedere una parametrizzazione ortogonale. Questo implica che, da un lato, il parametro d'interesse  $\psi$  deve essere scalare e che, dall'altro, la pseudo-verosimiglianza per determinati modelli non risulta semplice da derivare.

Inoltre, la  $l_{CA}(\psi)$  non gode della proprietà di invarianza rispetto a riparametrizzazioni che non alterano l'interesse (si veda Pace e Salvan, 1996, § 4.7).

**Esempio 2.1 (continuazione)**

Si consideri un campione casuale semplice proveniente dalla distribuzione  $N(\lambda, \psi)$ , dove la varianza  $\psi$  gioca il ruolo di parametro d'interesse, mentre la media  $\lambda$  è il parametro di disturbo. La log-verosimiglianza profilo risulta proporzionale a

$$l_p(\psi) = -\frac{n}{2} \log \psi - \frac{n\hat{\psi}}{2\psi}, \quad (1.23)$$

dove la s.m.v.  $\hat{\psi}$  è data dalla varianza campionaria  $\hat{\sigma}^2$ .

Poiché l'informazione attesa di Fisher è diagonale, questa parametrizzazione del modello soddisfa le condizioni di ortogonalità. Sotto queste ipotesi, la log-verosimiglianza condizionata approssimata è pari a

$$l_{CA}(\psi) = l_p(\psi) - \frac{1}{2} \log |j_{\lambda\lambda}(\hat{\lambda}_\psi, \psi)| = -\frac{n-1}{2} \log \psi - \frac{n\hat{\psi}}{2\psi}, \quad (1.24)$$

dove  $j_{\lambda\lambda}(\hat{\lambda}, \psi) = \frac{n}{\psi}$  e  $\hat{\lambda} = \hat{\lambda}_\psi = \frac{1}{n} \sum_i y_i = \bar{y}$ .

La (1.24) restituisce una stima di  $\psi$  non distorta, contrariamente a quanto avviene per la log-verosimiglianza profilo (1.23).

Il risultato espresso nella (1.24) può essere facilmente esteso al caso  $p > 1$  e  $k = 1$ . Si consideri il caso di un campione  $y = (y_1, \dots, y_n)$  di osservazioni indipendenti di variabili casuali normali  $N(x_i^T \beta, \psi)$ , dove  $\beta$  è un vettore  $h$ -dimensionale di coefficienti di regressione e  $x_i$  è un vettore di costanti note,  $i = 1, \dots, n$ . Sia  $\psi$  il parametro d'interesse. Sia  $X$  la matrice  $(n \times h)$  delle variabili esplicative, avente come riga  $i$ -esima  $x_i$ ,  $i = 1, \dots, n$ . Le quantità di verosimiglianza risultano

$$\begin{aligned} l(\beta, \psi) &= -\frac{n}{2} \log \psi - \frac{1}{2\psi} (y - X\beta)^T (y - X\beta), \\ l_\beta &= \frac{\partial l(\beta, \psi)}{\partial \beta} = -\frac{1}{\psi} (X^T X \beta - X^T y), \\ l_\lambda &= \frac{\partial l(\beta, \psi)}{\partial \psi} = -\frac{n}{2\psi} + \frac{1}{2\psi^2} (y - X\beta)^T (y - X\beta). \end{aligned}$$

L'informazione attesa è diagonale, ossia

$$i(\beta, \psi) = \begin{bmatrix} \frac{(X^T X)}{\psi} & 0^T \\ 0 & \frac{n}{2\psi^2} \end{bmatrix}.$$

La log-verosimiglianza condizionata approssimata per  $\psi$  risulta

$$l_{CA}(\psi) = -\frac{n-h}{2} \log \psi - \frac{1}{2\psi} \sum_{i=1}^n (y_i - x_i^T \hat{\beta})^2,$$

dove  $\hat{\beta}_\psi = \hat{\beta} = (X^T X)^{-1} (X^T y)$ . Dalla  $l_{CA}(\psi)$  si ottiene la stima della varianza non distorta, data da

$$s^2 = \frac{1}{n-h} \sum_{i=1}^n (y_i - x_i^T \hat{\beta})^2.$$

■

#### 1.4.5 Verosimiglianza profilo modificata

Per superare il problema della parametrizzazione ortogonale, necessaria per ricavare una verosimiglianza condizionata approssimata, esiste una funzione di pseudo-verosimiglianza che può essere calcolata in situazioni più generali (si vedano Pace e Salvan, 1996, cap. 11 e Severini, 2000, cap. 9).

La verosimiglianza e log-verosimiglianza profilo modificata sono, rispettivamente, definite come (Barndoff e Nielsen, 1983)

$$L_{MP}(\psi) = L_P(\psi) M(\psi) = L_P(\psi) |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{-\frac{1}{2}} \left| \frac{\partial \hat{\lambda}_\psi}{\partial \hat{\lambda}} \right|^{-1} \quad (1.25)$$

e

$$l_{MP}(\psi) = l_P(\psi) - \frac{1}{2} \log |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| - \log \left| \frac{\partial \hat{\lambda}_\psi}{\partial \hat{\lambda}} \right|, \quad (1.26)$$

dove  $M(\psi) = |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{-\frac{1}{2}} |\partial \hat{\lambda}_\psi / \partial \hat{\lambda}|^{-1}$  rappresenta un opportuno fattore di aggiustamento. Tale termine di aggiustamento può anche essere riscritto come

$$M(\psi) = \frac{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{\frac{1}{2}}}{|l_{\lambda, \hat{\lambda}}(\psi, \hat{\lambda}_\psi; \hat{\psi}, \hat{\lambda}, a)|}, \quad (1.27)$$

dove  $l_{\lambda, \hat{\lambda}}(\cdot)$  rappresenta la derivata nello spazio campionario

$$l_{\lambda, \hat{\lambda}}(\psi, \hat{\lambda}_\psi; \hat{\psi}, \hat{\lambda}, a) = \frac{\partial l(\psi, \hat{\lambda}_\psi; \hat{\psi}, \hat{\lambda}, a)}{\partial \lambda \partial \hat{\lambda}^T},$$

con  $a$  statistica ancillare.

Per calcolare la log-verosimiglianza profilo modificata (1.26), è necessario conoscere la s.m.v.  $(\hat{\psi}, \hat{\lambda})$ , la stima vincolata  $\hat{\lambda}_\psi$ , il blocco relativo alla componente di disturbo della

matrice dell'informazione osservata  $j_{\lambda\lambda}(\cdot)$ , ed infine la derivata della log-verosimiglianza rispetto a  $\lambda$  ed a  $\hat{\lambda}$ , ovvero  $l_{\lambda, \hat{\lambda}}(\cdot)$ . Reperire l'ultima quantità risulta essere il principale limite della  $L_{MP}(\psi)$ , poiché può essere facilmente ricavabile solo per alcuni modelli statistici parametrici (ad esempio le famiglie esponenziali o i modelli di gruppo); per ovviare a questo inconveniente sono state introdotte diverse approssimazioni per tale quantità (si veda Severini, 2000, cap. 9).

La funzione di verosimiglianza profilo modificata è ottenibile come approssimazione di una verosimiglianza condizionata o la verosimiglianza marginale per  $\psi$ , quando tali pseudo-verosimiglianze esistono; per maggiori approfondimenti si vedano Severini (2000, cap. 9) e Pace e Salvan (2006).

Si presentano in seguito due casi in cui il calcolo della verosimiglianza profilo modificata risulta agevole.

Si consideri il caso in cui le osservazioni possono essere descritte da una legge appartenente alla famiglia esponenziale naturale multiparametrica (per semplicità si assume  $p = 2$  e  $k = 1$ ), con la coppia  $(t, s)$  di statistiche sufficienti minimali, con distribuzione

$$p_{T,S}(t, s; \psi, \lambda) = h(t, s) \exp\{\psi t + \lambda s - K(\psi, \lambda)\}.$$

La log-verosimiglianza risulta

$$l(\psi, \lambda) = \psi t + \lambda s - K(\psi, \lambda).$$

Le quantità di verosimiglianza sono definite da

$$\begin{cases} l_{\psi}(\psi, \lambda) = t - K_{\psi}(\psi, \lambda) \\ l_{\lambda}(\psi, \lambda) = s - K_{\lambda}(\psi, \lambda) \end{cases}'$$

$$\text{con } K_{\psi}(\psi, \lambda) = \frac{\partial K(\psi, \lambda)}{\partial \psi} \text{ e } K_{\lambda}(\psi, \lambda) = \frac{\partial K(\psi, \lambda)}{\partial \lambda},$$

$$i(\psi, \lambda) = j(\psi, \lambda) = \begin{bmatrix} K_{\psi\psi}(\psi, \lambda) & K_{\psi\lambda}(\psi, \lambda) \\ K_{\lambda\psi}(\psi, \lambda) & K_{\lambda\lambda}(\psi, \lambda) \end{bmatrix},$$

$$\text{con } K_{\psi\psi}(\psi, \lambda) = \frac{\partial^2 K(\psi, \lambda)}{\partial \psi^2}, K_{\psi\lambda}(\psi, \lambda) = \frac{\partial^2 K(\psi, \lambda)}{\partial \psi \partial \lambda} \text{ e } K_{\lambda\lambda}(\psi, \lambda) = \frac{\partial^2 K(\psi, \lambda)}{\partial \lambda^2}.$$

La s.m.v.  $(\hat{\psi}, \hat{\lambda})$  è la soluzione di  $l_*(\psi, \lambda) = 0$ , e soddisfa  $(t, s) = (K_{\psi}(\hat{\psi}, \hat{\lambda}), K_{\lambda}(\hat{\psi}, \hat{\lambda}))$ .

La log-verosimiglianza per  $(\psi, \lambda)$  e la funzione punteggio per  $\lambda$  possono essere riscritte come

$$l(\psi, \lambda; \hat{\psi}, \hat{\lambda}) = \psi K_{\psi}(\hat{\psi}, \hat{\lambda}) + \lambda K_{\lambda}(\hat{\psi}, \hat{\lambda}) - K(\psi, \lambda)$$

e

$$l_{\lambda}(\psi, \lambda; \hat{\psi}, \hat{\lambda}) = K_{\lambda}(\hat{\psi}, \hat{\lambda}) - K_{\lambda}(\psi, \lambda).$$

La derivata nello spazio campionario  $l_{\lambda; \hat{\lambda}}$  risulta

$$l_{\lambda; \hat{\lambda}}(\psi, \lambda; \hat{\psi}, \hat{\lambda}) = K_{\lambda\lambda}(\hat{\psi}, \hat{\lambda}) = j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda}).$$

Tale quantità è funzione solo dei dati, e quindi può essere trascurata nel calcolo della log-verosimiglianza profilo modificata. La  $l_{MP}(\psi)$  risulta essere data da

$$l_{MP}(\psi) = l_P(\psi) + \frac{1}{2} \log |K_{\lambda\lambda}(\psi, \hat{\lambda}_{\psi})|.$$

Un risultato analogo si ha quando la stima del parametro di disturbo, tenendo fissato il parametro d'interesse, coincide con la s.m.v. per tale componente, ossia quando  $\hat{\lambda}_{\psi} = \hat{\lambda}$ .

In questo caso,  $|\partial \hat{\lambda}_{\psi} / \partial \hat{\lambda}|$  coincide con la matrice identità e pertanto

$$l_{MP}(\psi) = l_P(\psi) - \frac{1}{2} \log |j_{\lambda\lambda}(\psi, \hat{\lambda}_{\psi})| = l_{CA}(\psi).$$

In generale, se  $\psi$  e  $\lambda$  sono ortogonali, ma non si verifica  $\hat{\lambda}_{\psi} = \hat{\lambda}$ , la log-verosimiglianza profilo modificata coincide con la log-verosimiglianza condizionata approssimata a meno di un errore di ordine  $O_p(n^{-1})$ , poiché  $\hat{\lambda}_{\psi} = \hat{\lambda} + O_p(n^{-1})$ . Per approfondimenti si vedano Pace e Salvan (1996, § 4.7 e 11.6) e Severini (2000, § 9.5).

La differenza tra  $l_{MP}(\psi)$  e  $l_{CA}(\psi)$  sta nella facilità di calcolo. Per la log-verosimiglianza condizionata approssimata non si richiede la specificazione della statistica ancillare  $a$ . Nonostante ciò, necessita tuttavia di una parametrizzazione ortogonale, non sempre agevole da trovare, e non è invariante alla parametrizzazione, proprietà invece della log-verosimiglianza profilo modificata. Tuttavia in letteratura esistono numerose versioni del termine d'aggiustamento  $M(\psi)$  che non richiedono la specificazione della statistica ancillare; si veda, ad esempio, Severini (2000, cap. 9).

#### **Esempio 4.1**

Siano  $(x_1, \dots, x_n)$  e  $(y_1, \dots, y_n)$  due campioni casuali indipendenti da due distribuzioni normali con medie  $\mu_x$  e  $\mu_y$ , rispettivamente, e varianza unitaria. Si assuma  $\mu_x > 0$  e  $\mu_y > 0$ . Il parametro d'interesse è  $\psi = \mu_x / \mu_y$ , e sia  $\lambda = \sqrt{\mu_x^2 + \mu_y^2}$ , così che  $\psi$  e  $\lambda$  sono ortogonali (Severini, 2011, Ventura e Racugno, 2012).

La funzione di log-verosimiglianza per  $(\psi, \lambda)$  è

$$l(\psi, \lambda) = -\frac{n}{2} \left[ \left( \bar{x} - \frac{\lambda\psi}{\sqrt{1+\psi^2}} \right)^2 + \left( \bar{y} - \frac{\lambda}{\sqrt{1+\psi^2}} \right)^2 \right],$$

dove  $\bar{x}$  e  $\bar{y}$  sono le medie campionarie.

La funzione punteggio per  $\lambda$  e l'informazione osservata per  $(\lambda, \lambda)$  risultano, rispettivamente,

$$l_\lambda(\psi, \lambda) = \frac{n}{\sqrt{1+\psi^2}} \left[ \psi \left( \bar{x} - \frac{\lambda\psi}{\sqrt{1+\psi^2}} \right) + \left( \bar{y} - \frac{\lambda}{\sqrt{1+\psi^2}} \right) \right]$$

e

$$j_{\lambda\lambda}(\psi, \lambda) = n.$$

Per l'ortogonalità di  $\psi$  e  $\lambda$ , la log-verosimiglianza profilo modificata coincide con la log-verosimiglianza condizionata approssimata, e inoltre vale

$$l_{MP}(\psi) = l_P(\psi);$$

ossia, la log-verosimiglianza profilo modificata, in questo caso, è proporzionale alla log-verosimiglianza profilo, che risulta

$$l_P(\psi) = \frac{n}{2} \left[ \frac{(\psi\bar{x} + \bar{y})^2}{1 + \psi^2} \right] = \frac{n}{2} (\hat{\lambda}_\psi)^2$$

dove  $\hat{\lambda}_\psi = \frac{\psi\bar{x} + \bar{y}}{\sqrt{1 + \psi^2}}$ .

■

#### 1.4.6 Verosimiglianza integrata

L'ultima pseudo-verosimiglianza presentata in questo capitolo si basa sull'eliminazione della componente di disturbo tramite integrazione, invece che massimizzazione.

La verosimiglianza integrata trova naturale applicazione nella letteratura bayesiana. Infatti, lo strumento dell'integrazione si caratterizza come l'unico metodo di eliminazione del parametro di disturbo nell'ambito dell'inferenza bayesiana classica. L'utilizzo di verosimiglianze alternative sarà illustrato nel Capitolo 2.

Le tecniche bayesiane necessitano di un'opportuna distribuzione a priori per il parametro. Quando  $\theta = (\psi, \lambda)$ , la distribuzione a priori assume la forma

$$\pi(\psi, \lambda) = \pi(\lambda|\psi)\pi(\psi), \tag{1.28}$$

dove  $\pi(\psi)$  è la distribuzione a priori marginale per  $\psi$  e  $\pi(\lambda|\psi)$  è la distribuzione a priori condizionata di  $\lambda$  dato  $\psi$ .

La verosimiglianza integrata è ottenuta integrando la verosimiglianza completa  $L(\psi, \lambda)$  rispetto al parametro di disturbo, attraverso  $\pi(\lambda|\psi)$ , che funge da peso ai diversi valori che il parametro di disturbo può assumere. Essa è definita come (Severini 2000, cap. 8)

$$L_I(\psi) = \int_A L(\psi, \lambda) \pi(\lambda|\psi) d\lambda. \quad (1.29)$$

In pratica, il parametro di disturbo  $\lambda$  viene eliminato per integrazione dalla verosimiglianza per  $\theta$ , pesato con una opportuna distribuzione a priori per  $\lambda$  condizionata al valore di  $\psi$ . Analogamente si definisce la log-verosimiglianza integrata come

$$l_I(\psi) = \log L_I(\psi). \quad (1.30)$$

Al contrario delle tecniche di condizionamento e di marginalizzazione, che possono essere applicate solo in determinate circostanze, la funzione di verosimiglianza integrata è sempre calcolabile. La distribuzione a priori  $\pi(\lambda|\psi)$  consente di tenere conto dell'incertezza sul parametro  $\lambda$  e il risultato ottenuto può essere interpretato come una media di tutte le verosimiglianze condizionate dato il parametro di disturbo. Inoltre la verosimiglianza integrata è una verosimiglianza propria. Una rassegna completa su  $L_I(\psi)$  è data in Berger *et al.* (1999); si veda anche Severini (2000, cap. 8).

Un nodo molto importante, per la costruzione di  $L_I(\psi)$ , è legato alla scelta della distribuzione a priori. La forma della funzione di verosimiglianza integrata, infatti, dipende da  $\pi(\lambda|\psi)$ . Questo aspetto risulta essere il punto cruciale di questa metodologia. Una scelta ottimale non esiste, poiché non sussiste una tecnica che possa individuare l'alternativa più idonea per i diversi casi.

Una soluzione possibile quando le informazioni in mano al ricercatore sono scarse, oppure si vuole applicare al problema una tecnica che sia la più oggettiva possibile, consiste nel ricorrere alle distribuzioni a priori non informative.

La versione più semplice della verosimiglianza integrata è quella che fa uso di una distribuzione a priori uniforme per il parametro di disturbo, chiamata per questo integrata uniforme. Essa è data da

$$L_U(\psi) = \int_{\Lambda} L(\psi, \lambda) d\lambda, \quad (1.31)$$

in quanto  $\pi(\lambda|\psi) \propto 1$ .

Utilizzare una distribuzione a priori uniforme ha l'inconveniente di dipendere dalla parametrizzazione utilizzata. La verosimiglianza uniforme, infatti, non è invariante rispetto a riparametrazioni del parametro d'interesse. Per la scelta di altre distribuzioni a priori si veda Berger *et al.* (1999).

Un secondo problema derivante dall'uso della  $L_I(\psi)$  può essere il calcolo dell'integrale, in particolare quando la dimensione di  $\lambda$  è elevata. Per sopperire a questo inconveniente si può utilizzare un'utile approssimazione asintotica fornita dalla formula di Laplace; si vedano, ad esempio, Pace e Salvan (1996, cap. 9), Tierney e Kadane (1986) e Tierney *et al.* (1989).

Si fornisce nel seguito breve cenno di tale approssimazione per il caso univariato. Per la dimostrazione e l'estensione multivariata si veda Pace e Salvan (1996, § 9.3.3).

Si consideri un integrale del tipo

$$I = \int_{\mathbb{R}} h(y) \exp\{-ng(y)\} dy,$$

dove  $g(\cdot)$  è una funzione reale regolare dotata di minimo assoluto in  $\tilde{y}$ , ossia tale che  $g'(\tilde{y}) = 0$  e  $g''(\tilde{y}) > 0$ , e dove  $h(\cdot)$  è una funzione di ordine  $O(1)$ , tale che  $h(\tilde{y}) \neq 0$ . Sotto tali assunzioni, il comportamento asintotico di  $I$  è interamente determinato dal comportamento locale di  $g(\cdot)$  in  $\tilde{y}$ . Vale, infatti, lo sviluppo

$$I = \frac{\exp\{-n\tilde{g}\} \sqrt{2\pi}}{\sqrt{n\tilde{g}''}} \left\{ \tilde{h} + \frac{1}{2n} \left[ \frac{\tilde{h}''}{\tilde{g}''} - \frac{\tilde{h}'\tilde{g}'''}{(\tilde{g}'')^2} - \frac{1}{4} \frac{\tilde{h}\tilde{g}''''}{(\tilde{g}'')^2} + \frac{5}{12} \frac{\tilde{h}(\tilde{g}''')^2}{(\tilde{g}'')^3} \right] + O(n^{-2}) \right\},$$

dove  $g^r(y) = \partial^r g(y)/\partial y^r$ ,  $\tilde{g}^r = g^r(\tilde{y})$ ,  $h^r(y) = \partial^r h(y)/\partial y^r$  e  $\tilde{h}^r = h^r(\tilde{y})$  con  $r = 1, \dots, 4$ .

Questo risultato può tornare utile nell'approssimare l'integrale della verosimiglianza integrata. Infatti, si consideri un m.s.p. con funzione di log-verosimiglianza  $l(\psi, \lambda)$ , dove  $\psi$  è il parametro d'interesse e  $\lambda$  quello di disturbo. Si assuma che, per  $\psi$  fissato,  $l(\psi, \lambda)$  ammetta un unico massimo in corrispondenza di  $\hat{\lambda}_\psi$ . Si può allora dimostrare che la verosimiglianza integrata è approssimabile come (Tierney e Kadane, 1986)

$$\begin{aligned}
L_I(\psi) &= \int_{\Lambda} L(\psi, \lambda) \pi(\lambda|\psi) d\lambda = \int_{\Lambda} \exp\{l(\psi, \lambda)\} \pi(\lambda|\psi) d\lambda = \\
&= \frac{L_P(\psi) \pi(\hat{\lambda}_\psi|\psi)}{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2}} \{1 + O(n^{-1})\}, \tag{1.32}
\end{aligned}$$

e la log-verosimiglianza integrata come

$$l_I(\psi) = l_P(\psi) - \frac{1}{2} \log |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| + \log \pi(\hat{\lambda}_\psi|\psi) + O(n^{-1}). \tag{1.33}$$

Si ipotizzi ora che la densità a priori sia della forma  $\pi(\psi, \lambda) \propto \pi(\psi)\pi(\lambda)$  o, equivalentemente, che  $\psi$  e  $\lambda$  siano indipendenti (a priori), con  $\psi$  avente distribuzione marginale impropria uniforme e  $\lambda$  con distribuzione  $\pi(\lambda)$ . Allora la (1.33) diventa

$$l_I(\psi) = l_P(\psi) - \frac{1}{2} \log |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| + \log \pi(\hat{\lambda}_\psi) + O(n^{-1}). \tag{1.34}$$

Nel caso di  $\psi$  e  $\lambda$  ortogonali, si ha che  $\pi(\hat{\lambda}_\psi) = \pi(\hat{\lambda}) + O_p(n^{-1})$  e, pertanto, nell'approssimazione della log-verosimiglianza integrata il contributo della distribuzione a priori diventa trascurabile e risulta

$$l_I(\psi) = l_P(\psi) - \frac{1}{2} \log |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| + O(n^{-1}) = l_{CA}(\psi) + O(n^{-1}). \tag{1.35}$$

Sotto queste ipotesi, l'inferenza su  $\psi$  basata su  $l_{CA}(\psi)$  risulta pertanto un'approssimazione dell'inferenza basata su  $l_I(\psi)$  nell'ambito del paradigma bayesiano. Si noti che il risultato vale per qualunque distribuzione a priori  $\pi(\lambda)$  scelta.

### **Esempio 2.1 (cont.)**

Siano  $Y_1, Y_2, \dots, Y_n$  variabili casuali i.i.d. normali con media  $\mu$  e varianza  $\sigma^2$ . Il parametro d'interesse è  $\sigma^2$  e la media  $\mu$  della distribuzione risulta essere il parametro di disturbo, ossia  $\psi = \sigma^2$  e  $\lambda = \mu$ .

Si consideri  $L_U(\psi)$  per eliminare la componente di disturbo. Essendo in questo caso lo spazio del parametro di disturbo non limitato, la distribuzione uniforme risulta essere impropria. Si ottiene

$$L_U(\sigma^2) = \int_{-\infty}^{+\infty} L(\mu, \sigma^2) d\mu = \int_{-\infty}^{+\infty} (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right\} d\mu =$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \bar{y})^2\right\} \int_{-\infty}^{+\infty} \exp\left\{-\frac{n}{2\sigma^2} (\bar{y} - \mu)\right\} d\mu.$$

Considerando un cambio di variabile da  $\mu$  a  $t = (\bar{y} - \mu)/\sigma$ , si ottiene

$$\begin{aligned} L_U(\sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{n\hat{\sigma}^2}{2\sigma^2}\right\} \int_{-\infty}^{+\infty} \exp\left\{-\frac{n}{2}t^2\right\} \sigma dt \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \sigma \sqrt{\frac{2\pi}{n}} \exp\left\{-\frac{n\hat{\sigma}^2}{2\sigma^2}\right\} = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{n-1} \frac{1}{\sqrt{n}} \exp\left\{-\frac{n\hat{\sigma}^2}{2\sigma^2}\right\}. \end{aligned}$$

La funzione di log-verosimiglianza integrata risulta quindi

$$l_U(\sigma^2) = -\frac{(n-1)}{2} \log \sigma^2 - \frac{n\hat{\sigma}^2}{2\sigma^2}.$$

La funzione di verosimiglianza integrata calcolata con la distribuzione a priori uniforme per  $\mu$  risulta essere equivalente alla funzione di verosimiglianza profilo modificata. ■

Un problema che può sorgere con la verosimiglianza integrata è la possibilità che l'integrale, per una determinata distribuzione a priori su  $\lambda$ , non esista (Berger *et al.* 1999). Si prenda, ad esempio, un campione di numerosità  $n = 2$  da una variabile casuale  $N(\mu, \sigma^2)$ . In questo caso, posto  $\psi = \mu$  e  $\lambda = \sigma^2$ , si può dimostrare che utilizzando una distribuzione a priori uniforme per la varianza, l'integrale non è finito. Si ha

$$\begin{aligned} L_U(\mu) &= \int_0^{+\infty} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right\} d\sigma^2 = \\ &= \int_0^{+\infty} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{K}{2\sigma^2}\right\} d\sigma^2, \end{aligned}$$

con  $K = \sum_{i=1}^n (y_i - \mu)^2$ . Considerando un cambio di variabile da  $\sigma^2$  a  $\tau = K/(2\sigma^2)$ , risulta

$$\begin{aligned} L_U(\mu) &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int_0^{+\infty} \left(\frac{2\tau}{K}\right)^{\frac{n}{2}} \exp\{-\tau\} \left(\frac{K}{2\tau^2}\right) d\tau = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{2}{K}\right)^{\frac{n}{2}-1} \int_0^{+\infty} \tau^{\frac{n}{2}-2} \exp\{-\tau\} d\tau = \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{2}{K}\right)^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2} - 1\right). \end{aligned}$$

Quindi, la funzione di verosimiglianza integrata per la media è

$$L_U(\mu) = \left( \frac{1}{\sum_{i=1}^n (y_i - \mu)^2} \right)^{\frac{n}{2}-1} = \left( 1 + \frac{(\mu - \bar{y})}{\hat{\sigma}^2} \right)^{\frac{n}{2}-1}.$$

Naturalmente l'ultima espressione vale solo se  $\Gamma\left(\frac{n}{2} - 1\right) < \infty$ , e per  $n = 2$  l'integrale risulta essere pari ad infinito. Per ulteriori approfondimenti sulla scelta di altre distribuzioni a priori non informative quando  $\mu$  è il parametro d'interesse per una variabile casuale  $N(\mu, \sigma^2)$  si veda Severini (2000, § 8.4)

■

### 1.4.7 Alcuni esempi di confronto

Con riferimento alle diverse pseudo-verosimiglianze presentate in questo capitolo sono nel seguito proposti alcuni esempi che confrontano il comportamento dei differenti metodi di eliminazione dei parametri di disturbo.

#### Esempio 1.1 (cont.)

Siano  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  osservazioni i.i.d. da, rispettivamente, una distribuzione di Poisson di media  $\mu$  e da una distribuzione di Poisson di media  $\eta$ . Sia  $\psi = \log \frac{\mu}{\eta}$  il parametro d'interesse e sia  $\lambda = \log \eta$  il parametro di disturbo.

La funzione di log-verosimiglianza per  $(\psi, \lambda)$  è

$$l(\psi, \lambda) = \psi \sum_{i=1}^n x_i + \lambda \sum_{i=1}^n (x_i + y_i) - ne^\lambda (e^\psi + 1).$$

Data la coppia di statistiche sufficienti minimali  $(t, s) = (\sum x_i, \sum (x_i + y_i))$ , la distribuzione di  $T$  condizionata ad  $S$  non dipende dal parametro di disturbo  $\lambda$ . Infatti, vale

$$T|S = s \sim \text{Bin}(s, e^\psi / (e^\psi + 1)),$$

da cui si ottiene la log-verosimiglianza condizionata

$$l_c(\psi) \propto t\psi - s \log(e^\psi + 1).$$

Calcolando la derivata prima della funzione di  $l(\psi, \lambda)$  rispetto alla componente di disturbo, si ottiene

$$\frac{\partial l(\psi, \lambda)}{\partial \lambda} = s - ne^\lambda (e^\psi + 1),$$

da cui si ottiene

$$\hat{\lambda}_\psi = \log \left[ \frac{s}{n(e^\psi + 1)} \right].$$

La log-verosimiglianza profilo è perciò

$$\begin{aligned} l_p(\psi) &= t\psi + s\hat{\lambda}_\psi - ne^{\hat{\lambda}_\psi}(e^\psi + 1) = t\psi + s \log \left( \frac{s}{n(e^\psi + 1)} \right) = \\ &= t\psi - s \log(e^\psi + 1), \end{aligned}$$

che coincide con  $l_C(\psi)$ .

Per la verosimiglianza integrata con a priori uniforme su  $\lambda$ , si ottiene

$$\begin{aligned} L_U(\psi) &= \int_{-\infty}^{+\infty} \exp\{t\psi + s\lambda - ne^\lambda(e^\psi + 1)\} d\lambda = \\ &= e^{t\psi} \int_{-\infty}^{+\infty} \exp\{s\lambda - ne^\lambda(e^\psi + 1)\} d\lambda. \end{aligned}$$

Posto  $\tau = ne^\lambda(e^\psi + 1)$ , e considerando il cambio di variabile  $\lambda = \log[\tau/n(e^\psi + 1)]$ , si trova

$$\begin{aligned} L_U(\psi) &= \frac{e^{t\psi}}{(n(e^\psi + 1))^s} \int_0^{+\infty} \tau^{s-1} e^{-\tau} d\tau = \\ &= \frac{e^{t\psi} \Gamma(s)}{(n(e^\psi + 1))^s} \int_0^{+\infty} \frac{\tau^{s-1} e^{-\tau}}{\Gamma(s)} d\tau = \\ &= \frac{e^{t\psi} (s-1)!}{(n(e^\psi + 1))^s}. \end{aligned}$$

La funzione di log-verosimiglianza integrata uniforme è quindi

$$l_U(\psi) = t\psi - s \log(e^\psi + 1),$$

che coincide con  $l_C(\psi)$  e  $l_P(\psi)$ .

La s.m.v. è  $(\hat{\psi}, \hat{\lambda}) = (\log(t/(s-t)), \log((s-t)/n))$ . Poiché le s.m.v. sono funzione delle statistiche sufficienti minimali, la funzioni di log-verosimiglianza può essere riscritta come

$$l(\psi, \lambda; \hat{\psi}, \hat{\lambda}) = n\psi e^{\hat{\psi} + \hat{\lambda}} + n\lambda e^{\hat{\lambda}}(e^{\hat{\psi}} + 1) - ne^\lambda(e^\psi + 1).$$

Derivando la funzione di log-verosimiglianza rispetto a  $\lambda$  e a  $\hat{\lambda}$ , si ottiene

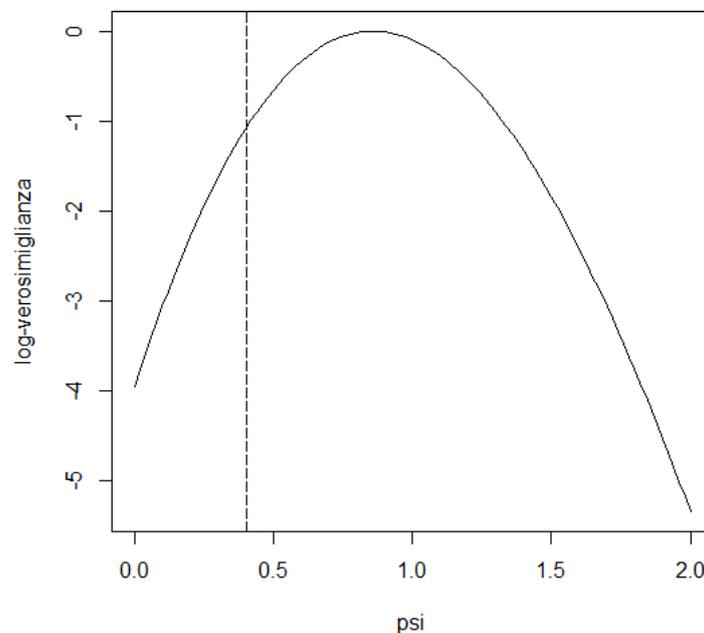
$$l_{\lambda; \hat{\lambda}}(\psi, \lambda; \hat{\psi}, \hat{\lambda}) = ne^{\hat{\lambda}}(e^{\hat{\psi}} + 1),$$

che dipende solo dai dati. Quindi tale quantità può essere trascurata nel calcolo della log-verosimiglianza profilo modificata. L'informazione osservata per  $\lambda$  calcolata in  $(\psi, \hat{\lambda}_\psi)$  è pari a

$$j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi) = ne^{\hat{\lambda}_\psi}(e^\psi + 1) = s.$$

Quindi la log-verosimiglianza profilo modifica è equivalente alla log-verosimiglianza profilo, ossia

$$l_{MP}(\psi) = l_P(\psi).$$



**Figura 1.1. Esempio 1.1: Andamento della log-verosimiglianza condizionata con  $\psi = 0.4$  e  $n = 10$ .**

La Figura 1.1 illustra l'andamento della funzione di log-verosimiglianza condizionata per  $\psi$ , per un campione di numerosità modesta ( $n = 10$ ).

■

### **Esempio 2.1 (cont.)**

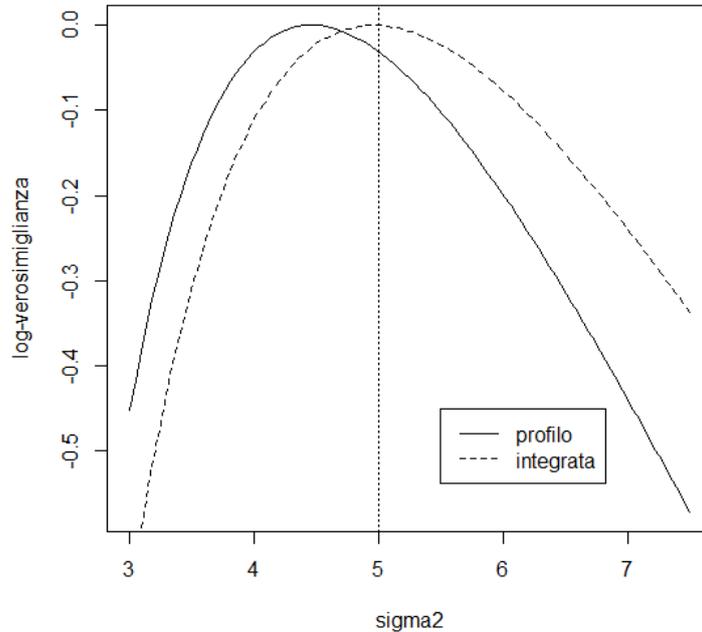
Riprendendo l'esempio di un campione i.i.d. da una  $N(\mu, \sigma^2)$ , con parametro d'interesse  $\sigma^2$ , le funzioni di log-verosimiglianza profilo, profilo modificata e integrata differiscono per il peso dato dalla numerosità campionaria. Infatti, si ha

$$l_P(\sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{n\hat{\sigma}^2}{2\sigma^2}$$

e

$$l_{MP}(\sigma^2) = l_U(\sigma^2) = -\frac{(n-1)}{2} \log \sigma^2 - \frac{n\hat{\sigma}^2}{2\sigma^2}.$$

Inoltre, poiché i parametri  $(\mu, \sigma^2)$  sono ortogonali, vale  $l_{MP}(\sigma^2) = l_{CA}(\sigma^2)$ .



**Figura 2.1 Esempio 3.1: Confronto delle diverse pseudo-verosimiglianze per  $\sigma^2$ , con  $\mu = 1$ ,  $\sigma^2 = 5$  e  $n = 10$ .**

La Figura 2.1 illustra l'andamento delle funzioni di log-verosimiglianza profilo e integrata normalizzate, per un campione casuale di  $n = 10$  osservazioni estratto da una  $N(1,5)$ . Le due curve evidenziano delle differenze rilevanti in termini di massimo. Infatti, se per la funzione di log-verosimiglianza profilo il massimo viene raggiunto con la varianza campionaria, per la log-verosimiglianza integrata il massimo coincide con la varianza campionaria corretta.

■

### Esempio 5.1

Siano  $X_1, X_2, \dots, X_n$  variabili casuali indipendenti con distribuzione esponenziale di media  $\mu$ , e siano  $Y_1, Y_2, \dots, Y_n$  variabili indipendenti con distribuzione esponenziale di media  $\eta$ , dove  $X_i$  è indipendente da  $Y_j$ , per ogni  $i, j = 1, \dots, n$ .

La funzione di verosimiglianza per  $(\mu, \eta)$  è

$$L(\mu, \eta) = \mu^{-n} \eta^{-n} \exp\left\{-\frac{n\bar{x}}{\mu} - \frac{n\bar{y}}{\eta}\right\} = L(\mu; \bar{x})L(\eta; \bar{y}).$$

La s.m.v.  $(\hat{\mu}, \hat{\eta})$  può essere trovata massimizzando singolarmente le due verosimiglianze, poiché i parametri del modello sono separabili tra loro. Si ha  $(\hat{\mu}, \hat{\eta}) = (\bar{x}, \bar{y})$ .

Sia  $\psi = \mu/\eta$  il parametro d'interesse e sia  $\lambda = \eta$  il parametro di disturbo. Le funzioni di verosimiglianza e log-verosimiglianza nella nuova parametrizzazione assumono, rispettivamente, le espressioni

$$L(\psi, \lambda) = \psi^{-n} \lambda^{-2n} \exp \left\{ -\frac{1}{\lambda} \sum_{i=1}^n \left( \frac{x_i}{\psi} + y_i \right) \right\}$$

e

$$\begin{aligned} l(\psi, \lambda) &= -n \log \psi - 2n \log \lambda - \frac{1}{\lambda} \sum_{i=1}^n \left( \frac{x_i}{\psi} + y_i \right) = \\ &= -\frac{1}{\lambda} \left( \frac{t}{\psi} + s \right) - n \log \psi - 2n \log \lambda, \end{aligned}$$

con  $(t, s) = (\sum_i x_i, \sum_i y_i)$  coppia di statistiche sufficienti minimali. La funzione punteggio per il parametro di disturbo è

$$l_\lambda(\psi, \lambda) = \frac{\partial l(\psi, \lambda)}{\partial \lambda} = -\frac{2n}{\lambda} + \frac{1}{\lambda^2} \left( \frac{t}{\psi} + s \right),$$

da cui si ottiene

$$\hat{\lambda}_\psi = \frac{1}{2n} \left( \frac{t}{\psi} + s \right).$$

La log-verosimiglianza profilo risulta

$$l_p(\psi) = -n \log \psi - 2n \log(\bar{x} + \psi \bar{y}).$$

Poiché  $(\hat{\psi}, \hat{\lambda}) = (t/s, s/n)$ , la funzione di log-verosimiglianza può essere riscritta come

$$l(\psi, \lambda; \hat{\psi}, \hat{\lambda}) = -\frac{n\hat{\psi}\hat{\lambda}}{\psi\lambda} - \frac{n\hat{\lambda}}{\lambda} - n \log \psi - 2n \log \lambda,$$

con

$$\hat{\lambda}_\psi = \frac{1}{2n} \left( s + \frac{t}{\psi} \right) = \frac{\hat{\lambda}}{2} \left( 1 + \frac{\hat{\psi}}{\psi} \right).$$

Per calcolare la log-verosimiglianza profilo modificata sono necessarie la derivata di  $\hat{\lambda}_\psi$  rispetto a  $\hat{\lambda}$  e l'elemento  $(\lambda, \lambda)$  della matrice di informazione osservata valutata in  $(\psi, \hat{\lambda}_\psi)$ , ossia  $j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi; \hat{\psi}, \hat{\lambda})$ . Si ha

$$\left| \frac{\partial \hat{\lambda}_\psi}{\partial \hat{\lambda}} \right| = \frac{1}{2} \left( 1 + \frac{\hat{\psi}}{\psi} \right)$$

e

$$j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi; \hat{\psi}, \hat{\lambda}) = \frac{2n}{\hat{\lambda}_\psi^2} \left[ \left( \frac{\hat{\psi}}{\psi} + 1 \right) \frac{\hat{\lambda}}{\hat{\lambda}_\psi} + 1 \right] = \frac{4n}{\hat{\lambda}_\psi^2}.$$

Si noti che  $\frac{\hat{\lambda}}{\hat{\lambda}_\psi} = \left( 1 + \frac{\hat{\psi}}{\psi} \right)^{-1}$ .

La log-verosimiglianza profilo modificata risulta quindi data da

$$\begin{aligned} l_{MP}(\psi) &= l_P(\psi) - \frac{1}{2} \log |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| - \log \left| \frac{\partial \hat{\lambda}_\psi}{\partial \hat{\lambda}} \right| = \\ &= l_P(\psi) + \log(\hat{\lambda}_\psi) - \log(\hat{\lambda}_\psi \hat{\lambda}) = l_P(\psi); \end{aligned}$$

ossia, la log-verosimiglianza profilo modificata coincide con la log-verosimiglianza profilo.

La verosimiglianza integrata con distribuzione a priori uniforme su  $\lambda$  risulta data da

$$L_U(\psi) = \int_0^{+\infty} L(\psi, \lambda) d\lambda = \int_0^{+\infty} \psi^{-n} \lambda^{-2n} \exp \left\{ -\frac{1}{\lambda} \left( \frac{t}{\psi} + s \right) \right\} d\lambda.$$

Posto  $\tau = \left( \frac{t}{\psi} + s \right) / \lambda$ , ed effettuando un cambio di variabile, si ottiene

$$\begin{aligned} L_U(\psi) &= \psi^{-n} \left( \frac{t}{\psi} + s \right)^{-2n+1} \int_0^{+\infty} \tau^{2n-2} \exp\{-\tau\} d\tau = \\ &= \psi^{-n} \left( \frac{t}{\psi} + s \right)^{-2n+1} \Gamma(2n-1) \end{aligned}$$

e quindi

$$l_U(\psi) = (n-1) \log \psi - (2n-1) \log(\bar{x} + \psi \bar{y}).$$

La log-verosimiglianza profilo e la log-verosimiglianza integrata differiscono per un fattore moltiplicativo, che dipende dalla numerosità campionaria.

Il punto che massimizza la log-verosimiglianza profilo si ha in corrispondenza della s.m.v. per  $\psi$ , data dal rapporto tra le medie dei due campioni, ossia  $\hat{\psi} = \bar{x} / \bar{y}$ . Per quanto riguarda la log-verosimiglianza integrata, il massimo viene raggiunto in un punto molto vicino alla stima precedente, ossia

$$\hat{\psi}_U = \left( \frac{n-1}{n} \right) \frac{\bar{x}}{\bar{y}} = \left( \frac{n-1}{n} \right) \hat{\psi}.$$

Si noti che per lo stimatore di massima verosimiglianza  $\hat{\psi} = \bar{X}/\bar{Y}$  vale

$$\hat{\psi} = \frac{\bar{X}}{\bar{Y}} \sim \psi \frac{\chi_{2n}^2/2n}{\chi_{2n}^2/2n} = \psi F_{2n,2n},$$

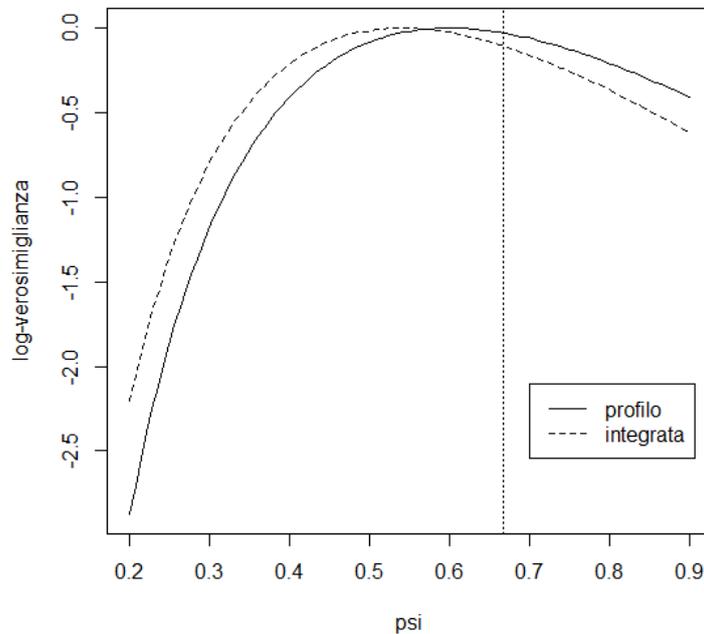
in quanto

$$\bar{X} = \frac{1}{n} \sum X_i \sim \frac{\lambda\psi}{2n} \chi_{2n}^2 \text{ e } \bar{Y} = \frac{1}{n} \sum Y_i \sim \frac{\lambda}{2n} \chi_{2n}^2.$$

Si può quindi ottenere la log-verosimiglianza marginale

$$l_M(\psi) = n \log \psi - 2n \log(\bar{x} + \psi\bar{y}),$$

che coincide con la log-verosimiglianza profilo e profilo modificata. La Figura 3.1 illustra l'andamento di  $l_P(\psi)$  e  $l_U(\psi)$ , per un campione di numerosità  $n = 10$ .



**Figura 3.1. Esempio 5.1: Confronto tra le diverse pseudo-verosimiglianze per  $\psi$ , con  $\lambda = 3$ ,  $\psi = 0.67$  e  $n = 10$ .**

■

Gli esempi riportati, hanno mostrato come non esista, in generale, una funzione di pseudo-verosimiglianza da preferire rispetto alle altre. La scelta di quale utilizzare, dipende dal parametro e dal tipo di paradigma che si desidera adottare.

Si è fatto notare, all'interno del paradigma fisheriano, che nelle situazioni dove esiste la possibilità di reperire una funzione di verosimiglianza marginale o condizionata per il

parametro d'interesse, essa è la soluzione da prediligere, in quanto è una verosimiglianza propria.

Una soluzione accettabile, tuttavia, è la funzione di verosimiglianza profilo. Tale pseudo-verosimiglianza ha il vantaggio di essere, nei problemi regolari, facilmente ricavabile, ma ha il difetto di non comportarsi come una verosimiglianza propria. Per rimediare a questo problema, le funzioni di verosimiglianza condizionata approssimata e profilo modificata sono valide alternative, in quanto sono versioni penalizzate della profilo, ma necessitano di condizioni di non sempre facile derivazione.

La soluzione al problema dell'eliminazione dei parametri di disturbo in ambito bayesiano, ossia la verosimiglianza integrata, produce una verosimiglianza propria, ma richiede sia di definire una distribuzione a priori per il parametro di disturbo e sia di risolvere un integrale, aspetti che in determinati casi possono non essere semplice, in particolare quando la dimensione di  $\lambda$  è elevata. Si può ricorrere, in generale, all'approssimazione di Laplace per il calcolo dell'integrale, che fornisce una versione penalizzata della funzione di verosimiglianza profilo.

Nel caso in cui, per il problema in esame, si riesca a ricavare una parametrizzazione ortogonale per il parametro d'interesse, la funzione di verosimiglianza condizionata approssimata è lo strumento più adatto per l'inferenza, in quanto è facilmente derivabile e può essere interpretato, per alcuni modelli statistici parametrici, come un'approssimazione delle altre pseudo-verosimiglianze, in particolare della verosimiglianza profilo modificata ed integrata (Pace e Salvan, 1996, § 9.3.3-10.10.2-11.6).

## **1.5 Conclusioni**

In questo capitolo è stata sinteticamente introdotta la teoria della verosimiglianza e dell'inferenza bayesiana, in particolare soffermandosi sul problema dell'eliminazione dei parametri di disturbo. L'obiettivo è quello di fornire gli strumenti che verranno poi utilizzati nel seguito della tesi.

Il Capitolo 2 si concentrerà inizialmente sul paradigma bayesiano, e cercherà, con tecniche più moderne, di presentare strumenti più sofisticati per la derivazione della distribuzione a posteriori per il solo parametro d'interesse. In particolare, ci si

concentrerà sulle sostituzioni della funzione di verosimiglianza integrata con un'altra pseudo-verosimiglianza illustrata nel presente capitolo.

Inoltre, nell'ambito frequentista verrà definita la funzione di verosimiglianza profilo modificata penalizzata con un'opportuna distribuzione a priori.

## Capitolo 2

### Metodi ibridi per l'eliminazione dei parametri di disturbo

Introdotta le nozioni di base dell'inferenza statistica in ambito fisheriano e bayesiano e presentate le tecniche classiche utilizzate per l'eliminazione dei parametri di disturbo, in questo capitolo viene illustrata una recente procedura, detta ibrida, per l'inferenza sui parametri d'interesse, che può essere utilizzata sia nell'approccio frequentista e sia in quello bayesiano.

Dopo una breve sintesi sulla metodologia classica nel paradigma bayesiano per l'eliminazione dei parametri di disturbo, nel Paragrafo 2.2 viene presentata una soluzione per il calcolo della funzione di verosimiglianza integrata e della distribuzione a posteriori marginale per il parametro d'interesse. Alcuni lavori di riferimento sono, ad esempio, Racugno *et al.* (2010) e Ventura *et al.* (2009, 2013).

Il Paragrafo 2.3 illustra una particolare classe di distribuzioni a priori non informative, la classe delle *matching priors*, che sono poi utilizzate nel Capitolo 3. Queste sono delle distribuzioni a priori per cui vi è un accordo tra inferenza frequentista e bayesiana. Viene poi presentata una versione penalizzata della funzione di verosimiglianza profilo modificata ottenuta mediante l'utilizzo delle *matching priors*, con le relative proprietà nel paradigma frequentista ed alcuni esempi illustrativi. Per una trattazione più approfondita sulle *matching priors* si vedano, ad esempio, Levine e Casella (2003), Datta e Mukerjee (2004) e Ventura *et al.* (2009, 2013).

Infine, nel Paragrafo 2.4 viene presentata un'approssimazione per la distribuzione a posteriori per il parametro d'interesse, in particolare illustrando un'approssimazione per la funzione di ripartizione (*tail area approximation*). Questo risultato viene utilizzato per derivare un metodo di simulazione, l'algoritmo di simulazione HOTA (Ruli *et al.*,

2013), che genera valori indipendenti dalla distribuzione marginale a posteriori. Per una più approfondita trattazione sui metodi asintotici si rimanda, ad esempio, a Reid (2003), Robert e Casella (2004), Brazzale *et al.* (2007), Ruli *et al.* (2013) e Ventura *et al.* (2013).

## **2.1 Distribuzione a posteriori marginale**

Come si è visto nel Capitolo 1, nell'approccio bayesiano lo strumento principe per l'inferenza sul parametro  $\theta$  è la distribuzione a posteriori, data da

$$\pi(\theta|y) \propto \pi(\theta)L(\theta; y),$$

dove  $\pi(\theta)$  è un'opportuna distribuzione a priori sul parametro  $\theta$  e  $L(\theta; y)$  è la funzione di verosimiglianza.

Quando  $\theta = (\psi, \lambda)$ , con  $\psi$  parametro d'interesse e  $\lambda$  parametro di disturbo, l'inferenza su  $\psi$  risulta più complicata e il problema dell'eliminazione del parametro di disturbo diventa centrale. Nell'inferenza bayesiana, la tecnica per l'eliminazione di  $\lambda$  richiede una distribuzione a priori della forma  $\pi(\psi, \lambda) = \pi(\psi)\pi(\lambda|\psi)$ , dove  $\pi(\psi)$  è la densità a priori marginale per  $\psi$  e  $\pi(\lambda|\psi)$  è la distribuzione a priori condizionata di  $\lambda$  dato  $\psi$ . Quindi si considera la distribuzione a posteriori marginale per  $\psi$ , data da

$$\begin{aligned} \pi(\psi|y) &\propto \int_{\Lambda} L(\psi, \lambda) \pi(\psi, \lambda) d\lambda \\ &\propto \pi(\psi)L_I(\psi), \end{aligned} \tag{2.1}$$

dove  $L_I(\psi)$  denota la funzione di verosimiglianza integrata per  $\psi$ , definita come (cfr. § 1.4.6)

$$L_I(\psi) = \int_{\Lambda} L(\psi, \lambda) \pi(\lambda|\psi) d\lambda.$$

Il calcolo della (2.1) può presentare due difficoltà, in particolare quando la dimensione del parametro di disturbo è elevata. Queste sono:

- (a) la necessità di elicitarre  $\pi(\psi, \lambda)$ , che può essere complicato sia nell'approccio bayesiano oggettivo che soggettivo;
- (b) il calcolo dell'integrale nella  $L_I(\psi)$ .

Un esempio in cui entrambi gli inconvenienti possono sorgere sono i modelli di regressione, quando il parametro d'interesse  $\psi$  è un coefficiente di regressione, e il parametro di disturbo  $\lambda$  è dato dai rimanenti parametri di regressione e da parametri di varianza.

## **2.2 Uso di pseudo-verosimiglianze alternative a $L_I(\psi)$**

Per evitare le difficoltà (a) e (b) è possibile ricorrere ad un approccio ibrido che prevede la sostituzione della funzione di verosimiglianza integrata  $L_I(\psi)$ , nella distribuzione a posteriori marginale  $\pi(\psi|y)$ , con una opportuna pseudo-verosimiglianza alternativa per l'eliminazione dei parametri di disturbo.

In questo modo viene definita una pseudo-distribuzione a posteriori (*pseudo-posterior*) per il parametro d'interesse  $\psi$ , data da (si veda, ad esempio, Racugno *et al.*, 2010)

$$\pi_{ps}(\psi|y) \propto \pi(\psi)L_{ps}(\psi), \quad (2.2)$$

dove  $L_{ps}(\psi)$  denota, ad esempio, una funzione di verosimiglianza condizionata, marginale, profilo, approssimata condizionata oppure profilo modificata (cfr. §1.4), e  $\pi(\psi)$  è una opportuna distribuzione a priori su  $\psi$  solamente.

Sebbene la (2.2) non possa essere considerata una distribuzione a posteriori ortodossa, esiste una vasta letteratura sull'uso della  $\pi_{ps}(\psi|y)$  in molti problemi d'interesse applicativo; si vedano, ad esempio, Fraser *et al.* (2003), Chang e Mukerjee (2006), Ventura *et al.* (2009, 2013), Racugno *et al.* (2010) e Cabras *et al.* (2012).

La (2.2) presenta i vantaggi di non dover né elicitarne la densità a priori  $\pi(\lambda|\psi)$  né calcolare l'integrale rispetto a  $\lambda$ . Infatti, per utilizzare questa procedura deve essere elicitata una distribuzione a priori  $\pi(\psi)$  per il solo parametro d'interesse  $\psi$ .

La scelta della distribuzione a priori su  $\psi$  rimane un tema cruciale. In letteratura sono state proposte numerose soluzioni come, ad esempio, le distribuzioni coniugate, le *matching priors*, le *reference priors*, le *conditional reference priors* e la distribuzione a priori non informativa di Jeffreys. Per una rassegna si vedano, ad esempio, Bernardo e Berger (1992), Berger *et al.* (1999), Datta e Mukerjee (2004) e Liseo (2008).

In questa tesi ci si concentra sulla classe delle *matching priors*, ossia una classe di distribuzioni a priori per cui vi è un accordo tra l'inferenza frequentista e quella

bayesiana. Sarà fatto anche un breve accenno alle *conditional reference priors* nel Capitolo 3.

### **2.3 Matching priors**

Si definisce *matching prior* una distribuzione a priori per la quale degli intervalli di credibilità a posteriori hanno copertura frequentista ad un certo ordine di approssimazione. In altre parole, la copertura degli intervalli costruiti mediante una distribuzione a posteriori che discende da una *matching prior* è pari, asintoticamente, al livello di confidenza nominale (Datta e Mukerjee, 2004). Più precisamente, le *matching priors* sono tali per cui (si veda, ad esempio, Staicu e Reid, 2008)

$$\mathbb{P}_\theta[\theta \leq \theta_\alpha] = \mathbb{P}_{\theta|Y}[\theta \leq \theta_\alpha] + O_p(n^{-1}), \quad (2.3)$$

dove  $\mathbb{P}_{\theta|Y}[\cdot]$  denota la probabilità a posteriori,  $\mathbb{P}_\theta[\cdot]$  rappresenta la probabilità sotto  $p(y; \theta)$ ,  $\theta_\alpha$  indica il quantile di livello  $\alpha$  della distribuzione a posteriori e  $O_p(n^{-1})$  è l'ordine di errore. Per approfondimenti si veda, ad esempio, Datta e Mukerjee (2004).

Per  $\theta$  scalare, la distribuzione *matching prior*, ossia la distribuzione a priori che soddisfa la (2.3), è data dalla distribuzione a priori di Jeffreys (cfr. §1.3), ossia

$$\pi(\theta) \propto i(\theta)^{1/2}.$$

In presenza di parametri di disturbo, bisogna distinguere il caso in cui sia presente una parametrizzazione ortogonale, da quello in cui non ci sia. Per approfondimenti si vedano, ad esempio, Tibshirani (1989) e Ventura *et al.* (2009).

#### **2.3.1 Parametrizzazione ortogonale**

In presenza di una parametrizzazione ortogonale, per l'eliminazione dei parametri di disturbo si può ricorrere alla funzione di verosimiglianza condizionata approssimata (cfr. §1.4.4), data da

$$L_{CA}(\psi) = L_P(\psi) |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{-\frac{1}{2}}.$$

Tibshirani (1989) e Nicolau (1993) mostrano che la classe delle distribuzioni a priori per  $(\psi, \lambda)$  che soddisfano la relazione (2.3), è data da

$$\pi(\psi, \lambda) \propto g(\lambda) i_{\psi\psi}(\psi, \lambda)^{1/2}, \quad (2.4)$$

dove  $g(\lambda)$  è un'arbitraria funzione positiva e  $i_{\psi\psi}(\psi, \lambda)$  è il blocco  $(\psi, \psi)$  dell'informazione di Fisher  $i(\psi, \lambda)$  basata su  $L(\psi, \lambda)$ .

Utilizzando la (2.4) e sfruttando l'approssimazione di Laplace (cfr. § 1.4.6) è possibile mostrare che la distribuzione a posteriori marginale è approssimabile come

$$\begin{aligned} \pi(\psi|y) &\propto \pi(\psi, \hat{\lambda}_\psi) L_{CA}(\psi) \\ &\propto g(\hat{\lambda}_\psi) i_{\psi\psi}(\psi, \hat{\lambda}_\psi)^{1/2} L_{CA}(\psi). \end{aligned}$$

In presenza di una parametrizzazione ortogonale si ha che  $\hat{\lambda}_\psi = \hat{\lambda} + O_p(n^{-1})$ , e quindi  $g(\hat{\lambda}_\psi) = g(\hat{\lambda}) + O_p(n^{-1})$ . Di conseguenza, la distribuzione *matching prior* per  $\psi$  associata a  $L_{CA}(\psi)$  risulta essere (Ventura *et al.*, 2009)

$$\pi_{mp}(\psi) \propto i_{\psi\psi}(\psi, \hat{\lambda}_\psi)^{1/2}. \quad (2.5)$$

La corrispondente pseudo-distribuzione a posteriori diviene

$$\pi_{ps}(\psi|y) \propto i_{\psi\psi}(\psi, \hat{\lambda}_\psi)^{1/2} L_{CA}(\psi). \quad (2.6)$$

Si può dimostrare che (Ventura *et al.*, 2009) tale distribuzione a posteriori è propria.

### 2.3.1 Parametrizzazione non ortogonale

La verosimiglianza condizionata approssimata è un valido strumento inferenziale per l'eliminazione dei parametri di disturbo, ma ha l'inconveniente di richiedere una parametrizzazione ortogonale, non sempre disponibile, e di non essere invariante rispetto a riparametrizzazioni che non alterano l'interesse.

Nel caso in cui si possa ricavare per  $\psi$  una funzione di verosimiglianza marginale o condizionata, la corrispondente *matching prior* è (Ventura *et al.*, 2009)

$$\pi_{mp}(\psi) \propto i_{ps}(\psi)^{1/2}, \quad (2.7)$$

dove  $i_{ps}(\psi)$  rappresenta il valore atteso della pseudo-informazione osservata, ossia dell'informazione basata sulla funzione di verosimiglianza marginale o condizionata del

modello ridotto. Si può, quindi, ottenere la corrispondente pseudo-distribuzione a posteriori

$$\pi_{ps}(\psi|y) \propto i_{ps}(\psi)^{1/2} L_{ps}(\psi), \quad (2.8)$$

dove  $L_{ps}(\psi)$  denota una funzione di verosimiglianza marginale o condizionata per  $\psi$ . Anche in questo caso, la (2.8) è una distribuzione a posteriori propria.

Le funzioni di verosimiglianza condizionata o marginale possono tuttavia essere derivate, in modo agevole, solo per alcuni m.s.p., in particolare per le famiglie esponenziali o di gruppo (cfr. § 1.4.1 e 1.4.2). Un'alternativa, quando non è possibile derivare un funzione di pseudo-verosimiglianza a partire da un modello ridotto basato solo sul parametro d'interesse, è utilizzare la funzione di verosimiglianza profilo modificata (cfr. § 1.4.5), data da

$$L_{MP}(\psi) = L_P(\psi)M(\psi),$$

dove  $M(\psi)$  è un termine d'aggiustamento opportuno (cfr. § 1.4.5 e Severini 2000, cap.8, Pace e Salvan 2006).

Si può dimostrare che (Ventura *et al.*, 2009) la distribuzione *matching prior* associata alla  $L_{MP}(\psi)$  è

$$\pi_{mp}(\psi) \propto i_{\psi\psi,\lambda}(\psi, \hat{\lambda}_\psi)^{1/2}, \quad (2.9)$$

dove  $i_{\psi\psi,\lambda}(\psi, \lambda) = i_{\psi\psi}(\psi, \lambda) - i_{\psi\lambda}(\psi, \lambda)i_{\lambda\lambda}(\psi, \lambda)^{-1}i_{\lambda\psi}(\psi, \lambda)$  è l'informazione parziale per  $\psi$ , con  $i_{\psi\lambda}(\psi, \lambda)$ ,  $i_{\lambda\lambda}(\psi, \lambda)$  e  $i_{\lambda\psi}(\psi, \lambda)$  blocchi della matrice  $i(\psi, \lambda)$ . La corrispondente distribuzione a posteriori è data da

$$\pi_{ps}(\psi|y) \propto i_{\psi\psi,\lambda}(\psi, \hat{\lambda}_\psi)^{1/2} L_{MP}(\psi), \quad (2.10)$$

che risulta una distribuzione a posteriori propria (Ventura *et al.*, 2009).

Si noti che nel caso in cui i parametri sono ortogonali vale  $i_{\psi\psi,\lambda}(\psi, \lambda) = i_{\psi\psi}(\psi, \lambda)$  e  $L_{MP}(\psi)$  coincide con  $L_{CA}(\psi)$ , riconducendosi alla (2.6).

#### **Esempio 4.1 (cont.)**

Siano  $(x_1, \dots, x_n)$  e  $(y_1, \dots, y_n)$  due campioni casuali indipendenti da due distribuzioni normali con medie  $\mu_x$  e  $\mu_y$ , rispettivamente, e varianza unitaria. Si assuma  $\mu_x > 0$  e

$\mu_y > 0$ . Il parametro d'interesse è  $\psi = \mu_x/\mu_y$  e sia  $\lambda = \sqrt{\mu_x^2 + \mu_y^2}$ . Con questa parametrizzazione  $\psi$  e  $\lambda$  risultano ortogonali (Severini, 2011, Ventura e Racugno, 2012).

Si è visto che (cfr. § 1.4.5) in questo esempio la funzione di verosimiglianza profilo modificata coincide sia con  $L_{CA}(\psi)$  sia con  $L_P(\psi)$ , poiché  $j_{\lambda\lambda}(\psi, \lambda) = n$ .

La matrice di informazione attesa di Fisher è

$$i(\psi, \lambda) = \begin{bmatrix} \frac{n\lambda^2}{(1 + \psi^2)^2} & 0 \\ 0 & n \end{bmatrix}$$

e la distribuzione *matching prior* per il parametro d'interesse  $\psi$  risulta pertanto data da

$$\pi_{mp}(\psi) \propto i_{\psi\psi}(\psi, \hat{\lambda}_\psi)^{1/2} = \left[ \frac{\hat{\lambda}_\psi^2}{(1 + \psi^2)^2} \right]^{1/2} = \left[ \frac{(\psi\bar{x} + \bar{y})^2}{(1 + \psi^2)^3} \right]^{1/2}.$$

Si noti che la *matching prior* per  $\psi$  è una distribuzione propria.

La corrispondente pseudo-distribuzione a posteriori assume la forma

$$\pi_{ps}(\psi|x, y) \propto i_{\psi\psi}(\psi, \hat{\lambda}_\psi)^{1/2} L_{CA}(\psi) = \left[ \frac{(\psi\bar{x} + \bar{y})^2}{(1 + \psi^2)^3} \right]^{1/2} \exp \left[ \frac{n(\psi\bar{x} + \bar{y})^2}{2(1 + \psi^2)} \right], \quad (2.11)$$

che è una distribuzione a posteriori propria, poiché è espressa in termini di  $L_{CA}(\psi)$  e di  $i_{\psi\psi}(\psi, \hat{\lambda}_\psi)^{1/2}$ ; per approfondimenti si veda Ventura *et al.* (2009).

La Figura 1.2 confronta la pseudo-distribuzione a posteriori marginale (2.11) (linea continua) con la *matching prior* (linea tratteggiata), per un campione simulato con  $n = 5$ ,  $\lambda = \sqrt{101}$  e  $\psi = 10$ .

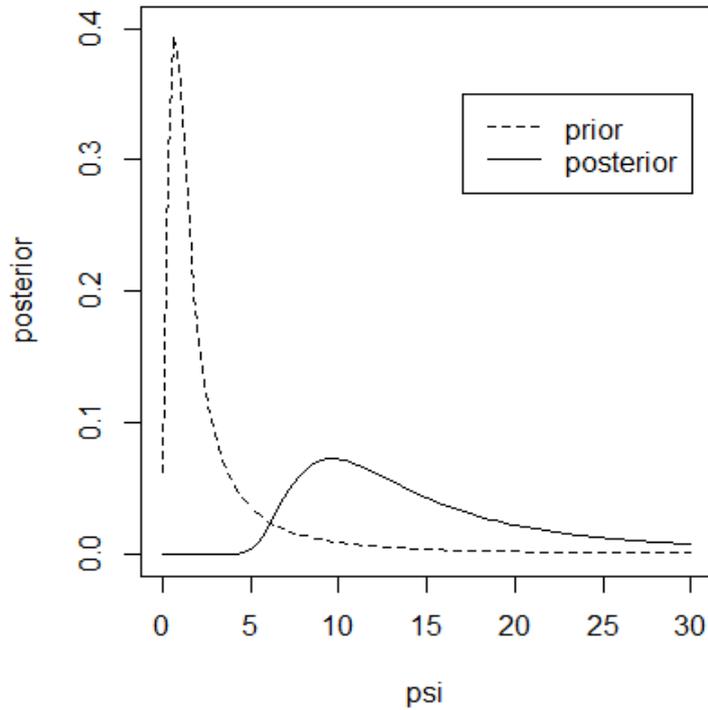


Figura 1.2. Esempio 4.1: Confronto tra  $\pi_{mp}(\psi)$  e  $\pi_{ps}(\psi|x, y)$  con  $n = 5$ ,  $\lambda = \sqrt{101}$  e  $\psi = 10$ .

### Esempio 1.2

Siano  $(x_1, \dots, x_n)$  e  $(y_1, \dots, y_n)$  due campioni casuali indipendenti da due distribuzioni di Poisson di medie  $\psi\lambda$  e  $\lambda$ , rispettivamente, con  $\psi > 0$  e  $\lambda > 0$ . Il parametro d'interesse  $\psi$  è il rapporto tra le medie delle due distribuzioni. La funzione di verosimiglianza per  $(\psi, \lambda)$  è

$$L(\psi, \lambda) = \exp\{n\bar{x}\log\psi + n(\bar{x} + \bar{y})\log\lambda - n\lambda(\psi + 1)\},$$

dove  $\bar{x}$  e  $\bar{y}$  denotano le medie campionarie.

Fissato  $\psi$ ,  $n(\bar{x} + \bar{y})$  è una statistica sufficiente. Si può dimostrare che (Severini, 2000, Cap. 8) la funzione di verosimiglianza condizionata per  $\psi$  è data da

$$L_C(\psi) = \frac{\psi^{n\bar{x}}}{(\psi + 1)^{n(\bar{x} + \bar{y})}}.$$

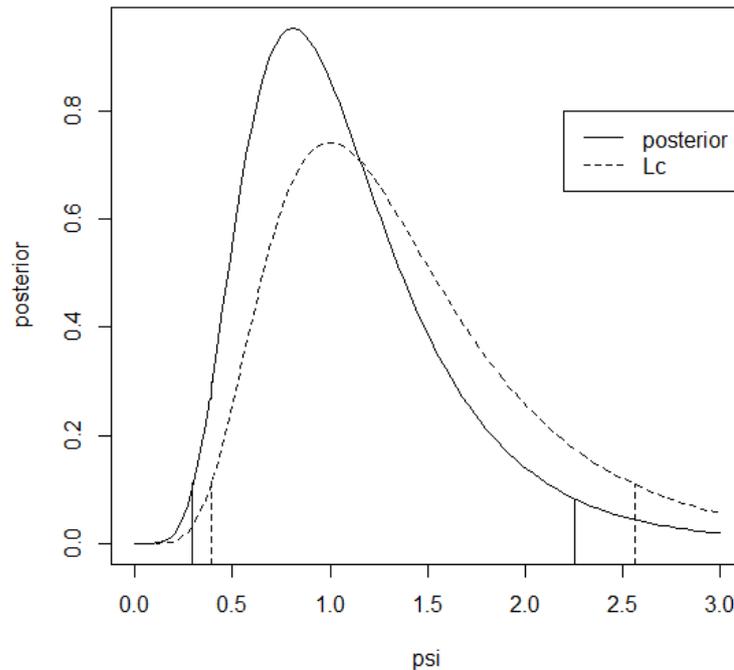
La *matching prior* risulta

$$\pi_{mp}(\psi) \propto i_C(\psi)^{1/2} = \frac{1}{\psi^{1/2}(\psi + 1)}$$

e la corrispondente pseudo-distribuzione a posteriori è (Ventura *et al.*, 2009)

$$\pi_{ps}(\psi|y) \propto i_C(\psi)^{1/2} L_C(\psi) = \frac{\psi^{n\bar{x}-\frac{1}{2}}}{(\psi+1)^{n(\bar{x}+\bar{y})+1}}.$$

Si noti che la *matching prior* è una distribuzione propria, in quanto è una funzione positiva il cui integrale è finito.



**Figura 2.2. Esempio 1.2: Confronto tra  $\pi_{ps}(\psi|y)$  e  $L_C(\psi)$  normalizzata con  $n = 5$ ,  $\lambda = 1$  e  $\psi = 2$ .**

La Figura 2.2, ottenuta da due distribuzioni Poisson con  $\psi = 2$ ,  $\lambda = 1$  e  $n = 5$ , confronta  $\pi_{ps}(\psi|y)$  (linea continua) con la  $L_C(\psi)$  normalizzata (linea tratteggiata), riportando per la prima un intervallo *HPD* (Highest Posterior Density) di livello  $1 - \alpha = 0.95$ , ovvero un intervallo che raccoglie tutti i valori di  $\psi$  la cui densità a posteriori risulta più elevata, e per la seconda un intervallo di confidenza, sempre di livello  $1 - \alpha = 0.95$  basato sul log-rapporto di verosimiglianza condizionata (cfr. § 1.2). Si noti che l'intervallo *HPD* risulta più stretto di quello frequentista. Inoltre, la moda a posteriori è minore rispetto alla s.m.v. condizionata.

■

### Esempio 2.2

Siano  $(x_1, \dots, x_n)$  e  $(y_1, \dots, y_n)$  due campioni casuali indipendenti provenienti da due distribuzioni esponenziali con tasso, rispettivamente,  $\alpha$  e  $\beta$ . Si consideri la riparametrizzazione da  $\theta = (\alpha, \beta)$  a  $\theta = (\psi, \lambda)$ , con  $\psi = \alpha/(\alpha + \beta)$  parametro

d'interesse e  $\lambda = (\alpha + \beta)$  parametro di disturbo. Si noti che  $\psi = \mathbb{P}_\theta[X < Y]$  (Kotz *et al.*, 2003).

Le funzioni di verosimiglianza e di log-verosimiglianza per  $(\psi, \lambda)$  sono, rispettivamente,

$$L(\psi, \lambda) = \lambda^{2n} [\psi(1 - \psi)]^n \exp\{-n\lambda\psi\bar{x} - n\lambda(1 - \psi)\bar{y}\}$$

e

$$l(\psi, \lambda) = 2n \log \lambda + n \log \psi(1 - \psi) - n\lambda\psi\bar{x} - n\lambda(1 - \psi)\bar{y},$$

dove  $\bar{x}$  e  $\bar{y}$  denotano le medie campionarie.

La s.m.v. di  $\lambda$  con  $\psi$  fissato è

$$\hat{\lambda}_\psi = \frac{2}{\psi\bar{x} + (1 - \psi)\bar{y}}$$

e la funzione di verosimiglianza profilo risulta

$$L_P(\psi) = \hat{\lambda}_\psi^{2n} [\psi(1 - \psi)]^n.$$

La funzione di verosimiglianza profilo modificata coincide con la verosimiglianza profilo. Infatti, si ha

$$j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi) = 2n(\hat{\lambda}_\psi)^{-2}$$

e

$$\left| \frac{\partial \hat{\lambda}_\psi}{\partial \hat{\lambda}} \right| = \frac{\hat{\lambda}_\psi}{\hat{\lambda}},$$

ottenendo quindi

$$M(\psi) = |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{-\frac{1}{2}} \left| \frac{\partial \hat{\lambda}_\psi}{\partial \hat{\lambda}} \right|^{-1} = \frac{\hat{\lambda}}{\sqrt{2n}},$$

che dipende solo dai dati, dove  $(\hat{\psi}, \hat{\lambda}) = (\bar{y}/(\bar{x} + \bar{y}), (\bar{x} + \bar{y})/\bar{x}\bar{y})$  è la s.m.v. di  $(\psi, \lambda)$ .

La *matching prior* per  $\psi$  è data da

$$\pi_{mp}(\psi) \propto i_{\psi\psi,\lambda}(\psi, \hat{\lambda}_\psi)^{1/2} = [\psi(1 - \psi)]^{-1},$$

e la pseudo-distribuzione a posteriori risulta

$$\pi_{ps}(\psi|x, y) \propto i_{\psi\psi,\lambda}(\psi, \hat{\lambda}_\psi)^{1/2} L_{MP}(\psi) = \hat{\lambda}_\psi^{2n} [\psi(1 - \psi)]^{n-1}. \quad (2.12)$$

Si può dimostrare che (Ventura e Racugno, 2011) la (2.12) coincide con la distribuzione a posteriori utilizzando la distribuzione a priori congiunta non informativa di Jeffreys.

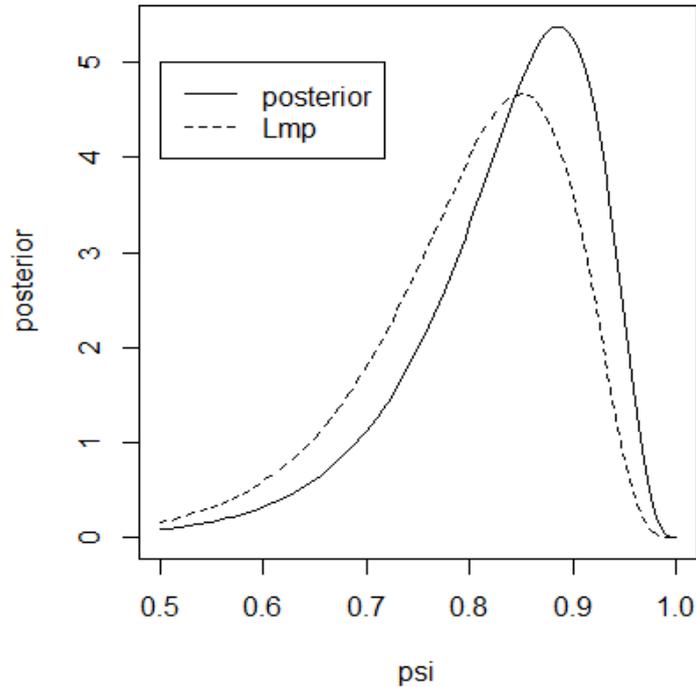


Figura 3.2. Esempio 2.2: Confronto tra  $\pi_{ps}(\psi|x, y)$  e  $L_{MP}(\psi)$  normalizzata con  $n = 5$ ,  $\lambda = 20$  e  $\psi = 0.85$ .

La Figura 3.2 confronta la pseudo-distribuzione a posteriori marginale con la funzione di verosimiglianza profilo modificata normalizzata per un campione simulato con  $n = 5$ ,  $\lambda = 20$  e  $\psi = 0.85$ .

■

### 2.3.2 Interpretazione frequentista delle *matching priors*

Un'altra interpretazione della *matching prior* è quella di rappresentare una penalizzazione per la funzione di verosimiglianza profilo modificata, allo scopo di migliorare le proprietà delle stime e degli intervalli di confidenza per il parametro d'interesse.

Infatti, a partire dalla *matching prior* (2.9), si può definire la funzione di verosimiglianza profilo modificata penalizzata (Ventura e Racugno, 2012)

$$L_{MP}^*(\psi) = L_{MP}(\psi)\pi_{mp}(\psi), \quad (2.13)$$

dove  $L_{MP}(\psi)$  è la funzione di verosimiglianza profilo modificata (cfr. § 1.4.5) e  $\pi_{mp}(\psi)$  è la *matching prior* corrispondente.

Si può dimostrare che la (2.13) gode delle seguenti proprietà (Ventura e Racugno, 2012):

- (1) la lunghezza attesa degli intervalli di confidenza basati su  $L_{MP}^*(\psi)$  è minore rispetto a quelli basati su  $L_{MP}(\psi)$ ;
- (2) la s.m.v. ottenuta dalla (2.13), ovvero il valore  $\hat{\psi}_{MP}^*$  che massimizza  $L_{MP}^*(\psi)$ , e la s.m.v.  $\hat{\psi}_{MP}$  di  $L_{MP}(\psi)$  sono tali che l'errore quadratico medio di  $\hat{\psi}_{MP}^*$  è minore rispetto a quello di  $\hat{\psi}_{MP}$ .

Inoltre, gli intervalli di confidenza basati su  $L_{MP}^*(\psi)$  forniscono degli intervalli di credibilità HPD bayesiani di ordine  $O_p(n^{-1})$  e  $\hat{\psi}_{MP}^*$  coincide con la moda a posteriori della pseudo-distribuzione marginale (2.10).

#### Esempio 4.1 (cont.)

La funzione di verosimiglianza profilo modificata penalizzata  $L_{MP}^*(\psi)$  coincide con il nucleo della pseudo-distribuzione a posteriori ed è pari a

$$L_{MP}^*(\psi) = \left[ \frac{(\psi\bar{x} + \bar{y})^2}{(1 + \psi^2)^3} \right]^{1/2} \exp \left[ \frac{n(\psi\bar{x} + \bar{y})^2}{2(1 + \psi^2)} \right]. \quad (2.14)$$

La funzione di log-verosimiglianza profilo modificata penalizzata è

$$l_{MP}^*(\psi) = \frac{1}{2} \log \left[ \frac{(\psi\bar{x} + \bar{y})^2}{(1 + \psi^2)^3} \right] + \frac{n(\psi\bar{x} + \bar{y})^2}{2(1 + \psi^2)}.$$

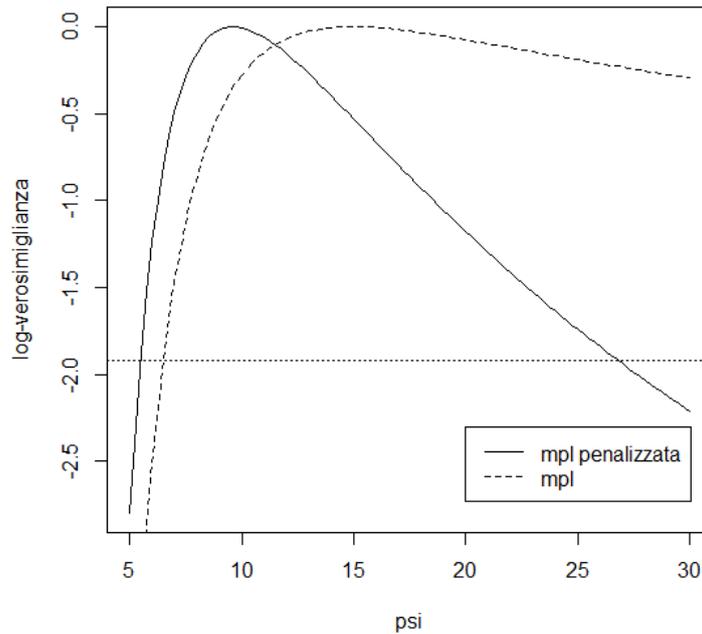


Figura 4.2. Esempio 4.1: Confronto tra  $l_{MP}^*(\psi)$  e  $l_{MP}(\psi)$  normalizzate con  $n = 5$ ,  $\lambda = \sqrt{101}$  e  $\psi = 10$ .

La Figura 4.2 illustra l'andamento delle funzioni di log-verosimiglianza profilo modificata penalizzata e profilo modificata normalizzate. Si noti che quando  $\psi \rightarrow \infty$ ,  $l_{MP}(\psi)$  può non tendere a  $-\infty$ , al contrario di  $l_{MP}^*(\psi)$ ; questo comporta che l'intervallo di confidenza basato sul log-rapporto di verosimiglianza profilo modificata (cfr. § 1.2) può non essere finito. Nel caso specifico della Figura 4.2, la retta orizzontale tratteggiata, che rappresenta  $-\chi_{1;0.95}^2/2$ , interseca solo una volta la funzione di log-verosimiglianza profilo modificata. Il campione utilizzato è stato simulato con  $n = 5$  da due distribuzioni normali con medie, rispettivamente,  $\mu_x = 10$  e  $\mu_y = 1$ , ossia con  $\psi = 10$  e  $\lambda = \sqrt{101}$ .

■

### Esempio 2.2 (cont.)

Le funzioni di verosimiglianza e di log-verosimiglianza profilo modificata penalizzata per  $\psi = \mathbb{P}_\theta[X < Y]$  sono, rispettivamente,

$$L_{MP}^*(\psi) = \hat{\lambda}_\psi^{2n} [\psi(1 - \psi)]^{n-1} \quad (2.15)$$

e

$$l_{MP}^*(\psi) = 2n \log \hat{\lambda}_\psi + (n - 1) \log[\psi(1 - \psi)].$$

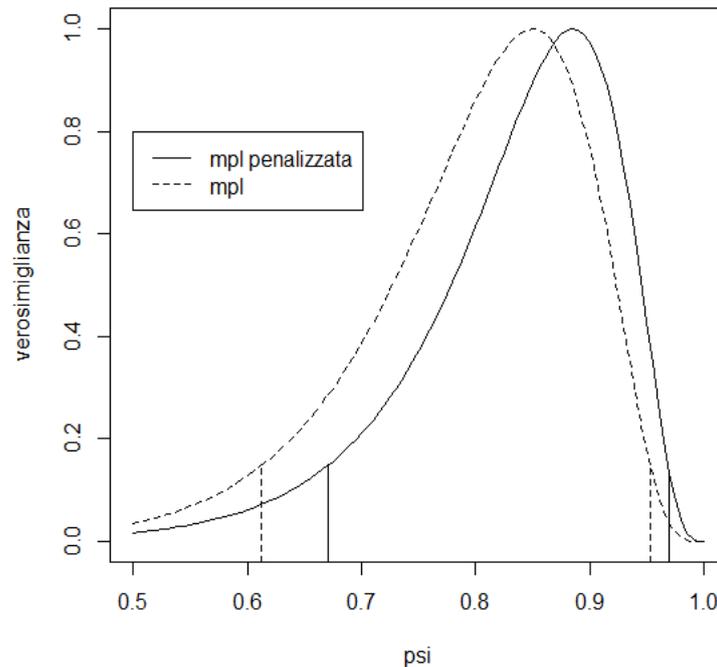


Figura 5.2. Esempio 2.2: Confronto tra  $L_{MP}^*(\psi)$  e  $L_{MP}(\psi)$  normalizzate, con  $n = 5$ ,  $\lambda = 20$  e  $\psi = 0.85$ .

La Figura 5.2 illustra le funzioni di verosimiglianza profilo modificata penalizzata e profilo modificata normalizzate. Nel campione simulato, si può notare che l'intervallo basato sulla statistica log-rapporto di verosimiglianza profilo modificata penalizzata, barre verticali continue, risulta più corto di quello costruito con  $l_{MP}(\psi)$ , barre verticali tratteggiate. Le due funzioni sono state ottenute da un campione simulato di  $n = 5$  osservazioni, da due distribuzioni esponenziali di parametri, rispettivamente,  $\alpha = 17$  e  $\beta = 3$ , ossia con  $\psi = 0.85$  e  $\lambda = 20$ .

■

## 2.4 Approssimazione della funzione di ripartizione di $\pi(\psi|y)$

In questo paragrafo si presenta la *tail area approximation* per il parametro d'interesse  $\psi$ , ovvero l'approssimazione asintotica di

$$\mathbb{P}_{\psi|Y}[\psi \leq \psi_0] = \int_{-\infty}^{\psi_0} \pi(\psi|y) d\psi. \quad (2.16)$$

Sotto determinate condizioni di regolarità (per approfondimenti si veda Kass *et al.*, 1990), la distribuzione a posteriori marginale (2.1) può essere approssimata attraverso la formula di Laplace (cfr. § 1.4.6) come

$$\pi(\psi|y) \doteq \frac{1}{\sqrt{2\pi}} j_P(\hat{\psi})^{1/2} \exp\{l_P(\psi) - l_P(\hat{\psi})\} \frac{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2}}{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2}} \frac{\pi(\psi, \hat{\lambda}_\psi)}{\pi(\hat{\psi}, \hat{\lambda})}, \quad (2.17)$$

dove il simbolo " $\doteq$ " indica che l'ordine di approssimazione è  $O(n^{-3/2})$  (Tierney e Kadane, 1986). Si può mostrare che, a partire dalla (2.17), la *tail area approximation* risulta data da (Reid, 2003, Brazzale *et al.*, 2007, e Ruli *et al.*, 2013)

$$\int_{-\infty}^{\psi_0} \pi(\psi|y) d\psi \doteq \Phi \left\{ r_P(\psi_0) + \frac{1}{r_P(\psi_0)} \log \frac{q_B(\psi_0)}{r_P(\psi_0)} \right\} = \Phi \{ r_B^*(\psi_0) \}, \quad (2.18)$$

dove  $r_P(\psi) = \text{sgn}(\hat{\psi} - \psi) \sqrt{W_P(\psi)}$  è la radice con segno del log-rapporto di verosimiglianza profilo (cfr. § 1.4.3) e

$$q_B(\psi) = l'_P(\psi) j_P(\psi)^{-1/2} \frac{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2}}{|j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2}} \frac{\pi(\hat{\psi}, \hat{\lambda})}{\pi(\psi, \hat{\lambda}_\psi)}, \quad (2.19)$$

con  $l'_p(\psi) = \partial l_p(\psi) / \partial \psi$  funzione punteggio profilo. L'approssimazione (2.18) fornisce risultati accurati anche nel caso in cui la dimensione di  $\lambda$  sia elevata, con  $n$  piccolo (Ruli *et al.*, 2013).

A partire dalla (2.18) si può costruire un intervallo di credibilità bilaterale (“*equi-tailed*”), ossia che lascia uguale probabilità a sinistra dell'estremo inferiore e a destra dell'estremo superiore di livello approssimato  $(1 - \alpha)$  come (Ventura *et al.*, 2013)

$$\{\psi: |r_B^*(\psi)| \leq z_{1-\alpha/2}\}. \quad (2.20)$$

Inoltre, è possibile ottenere la mediana a posteriori come la soluzione in  $\psi$  dell'equazione  $r_B^*(\psi) = 0$ .

Se nella (2.18) si utilizza la *matching prior* (2.9), il termine (2.19) di  $r_B^*(\psi)$  assume la forma (Ventura e Racugno, 2011 e Ventura *et al.*, 2013)

$$q_p(\psi) = l'_p(\psi) j_P(\hat{\psi})^{-1/2} \frac{|i_{\psi\psi,\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2}}{|i_{\psi\psi,\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2}} \frac{1}{M(\psi)}. \quad (2.21)$$

In questo caso  $r_B^*(\psi)$  coincide con la statistica frequentista di Barndorff-Nielsen e Chamberlin (1994) e quindi la (2.20) fornisce un intervallo di confidenza frequentista.

Infine, in Ventura *et al.* (2013) si dimostra che la pseudo-distribuzione a posteriori basata sulla *matching prior* (2.10) può anche essere riscritta come

$$\pi_{ps}(\psi|y) \propto \exp\left(-\frac{1}{2} r_p^*(\psi)^2\right) \left| \frac{s_p(\psi)}{r_p(\psi)} \right|, \quad (2.22)$$

dove  $s_p(\psi) = l'_p(\psi) j_P(\hat{\psi})^{-1/2}$  è la statistica punteggio profilo (Pace e Salvani, 1996, § 4.6) e  $r_p^*(\psi)$  è la statistica frequentista di Barndorff-Nielsen e Chamberlin (1994).

#### 2.4.1 Simulazione attraverso la *tail area approximation*

L'espressione (2.18) fornisce un'approssimazione accurata per i quantili della distribuzione a posteriori marginale e per gli intervalli di credibilità bilaterali. Tuttavia, in generale, non è possibile ottenere intervalli di credibilità *HPD* o momenti della distribuzione a posteriori (ad esempio, media e varianza a posteriori).

In questo paragrafo viene presentato lo schema di simulazione HOTA (Higher-Order Tail Area), che si basa sull'approssimazione (2.18). Per maggiori approfondimenti si veda Ruli *et al.* (2013). Lo schema di simulazione HOTA ha la caratteristica di generare

campioni indipendenti dalla distribuzione a posteriori marginale, al contrario dei metodi MCMC che, in generale, producono campioni dipendenti (Ruli *et al.*, 2013). La simulazione tramite HOTA, come per tutti gli altri metodi, è soggetta ad un errore Monte Carlo di ordine  $O_p(T^{-1/2})$ , dove  $T$  è il numero di valori simulati. È importante sottolineare che l'accuratezza del campione generato della distribuzione a posteriori mediante lo schema HOTA dipende dalla numerosità campionaria; tuttavia, l'approssimazione è tipicamente accurata anche per campioni di numerosità piccola (Ruli *et al.*, 2013).

Lo schema di simulazione HOTA usa la statistica  $r_B^*(\psi)$  definita nella (2.18). Per versioni alternative e maggiori approfondimenti si veda Ruli *et al.* (2013).

L'algoritmo può essere sintetizzato nei seguenti passi:

- (1) generare  $T$  valori  $(z_1, \dots, z_T)$  da una  $N(0,1)$ ;
- (2) trovare la soluzione in  $\psi_t$  dell'equazione  $r_B^*(\psi_t) = z_t$ , per  $t = 1, \dots, T$ ;
- (3)  $\psi = (\psi_1, \dots, \psi_T)$  è un campione indipendente dalla distribuzione a posteriori marginale  $\pi(\psi|y)$ .

La simulazione HOTA basata su  $r_B^*(\psi)$  fornisce buoni risultati quando la moda della distribuzione a posteriori è vicina alla s.m.v.. Per estensioni dello schema di simulazione HOTA si rinvia a Ruli *et al.* (2013).

### Esempio 2.2 (cont.)

Si consideri la pseudo-distribuzione a posteriori (2.12). Per applicare lo schema di simulazione HOTA è necessario calcolare le funzioni  $q_p(\psi)$  e  $r_p(\psi)$ , per ottenere  $r_p^*(\psi)$ .

Partendo dalla funzione di verosimiglianza profilo, la radice con segno del log-rapporto di verosimiglianza profilo assume la forma

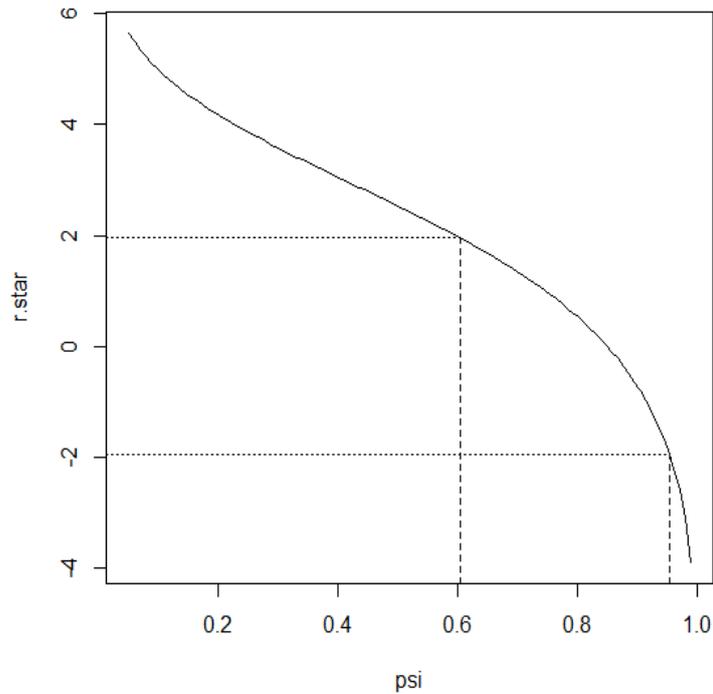
$$r_p(\psi) = \text{sgn}(\hat{\psi} - \psi) \sqrt{4n \log \frac{\hat{\lambda}}{\hat{\lambda}_\psi} + 2n \log \frac{[\hat{\psi}(1 - \hat{\psi})]}{[\psi(1 - \psi)]}}$$

Inoltre, si ha

$$q_p(\psi) = \frac{l'_p(\psi)}{j_p(\hat{\psi})^{1/2}} \frac{\hat{\psi}(1 - \hat{\psi})}{\psi(1 - \psi)},$$

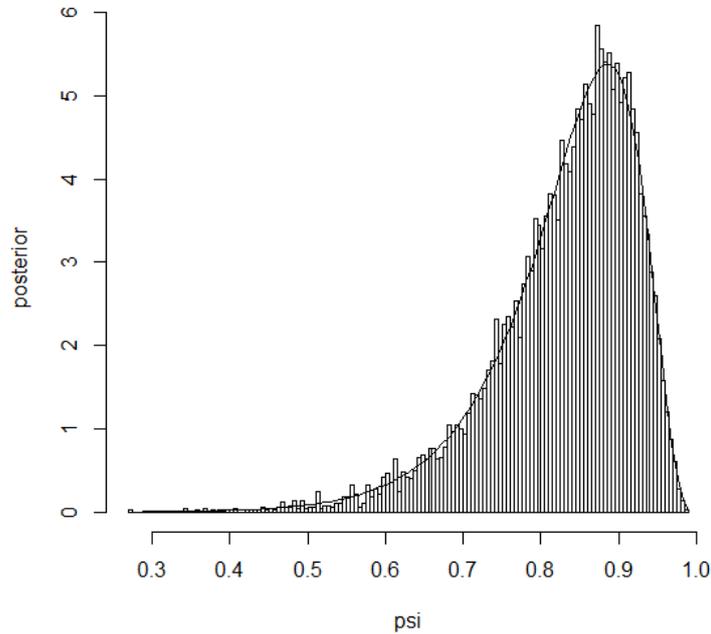
con

$$l'_p(\psi) = -n\hat{\lambda}_\psi(\bar{x} - \bar{y}) + n\frac{1 - 2\psi}{\psi(1 - \psi)}.$$



**Figura 6.2. Esempio 2.2: Statistica  $r_p^*(\psi)$  con  $n = 5$ ,  $\lambda = 20$  e  $\psi = 0.85$ .**

La Figura 6.2 illustra l'andamento della statistica  $r_p^*(\psi)$  per un campione generato con  $n = 5$ ,  $\lambda = 20$  e  $\psi = 0.85$ . Inoltre, nel grafico è riportato un intervallo di credibilità bilaterale di livello  $1 - \alpha = 0.95$ , costruito a partire dalla statistica  $r_p^*(\psi)$ , tale intervallo, quindi, è anche un intervallo di confidenza asintotico di ordine elevato. L'intervallo è pari a  $(0.604, 0.955)$ .



**Figura 7.2. Esempio 2.2: Istogramma dei valori simulati tramite lo schema HOTA con  $n = 5$ ,  $\lambda = 20$  e  $\psi = 0.85$ , dove la linea continua rappresenta la pseudo-distribuzione a posteriori marginale (2.11).**

La Figura 7.2 riporta l'istogramma dei valori simulati mediante lo schema HOTA, dove è stata aggiunta la curva della pseudo-distribuzione a posteriori marginale (2.11).

Dalla distribuzione calcolata con HOTA è possibile derivare un intervallo *HPD* per  $\psi$ , che risulta pari a (0.646, 0.968).

Inoltre, poiché si sta utilizzando la *matching prior* (2.9) lo schema HOTA fornisce anche un campione simulato da  $L_{MP}^*(\psi)$ , che potrebbe essere utilizzato nelle analisi frequentiste, ad esempio, per calcolare la s.m.v. profilo modificata penalizzata che coincide con la moda della distribuzione a posteriori.

■

## 2.5 Conclusioni

In questo capitolo sono stati presentati alcuni recenti sviluppi ibridi per l'eliminazione dei parametri di disturbo. In particolare, si è mostrato come in recenti studi si faccia ricorso a procedure ibride, ossia che accompagnano alle quantità inferenziali bayesiane elementi frequentisti. Queste tecniche ibride, infatti, permettono di ottenere funzioni del solo parametro d'interesse, che possono essere interpretate in ambito frequentista come

una versione penalizzata della funzione di pseudo-verosimiglianza tramite *priors*, e nel contesto bayesiano come pseudo-distribuzioni a posteriori marginali del parametro d'interesse. Infine, sono stati illustrati alcuni risultati asintotici per le pseudo-distribuzioni a posteriori, fornendo uno schema di simulazione (HOTA) che si basa sull'approssimazione della distribuzione a posteriori marginale.

Nel prossimo capitolo si studierà un problema relativo all'inferenza per un determinato parametro d'interesse nell'ambito dei modelli di regressione, applicando queste procedure inferenziali ibride.

## Capitolo 3

### Inferenza sul rapporto di coefficienti in un modello di regressione normale

In questo capitolo si desidera applicare i risultati illustrati nei capitoli precedenti sull'eliminazione dei parametri di disturbo, al caso in cui è di interesse l'inferenza sul rapporto di due coefficienti in un modello di regressione normale.

Dopo aver presentato il problema dell'inferenza sul rapporto tra due coefficienti di regressione (Ghosh *et al.*, 2003 e 2006), nel Paragrafo 3.2 si presenta un'opportuna riparametrizzazione ortogonale, che viene usata per calcolare le funzioni di verosimiglianza condizionata approssimata e profilo modificata penalizzata per il parametro d'interesse; per maggiori approfondimenti su tale riparametrizzazione si rinvia a Ghosh *et al.* (2003, 2006). Il Paragrafo 3.3 illustra tali funzioni di pseudo-verosimiglianza e in particolare si ottengono, per il parametro d'interesse, le funzioni di verosimiglianza profilo, condizionata approssimata e due diverse versioni della funzione di verosimiglianza integrata, ottenute scegliendo due alternative distribuzioni a priori condizionate. I principali risultati sono presentati in Ghosh *et al.* (2006), mentre per la scelta delle distribuzioni a priori condizionate si rimanda, ad esempio, a Berger e Bernardo (1992) e Berger *et al.* (1999).

Il Paragrafo 3.4 si focalizza sulla derivazione delle pseudo-distribuzioni a posteriori per il rapporto tra due coefficienti di regressione, applicando le tecniche ibride presentate nel Capitolo 2. Nello specifico, viene derivata una pseudo-distribuzione a posteriori utilizzando la funzione di verosimiglianza condizionata approssimata. Come visto nel Capitolo 2 tale distribuzione può essere interpretata, nel paradigma frequentista, come una versione penalizzata della verosimiglianza profilo modificata.

Nel Paragrafo 3.5 si presenta la costruzione di intervalli di confidenza per il parametro d'interesse basati sulle statistiche log-rapporto di pseudo-verosimiglianza. In particolare, si evidenzia come per le funzioni di verosimiglianza profilo e condizionata approssimata l'intervallo di confidenza per il rapporto di due coefficienti, sotto opportune condizioni che coinvolgono l'estremo superiore della statistica che si utilizza per costruire l'intervallo, può non essere finito, mentre questo non si verifica se si utilizzano le funzioni di verosimiglianza integrata (Paragrafo 3.3) o la funzione di verosimiglianza profilo modificata penalizzata (Paragrafo 3.4). Inoltre, gli intervalli di confidenza basati sul log-rapporto di pseudo-verosimiglianza profilo modificata penalizzata forniscono anche un intervallo *HPD* bayesiano.

Nel Paragrafo 3.6 vengono illustrati alcuni esempi, basati su dati simulati e reali. Infine, nel Paragrafo 3.7 vengono analizzate, attraverso alcuni studi di simulazione, le proprietà delle funzioni di pseudo-verosimiglianza introdotte nei Paragrafi 3.3 e 3.4. In particolare, vengono studiate l'ampiezza media e la copertura degli intervalli di confidenza costruiti a partire dalle statistiche log-rapporto di pseudo-verosimiglianza e la distorsione e l'errore quadratico medio delle stime di massima pseudo-verosimiglianza.

### **3.1 Il modello**

In ambito bayesiano l'inferenza sui parametri d'interesse può risultare difficoltosa quando, in particolare, la dimensione del parametro di disturbo è elevata, sia perché bisogna elicitarne una distribuzione a priori su tutto il parametro sia perché è necessario calcolare un integrale multidimensionale (cfr. § 2.1). Nel Capitolo 2 si è mostrato come queste difficoltà possono essere evitate applicando una metodologia ibrida, che si serve sia di elementi frequentisti (le pseudo-verosimiglianze) sia di elementi bayesiani (le *matching priors*). Un modello dove tale tecnica può essere applicata è il modello di regressione, quando il parametro d'interesse  $\psi$  è ad esempio un singolo coefficiente o una sua trasformazione, e il parametro di disturbo  $\lambda$  è dato dai rimanenti parametri. Nel modello di regressione è usuale che la dimensione di  $\lambda$  sia considerevole.

In letteratura esiste un'ampia classe di problemi statistici che possono essere ricondotti all'inferenza per una trasformata dei coefficienti di regressione, nello specifico il

rapporto tra due coefficienti. Questa classe include, ad esempio, i problemi di calibrazione ed il rapporto tra due medie (si veda, ad esempio, Ventura e Racugno, 2012). E come nel caso del rapporto tra due medie, anche l'inferenza sul rapporto di due coefficienti di regressione può incontrare alcuni problemi. In particolare, può accadere che (Ghosh *et al.*, 2006) l'intervallo di confidenza per il parametro d'interesse risulti non finito.

Nel caso in cui il parametro d'interesse è il rapporto tra medie di due gruppi nell'ambito di osservazioni i.i.d. con distribuzione normale, si introduce un vincolo aggiuntivo sui parametri che identificano le medie dei due campioni, secondo cui entrambi i parametri devono essere positivi. Questo esempio è un naturale sottocaso del rapporto di coefficienti in un modello di regressione normale ed è stato sviluppato, nel caso di varianza nota, nell'Esempio 4.1; per maggiori approfondimenti si vedano, ad esempio, Severini (2011) e Ventura e Racugno (2012).

Partendo dai risultati di Ghosh *et al.* (2006), in questo capitolo si applica al problema in esame la metodologia ibrida discussa nel Capitolo 2. Si consideri il modello di regressione lineare con errori normali, dato da

$$y_i = x_i^T \beta + \varepsilon_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \dots, n, \quad (3.1)$$

dove gli  $\varepsilon_i$  sono errori indipendenti  $N(0, \sigma^2)$ ,  $x_i$  è un vettore  $p$ -dimensionale di costanti note,  $i = 1, \dots, n$ ,  $\beta_j \in \mathbb{R}$ , per  $j \neq r$ ,  $\beta_r \in \mathbb{R} - \{0\}$  e  $\sigma^2 > 0$  è l'ignota varianza dell'errore.

Il modello (3.1) può essere riscritto in forma matriciale come

$$y = X\beta + \varepsilon, \quad (3.2)$$

dove  $y = (y_1, \dots, y_n)^T$ ,  $x_i = (x_{i1}, \dots, x_{ip})^T$ ,  $i = 1, \dots, n$ ,  $X^T = (x_1, \dots, x_n)$ ,  $\beta = (\beta_1, \dots, \beta_p)^T$  e  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ . Si assume che  $\text{rango}(X) = p < n$ .

È di interesse fare inferenza sul parametro

$$\psi = \frac{\beta_j}{\beta_r}$$

e il parametro di disturbo risulta pertanto dato da

$$\eta = (\beta^*, \sigma),$$

dove  $\beta^*$  rappresenta il vettore  $\beta$  dei coefficienti di regressione con  $\beta_j$  escluso.

Il parametro  $\psi$  può essere utilizzato, ad esempio, per verificare l'ipotesi nulla  $H_0: \psi = 1$ . Nel seguito si utilizza, per semplicità e senza perdita di generalità, la notazione adottata da Ghosh *et al.* (2006), per la quale  $\psi = \beta_1/\beta_2$  e  $\eta = (\beta_2, \dots, \beta_p, \sigma)$ .

### **3.2 Parametrizzazione ortogonale**

Per poter calcolare la funzione di verosimiglianza condizionata approssimata per  $\psi$  è necessario introdurre una parametrizzazione ortogonale, presentata in Ghosh *et al.* (2006).

Sia  $S = n^{-1}X^T X$  la matrice  $p \times p$  di rango pieno, partizionata come

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix},$$

dove  $S_{11} = [s_{ij}]$ , con  $i, j = 1, 2$ ,  $S_{12} = S_{21}^T = [s_{ij}]$ , con  $i = 1, 2$  e  $j = 3, \dots, p$ , e  $S_{22} = [s_{ij}]$ , con  $i, j = 3, \dots, p$ .

Si consideri la matrice  $2 \times 2$  data da  $C = S_{11} - S_{12}A$ , con  $A = S_{22}^{-1}S_{21}$ , dove  $C = [c_{ij}]$ , con  $i, j = 1, 2$ , e  $A = [a_{ij}]$ , con  $i = 1, \dots, p$  e  $j = 1, 2$ . Poiché  $X$  è di rango pieno, allora  $S$ ,  $S_{22}$  e  $C$  sono matrici definite positive. Si definisce infine la quantità  $Q(\psi) = c_{11}\psi^2 + 2c_{12}\psi + c_{22}$ , che è una funzione parabolica positiva per ogni  $\psi$ .

Si consideri, quindi, la riparametrizzazione da  $(\beta^*, \sigma)$  a  $(\psi, \lambda)$ , con

$$\begin{aligned} \psi &= \beta_1/\beta_2, \\ \theta_2 &= \beta_2 Q^{1/2}(\psi), \\ \theta_j &= \beta_j + \beta_2(a_{j1}\psi + a_{j2}), \quad j = 3, \dots, p, \\ \theta_{p+1} &= \sigma, \end{aligned} \tag{3.3}$$

e  $\lambda = (\theta_2, \dots, \theta_{p+1})$ .

Si può dimostrare che (Ghosh *et al.*, 2003)  $\psi$  è ortogonale a  $\lambda$  e che la matrice di informazione attesa di Fisher è data da

$$i(\theta) = i(\psi, \lambda) = \frac{n}{\theta_{p+1}^2} \text{diag} \left( \frac{\theta_2^2 |C|}{Q^2(\psi)}, 1, S_{22}, 2 \right), \tag{3.4}$$

dove  $\text{diag}(\cdot)$  indica una matrice diagonale a blocchi.

### **3.3 Metodologia classica**

La funzione di verosimiglianza per  $(\beta, \sigma)$  è (Ghosh *et al.*, 2006)

$$\begin{aligned} L(\beta, \sigma) &= \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} (y - X\beta)^T (y - X\beta) \right\} \\ &= \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \left[ SSE + (\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta}) \right] \right\}, \end{aligned} \quad (3.5)$$

dove  $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)^T = (X^T X)^{-1} X^T y$  è la s.m.v. di  $\beta$  e  $SSE = (y - X\hat{\beta})^T (y - X\hat{\beta})$ .

La funzione di verosimiglianza nella s.m.v. vale

$$L(\hat{\beta}, \hat{\sigma}^2) = (\hat{\sigma})^{-n}, \quad (3.6)$$

e quindi

$$L(\hat{\theta}) = (\hat{\theta}_{p+1})^{-n} = (SSE)^{-\frac{n}{2}}, \quad (3.7)$$

dove  $\hat{\sigma} = \sqrt{(SSE/n)} (= \hat{\theta}_{p+1})$  è la s.m.v. di  $\sigma$  (e  $\theta_{p+1}$ ).

#### **3.3.1 Verosimiglianza profilo**

Per ottenere la funzione di verosimiglianza profilo per  $\psi$  è necessario massimizzare la (3.5) rispetto a  $\eta = (\beta_2, \dots, \beta_p, \sigma)$  con  $\psi$  fissato. Si può dimostrare che (Ghosh *et al.*, 2006) la funzione di verosimiglianza profilo per  $\psi$  è

$$\begin{aligned} L_p(\psi) &\propto \left\{ SSE + \frac{n|C|(\hat{\beta}_2\psi - \hat{\beta}_1)^2}{Q(\psi)} \right\}^{-\frac{n}{2}} \\ &\propto \left\{ \frac{SSE}{n} + \frac{|C|(\hat{\beta}_2\psi - \hat{\beta}_1)^2}{Q(\psi)} \right\}^{-\frac{n}{2}} = (\tilde{\sigma})^{-n}, \end{aligned} \quad (3.8)$$

dove

$$\tilde{\sigma} = \left\{ \frac{SSE}{n} + \frac{|C|(\hat{\beta}_2\psi - \hat{\beta}_1)^2}{Q(\psi)} \right\}^{\frac{1}{2}}$$

è la s.m.v. di  $\sigma$  con  $\psi$  fissato.

Nella parametrizzazione ortogonale, la (3.8) può essere riscritta come

$$L_P(\psi) \propto (\tilde{\theta}_{p+1})^{-n}, \quad (3.9)$$

dove  $\tilde{\theta}_{p+1}$  è la s.m.v. di  $\theta_{p+1}$  con  $\psi$  fissato. Inoltre, vale  $\tilde{\theta}_{p+1} = \tilde{\sigma}$ .

La funzione di verosimiglianza profilo (3.8) ha due punti stazionari:

1. il massimo assoluto  $\psi_1$ , pari alla s.m.v.

$$\psi_1 = \hat{\psi} = \frac{\hat{\beta}_1}{\hat{\beta}_2}; \quad (3.10)$$

2. il minimo assoluto  $\psi_2$ , dato da

$$\psi_2 = -\frac{c_{22}\hat{\beta}_2 + c_{12}\hat{\beta}_1}{c_{12}\hat{\beta}_2 + c_{11}\hat{\beta}_1}. \quad (3.11)$$

Si noti che, quando  $|\psi| \rightarrow \infty$ , si ha  $(\hat{\beta}_2\psi - \hat{\beta}_1)^2/Q(\psi) \rightarrow \hat{\beta}_2^2/c_{11}$ . La funzione di verosimiglianza profilo, perciò, come  $\tilde{\sigma}$  (e  $\tilde{\theta}_{p+1}$ ) non tende a 0 quando il modulo di  $\psi$  tende ad infinito.

### 3.3.2 Verosimiglianza condizionata approssimata

Per calcolare la funzione di verosimiglianza condizionata approssimata (cfr. § 1.4.4) si utilizza la parametrizzazione ortogonale (3.3). Come definito nel Capitolo 1, la  $L_{CA}(\psi)$  è data da

$$L_{CA}(\psi) = L_P(\psi) |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{-1/2}.$$

Si può dimostrare che il blocco  $(\lambda, \lambda)$  della matrice di informazione osservata valutato nella s.m.v. di  $\lambda$  con  $\psi$  fissato è pari a (Ghosh *et al.*, 2006)

$$j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi) = \frac{n}{(\tilde{\theta}_{p+1})^2} \text{diag}(1, S_{22}, 2). \quad (3.12)$$

Quindi il fattore di penalizzazione nella  $L_{CA}(\psi)$  è proporzionale a

$$\left|(\tilde{\theta}_{p+1})^{-2}\right|^{-p/2} = (\tilde{\theta}_{p+1})^p. \quad (3.13)$$

La funzione di verosimiglianza condizionata approssimata è (Ghosh *et al.*, 2006)

$$L_{CA}(\psi) = \left\{ SSE + \frac{n|C|(\hat{\beta}_2\psi - \hat{\beta}_1)^2}{Q(\psi)} \right\}^{-\frac{n-p}{2}} = (\tilde{\theta}_{p+1})^{-(n-p)}. \quad (3.14)$$

Si noti che, come la funzione di verosimiglianza profilo (3.8), anche la (3.14) presenta gli stessi punti stazionari e dipende solo da  $\tilde{\theta}_{p+1}$ , che non tende a 0 quando  $|\psi| \rightarrow \infty$ .

Per derivare la funzione di verosimiglianza profilo modificata si può nuovamente sfruttare la parametrizzazione ortogonale. Siccome  $\psi$  e  $\lambda$  sono ortogonali, ma non si verifica che  $\hat{\lambda}_\psi = \hat{\lambda}$ , la log-verosimiglianza profilo modificata coincide con la log-verosimiglianza condizionata approssimata a meno di un errore di ordine  $O_p(n^{-1})$ , poiché  $\hat{\lambda}_\psi = \hat{\lambda} + O_p(n^{-1})$ . In base a questo risultato,  $L_{CA}(\psi)$  verrà nel seguito utilizzata come approssimazione di  $L_{MP}(\psi)$ . Per approfondimenti si vedano Pace e Salvan (1996, § 4.7 e 11.6) e Severini (2000, § 9.5).

### 3.3.3 Verosimiglianza integrata

Il calcolo della verosimiglianza integrata (cfr. § 1.4.6) necessita di una distribuzione a priori condizionata  $\pi(\lambda|\psi)$ . Ghosh *et al.* (2006) ricorrono alla *conditional reference prior* sviluppata da Berger *et al.* (1999), che è definita in generale come

$$\pi_{CR}(\lambda|\psi) = h(\psi)\pi^*(\lambda|\psi), \quad (3.15)$$

dove

$$\pi^*(\lambda|\psi) = \sqrt{i_{\lambda\lambda}(\psi, \lambda)} \quad (3.16)$$

è la distribuzione a priori di Jeffreys per  $\lambda$  dato  $\psi$  e  $h(\psi)$  è una funzione che dipende solo dal parametro d'interesse  $\psi$ . Per maggiori approfondimenti si veda Berger *et al.* (1999).

Nella parametrizzazione originale  $(\psi, \beta_2, \dots, \beta_p, \sigma)$  si può dimostrare che (Ghosh *et al.*, 2006) la *conditional reference prior* (3.15)

$$\pi_{CR}(\beta_2, \dots, \beta_p, \sigma|\psi) \propto \sigma^{-p}, \quad (3.17)$$

dove la distribuzione a priori di Jeffreys di  $(\beta_2, \dots, \beta_p, \sigma)$  dato  $\psi$  è data da

$$\pi^*(\beta_2, \dots, \beta_p, \sigma|\psi) \propto Q^{1/2}(\psi)\sigma^{-p}. \quad (3.18)$$

La funzione di verosimiglianza integrata, utilizzando come distribuzione a priori condizionata la (3.17), è quindi (Ghosh *et al.*, 2006)

$$\begin{aligned}
L_I^{CR}(\psi) &= \int_{\mathbb{R}^{p-1}} \int_0^{+\infty} L(\psi, \beta_2, \dots, \beta_p, \sigma) \pi_{CR}(\beta_2, \dots, \beta_p, \sigma | \psi) d\beta d\sigma \\
&\propto Q^{-1/2}(\psi) L_P(\psi) = Q^{-1/2}(\psi) (\tilde{\sigma})^{-n}.
\end{aligned} \tag{3.19}$$

Contrariamente a quanto avviene per le funzioni di verosimiglianza profilo e condizionata approssimata,  $L_I^{CR}(\psi) \rightarrow 0$  quando  $|\psi| \rightarrow \infty$ , poiché  $Q^{-1/2}(\psi) \rightarrow 0$  al crescere di  $|\psi|$ .

In alternativa alla *conditional reference prior* (3.19), Ghosh *et al.* (2006) propongono, nella parametrizzazione originale, una diversa distribuzione a priori condizionata, data da

$$\pi_{OCR}(\beta_2, \dots, \beta_p, \sigma | \psi) \propto \sigma^{-1}, \tag{3.20}$$

dove la sigla "OCR" è l'acronimo di "one-at-a-time *conditional reference prior*" (Ghosh *et al.*, 2006). Per maggiori approfondimenti si veda Berger e Bernardo (1992).

Con la distribuzione a priori (3.20), la funzione di verosimiglianza integrata risulta

$$L_I^{OCR}(\psi) = Q^{-1/2}(\psi) (\tilde{\sigma})^{-[n-(p-1)]}. \tag{3.21}$$

Come la (3.19), anche la (3.21) è caratterizzata dal fatto che tende a 0 quando  $|\psi| \rightarrow \infty$ . Infine, una soluzione più semplice e meno onerosa rispetto alle (3.17) e (3.20) è la distribuzione a priori condizionata di Jeffreys (3.18) che, nella parametrizzazione originale, risulta proporzionale a

$$\pi_J(\beta_2, \dots, \beta_p, \sigma | \psi) \propto \sigma^{-p} Q^{1/2}(\psi). \tag{3.22}$$

Si può dimostrare che la corrispondente funzione di verosimiglianza integrata è proporzionale, a meno di un fattore moltiplicativo che non dipende dal parametro d'interesse  $\psi$ , alla funzione di verosimiglianza profilo (3.8), presentando perciò gli stessi svantaggi che caratterizzano la  $L_P(\psi)$ .

### **3.4 Metodologia ibrida**

In alternativa alla metodologia classica per l'inferenza sul rapporto tra due coefficienti di regressione nel modello con errori normali, è possibile applicare le tecniche illustrate nel Capitolo 2, adattando perciò un approccio ibrido, che si serve contemporaneamente sia di elementi frequentisti che bayesiani.

Per calcolare la pseudo-distribuzione a posteriori basata sulla funzione di verosimiglianza condizionata approssimata (2.5) è necessario calcolare la corrispondente *matching prior*, data da  $\pi(\psi) \propto i_{\psi\psi}(\psi, \hat{\lambda}_\psi)^{1/2}$ . Da tale distribuzione si ottiene la funzione di verosimiglianza profilo modificata penalizzata (2.10).

Con semplici passaggi, si dimostra che la *matching prior* risulta data da

$$\pi(\psi) \propto \left( \frac{(\tilde{\theta}_2)^2}{Q^2(\psi)(\tilde{\theta}_{p+1})^2} \right)^{1/2}, \quad (3.23)$$

dove

$$\tilde{\theta}_2 = \frac{\psi(c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2) + c_{12}\hat{\beta}_1 + c_{22}\hat{\beta}_2}{Q^{1/2}(\psi)} \quad (3.24)$$

è la s.m.v. di  $\theta_2$  con  $\psi$  fissato. Si noti che la (3.23) è una quantità che tende a 0 quando  $|\psi| \rightarrow \infty$  ed è una distribuzione propria.

Utilizzando la (3.14) e la (3.23), la pseudo-distribuzione a posteriori basata sulla funzione di verosimiglianza condizionata approssimata e la corrispondente *matching prior* è

$$\begin{aligned} \pi_{ps}(\psi|y) &\propto i_{\psi\psi}(\psi, \hat{\lambda}_\psi)^{1/2} L_{CA}(\psi) \\ &\propto \frac{|\tilde{\theta}_2|(\tilde{\theta}_{p+1})^{-[n-(p-1)]}}{Q(\psi)}. \end{aligned} \quad (3.25)$$

Si osservi che anche la (3.25) è una distribuzione a posteriori propria in quanto è una funzione positiva, tende a 0 se  $|\psi| \rightarrow \infty$  e l'integrale su tutto il supporto di  $\psi$  è finito.

La pseudo-distribuzione a posteriori ottenuta a partire dalla  $L_{CA}(\psi)$ , e quindi per l'ortogonalità dei parametri da  $L_{MP}(\psi)$ , può essere interpretata anche come una versione penalizzata della funzione di verosimiglianza profilo modificata (cfr. § 2.3.2).

Inoltre, la moda a posteriori della (3.25) coincide con la s.m.v. profilo modificata penalizzata.

Attraverso la distribuzione (3.25) è possibile applicare lo schema di simulazione HOTA (cfr. § 2.4.1) per ottenere intervalli di confidenza *HPD*, oppure alcuni momenti della distribuzione a posteriori (ad esempio, media e varianza a posteriori).

### **3.5 Intervalli di confidenza**

Il problema centrale per l'inferenza sul rapporto di due coefficienti di regressione nel modello con errori normali è che, per alcune pseudo-verosimiglianze, l'intervallo di confidenza per  $\psi$  basato sulla statistica del log-rapporto di pseudo-verosimiglianza (cfr. § 1.2) non è finito.

La statistica log-rapporto di verosimiglianza profilo (cfr. § 1.4.4) è data da

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

dove la s.m.v.  $\hat{\psi}$  di  $\psi$  è pari al rapporto tra le s.m.v. dei due coefficienti di regressione, ovvero

$$\hat{\psi} = \frac{\hat{\beta}_1}{\hat{\beta}_2}. \quad (3.26)$$

Per la (3.7) e la (3.8), la statistica  $W_p(\psi)$  assume l'espressione

$$W_p(\psi) = n \log \left\{ 1 + \frac{n|C|(\hat{\beta}_2\psi - \hat{\beta}_1)^2}{Q(\psi)SSE} \right\}. \quad (3.27)$$

A partire dalla (3.27) è possibile costruire un intervallo di confidenza di livello approssimato  $(1 - \alpha)$  come

$$\{\psi: W_p(\psi) \leq \chi_{1;1-\alpha}^2\} \quad (3.28)$$

o, in modo alternativo, utilizzando la statistica radice con segno profilo  $r_p(\psi) = \text{sgn}(\hat{\psi} - \psi)\sqrt{W_p(\psi)}$ , come

$$\{\psi: |r_p(\psi)| \leq z_{1-\alpha/2}\}. \quad (3.29)$$

Si noti che la statistica  $W_p(\psi)$  non tende a  $+\infty$  quando  $|\psi| \rightarrow \infty$ , e questo fatto può portare a intervalli di confidenza non finiti.

La statistica log-rapporto di verosimiglianza profilo (3.27) è una funzione monotona crescente di

$$F = \frac{n|C|(\hat{\beta}_2\psi - \hat{\beta}_1)^2 Q^{-1}(\psi)}{MSE}, \quad (3.30)$$

dove  $MSE = SSE/(n - p)$ . Ghosh *et al.* (2006) mostrano che la (3.30) si distribuisce come una  $F_{1,n-p}$ , e l'intervallo di confidenza di livello  $(1 - \alpha)$  esatto è dato da

$$\{\psi: F \leq F_{1,n-p;1-\alpha}\}, \quad (3.31)$$

dove  $F_{1,n-p;1-\alpha}$  è il quantile  $(1 - \alpha)$  della distribuzione  $F_{1,n-p}$ .

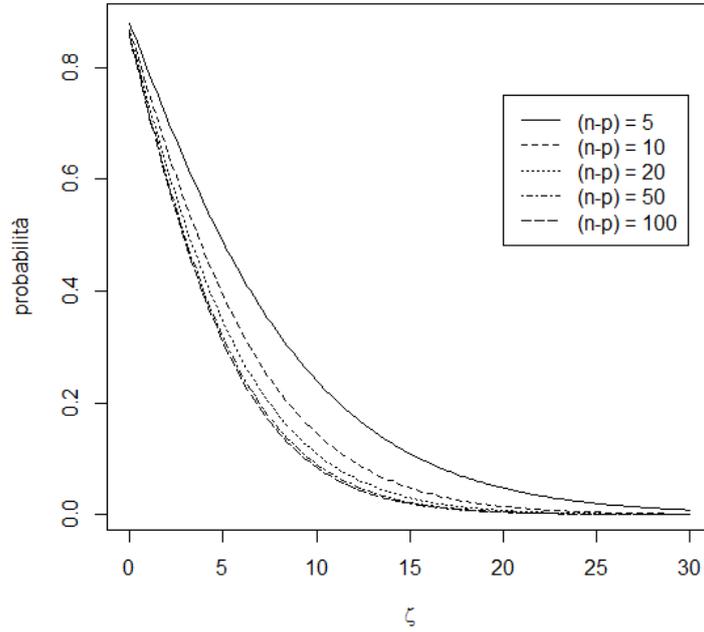
A causa della forma della funzione di verosimiglianza profilo, l'intervallo di confidenza (3.31) può risultare non finito. In Ghosh *et al.* (2006) viene discusso sotto quali condizioni questo può accadere, dimostrando inizialmente che  $\sup_{\psi} F = n\hat{\gamma}^T C\hat{\gamma}/MSE$ , con  $\hat{\gamma} = (\hat{\beta}_1, \hat{\beta}_2)^T$ . In base a questo risultato, l'intervallo di confidenza (3.31) diventa la retta reale se e solo se

$$\frac{n\hat{\gamma}^T C\hat{\gamma}}{MSE} \leq F_{1,n-p;1-\alpha} \quad (3.32)$$

Dato che  $n\hat{\gamma}^T C\hat{\gamma}/(2MSE)$  è distribuito come una variabile casuale  $F_{2,n-p}(\zeta)$ , con parametro di non centralità  $\zeta = n\gamma^T C\gamma/(2\sigma^2)$ , la probabilità che la (3.31) sia la retta reale è data da  $\mathbb{P}_{\psi}[F_{2,n-p}(\zeta) \leq F_{1,n-p;1-\alpha}/2]$ .

Ghosh *et al.* (2006) mostrano che la probabilità che l'intervallo di confidenza diventi la retta reale cresce al diminuire del parametro di non centralità. Quindi più il modello si adatta bene ai dati, riducendo l'errore quadratico medio, con maggiore probabilità l'intervallo di confidenza (3.31) è finito. Al contrario, la probabilità che l'intervallo esatto costruito con la (3.30) non sia finito diventando la retta reale diminuisce al crescere di  $(n - p)$  e di  $\alpha$ .

La Figura 1.3 illustra l'andamento della probabilità che l'intervallo di confidenza costruito mediante la statistica  $F$  sia pari alla retta reale, al variare di  $\zeta$  e di  $n - p$ , con  $(1 - \alpha) = 0.95$ . Si nota che, al crescere di  $\zeta$ , la probabilità tende a ridursi.



**Figura 1.3: Probabilità di ottenere un intervallo di confidenza di ampiezza pari alla retta reale al variare di  $\zeta$  e di  $n - p$ .**

Un risultato analogo si verifica per l'intervallo di confidenza costruito con la statistica log-rapporto di verosimiglianza condizionata approssimata. Tale statistica risulta

$$W_{CA}(\psi) = 2\{l_{CA}(\hat{\psi}) - l_{CA}(\psi)\} = (n - p)\log\left\{1 + \frac{n|C|(\hat{\beta}_2\psi - \hat{\beta}_1)^2}{Q(\psi)SSE}\right\} \quad (3.33)$$

e il corrispondente intervallo di confidenza di livello approssimato  $(1 - \alpha)$  è

$$\{\psi: W_{CA}(\psi) \leq \chi_{1;1-\alpha}^2\}. \quad (3.34)$$

L'intervallo di confidenza (3.34), come accade per gli intervalli (3.28) e (3.31), può non essere finito e questo avviene perché  $W_{CA}(\psi)$  non tende a  $+\infty$  quando  $|\psi| \rightarrow \infty$ .

Si può dimostrare che si ottengono intervalli di confidenza non finiti se e solo se  $(n - p)\log\left\{1 + n|C|(\hat{\beta}_2)^2/(SSEc_{11})\right\} \leq \chi_{1;1-\alpha}^2$ .

Contrariamente a quanto accade con le funzioni di verosimiglianza profilo e condizionata approssimata, gli intervalli di confidenza costruiti con la statistica log-rapporto di verosimiglianza integrata sono sempre finiti. Questo perché  $L_I^{CR}(\psi)$  e  $L_I^{OCR}(\psi)$  sono caratterizzate dalla presenza di un fattore moltiplicativo  $Q(\psi)$ , dipendente dal parametro d'interesse  $\psi$ , che tende all'infinito al crescere di  $\psi$ .

Presentando per sintesi solo l'intervallo di confidenza basato su  $L_I^{CR}(\psi)$  e definendo con  $\hat{\psi}_{CR}$  la s.m.v. integrata, ossia il valore di  $\psi$  che massimizza la funzione di verosimiglianza integrata costruita mediante la *reference prior* (3.17), l'intervallo di confidenza di livello approssimato  $(1 - \alpha)$  basato sulla statistica log-rapporto di verosimiglianza integrata è

$$\{\psi: W_I^{CR}(\psi) \leq \chi_{1;1-\alpha}^2\}, \quad (3.35)$$

dove  $W_I^{CR}(\psi) = 2\{l_I^{CR}(\hat{\psi}_{CR}) - l_I^{CR}(\psi)\}$  e  $l_I^{CR}(\psi) = \log L_I^{CR}(\psi)$ . Si noti che  $W_I^{CR}(\psi) \rightarrow +\infty$  se  $|\psi| \rightarrow \infty$ .

Si consideri ora la funzione di verosimiglianza profilo modificata penalizzata. Essa coincide con il nucleo della pseudo-distribuzione a posteriori (3.25) ed è data da

$$L_{MP}^*(\psi) = \frac{|\tilde{\theta}_2|(\tilde{\theta}_{p+1})^{-[n-(p-1)]}}{Q(\psi)}. \quad (3.36)$$

A partire da  $L_{MP}^*(\psi)$  è agevole ricavare un intervallo di confidenza basato sul log-rapporto di verosimiglianza profilo modificata penalizzata. Sia  $\hat{\psi}_{MP}^*$  la s.m.v. di  $L_{MP}^*(\psi)$ . La statistica log-rapporto di verosimiglianza profilo modificata penalizzata è

$$W_{MP}^*(\psi) = 2\{l_{MP}^*(\hat{\psi}_{MP}^*) - l_{MP}^*(\psi)\},$$

dove  $l_{MP}^*(\psi) = \log L_{MP}^*(\psi)$ .

Il corrispondente intervallo di confidenza di livello approssimato  $(1 - \alpha)$  risulta dato da

$$\{\psi: W_{MP}^*(\psi) \leq \chi_{1;1-\alpha}^2\}. \quad (3.37)$$

Si noti che  $W_{MP}^*(\psi) \rightarrow +\infty$  quando  $|\psi| \rightarrow \infty$ , e quindi la (3.37) fornisce intervalli di confidenza per  $\psi$  finiti. Inoltre, l'intervallo (3.37) fornisce anche un intervallo di credibilità *HPD* di ordine  $O(n^{-1})$ .

In alternativa, seguendo un'ottica bayesiana, a partire dalla pseudo-distribuzione a posteriori (3.25) si può costruire un intervallo di credibilità bilaterale per il parametro d'interesse  $\psi$ . Tale intervallo coincide con l'intervallo di confidenza frequentista basato sulla statistica  $r_p^*(\psi)$  (cfr. § 2.4).

### 3.6 Esempi

#### **Esempio 1.3: Dati simulati.**

Sia dato il seguente modello di regressione

$$y_i = \begin{cases} \beta_1 + \varepsilon_i, & i = 1, \dots, 5 \\ \beta_2 + \varepsilon_i, & i = 6, \dots, 10 \end{cases}$$

dove  $(\varepsilon_1, \dots, \varepsilon_{10})$  sono variabili aleatorie i.i.d.  $N(0,1)$ .

I dati sono stati generati da una variabile casuale  $N(10,1)$  per le prime cinque osservazioni e da una  $N(1,1)$  per le restanti cinque, in modo tale da avere  $\psi = 10$ .

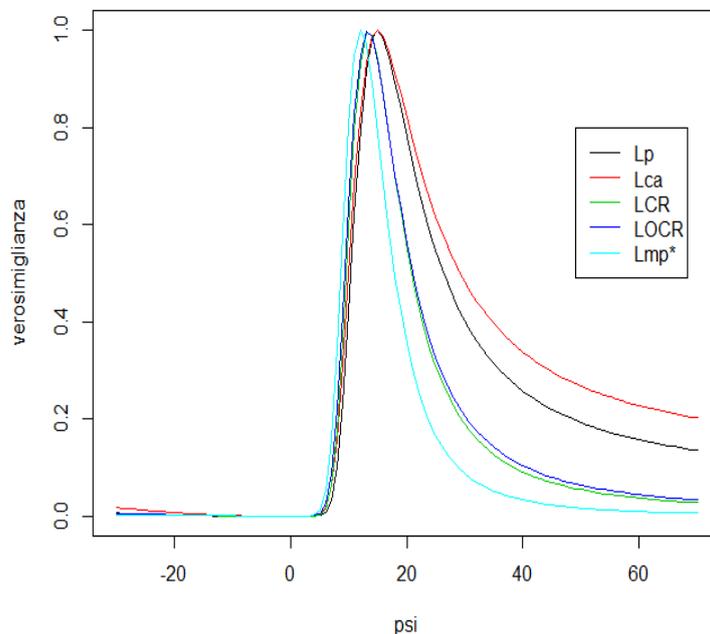
La matrice di disegno  $X$  è definita come

$$X = \begin{bmatrix} 1_5 & 0_5 \\ 0_5 & 1_5 \end{bmatrix}.$$

La s.m.v. per  $(\beta, \sigma)$  è

$$(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}) = (10.193, 0.683, 0.542),$$

e perciò  $\hat{\psi} = \hat{\beta}_1 / \hat{\beta}_2 = 14.920$ .



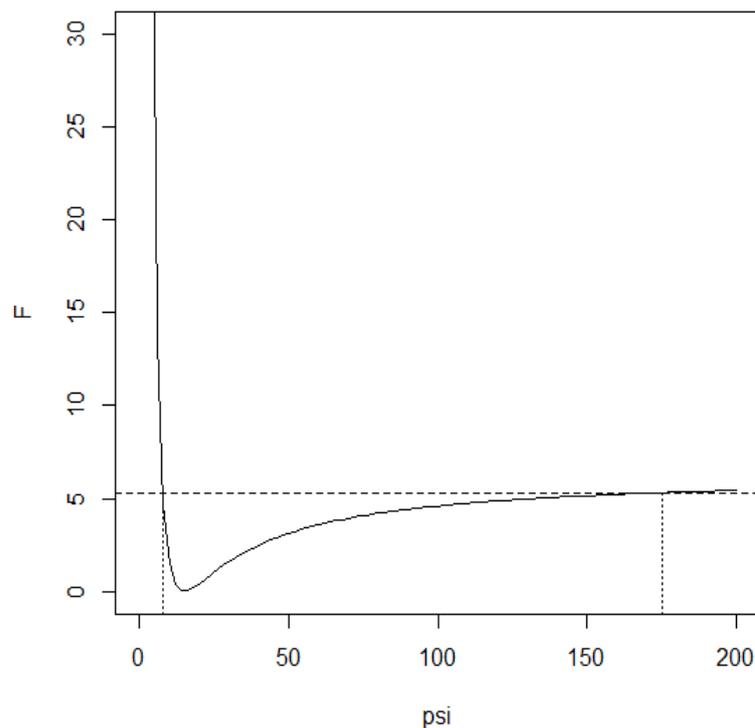
**Figura 2.3. Esempio 1.3: Confronto tra le diverse pseudo-verosimiglianze normalizzate per  $\psi$ .**

La Figura 2.3 mostra l'andamento delle diverse funzioni di pseudo-verosimiglianza normalizzate per  $\psi$ . Si può notare che le funzioni di verosimiglianza profilo e

condizionata approssimata, al contrario di  $L_I^{CR}(\psi)$ ,  $L_I^{OCR}(\psi)$  e  $L_{MP}^*(\psi)$ , non tendono a 0 quando  $|\psi| \rightarrow \infty$ . Il punto di minimo di  $l_P(\psi)$  e  $l_{CA}(\psi)$  si ha in corrispondenza di  $\psi_2 = -0.067$ , che in questo caso coincide con  $-1/\hat{\psi}$ .

Dalla Figura 2.3 si nota come il punto di massimo delle funzioni di verosimiglianza integrata e profilo modificata penalizzata sia inferiore rispetto alla s.m.v. ottenuta a partire da  $L_P(\psi)$ . Infatti, si ha  $(\hat{\psi}, \hat{\psi}_I^{CR}, \hat{\psi}_I^{OCR}, \hat{\psi}_{MP}^*) = (14.920, 13.391, 13.252, 12.067)$ .

Il  $\sup_{\psi} F$  è pari a 1421 ed è, quindi maggiore di  $F_{1,8;0.95} = 5.318$ . Questo implica che l'intervallo di confidenza per  $\psi$  costruito con la statistica  $F$  risulta finito. Tuttavia, la forma della funzione di verosimiglianza profilo porta ad avere un intervallo di confidenza molto ampio.



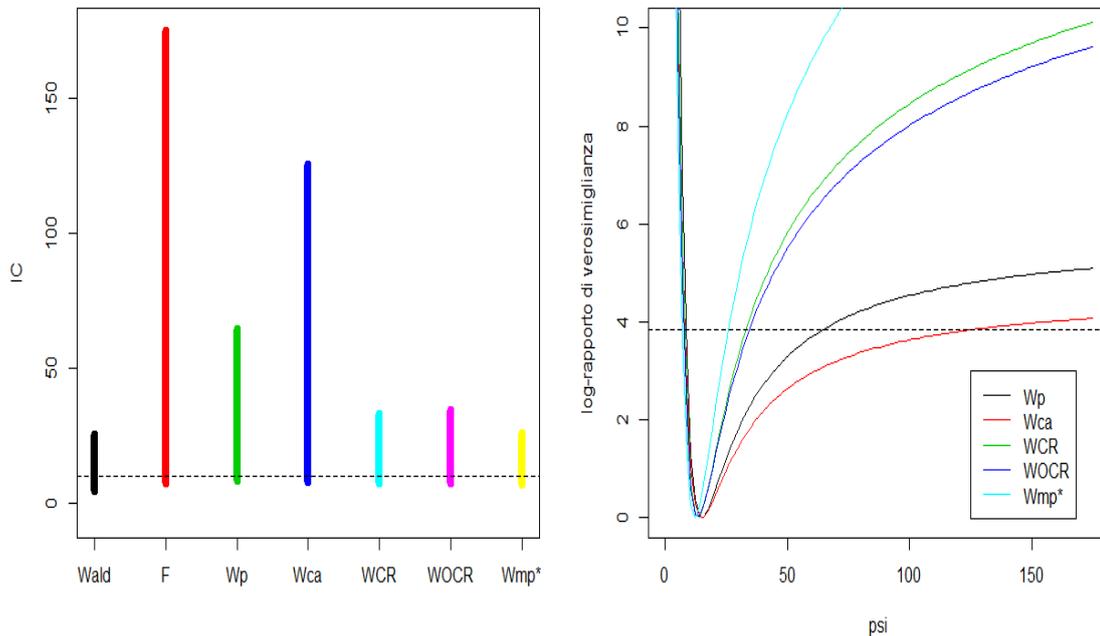
**Figura 3.3. Esempio 1.3: Statistica  $F$ .**

Dalla Figura 3.3 si può vedere che l'intervallo di confidenza di livello esatto  $1 - \alpha = 0.95$  costruito tramite la statistica  $F$  è molto ampio ed è pari a  $(7.761, 175.237)$ , con un'ampiezza di 167.476.

$\alpha$	0.01		0.05		0.1	
	Intervallo (ampiezza)		Intervallo (ampiezza)		Intervallo (ampiezza)	
$Wald$	1.253	28.588 (27.336)	4.520	25.320 (20.800)	6.192	23.648 (17.456)
$F$			7.761	175.237 (167.476)	8.561	56.912 (48.351)
$W_P$			8.414	64.318 (55.904)	9.158	39.857 (30.698)
$W_{CA}$			7.903	125.358 (117.455)	8.692	51.797 (43.105)
$W_I^{CR}$	6.564	60.595 (54.032)	7.800	33.053 (25.253)	8.490	26.682 (18.192)
$W_I^{OCR}$	6.243	66.714 (60.471)	7.515	34.472 (26.957)	8.223	27.390 (19.167)
$W_{MP}^*$	5.689	38.781 (33.092)	6.917	25.848 (18.931)	7.590	21.972 (14.382)

Tabella 1.3. Esempio 1.3: Intervalli di confidenza per  $\psi$  con  $\alpha = 0.01, 0.05$  e  $0.10$  e  $n = 10$ .

La Tabella 1.3 riporta gli intervalli di confidenza costruiti a partire dalle diverse funzioni di pseudo-verosimiglianza per  $\psi$ . Si noti che per  $\alpha = 0.01$  gli intervalli costruiti mediante le statistiche  $F$ ,  $W_P(\psi)$  e  $W_{CA}(\psi)$  non hanno estremo superiore, risultando perciò non finiti. Inoltre, l'intervallo di confidenza costruito mediante il log-rapporto di verosimiglianza profilo modificata penalizzata è quello meno ampio per ogni livello di confidenza.

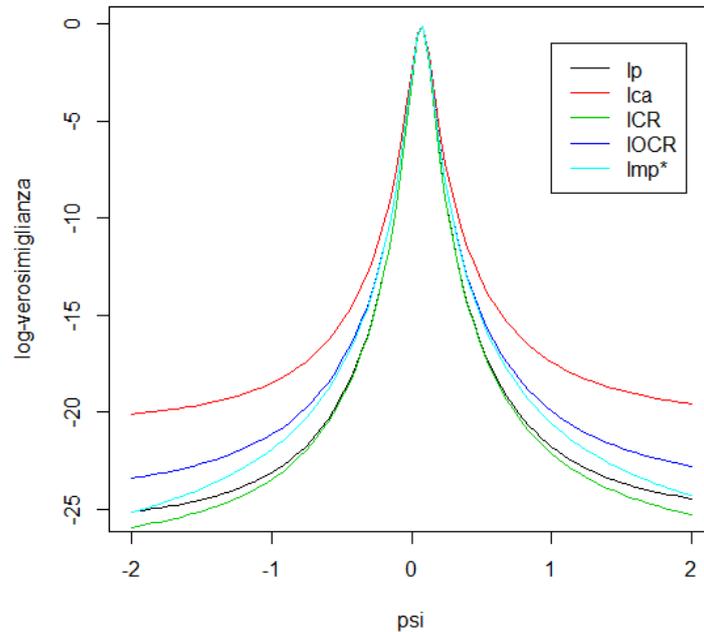


**Figura 4.3. Esempio 1.3: Confronto tra i diversi intervalli di confidenza con livello  $1 - \alpha = 0.95$  (sinistra), andamento delle statistiche log-rapporto di pseudo-verosimiglianza (destra).**

Nel grafico a sinistra della Figura 4.3 sono stati riportati gli IC per  $\psi$  di livello  $1 - \alpha = 0.95$  e  $n = 10$ . Si noti che tutti comprendono il vero valore del parametro (linea orizzontale tratteggiata) e che l'intervallo costruito tramite la statistica  $F$  è quello più ampio. Il grafico a destra nella Figura 4.3 illustra l'andamento delle statistiche log-rapporto di pseudo-verosimiglianza, dove la linea orizzontale tratteggiata indica il quantile  $\chi_{1;0.95}^2$ . L'intervallo costruito con la funzione di verosimiglianza profilo modificata penalizzata sia il più corto. Tale risultato è discusso in Ventura e Racugno (2012) dove, mediante uno studio di simulazione sul rapporto di medie in un modello normale, si è riscontrato che l'ampiezza media dell'intervallo ottenuto mediante  $W_{MP}^*(\psi)$  è inferiore rispetto all'ampiezza media dell'intervallo di confidenza conseguito mediante la statistica log-rapporto di verosimiglianza condizionata approssimata (cfr. § 2.3.2). Questa conclusione è conseguenza di un risultato teorico generale presentato in Ventura e Racugno (2012), secondo il quale l'ampiezza media degli intervalli di confidenza costruiti mediante la statistica log-rapporto di verosimiglianza profilo modificata penalizzata sono inferiori a quelli costruiti attraverso la statistica log-rapporto di verosimiglianza profilo modificata.

Si consideri ora il caso, con gli stessi dati simulati, in cui il parametro d'interesse sia  $\psi = \beta_2/\beta_1$ , al fine di ottenere  $\psi = 0.10$ . In questo caso  $|\psi| < 1$ .

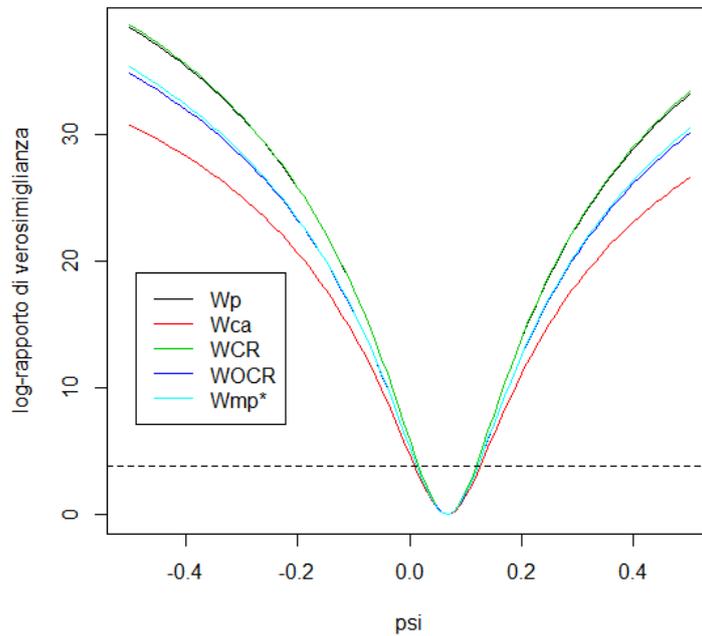
Le s.m.v. per i coefficienti di regressione e per la varianza dell'errore rimangono uguali, mentre  $\hat{\psi} = \hat{\beta}_2/\hat{\beta}_1 = 0.067$ .



**Figura 5.3. Esempio 1.3: Funzioni di pseudo-verosimiglianza normalizzate per  $\psi = \beta_2/\beta_1$ .**

Nella Figura 5.3 vengono confrontate le diverse funzioni di pseudo-verosimiglianza per  $\psi$ . Il valore di  $\psi$  che massimizza le differenti funzioni è molto simile; infatti  $(\hat{\psi}, \hat{\psi}_I^{CR}, \hat{\psi}_I^{OCR}, \hat{\psi}_{MP}^*) = (0.0670, 0.0670, 0.0670, 0.0669)$ .

Nella Figura 6.3 si nota che l'intervallo di confidenza costruito con le statistiche log-rapporto di verosimiglianza profilo e condizionata approssimata risulta finito sia a livello approssimato  $1 - \alpha = 0.95$  (linea tratteggiata orizzontale), sia a livello approssimato  $1 - \alpha = 0.99$ .



**Figura 6.3. Esempio 1.3: Statistiche log-rapporto di pseudo-verosimiglianza.**

La forma delle diverse curve, concentrate nelle s.m.v., porta ad avere intervalli di confidenza con un'ampiezza limitata, contrariamente a quanto ottenuto nel caso precedente, dove gli intervalli costruiti con le statistiche  $W_P(\psi)$  e  $W_{CA}(\psi)$  presentavano ampiezze elevate.

Come per le statistiche log-rapporto di pseudo-verosimiglianza, anche la statistica  $F$  mostra un andamento regolare intorno alla s.m.v.. La Figura 7.3 presenta l'andamento della statistica  $F$  e riporta un intervallo di confidenza di livello esatto  $1 - \alpha = 0.95$  pari a  $(0.0057, 0.1288)$ .

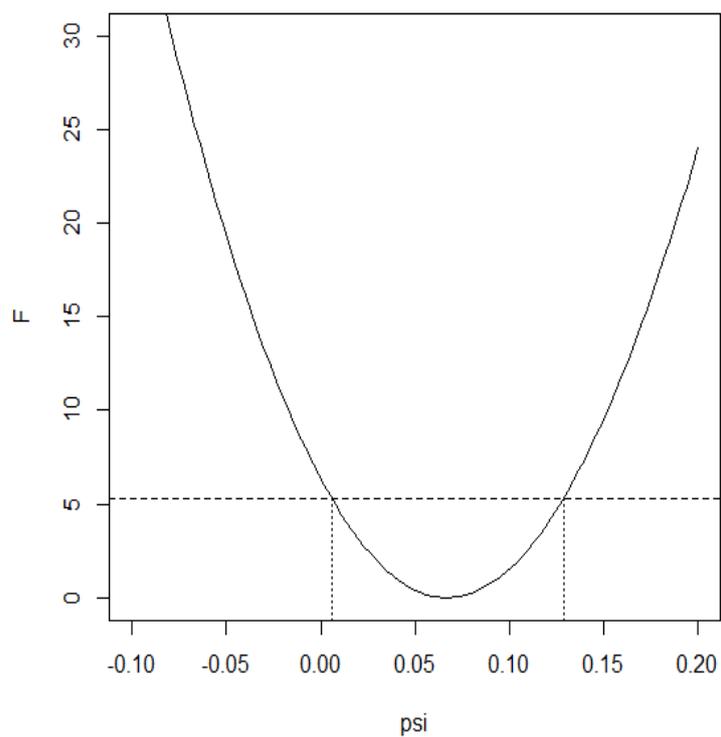


Figura 7.3: Esempio 1.3. Statistica  $F$ .

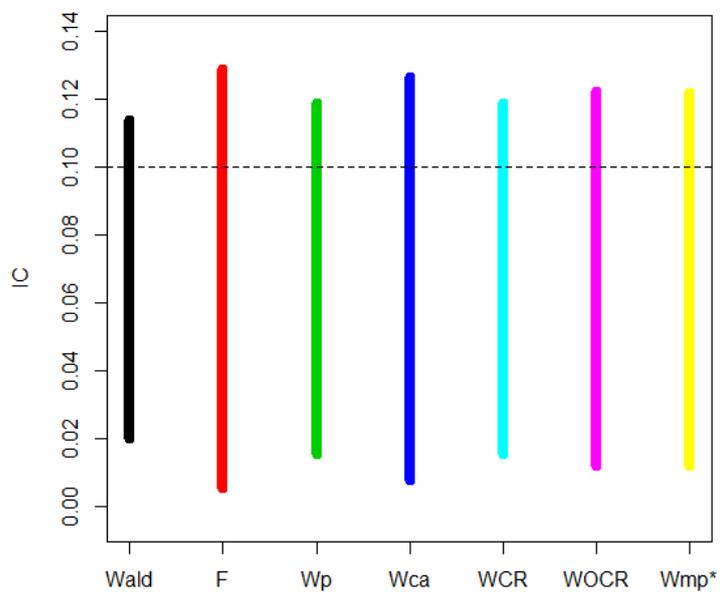


Figura 8.3. Esempio 1.3: Confronto tra i diversi intervalli di confidenza con livello  $1 - \alpha = 0.95$ .

La Figura 8.3 confronta i diversi intervalli di confidenza di livello  $1 - \alpha = 0.95$  costruiti con le statistiche di *Wald*, *F* e log-rapporto di pseudo-verosimiglianza, dove è stata aggiunta la retta tratteggiata che indica il vero valore del parametro.

$\alpha$	0.01		0.05		0.1	
	Intervallo (ampiezza)		Intervallo (ampiezza)		Intervallo (ampiezza)	
<i>Wald</i>	0.006	0.128 (0.122)	0.020	0.114 (0.094)	0.028	0.106 (0.078)
<i>F</i>	-0.022	0.157 (0.179)	0.006	0.129 (0.123)	0.018	0.117 (0.099)
$W_P$	-0.006	0.141 (0.147)	0.016	0.119 (0.103)	0.025	0.109 (0.084)
$W_{CA}$	-0.018	0.153 (0.172)	0.008	0.127 (0.119)	0.019	0.115 (0.096)
$W_I^{CR}$	-0.006	0.141 (0.147)	0.016	0.119 (0.103)	0.025	0.109 (0.084)
$W_I^{OCR}$	-0.012	0.146 (0.158)	0.012	0.122 (0.110)	0.022	0.112 (0.089)
$W_{MP}^*$	-0.012	0.146 (0.158)	0.012	0.122 (0.110)	0.022	0.112 (0.089)

Tabella 2.3. Esempio 1.3: Intervalli di confidenza per  $\psi$  con  $\alpha = 0.01, 0.05$  e  $0.10$  e  $n = 10$ .

Si noti che tutti gli intervalli riportati sulla Tabella 2.3 comprendono il vero valore del parametro. ■

### Esempio 2.3: Dati reali.

Per un'illustrazione basata su dati reali, viene presentato un esempio tratto da Finney (1978). Si consideri un esperimento in cui  $k$  dosi ( $x_{11}, \dots, x_{1k}$ ) di un farmaco standard  $S$  sono analizzate  $m$  volte e  $q$  dosi ( $x_{21}, \dots, x_{2q}$ ) di un farmaco test  $T$  vengono analizzate  $u$  volte. Si ottiene in questo modo un campione  $\{y_{1ih}, i = 1, \dots, k; h = 1, \dots, m$  e  $y_{2jh}, j = 1, \dots, q; h = 1, \dots, u\}$  di numerosità  $n = km + qu$ .

Il modello assume la forma

$$\begin{aligned}
 y_{1ih} &= \beta_1 + \beta_2 x_{1i} + \varepsilon_{1ih}, & i = 1, \dots, k; h = 1, \dots, m \\
 y_{2jh} &= \beta_1 + \beta_2 x_{2j} + \beta_3 + \varepsilon_{2jh}, & j = 1, \dots, q; h = 1, \dots, u
 \end{aligned}
 \tag{3.38}$$

o in modo equivalente

$$y_{1ih} = \beta_1 + \beta_2 x_{1i} + \varepsilon_{1ih}, \quad i = 1, \dots, k; h = 1, \dots, m$$

$$y_{2jh} = \beta_1 + \beta_2(x_{2j} + \psi) + \varepsilon_{2jh}, \quad j = 1, \dots, q; h = 1, \dots, u. \quad (3.39)$$

Il parametro d'interesse è  $\psi = \beta_3/\beta_2$  e  $(\varepsilon_{111}, \dots, \varepsilon_{1km}, \varepsilon_{211}, \dots, \varepsilon_{2qu})$  sono variabili casuali i.i.d.  $N(0, \sigma^2)$ .

La matrice di disegno  $X$  è

$$X^T = \begin{bmatrix} 1 & \dots & 1 & \dots & 1 & \dots & 1 & 1 & \dots & 1 & \dots & 1 & \dots & 1 \\ x_{11} & \dots & x_{11} & \dots & x_{1k} & \dots & x_{1k} & x_{21} & \dots & x_{21} & \dots & x_{2q} & \dots & x_{2q} \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 & 1 & \dots & 1 & \dots & 1 & \dots & 1 \end{bmatrix}. \quad (3.40)$$

Si assume, inoltre, che  $k = 3, q = 3, m = 6, u = 6$  e  $n = 36$ .

Farmaco	Standard			Test		
Dose	-1.00	0	1.00	-1.00	0	1.00
	0.96	1.06	1.17	0.91	1.09	1.15
	0.91	1.07	1.14	0.93	1.04	1.15
	0.92	0.99	1.14	0.98	0.97	1.14
	0.76	0.86	1.13	0.96	1.06	1.16
	1.03	1.06	1.13	0.89	1.04	1.10
	0.93	1.02	1.15	1.01	1.02	1.15

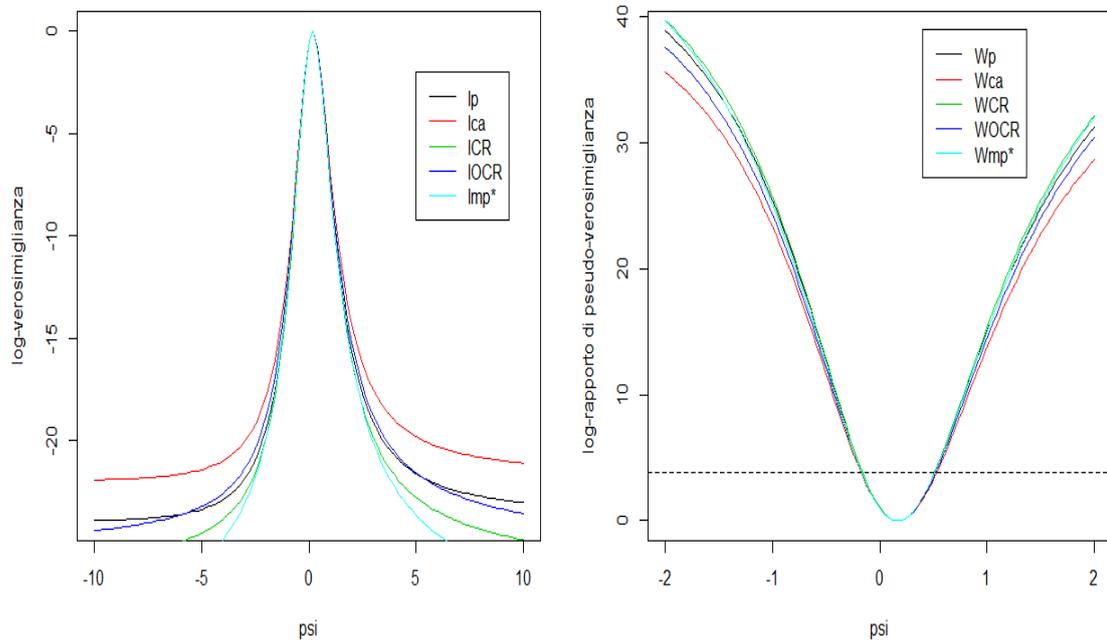
**Tabella 3.3. Esempio 2.3: Dati di Finney (1978).**

I dati riportati nella Tabella 3.3, originariamente analizzati da Finney (1978), si riferiscono a misurazioni torbidimetriche sulla risposta di crescita del *Lactobacillus leichmannii* nella vitamina  $B_{12}$ .

La s.m.v. per  $(\beta, \sigma)$  è

$$(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\sigma}) = (1.024, 0.105, 0.018, 0.055),$$

per cui  $\hat{\psi} = \hat{\beta}_3/\hat{\beta}_2 = 0.169$ .



**Figura 9.3. Esempio 2.3: Confronto tra le diverse funzioni di log-pseudo-verosimiglianza normalizzata (sinistra) e tra le statistiche di log-rapporto di pseudo-verosimiglianza (destra) per i dati di Finney (1978) con  $n = 36$ .**

La Figura 9.3 mostra l'andamento delle diverse funzioni di log-pseudo-verosimiglianza (sinistra) e delle differenti statistiche log-rapporto di pseudo-verosimiglianza (destra). Si noti che tutte le curve hanno un andamento simile in un intorno della propria s.m.v.. Le s.m.v. di  $\psi$ , ottenute massimizzando le diverse funzioni di pseudo-verosimiglianza, sono  $(\hat{\psi}, \hat{\psi}_I^{CR}, \hat{\psi}_I^{OCR}, \hat{\psi}_{MP}^*) = (0.169, 0.168, 0.168, 0.166)$ .

Dal grafico a destra nella Figura 9.3 si può osservare che tutti gli intervalli di livello approssimato  $1 - \alpha = 0.95$ , costruiti con le statistiche log-rapporto di pseudo-verosimiglianza, risultano finiti.

Nella Tabella 4.3 sono riportati gli intervalli di confidenza costruiti a partire dalle statistiche *Wald*, *F* e log-rapporto di pseudo-verosimiglianza.

Si può osservare che gli intervalli risultano abbastanza simili e concentrati attorno alla propria stima.

$\alpha$	0.01		0.05		0.1	
	Intervallo (ampiezza)		Intervallo (ampiezza)		Intervallo (ampiezza)	
$Wald$	-0.255	0.594	-0.154	0.492	-0.102	0.441
	(0.849)		(0.646)		(0.542)	
$F$	-0.307	0.676	-0.181	0.536	-0.121	0.471
	(0.983)		(0.717)		(0.592)	
$W_P$	-0.280	0.645	-0.163	0.516	-0.106	0.455
	(0.925)		(0.678)		(0.561)	
$W_{CA}$	-0.303	0.671	-0.178	0.533	-0.119	0.469
	(0.974)		(0.711)		(0.588)	
$W_I^{CR}$	-0.279	0.640	-0.163	0.511	-0.107	0.451
	(0.919)		(0.674)		(0.558)	
$W_I^{OCR}$	-0.294	0.656	-0.173	0.522	-0.115	0.460
	(0.950)		(0.695)		(0.575)	
$W_{MP}^*$	-0.290	0.646	-0.171	0.516	-0.114	0.454
	(0.936)		(0.687)		(0.568)	

Tabella 4.3. Esempio 2.3: Intervalli di confidenza per  $\psi$  con  $\alpha = 0.01, 0.05$  e  $0.10$  e  $n = 36$ .

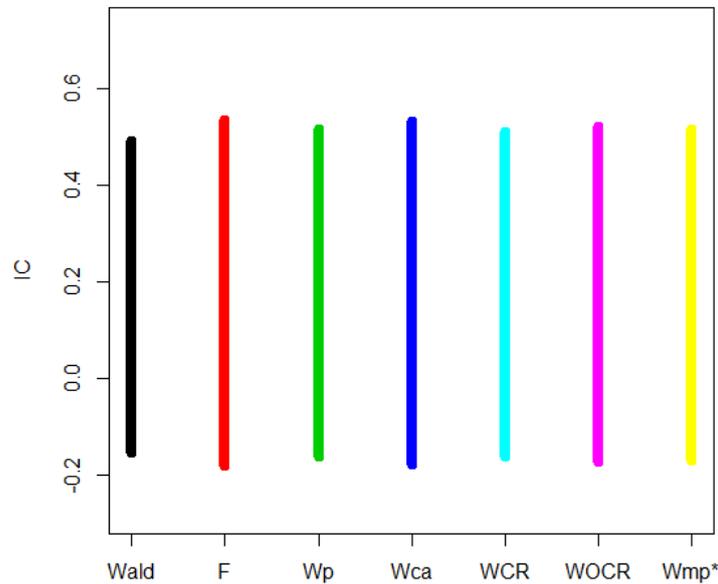


Figura 10.3. Esempio 2.3: Ampiezza degli intervalli di confidenza con livello  $1 - \alpha = 0.95$  e  $n = 36$ .

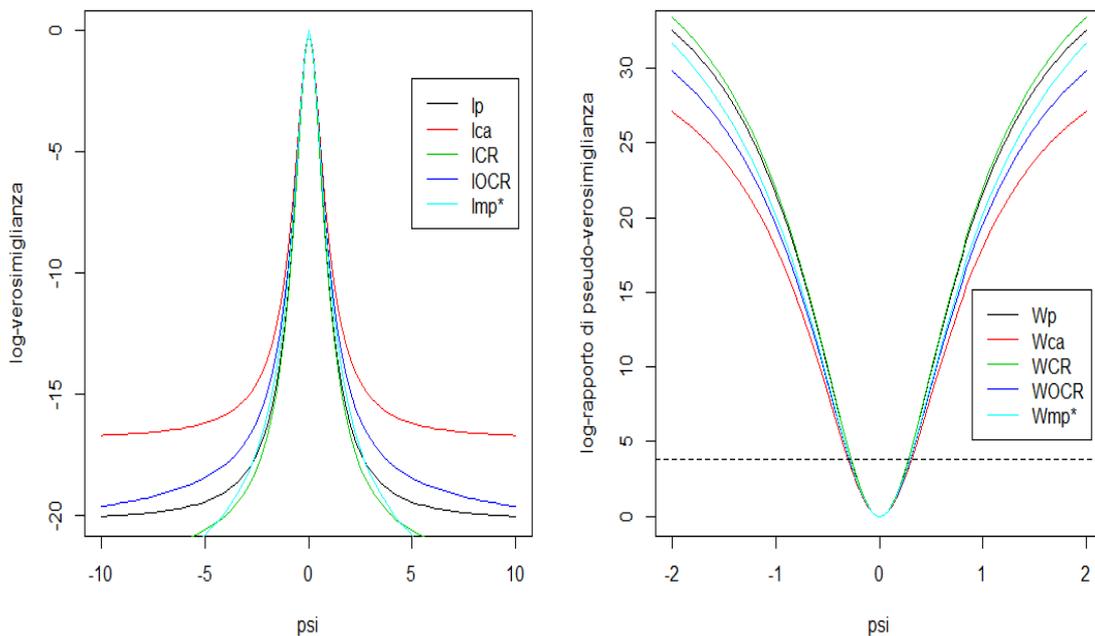
Nella Figura 10.3 sono stati riportati gli intervalli di confidenza di livello  $1 - \alpha = 0.95$  con  $n = 36$ . Si può notare che l'ampiezza dei vari intervalli è molto simile.

Si consideri ora lo stesso modello (3.39) dove però vengono analizzate solo le prime 18 osservazioni, 9 per il farmaco standard e 9 per il farmaco test.

La s.m.v. per  $(\beta, \sigma)$  è

$$(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\sigma}) = (1.040, 0.107, 0, 0.030),$$

per cui  $\hat{\psi} = \hat{\beta}_3 / \hat{\beta}_2 = 0$ . Si noti che con questi dati la s.m.v. per  $\beta_3$  è circa 0.



**Figura 11.3. Esempio 2.3: Confronto tra le diverse funzioni di log-pseudo-verosimiglianza normalizzata (sinistra) e tra le statistiche di log-rapporto di pseudo-verosimiglianza (destra) per i dati di Finney (1978) con  $n = 18$ .**

La Figura 11.3 mostra l'andamento delle funzioni di log-pseudo-verosimiglianza (sinistra) e delle funzioni di log-rapporto di pseudo-verosimiglianza (destra) quando  $n = 18$ . Anche in questo caso l'andamento delle diverse curve è abbastanza simile vicino alla propria s.m.v.; si noti, poi, che le diverse stime sono molto simili, infatti  $\hat{\psi}, \hat{\psi}_I^{CR}, \hat{\psi}_I^{OCR}$  e  $\hat{\psi}_{MP}^*$  assumono un valore molto vicino a 0.

Nel grafico a destra della Figura 11.3 è stata aggiunta una linea orizzontale tratteggiata che identifica il quantile  $\chi_{1;0.95}^2$ .

$\alpha$	0.01		0.05		0.1	
	Intervallo (ampiezza)		Intervallo (ampiezza)		Intervallo (ampiezza)	
$Wald$	-0.340	0.340	-0.259	0.259	-0.217	0.217
	(0.680)		(0.518)		(0.434)	
$F$	-0.441	0.442	-0.314	0.314	-0.257	0.257
	(0.883)		(0.628)		(0.513)	
$W_P$	-0.384	0.384	-0.277	0.277	-0.228	0.228
	(0.768)		(0.554)		(0.456)	
$W_{CA}$	-0.432	0.432	-0.308	0.308	-0.252	0.252
	(0.864)		(0.616)		(0.504)	
$W_I^{CR}$	-0.382	0.382	-0.276	0.276	-0.227	0.227
	(0.765)		(0.552)		(0.454)	
$W_I^{OCR}$	-0.412	0.412	-0.295	0.295	-0.242	0.242
	(0.824)		(0.590)		(0.484)	
$W_{MP}^*$	-0.407	0.407	-0.292	0.292	-0.240	0.240
	(0.814)		(0.585)		(0.480)	

**Tabella 5.3. Esempio 2.3: Intervalli di confidenza per  $\psi$  con  $\alpha = 0.01, 0.05$  e  $0.10$  e  $n = 18$ .**

Nella Tabella 5.3 sono riportati gli intervalli di confidenza costruiti a partire dalle diverse statistiche con  $n = 18$ . Si noti che tutti gli intervalli sono abbastanza simmetrici intorno allo 0.

■

### **3.7 Studi di simulazione**

Per analizzare il comportamento delle stime di massima pseudo-verosimiglianza e degli intervalli di confidenza costruiti con la statistica  $F$  e le statistiche log-rapporto di pseudo-verosimiglianza, vengono presentati alcuni studi di simulazione per valutare la distorsione e l'errore quadratico medio delle stime, le probabilità che gli intervalli di confidenza costruiti con le statistiche  $F$ ,  $W_P(\psi)$  e  $W_{CA}(\psi)$  siano finiti, l'ampiezza media e la copertura degli intervalli.

### 3.7.1 Primo studio di simulazione.

Sia dato il seguente modello

$$y_i = \beta_1 + \varepsilon_i, \quad i = 1, \dots, m \tag{3.41}$$

$$y_i = \beta_2 + \varepsilon_i, \quad i = m + 1, \dots, n,$$

dove  $(\varepsilon_1, \dots, \varepsilon_n)$  sono variabili casuali i.i.d.  $N(0, \sigma^2)$ .

In questa simulazione si è utilizzato un disegno bilanciato, tale per cui  $n = 2m$ .

La matrice di disegno  $X$  è definita come

$$X = \begin{bmatrix} \mathbf{1}_m & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{1}_m \end{bmatrix}$$

e le s.m.v. sono  $(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}) = (\bar{y}_1, \bar{y}_2, \sqrt{SSE/n})$ , dove  $\bar{y}_1 = \sum_{i=1}^m y_i$ ,  $\bar{y}_2 = \sum_{i=m+1}^n y_i$  e  $SSE = \sum_{i=1}^m (y_i - \bar{y}_1)^2 + \sum_{i=m+1}^n (y_i - \bar{y}_2)^2$ .

Per la simulazione si pone  $m = 5, 10, 25$  e  $\sigma = 1$ .

Le Tabelle 6.3 e 7.3 presentano i risultati di uno studio di simulazione sulla distorsione e l'errore quadratico medio (MSE) delle stime ottenute massimizzando le diverse pseudo-verosimiglianze, con  $\psi$  pari a 1 e 5, basato su 10000 repliche Monte Carlo. È stata analizzata per ogni funzione di pseudo-verosimiglianza la stima di massima pseudo-verosimiglianza, ovvero il valore di  $\psi$  che massimizza la funzione di pseudo-verosimiglianza (nello specifico  $\hat{\psi}$ ,  $\hat{\psi}_I^{CR}$ ,  $\hat{\psi}_I^{OCR}$  e  $\hat{\psi}_{MP}^*$ ).

	$(\beta_1, \beta_2)$	$m$		$\hat{\psi}$	$\hat{\psi}_I^{CR}$	$\hat{\psi}_I^{OCR}$	$\hat{\psi}_{MP}^*$
$\psi = 1$	(1,1)	5	Bias	0.786	-0.085	-0.103	-0.209
			(MSE)	(>1000)	(0.524)	(0.490)	(0.344)
		10	Bias	0.176	-0.009	-0.015	-0.098
			(MSE)	(2.730)	(0.225)	(0.219)	(0.163)
		25	Bias	0.045	-0.003	-0.004	-0.043
			(MSE)	(0.125)	(0.087)	(0.087)	(0.074)
	(5,5)	5	Bias	0.009	0.002	0.002	-0.006
			(MSE)	(0.016)	(0.016)	(0.016)	(0.015)
		10	Bias	0.003	-0.001	-0.001	-0.005
			(MSE)	(0.008)	(0.008)	(0.008)	(0.008)
		25	Bias	0.002	<0.001	<0.001	-0.002
			(MSE)	(0.003)	(0.003)	(0.003)	(0.003)

Tabella 6.3: Distorsione e MSE per le stime di pseudo-verosimiglianza con  $\psi = 1$  e  $m = 5, 10, 25$ .

	$(\beta_1, \beta_2)$	$m$	$\hat{\psi}$	$\hat{\psi}_I^{CR}$	$\hat{\psi}_I^{OCR}$	$\hat{\psi}_{MP}^*$	
$\psi = 5$	(5,1)	5	Bias	1.194	-0.301	-0.383	-0.980
			(MSE)	(>1000)	(7.582)	(6.991)	(4.561)
		10	Bias	0.633	-0.049	-0.074	-0.491
			(MSE)	(195.065)	(3.053)	(2.951)	(1.971)
		25	Bias	0.231	-0.002	-0.006	-0.200
			(MSE)	(1.651)	(1.141)	(1.135)	(0.922)
	(25,5)	5	Bias	0.043	0.010	0.006	-0.030
			(MSE)	(0.226)	(0.215)	(0.214)	(0.207)
		10	Bias	0.017	-0.001	-0.002	-0.021
			(MSE)	(0.104)	(0.102)	(0.102)	(0.100)
		25	Bias	0.007	-0.001	-0.001	-0.009
			(MSE)	(0.041)	(0.041)	(0.041)	(0.041)

**Tabella 7.3: Distorsione e MSE per le stime di pseudo-verosimiglianza con  $\psi = 5$  e  $m = 5, 10, 25$ .**

Come si può osservare dalle Tabelle 6.3 e 7.3, la distorsione e l'MSE differiscono in base al valore di  $\psi$ , secondo il valore assunto dai coefficienti di regressione e dalla numerosità campionaria. Si noti, infatti, che la distorsione e l'MSE sono minori quando il vero valore del parametro è pari a  $\psi = 1$ , mentre tendono a essere più grandi quando  $\psi$  assume valori maggiori di 1, a parità di numerosità campionaria. Inoltre, le due quantità dipendono sia dalla numerosità campionaria sia al valore che assumono i coefficienti di regressione, in particolare  $\beta_2$ . Al diminuire di quest'ultimo, le stime mostrano maggiore variabilità e si allontanano maggiormente dal vero valore, a parità di numerosità campionaria.

Dalle tabelle si può anche osservare che il minor valore dell'MSE viene sempre presentato dalla s.m.v. profilo modificata penalizzata; questa conclusione conferma il risultato teorico presentato da Mukerjee e Reid (1999) e da Ventura e Racugno (2012), secondo il quale l'MSE di  $\hat{\psi}_{MP}^*$ , ottenuto massimizzando in  $\psi$  la funzione di log-verosimiglianza profilo modificata penalizzata, risulta inferiore rispetto all'MSE di  $\hat{\psi}_{MP}$  che in questo modello coincide con  $\hat{\psi}$ . Inoltre, si noti che l'MSE della s.m.v. profilo tende ad essere simile a quello delle stime di  $L_I^{CR}(\psi)$ ,  $L_I^{OCR}(\psi)$  e  $L_{MP}^*(\psi)$ , solo quando la numerosità campionaria è elevata.

Dalle Tabelle 6.3 e 7.3 si nota che  $\hat{\psi}$  tende ad avere quasi sempre la distorsione maggiore, mentre  $\hat{\psi}_I^{CR}$  e  $\hat{\psi}_I^{OCR}$  hanno distorsione minore anche per  $m$  piccolo rispetto a  $\hat{\psi}_{MP}^*$ .

La Tabella 8.3 presenta le probabilità stimate di osservare un intervallo di confidenza di livello  $1 - \alpha = 0.95$  finito, basato sulle statistiche  $F$ ,  $W_P(\psi)$  e  $W_{CA}(\psi)$ , per differenti numerosità campionarie e valori di  $\psi$ . I risultati sono stati ottenuti attraverso uno studio di simulazione basato su 10000 replicazioni Monte Carlo. Si noti che la probabilità stimata di osservare un intervallo di confidenza finito tende a crescere all'aumentare della numerosità campionaria e che, a parità di numerosità campionaria, la maggiore probabilità stimata viene sempre mostrata dalla statistica  $W_P(\psi)$ .

			$m = 5$	$m = 10$	$m = 15$	$m = 20$	$m = 25$
$\psi = 1$	$(1, 1)$	$F$	0.4864	0.8412	0.9605	0.9922	0.9987
		$W_P$	0.6187	0.8737	0.9666	0.9935	0.9990
		$W_{CA}$	0.5199	0.8505	0.9613	0.9926	0.9987
$\psi = 5$	$(5, 1)$	$F$	0.4972	0.8456	0.9574	0.9910	0.9982
		$W_P$	0.6217	0.8768	0.9648	0.9925	0.9984
		$W_{CA}$	0.5249	0.8535	0.9589	0.9913	0.9983

**Tabella 8.3: Probabilità stimata che gli IC di livello  $1 - \alpha = 0.95$  costruiti con le statistiche  $F$ ,  $W_P(\psi)$  e  $W_{CA}(\psi)$  risultino finiti.**

La Tabella 9.3 mostra i risultati di uno studio di simulazione sull'ampiezza e la copertura degli intervalli di confidenza costruiti con la statistica  $F$  e le statistiche log-rapporto di pseudo-verosimiglianza. Gli intervalli hanno livello  $1 - \alpha = 0.95$ , per due diversi valori di  $\psi$ . I risultati ottenuti si basano su 10000 replicazioni Monte Carlo.

$\psi$	$(\beta_1, \beta_2)$	$m$		$F$	$W_P$	$W_{CA}$	$W_I^{CR}$	$W_I^{OCR}$	$W_{MP}^*$	
1	(1, 1)	5	Ampiezza media (copertura)	13.743 (93.8)	17.935 (90.6)	13.994 (93.1)	5.809 (91.8)	6.047 (93.1)	2.779 (91.0)	
		10	Ampiezza media (copertura)	13.378 (95.9)	9.458 (94.4)	10.943 (95.6)	2.949 (93.8)	3.033 (94.5)	1.967 (93.3)	
		25	Ampiezza media (copertura)	1.704 (95.1)	1.625 (94.6)	1.724 (95.0)	1.322 (94.5)	1.338 (94.8)	1.188 (94.2)	
	(5, 5)	5	Ampiezza media (copertura)	0.596 (94.4)	0.494 (90.5)	0.572 (93.6)	0.486 (90.7)	0.519 (92.1)	0.509 (92.1)	
		10	Ampiezza media (copertura)	0.379 (95.0)	0.351 (93.3)	0.373 (94.7)	0.349 (93.2)	0.359 (94.1)	0.356 (94.0)	
		25	Ampiezza media (copertura)	0.228 (95.2)	0.222 (94.5)	0.227 (95.0)	0.222 (94.5)	0.224 (94.8)	0.223 (94.9)	
	5	(5, 1)	5	Ampiezza media (copertura)	171.029 (95.2)	90.155 (92.6)	66.602 (94.7)	22.054 (91.3)	22.571 (92.7)	9.422 (89.1)
			10	Ampiezza media (copertura)	35.428 (97.0)	42.355 (95.7)	37.058 (96.8)	11.414 (93.7)	11.677 (94.4)	6.937 (92.4)
			25	Ampiezza media (copertura)	6.421 (95.3)	6.331 (94.5)	7.247 (95.2)	4.859 (94.3)	4.915 (94.6)	4.240 (93.9)
(25, 5)		5	Ampiezza media (copertura)	2.203 (94.8)	1.820 (90.8)	2.113 (94.2)	1.771 (90.8)	1.890 (92.6)	1.842 (92.3)	
		10	Ampiezza media (copertura)	1.380 (94.8)	1.279 (93.1)	1.359 (94.4)	1.265 (93.2)	1.301 (93.8)	1.287 (93.8)	
		25	Ampiezza media (copertura)	0.825 (95.0)	0.803 (94.3)	0.820 (94.9)	0.800 (94.4)	0.808 (94.6)	0.805 (94.5)	

**Tabella 9.3: Ampiezza media e copertura stimata per gli IC di livello  $1 - \alpha = 0.95$ , con  $m = 5, 10, 25$  e  $\psi = 1$  e  $5$ .**

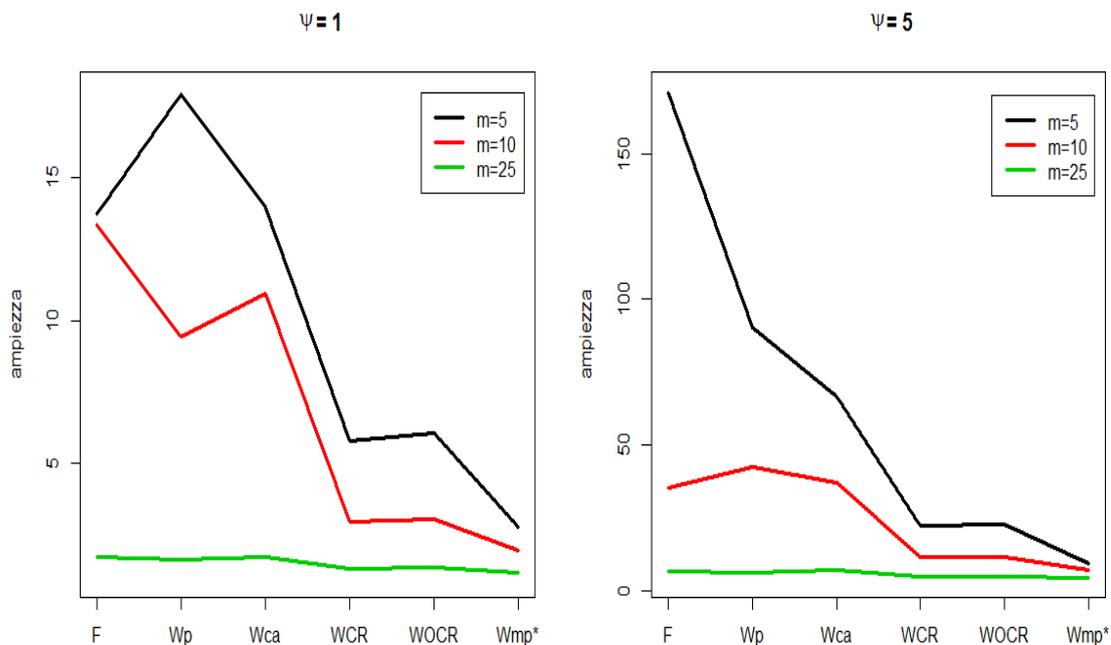
Dalla Tabella 9.3 si può notare che per tutte le numerosità campionarie e per ogni valore del parametro d'interesse, gli intervalli di confidenza per  $\psi$  ottenuti dal modello con il valore di  $\beta_2$  maggiore risultano più stretti in media rispetto a quelli ottenuti dal campione simulato con il valore di  $\beta_2$  minore.

Si può osservare inoltre che l'ampiezza degli intervalli costruiti con la statistica  $F$  è quasi sempre la più grande, mentre, in generale, quelli ottenuti tramite  $W_{MP}^*(\psi)$  risultano i meno ampi. In particolare, tale intervallo è sempre più stretto se confrontato con l'intervallo costruito con la statistica log-rapporto di verosimiglianza condizionata approssimata; questo risultato è confermato in Ventura e Racugno (2012).

I risultati riportati nella Tabella 9.3 evidenziano che, quando la numerosità campionaria cresce, l'ampiezza tende a ridursi e la copertura stimata tende ad avvicinarsi al valore

teorico del 95%. Inoltre, si nota che, quando il campione viene generato con un valore di  $\beta_2$  grande, l'ampiezza degli intervalli tende ad essere simile. Questo risultato è particolarmente evidente per  $m$  grande.

Infine, si noti che quando  $m = 25$  la copertura stimata degli intervalli tende ad essere molto vicina al 95% teorico.



**Figura 12.3:** Andamento ampiezza media degli intervalli di confidenza, con  $\psi = 1$  e 5, con  $(\beta_1, \beta_2)$  pari, rispettivamente, a  $(1, 1)$  e  $(5, 1)$  e con  $m = 5, 10, 25$ .

Dalla Figura 12.3 si può osservare che gli intervalli derivati dalla statistica  $W_I^{CR}(\psi)$  sono sempre meno ampi di quelli basati su  $W_I^{OCR}(\psi)$ , per ogni numerosità campionaria analizzata.

■

### 3.7.2 Secondo studio di simulazione.

Il secondo studio di simulazione è basato sull'Esempio 2.3, ossia sull'esperimento analizzato da Finney (1978).

I dati vengono generati utilizzando la matrice di disegno (3.40), con  $n = 18$  e  $36$ ; per l'analisi si pone, seguendo la simulazione condotta da Ghosh *et al.* (2006),  $(\beta_1, \beta_2, \beta_3, \sigma) = (1, 0.25, 0.25, \sqrt{3})$ , in modo da ottenere  $\psi = 1$ , e  $(\beta_1, \beta_2, \beta_3, \sigma) = (1, 1, 0.25, \sqrt{3})$ , così che  $\psi = 0.25$ .

	$n$		$\hat{\psi}$	$\hat{\psi}_I^{CR}$	$\hat{\psi}_I^{OCR}$	$\hat{\psi}_{MP}^*$
$\psi = 1$	18	Bias	-2.595	-0.894	-0.898	-0.908
		(MSE)	(>10000)	(2.142)	(2.002)	(1.798)
	36	Bias	-0.173	-0.794	-0.799	-0.820
		(MSE)	(>1000)	(1.838)	(1.778)	(1.598)
$\psi = 0.25$	18	Bias	-0.266	-0.044	-0.050	-0.072
		(MSE)	(>1000)	(0.621)	(0.578)	(0.462)
	36	Bias	0.061	-0.025	-0.027	-0.047
		(MSE)	(8.058)	(0.349)	(0.341)	(0.279)

**Tabella 10.3: Distorsione e MSE per le stime di massima pseudo-verosimiglianza con  $\psi = 1$  e  $0.25$  per la simulazione con i dati Finney (1978).**

Nella Tabella 10.3 sono stati riportati la distorsione e MSE per le stime di massima pseudo-verosimiglianza, ottenuti mediante una simulazione basata su 10000 replicazioni Monte Carlo, per due valori di  $\psi$  analizzati e per  $n = 18$  e  $36$ .

Si può notare che, per entrambe le numerosità campionarie, quando  $\psi = 1$  le stime tendono ad essere abbastanza distorte, in particolare la s.m.v. profilo. L'MSE minore viene mostrato, per entrambi i valori di  $\psi$  e di  $n$ , dalla s.m.v. profilo modificata penalizzata. La performance non ottima, in termini di MSE per  $\hat{\psi}$  e di distorsione per  $\hat{\psi}_I^{CR}$ ,  $\hat{\psi}_I^{OCR}$  e  $\hat{\psi}_{MP}^*$ , quando  $\psi = 1$ , può dipendere dal valore di  $\beta_2$ , molto vicino a 0. Infatti, come evidenziato nella precedente simulazione, a parità di numerosità campionaria, si ottengono risultati migliori nelle simulazioni che considerano valori di  $\beta_2$  lontani da 0.

La stima che mostra il più piccolo MSE è, come visto nella precedente simulazione,  $\hat{\psi}_{MP}^*$ , mentre l'MSE di  $\hat{\psi}_I^{CR}$  risulta maggiore rispetto a quello di  $\hat{\psi}_I^{OCR}$ , per entrambi i valori di  $\psi$  analizzati.

Dalla Tabella 10.3 è inoltre possibile notare che, a parità di numerosità campionaria, quando  $\psi = 0.25$  tutte le stime analizzate mostrano un valore dell'MSE minore rispetto al caso in cui  $\psi = 1$ . Questo risultato può derivare dal fatto che il valore di  $\beta_2$  è più lontano da 0, ed in parte conferma i risultati ottenuti nella prima simulazione, dove si evidenziava la riduzione dell'MSE per valori di  $\psi$  piccoli.

Infine, all'aumentare di  $n$ , per entrambi i valori di  $\psi$ , le stime mostrano risultati migliori sia in termini di distanza dal vero valore del parametro, sia in termini di variabilità.

$\psi$	$n$		$F$	$W_P$	$W_{CA}$	$W_I^{CR}$	$W_I^{OCR}$	$W_{MP}^*$
1	18	Ampiezza media	27.727	31.938	13.118	19.907	20.103	6.746
		(copertura)	(83.1)	(80.4)	(82.7)	(94.6)	(96.1)	(81.9)
		% IC finiti	7.5	11.0	7.9			
	36	Ampiezza media	21.750	21.037	16.430	18.720	18.838	6.440
		(copertura)	(86.4)	(84.9)	(86.4)	(95.6)	(96.3)	(84.9)
		% IC finiti	10.7	12.6	11.0			
0.25	18	Ampiezza media	10.433	11.651	15.941	7.652	8.070	3.704
		(copertura)	(93.5)	(90.6)	(92.9)	(93.8)	(95.3)	(93.9)
		% IC finiti	47.5	56.8	49.0			
	36	Ampiezza media	7.855	9.430	6.858	4.143	4.293	2.665
		(copertura)	(95.0)	(93.6)	(94.8)	(95.2)	(95.9)	(95.2)
		% IC finiti	78.5	81.4	79.1			

**Tabella 11.3: Ampiezza media per gli intervalli di confidenza con livello  $1 - \alpha = 0.95$ , copertura stimata e percentuale di intervalli finiti per  $\psi = 1$  e  $0.25$  e  $n = 18, 36$ .**

La Tabella 11.3 presenta l'ampiezza media per gli intervalli di confidenza di livello  $1 - \alpha = 0.95$  costruiti con le diverse statistiche, la copertura stimata degli intervalli e la percentuale di intervalli finiti al variare di  $\psi$  con  $n = 18$  e  $36$ . Si noti come la percentuale di IC finiti per le statistiche  $F$ ,  $W_P(\psi)$  e  $W_{CA}(\psi)$  sia molto basso quando  $\psi = 1$  per entrambe le numerosità campionarie analizzate e di come questo valore decresca al diminuire di  $n$ .

I risultati, ottenuti tramite uno studio di simulazione basato su 10000 repliche Monte Carlo, sono coerenti con le conclusioni ottenute da Ghosh *et al.* (2006), dove si mostra che l'ampiezza media dell'intervallo costruito con  $W_I^{CR}(\psi)$  è inferiore rispetto a quella conseguita con  $W_I^{OCR}(\psi)$ . In particolare, si può osservare che l'ampiezza media dell'intervallo basato su  $W_{MP}^*(\psi)$  è quella che presenta il valore più piccolo per entrambi i valori di  $\psi$  e di  $n$  analizzati.

In aggiunta, è evidente che, anche per le quantità analizzate nella Tabella 10.3, quando il valore di  $\beta_2$  si allontana da 0, la performance degli intervalli di confidenza migliora a parità di  $n$ . Infatti, fissata la numerosità campionaria, l'ampiezza media degli intervalli presentati si riduce notevolmente quando  $\beta_2 = 1$ , il livello di copertura stimato degli IC si avvicina maggiormente al valore teorico del 95%, infine, la percentuale di intervalli di confidenza finiti quando  $\psi = 0.25$  è maggiore rispetto al caso in cui  $\psi = 1$ .

La performance della copertura stimata quando  $\psi = 1$  può dipendere da un lato dalla numerosità campionaria modesta, dall'altro che dal valore di  $\beta_2$  vicino a 0.

In aggiunta, dalla Tabella 11.3 si può osservare che, come visto nella precedente simulazione, l'ampiezza media di tutti gli intervalli di confidenza considerati decresce all'aumentare della numerosità campionaria per ogni valore di  $\psi$  analizzato. Un risultato analogo è mostrato dalla probabilità stimata di ottenere IC finiti per le statistiche  $F$ ,  $W_P(\psi)$  e  $W_{CA}(\psi)$ , mentre la copertura stimata tende ad avvicinarsi maggiormente al 95%, per entrambi i valori di  $\psi$ .

Infine, i risultati presentati nelle Tabelle 9.3 e 10.3 confermano i risultati ottenuti negli Esempi 1.3 e 2.3. Pertanto, come si è notato nella prima simulazione, se  $\beta_2$  o  $n$  assumono un valore più grande e al decrescere di  $\psi$ , le stime e gli intervalli di confidenza sono soggetti ad una minore variabilità.

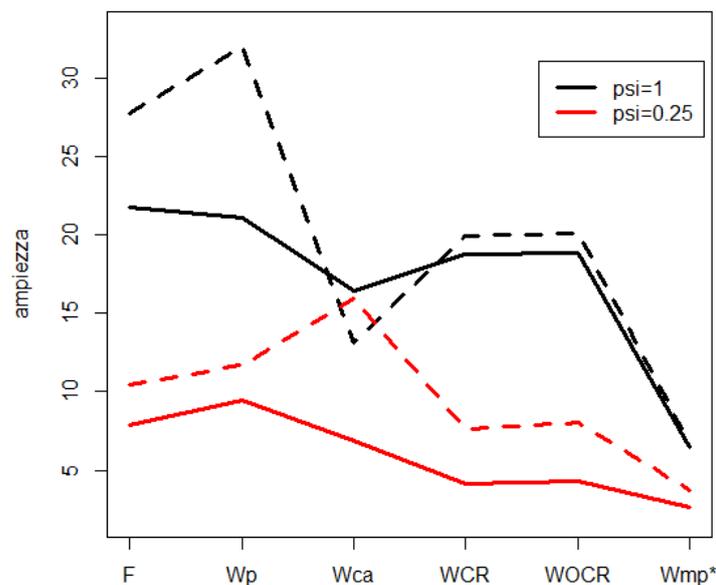


Figura 13.3: Confronto tra l'ampiezza media degli intervalli di confidenza di livello  $1 - \alpha = 0.95$  quando  $\psi = 1$  (linea nera) e  $0.25$  (linea rossa) con  $n = 18$  (linea tratteggiata) e  $36$  (linea continua).

La Figura 13.3 confronta l'ampiezza media degli IC ottenuti con le diverse statistiche con livello  $1 - \alpha = 0.95$ , quando  $\psi = 1$  (linea nera) e  $0.25$  (linea rossa) e con  $n = 18$  (linea tratteggiata) e  $36$  (linea continua). Dal grafico si nota che, per tutti gli intervalli

analizzati, l'ampiezza media si riduce all'aumentare di  $n$  o al decrescere di  $\psi$ . Inoltre, si nota che l'intervallo ottenuto a partire da  $W_{MP}^*(\psi)$  risulta essere, a parità di  $n$  e per entrambi i valori di  $\psi$ , quello più corto, e che l'ampiezza degli IC si riduce quando il valore di  $\psi$  decresce o la numerosità campionaria cresce.

■

### 3.7.3 Terzo studio di simulazione.

Il terzo studio di simulazione si basa su un modello che, rispetto ai casi precedenti, ha un numero di parametri di regressione maggiore. In particolare, si assume il seguente modello di regressione normale

$$y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_J x_{iJ} + \varepsilon_i$$

o, in modo equivalente,

$$y_i = \beta_1(1 + \psi x_{i2}) + \dots + \beta_J x_{iJ} + \varepsilon_i, \quad i = 1, \dots, n, \quad (3.42)$$

dove  $\psi = \beta_2/\beta_1$  è il parametro d'interesse e  $(\varepsilon_1, \dots, \varepsilon_n)$  sono variabili casuali i.i.d.  $N(0, \sigma^2)$ . Inoltre, le variabili  $x_{ij}$  sono variabili dicotomiche che assumono valore 1 se l'osservazione  $i$ -esima appartiene al gruppo  $j$ -esimo e 0 altrimenti,  $j = 2, \dots, J$  e  $i = 1, \dots, n$ . La numerosità campionaria è data dalla somma delle numerosità di ciascun gruppo, ossia  $n = \sum_{j=1}^J m_j$ , dove  $m_j$  è la numerosità del gruppo  $j$ -esimo,  $j = 1, \dots, J$ . In questo esempio si assume un disegno bilanciato, ossia tale che  $m_j = m$ ,  $j = 1, \dots, J$ ; in questo modo  $n = mJ$ .

La matrice di disegno  $X$  di dimensione  $mJ \times J$  è definita come

$$X = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \ddots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \otimes 1_m,$$

dove  $\otimes$  indica il prodotto di Kronecker. Tale matrice fa riferimento a un modello ANOVA. La s.m.v. è  $(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_J, \hat{\sigma}) = (\bar{y}_1, \bar{y}_2 - \bar{y}_1, \dots, \bar{y}_J - \bar{y}_1, \sqrt{SSE/n})$ , dove  $SSE = \sum_{j=1}^J \sum_{i=1}^m (y_{ij} - \bar{y}_j)^2$ .

Per la simulazione si pone  $J = 2, 5, 7$ ,  $m = 5, 10, 20, 30$ ,  $\sigma = 1$ ,  $\beta_2 = 1$  e  $\beta_2 = 5$  e  $(\beta_1, \beta_3, \dots, \beta_J) = (1, 1, \dots, 1)$ . In questo modo  $\psi = 1$  e  $\psi = 5$ .

Le Tabelle 12.3 e 13.3 presentano i risultati di uno studio di simulazione basato su 10000 replicazioni Monte Carlo sulla distorsione e l'MSE delle stime di massima

pseudo-verosimiglianza per il modello (3.42), al variare della numerosità di ogni gruppo e del numero di gruppi considerati.

$m$	$J$		$\hat{\psi}$	$\hat{\psi}_I^{CR}$	$\hat{\psi}_I^{OCR}$	$\hat{\psi}_{MP}^*$	
$\psi = 1$	5	Bias	0.147	-0.162	-0.194	-0.419	
		(MSE)	(>1000)	(1.329)	(1.236)	(0.855)	
	5	Bias	2.717	-0.160	-0.214	-0.443	
		(MSE)	(>1000)	(1.244)	(1.106)	(0.799)	
	7	Bias	0.382	-0.146	-0.206	-0.437	
		(MSE)	(>1000)	(1.211)	(1.064)	(0.769)	
	10	2	Bias	0.355	-0.025	-0.036	-0.203
			(MSE)	(17.617)	(0.568)	(0.551)	(0.385)
		5	Bias	-1.105	-0.020	-0.038	-0.208
			(MSE)	(>1000)	(0.588)	(0.558)	(0.386)
		7	Bias	0.286	-0.026	-0.045	-0.215
			(MSE)	(154.145)	(0.582)	(0.551)	(0.385)
	20	2	Bias	0.119	-0.002	-0.005	-0.100
			(MSE)	(0.420)	(0.268)	(0.266)	(0.213)
		5	Bias	0.129	0.005	0.001	-0.096
			(MSE)	(0.454)	(0.269)	(0.265)	(0.210)
		7	Bias	0.123	-0.001	-0.006	-0.102
			(MSE)	(0.438)	(0.271)	(0.267)	(0.211)
	30	2	Bias	0.067	-0.008	-0.009	-0.073
			(MSE)	(0.229)	(0.174)	(0.173)	(0.149)
		5	Bias	0.081	0.005	0.003	-0.063
			(MSE)	(0.230)	(0.177)	(0.176)	(0.150)
		7	Bias	0.071	-0.004	-0.006	-0.071
			(MSE)	(0.225)	(0.175)	(0.174)	(0.150)

**Tabella 12.3: Distorsione e MSE per le stime di massima pseudo-verosimiglianza con  $\psi = 1$  e per  $m = 5, 10, 20, 30$  e  $J = 2, 5, 7$ .**

$m$	$J$		$\hat{\psi}$	$\hat{\psi}_I^{CR}$	$\hat{\psi}_I^{OCR}$	$\hat{\psi}_{MP}^*$	
$\psi = 5$	5	Bias	0.057	-0.443	-0.536	-1.228	
		(MSE)	(>1000)	(10.285)	(9.528)	(6.442)	
	5	Bias	0.465	-0.455	-0.611	-1.302	
		(MSE)	(>1000)	(10.015)	(8.861)	(6.283)	
	7	Bias	0.093	-0.471	-0.640	-1.327	
		(MSE)	(>1000)	(9.572)	(8.395)	(6.060)	
	10	2	Bias	1.162	-0.027	-0.061	-0.567
			(MSE)	(360.279)	(4.508)	(4.337)	(2.845)
		5	Bias	0.799	-0.038	-0.090	-0.602
			(MSE)	(265.615)	(4.425)	(4.182)	(2.787)
		7	Bias	0.969	-0.037	-0.095	-0.609
			(MSE)	(43.850)	(4.143)	(3.949)	(2.644)
20	2	Bias	0.370	-0.009	-0.017	-0.299	
		(MSE)	(5.523)	(2.011)	(1.993)	(1.558)	
	5	Bias	0.325	-0.034	-0.046	-0.330	
		(MSE)	(3.325)	(1.986)	(1.957)	(1.535)	
	7	Bias	0.369	0.001	-0.013	-0.302	
		(MSE)	(3.464)	(2.051)	(2.018)	(1.556)	
30	2	Bias	0.225	0.001	-0.003	-0.197	
		(MSE)	(1.712)	(1.299)	(1.294)	(1.094)	
	5	Bias	0.216	-0.008	-0.014	-0.209	
		(MSE)	(1.672)	(1.261)	(1.254)	(1.059)	
	7	Bias	0.200	-0.023	-0.029	-0.224	
		(MSE)	(1.772)	(1.340)	(1.331)	(1.125)	

**Tabella 13.3: Distorsione e MSE per le stime di massima pseudo-verosimiglianza con  $\psi = 5$  e per  $m = 5, 10, 20, 30$  e  $J = 2, 5, 7$ .**

Dai risultati riportati nelle Tabelle 12.3 e 13.3 è possibile osservare che, a parità di numerosità  $m$ , il comportamento delle stime, valutate in termini di distorsione e di MSE, sembra abbastanza stabile al variare del numero di gruppi. Al contrario, si notano marcate differenze quando la numerosità all'interno dei singoli gruppi varia.

Quando la numerosità per gruppo è bassa (ossia  $m = 5$  o  $10$ ), si nota che la s.m.v. profilo presenta una variabilità molto elevata, mentre le s.m.v. integrata e profilo modificata penalizzata sembrano dipendere meno alla grandezza di  $m$ , esibendo un valore dell'MSE contenuto, in particolare quando  $\psi = 1$ .

Un'altra importante differenza è data dal valore del parametro. Come notato nella prima e seconda simulazione, al crescere di  $\psi$  la performance delle stime peggiora. Le s.m.v.

integrata sono quelle che presentano una minore distorsione, mentre  $\hat{\psi}_{MP}^*$  mostra sempre l'MSE più piccolo. Inoltre, si può osservare che, a parità di numerosità per gruppo e di numero di gruppi, l'MSE di  $\hat{\psi}_I^{OCR}$  è sempre maggiore di quello di  $\hat{\psi}_I^{OCR}$ . Questo risultato è particolarmente evidente quando  $m$  è piccolo.

Le Tabelle 14.3 e 15.3 mostrano i risultati di uno studio di simulazione basato su 10000 replicazioni Monte Carlo sull'ampiezza media e la copertura degli intervalli di confidenza costruiti con la statistica  $F$  e le statistiche log-rapporto di pseudo-verosimiglianza di livello  $1 - \alpha = 0.95$ , per due diversi valori di  $\psi$ .

$m$	$J$		$F$	$W_P$	$W_{CA}$	$W_I^{CR}$	$W_I^{OCR}$	$W_{MP}^*$	
$\psi = 1$	2	Ampiezza media	71.351	29.640	28.459	9.568	9.846	4.259	
		(copertura)	(94.2)	(91.7)	(93.6)	(91.4)	(92.9)	(89.9)	
	5	Ampiezza media	38.652	28.805	26.082	9.503	9.934	4.237	
		(copertura)	(95.8)	(93.5)	(95.7)	(92.2)	(94.1)	(91.0)	
	7	Ampiezza media	26.092	26.915	28.350	9.729	10.150	4.254	
		(copertura)	(95.2)	(93.5)	(95.1)	(91.8)	(94.0)	(90.6)	
	10	2	Ampiezza media	23.788	13.540	17.056	4.775	4.895	3.012
			(copertura)	(96.3)	(95.3)	(96.1)	(93.6)	(94.2)	(92.6)
		5	Ampiezza media	13.261	11.637	31.963	4.627	4.812	3.005
	7	(copertura)	(96.7)	(95.6)	(96.7)	(93.6)	(94.6)	(93.0)	
		Ampiezza media	18.471	14.639	12.897	4.750	4.947	3.025	
	(copertura)	(96.7)	(95.6)	(96.6)	(93.6)	(94.8)	(92.6)		
20	2	Ampiezza media	4.681	4.028	5.214	2.500	2.537	2.110	
		(copertura)	(95.5)	(94.7)	(95.4)	(94.4)	(94.8)	(93.9)	
	5	Ampiezza media	3.958	4.291	5.279	2.455	2.512	2.099	
(copertura)		(95.7)	(94.9)	(95.6)	(94.6)	(95.0)	(94.2)		
7	Ampiezza media	4.060	3.793	3.834	2.433	2.494	2.089		
	(copertura)	(95.8)	(95.2)	(95.7)	(94.7)	(95.2)	(94.3)		
30	2	Ampiezza media	2.293	2.179	2.263	1.877	1.894	1.700	
		(copertura)	(95.0)	(94.7)	(94.9)	(94.8)	(95.0)	(94.2)	
	5	Ampiezza media	2.256	2.157	2.243	1.863	1.891	1.703	
(copertura)		(95.3)	(94.9)	(95.3)	(94.9)	(95.3)	(94.5)		
7	Ampiezza media	2.189	2.119	2.183	1.845	1.875	1.692		
	(copertura)	(95.4)	(94.9)	(95.4)	(94.9)	(95.3)	(94.7)		

Tabella 14.3: Ampiezza media e copertura stimata degli IC di livello  $1 - \alpha = 0.95$  per  $\psi = 1$  e con  $m = 5, 10, 20, 30$  e  $J = 2, 5, 7$ .

$m$	$J$		$F$	$W_P$	$W_{CA}$	$W_I^{CR}$	$W_I^{OCR}$	$W_{MP}^*$	
$\psi = 5$	2	Ampiezza media	72.578	79.113	126.910	26.218	26.898	11.145	
		(copertura)	(94.4)	(92.2)	(94.0)	(91.4)	(92.6)	(89.0)	
	5	Ampiezza media	101.898	58.948	67.144	26.534	27.482	11.136	
		(copertura)	(95.6)	(93.6)	(95.5)	(91.5)	(93.7)	(89.6)	
	7	Ampiezza media	86.340	49.313	70.730	27.035	27.997	11.201	
		(copertura)	(96.0)	(94.0)	(95.8)	(92.1)	(94.0)	(90.0)	
	10	2	Ampiezza media	37.803	70.397	38.072	13.211	13.523	8.073
			(copertura)	(96.8)	(95.4)	(96.6)	(93.5)	(94.1)	(92.0)
		5	Ampiezza media	45.534	118.948	38.436	13.290	13.784	8.149
			(copertura)	(97.0)	(96.0)	(96.9)	(93.6)	(94.4)	(92.5)
		7	Ampiezza media	33.879	50.563	37.117	13.273	13.787	8.113
			(copertura)	(97.2)	(96.4)	(97.1)	(93.5)	(94.5)	(92.4)
	20	2	Ampiezza media	13.240	11.781	14.518	6.932	7.032	5.708
			(copertura)	(95.7)	(94.9)	(95.6)	(94.4)	(94.6)	(93.6)
		5	Ampiezza media	11.480	9.926	12.382	6.778	6.934	5.697
			(copertura)	(95.6)	(95.0)	(95.6)	(94.4)	(95.0)	(93.8)
		7	Ampiezza media	24.544	11.537	15.057	6.746	6.912	5.678
			(copertura)	(95.7)	(94.9)	(95.6)	(94.4)	(94.9)	(94.0)
30	2	Ampiezza media	6.833	6.069	6.403	5.098	5.146	4.594	
		(copertura)	(95.0)	(94.6)	(94.9)	(94.6)	(94.9)	(94.1)	
	5	Ampiezza media	6.046	5.866	6.029	5.040	5.115	4.570	
		(copertura)	(94.9)	(94.5)	(94.8)	(94.6)	(94.9)	(94.0)	
	7	Ampiezza media	6.008	5.843	5.997	5.047	5.127	4.589	
		(copertura)	(95.1)	(94.6)	(95.1)	(94.6)	(94.9)	(94.3)	

**Tabella 15.3: Ampiezza media e copertura stimata degli IC di livello  $1 - \alpha = 0.95$  per  $\psi = 5$  e con  $m = 5, 10, 20, 30$  e  $J = 2, 5, 7$ .**

Dalle Tabelle 14.3 e 15.3 si nota che l'ampiezza media di tutti gli intervalli decresce al crescere della numerosità all'interno dei singoli gruppi e che la copertura stimata tende ad avvicinarsi maggiormente al valore teorico del 95% all'aumentare di  $m$ . Inoltre, come visto nei precedenti studi di simulazione, l'intervallo costruito con la statistica  $W_{MP}^*(\psi)$  risulta essere sempre il più stretto. Si nota anche che, a parità di numerosità  $m$  in ogni gruppo, si ottengono sempre intervalli di confidenza più stretti quando il vero valore del parametro è  $\psi = 1$ , rispetto al caso in cui  $\psi = 5$ . Questo risultato conferma quelli ottenuti nelle precedenti simulazioni dove, al decrescere del valore di  $\psi$ , l'ampiezza degli IC ottenuti, a partire dalle diverse statistiche, mostrava un andamento decrescente.

Si noti che quando  $m$  è grande, il numero di gruppi introdotti nel modello non sembra influenzare particolarmente l'ampiezza degli intervalli. Questo è più evidente se si considerano gli IC costruiti con le statistiche log-rapporto di verosimiglianza integrata o profilo modificata penalizzata, mentre è meno evidente per gli altri.

I risultati della simulazione indicano che, a parità di numerosità all'interno dei gruppi e di gruppi considerati nel modello (3.42), l'ampiezza dell'intervallo basato su  $W_I^{CR}(\psi)$  è sempre inferiore a quello derivato da  $W_I^{OCR}(\psi)$ , in particolare quando  $m$  è piccolo, mentre tende ad essere simile quando questo cresce. In aggiunta, è possibile osservare che, quando  $m = 30$ , l'IC ottenuto con la statistica  $F$  mostra sempre l'ampiezza maggiore, per ogni valore di  $\psi$  considerato.

Come è stato evidenziato nelle precedenti simulazioni, l'ampiezza dell'intervallo basato su  $W_{CA}(\psi)$  presenta sempre un'ampiezza superiore a quello costruito a partire dalla statistica  $W_{MP}^*(\psi)$ .

Infine, si può osservare che, a parità di  $m$  e di  $J$ , la copertura degli IC ottenuti a partire da  $W_I^{CR}(\psi)$  è minore di quelli costruiti con  $W_I^{OCR}(\psi)$ . Inoltre, quando  $m = 30$  la copertura dei diversi intervalli è molto vicina al 95% teorico, in particolare quando  $\psi = 1$ .

Si consideri nuovamente il modello (3.42), ma con una numerosità  $m$  piccola, (pari a 5 e 10), mentre il numero di gruppi viene fatto crescere sensibilmente, assumendo i valori di  $J = 10, 30, 50$ . In questo modo si vuole analizzare il comportamento delle stime e degli intervalli di confidenza al crescere dei parametri di disturbo mantenendo una numerosità  $m$  esigua.

$m$	$J$		$\hat{\psi}$	$\hat{\psi}_I^{CR}$	$\hat{\psi}_I^{OCR}$	$\hat{\psi}_{MP}^*$	
$\psi = 1$	5	Bias	1.139	-0.174	-0.235	-0.461	
		(MSE)	(>1000)	(1.303)	(1.142)	(0.833)	
	5	Bias	1.546	-0.145	-0.215	-0.449	
		(MSE)	(>1000)	(1.171)	(1.010)	(0.746)	
	5	Bias	0.665	-0.157	-0.226	-0.456	
		(MSE)	(>100)	(1.127)	(0.976)	(0.735)	
	10	10	Bias	0.581	-0.020	-0.040	-0.212
			(MSE)	(>100)	(0.540)	(0.510)	(0.358)
		30	Bias	0.293	-0.014	-0.035	-0.209
			(MSE)	(13.533)	(0.567)	(0.532)	(0.370)
		50	Bias	0.326	-0.031	-0.052	-0.224
			(MSE)	(21.567)	(0.573)	(0.537)	(0.377)
$\psi = 5$	5	Bias	3.513	-0.431	-0.614	-1.313	
		(MSE)	(>10000)	(9.341)	(8.109)	(5.851)	
	5	Bias	5.135	-0.503	-0.696	-1.377	
		(MSE)	(>10000)	(9.653)	(8.348)	(6.174)	
	5	Bias	2.434	-0.469	-0.664	-1.347	
		(MSE)	(>10000)	(9.083)	(7.827)	(5.813)	
	10	10	Bias	0.630	-0.043	-0.102	-0.617
			(MSE)	(>1000)	(4.154)	(3.893)	(2.610)
		30	Bias	1.185	-0.042	-0.106	-0.622
			(MSE)	(>100)	(3.912)	(3.650)	(2.465)
		50	Bias	1.052	-0.032	-0.097	-0.615
			(MSE)	(83.504)	(4.177)	(3.887)	(2.588)

**Tabella 16.3: Distorsione e MSE per le stime di massima pseudo-verosimiglianza, con  $\psi = 1$  e  $5$ ,  $m = 5$  e  $10$  e  $J = 10, 30, 50$ .**

Nella Tabella 16.3 sono riportate la distorsione e l'MSE per le stime di massima pseudo-verosimiglianza ottenute da una simulazione basata su 10000 repliche Monte Carlo. Anche in questa simulazione il numero di parametri introdotti nel modello non sembra essere un fattore determinante per le performance delle stime di massima pseudo-verosimiglianza, quando la numerosità campionaria  $m$  rimane costante. Infatti, non si notano particolari differenze in termini di distorsione e di MSE al variare del numero di gruppi  $J$ . Tale fenomeno è particolarmente evidente per le s.m.v. integrata e profilo modificata penalizzata, dove, anche a fronte di una piccola numerosità  $m$ , le due grandezze considerate rimangono sostanzialmente invariate, per entrambi i valori di  $\psi$  analizzati. Al contrario, le maggiori differenze si notano nella distorsione di  $\hat{\psi}$ ; tuttavia,

tale stima presenta anche un valore dell'MSE particolarmente elevato, soprattutto quando  $m = 5$ .

Un'altra importante differenza, già notata nelle precedenti simulazioni, consiste nel miglioramento delle performance delle stime al decrescere del valore di  $\psi$ , a parità di numerosità  $m$  e  $J$ .

Si può quindi affermare che, per il modello (3.42), anche a fronte di un numero di gruppi elevato, la performance delle stime dipende soprattutto dalla numerosità per gruppo e dal valore di  $\psi$  e non sembra essere particolarmente influenzata dal numero dei parametri di disturbo.

	$m$	$J$		$F$	$W_P$	$W_{CA}$	$W_I^{CR}$	$W_I^{OCR}$	$W_{MP}^*$	
$\psi = 1$	5	10	Ampiezza media	18.961	17.028	19.867	10.305	10.691	4.368	
			(copertura)	(96.7)	(95.2)	(96.7)	(92.0)	(94.8)	(91.3)	
		30	Ampiezza media	20.961	32.476	14.509	9.213	9.688	4.145	
				(copertura)	(94.8)	(93.5)	(94.8)	(92.2)	(94.9)	(91.0)
		50	Ampiezza media	41.610	16.063	21.853	9.693	10.213	4.242	
			(copertura)	(95.9)	(94.1)	(95.9)	(92.0)	(94.7)	(91.1)	
	10	10	Ampiezza media	9.692	16.881	8.511	4.592	4.803	2.989	
			(copertura)	(97.5)	(96.3)	(97.4)	(94.5)	(95.2)	(93.5)	
		30	Ampiezza media	16.330	11.041	12.646	4.929	5.166	3.082	
		(copertura)	(97.4)	(96.8)	(97.4)	(94.3)	(95.9)	(94.0)		
	50	Ampiezza media	19.339	9.864	15.882	4.937	5.163	3.095		
		(copertura)	(96.9)	(96.2)	(96.9)	(93.8)	(95.0)	(93.0)		
$\psi = 5$	5	10	Ampiezza media	112.853	48.700	161.566	28.127	29.196	11.462	
			(copertura)	(96.9)	(95.7)	(96.9)	(93.3)	(95.7)	(91.2)	
		30	Ampiezza media	57.501	51.998	57.512	26.880	27.970	11.172	
			(copertura)	(96.7)	(95.4)	(96.7)	(93.4)	(95.6)	(90.5)	
		50	Ampiezza media	47.701	185.559	43.777	26.015	27.128	10.919	
			(copertura)	(96.6)	(95.0)	(96.6)	(92.6)	(95.1)	(90.8)	
	10	10	Ampiezza media	107.890	33.842	51.939	13.601	14.114	8.175	
			(copertura)	(97.1)	(96.1)	(97.1)	(93.1)	(93.8)	(91.9)	
		30	Ampiezza media	60.317	21.747	61.787	11.767	12.355	7.838	
		(copertura)	(97.2)	(96.5)	(97.1)	(94.9)	(95.5)	(93.2)		
	50	Ampiezza media	47.925	26.014	44.803	12.602	13.189	8.075		
		(copertura)	(97.2)	(96.5)	(97.2)	(94.5)	(95.1)	(92.4)		

Tabella 17.3: Ampiezza media coperture stimate per gli intervalli di confidenza di livello  $1 - \alpha = 0.95$  per  $\psi = 1$  e  $5$ ,  $m = 5, 10$  e  $J = 10, 30, 50$ .

La Tabella 17.3 riporta le ampiezze medie e le coperture stimate degli intervalli di confidenza di livello  $1 - \alpha = 0.95$  ottenute tramite uno studio di simulazione basato su

1000 replicazioni Monte Carlo. Dalla Tabella 17.3 è possibile osservare che, così come per distorsione e MSE, quando  $m$  cresce l'ampiezza media degli intervalli tende a decrescere a parità di numerosità  $J$ . Tale fenomeno è molto visibile per gli intervalli ottenuti a partire dalle statistiche  $W_I^{CR}(\psi)$ ,  $W_I^{OCR}(\psi)$  e  $W_{MP}^*(\psi)$ . Un altro risultato che emerge dalla Tabella 17.3, e già notato nelle simulazioni precedenti, è il diminuire dell'ampiezza degli IC al diminuire del valore di  $\psi$ , anche a fronte di un elevato numero di parametri di disturbo. Infatti, anche quando  $J = 50$ , l'ampiezza media di tutti gli intervalli mostra un valore più piccolo se  $\psi = 1$  rispetto al caso in cui  $\psi = 5$ .

Quando  $m = 10$  e  $\psi = 1$ , gli intervalli ottenuti a partire dalle statistiche  $W_I^{CR}(\psi)$ ,  $W_I^{OCR}(\psi)$  e  $W_{MP}^*(\psi)$  presentano un valore dell'ampiezza media abbastanza simile al variare di  $J$ . Questo risultato era già stato notato nelle Tabelle 14.3 e 15.3.

I dati presenti nella Tabella 17.3 mostrano che, anche in presenza di un elevato numero di parametri di disturbo, l'ampiezza media degli intervalli costruiti mediante  $W_{MP}^*(\psi)$  è sempre minore di quella degli IC basati su  $W_{CA}(\psi)$ . Si può quindi affermare che, al crescere di  $m$  o al diminuire del valore di  $\psi$ , l'ampiezza media degli intervalli di confidenza non sembra essere influenzata dal numero di parametri di disturbo, mostrando valori simili per i diversi IC. Al contrario, se  $m$  è piccolo e  $J$  è grande, l'ampiezza degli intervalli presenta un'elevata variabilità.

Infine, si può osservare che le coperture degli intervalli ottenuti a partire da  $W_I^{OCR}(\psi)$  tendono ad avvicinarsi maggiormente al valore teorico del 95%, per ogni valore di  $m$  e di  $J$  analizzati.

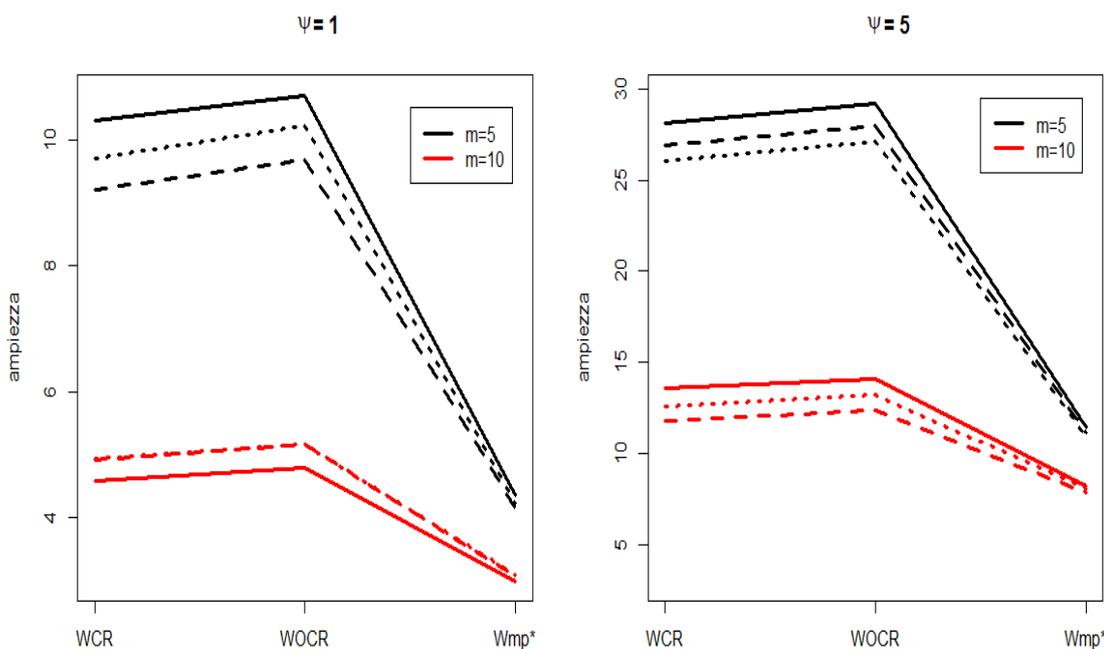


Figura 14.3: Ampiezza degli intervalli di confidenza costruiti a partire da  $W_I^{CR}(\psi)$ ,  $W_I^{OCR}(\psi)$  e  $W_{MP}^*(\psi)$  di livello  $1 - \alpha = 0.95$ , con  $\psi = 1$  e  $5$  per  $J = 10, 30, 50$  e  $m = 5$  e  $10$ .

La Figura 14.3 riporta l'ampiezza media degli intervalli ottenuti tramite le statistiche  $W_I^{CR}(\psi)$ ,  $W_I^{OCR}(\psi)$  e  $W_{MP}^*(\psi)$  di livello  $1 - \alpha = 0.95$ , quando  $J = 10$  (linea continua),  $30$  (linea tratteggiata) e  $50$  (linea a punti). Dalla Figura 14.3 è evidente che l'ampiezza media costruita con le tre statistiche dipende soprattutto dalla numerosità per gruppo e non dal numero di gruppi. Tale risultato è più marcato per  $m = 10$ . Inoltre, il grafico mostra che l'ampiezza media più piccola si ha per l'IC basato su  $W_{MP}^*(\psi)$  e che l'ampiezza media dell'intervallo basato su  $W_I^{OCR}(\psi)$  è sempre maggiore di quello ottenuto tramite  $W_I^{CR}(\psi)$ .

### 3.8 Conclusioni

Nel presente capitolo è stato presentato un problema inerente all'inferenza sul rapporto di coefficienti in un modello di regressione normale. In particolare, si è mostrato che gli intervalli di confidenza costruiti a partire da alcune statistiche del tipo log-rapporto di pseudo-verosimiglianza non sono utili, in quanto tali funzioni possono produrre intervalli non finiti per il parametro d'interesse. Questo perché sono caratterizzate dal fatto di non tendere a  $+\infty$  quando  $|\psi| \rightarrow \infty$ .

Al contrario, applicando la tecnica ibrida introdotta nel Capitolo 2, si è mostrato che gli intervalli basati sulla statistica log-rapporto di pseudo-verosimiglianza profilo modificata penalizzata sono sempre finiti.

Infine, si sono presentati i risultati di alcuni studi di simulazione in cui risultati confermano le proprietà teoriche discusse in Mukerjee e Reid (1999) e Ventura e Racugno (2012). Emerge che l'ampiezza media degli intervalli di confidenza e l'MSE delle stime dipendono dalla numerosità campionaria, dal valore assunto da  $\psi$  (e quindi dai coefficienti di regressione, in particolare di  $\beta_2$ ). Inoltre, per il modello ANOVA, quando la numerosità per gruppo è elevata o il valore di  $\psi$  è piccolo, sembra non esserci una relazione tra la performance delle stime e degli intervalli di confidenza e il numero di parametri di disturbo. Infatti, si è mostrato che tali quantità sono principalmente influenzate dal valore di  $\psi$  e dalla numerosità campionaria di ogni gruppo.

## Conclusioni

L'obiettivo principale della tesi è lo studio dei principali metodi per l'eliminazione dei parametri di disturbo, classici e moderni, in ambito fisheriano e bayesiano, per il problema dell'inferenza sul rapporto di coefficienti di regressione nel modello lineare.

I metodi classici per l'eliminazione dei parametri di disturbo, per l'approccio frequentista e bayesiano, sono opportune funzioni di pseudo-verosimiglianza, ossia la verosimiglianza condizionata, marginale, profilo, condizionata approssimata, profilo modificata per l'ambito frequentista e integrata per l'ambito bayesiano.

La metodologia ibrida unisce aspetti frequentisti (le funzioni di pseudo-verosimiglianza) ad elementi bayesiani (le *matching prior*). Il metodo ibrido nasce per superare alcune difficoltà in ambito bayesiano, quando la dimensione del parametro di disturbo è elevata. Tali problemi sono legati sia all'esigenza di dover elicitarne una distribuzione a priori su tutto il parametro sia alla necessità di dover calcolare un integrale multidimensionale. Mediante questa tecnica si possono costruire funzioni del solo parametro d'interesse che si prestano ad essere interpretate in ambito frequentista come funzioni penalizzate di alcune pseudo-verosimiglianze, ed in ambito bayesiano come pseudo-distribuzioni a posteriori marginali.

L'inferenza sul rapporto di coefficienti di regressione nel modello lineare risulta essere un problema non banale nel metodo classico, poiché in alcuni casi determinate funzioni di pseudo-verosimiglianza portano a conclusioni poco utili. Questo è dovuto al fatto che certi intervalli di confidenza basati sul log-rapporto di pseudo-verosimiglianza possono essere non finiti. Al contrario, l'utilizzo della funzione di verosimiglianza profilo modificata penalizzata non presenta tali difficoltà e risulta più efficace.

Alla fine del Capitolo 3 sono stati riportati alcuni studi di simulazione sulla distorsione e sull'MSE per le stime di massima pseudo-verosimiglianza e sull'ampiezza e la copertura degli intervalli di confidenza costruiti con la statistica  $F$  e le statistiche log-rapporto di pseudo-verosimiglianza. Questi studi hanno mostrato la bontà del metodo

ibrido, confermando le proprietà teoriche descritte da Ventura e Racugno (2012). In particolare, è emerso che la s.m.v. profilo modificata penalizzata presenta il più piccolo valore dell'MSE e che l'intervallo ottenuto a partire dal log-rapporto di pseudo-verosimiglianza profilo modificata penalizzata esibisce l'ampiezza più corta tra tutte quelle analizzate.

Inoltre, le quantità studiate sono apparse abbastanza sensibili alla numerosità campionaria e al valore assunto dal parametro d'interesse; infatti, si sono ottenute, mediamente, stime più precise e meno variabili ed ampiezze più contenute all'aumentare della dimensione del campione o al decrescere del valore di  $\psi$  per ogni modello analizzato.

## Bibliografia

Azzalini, A. (2008). *Inferenza statistica: una presentazione basata sul concetto di verosimiglianza*. Springer-Verlag Italia, Milano.

Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343-365.

Barndorff-Nielsen, O.E., Chamberlin, S.R. (1994). Stable and invariant adjusted directed likelihoods. *Biometrika*, **81**, 485-499.

Berger, J.O., Liseo, B., Wolpert, R. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, **1**, 1-28.

Berger, J. O., Bernardo, J.M. (1992). On the development of reference priors (with discussion). *Bayesian Statistics*, **4**, 35-60.

Brazzale, A.R., Davison, A.C., Reid, N., (2007). *Applied Asymptotic*. Cambridge University Press, Cambridge.

Cabras, S., Racugno, W., Castellanos M.E., Ventura, L. (2012). A matching prior for the shape parameter of the skew-normal distribution. *Scandinavian Journal of Statistics*, **39**, 236-246.

Cox, D.R., Reid, N., (1987). Orthogonal parameters and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society, Series B*, **49**, 1-39.

Chang, H., Mukerjee, R., (2006). Probability matching property of adjusted likelihoods. *Statistics and Probability Letters*, **76**, 838-842.

Datta, G.S., Mukerjee, R. (2004). *Probability Matching Priors: Higher-Order Asymptotics*. In: *Lecture Notes in Statistics*. Springer-Verlag, New York.

Finney, D.J. (1978). *Statistical method in biological assay*. Charles Griffin and Company, London.

Fraser, D.A.S., Reid, N., Wong, A., Yun, Y.G. (2003). Direct Bayes for interest parameters. *Bayesian Statistics*, **7**, 529–533

Ghosh, M., Yin, M., Kim, Y-H. (2003). Objective Bayesian inference for ratios of regression coefficients in linear models. *Statistica Sinica*, **13**, 409–422.

Ghosh, M., Datta, G.S., Kim, D., Sweeting, T.J. (2006). Likelihood-based inference for the ratios of regression coefficients in linear models. *Annals of the Institute of Statistical Mathematics*, **58**, 457-473.

Kass, R.E., Tierney, L., Kadane, J.B. (1990). The validity of posterior expansions based on Laplace's method. In Geisser, S., Hodges, J., Press, S., ZeUner, A. (eds.). *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard*, 473-487. North Holland.

Kotz S., Lumelskii, Y., Pensky, M., (2003). *The Stress-Strength Model and its Generalizations. Theory and Applications*, Singapore: World Scientific.

Levine, R.A., Casella G., (2003). Implementing matching priors for frequentist inference. *Biometrika*, **90**, 127-137.

Liseo, B. (2008). *Introduzione alla statistica bayesiana*. Roma.

Mukerjee, R., Reid, N. (1999). On a property of probability matching priors: matching the alternative coverage probabilities. *Biometrika*, **86**, 333-340.

Nicolau, A. (1993). Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. *Journal of the Royal Statistical Society, Series B*, **55**, 377-390.

Pace, L., Salvan, A. (1996). *Teoria della statistica. Metodi, modelli, approssimazioni asintotiche*. Cedam, Padova.

Pace, L., Salvan, A. (2001). *Introduzione alla statistica - II - Inferenza, verosimiglianza, modelli*. Cedam, Padova.

Pace, L., Salvan, A. (2006). Adjustments of the profile likelihood from a new perspective. *Journal of Statistical Planning and Inference*, **136**, 3554–3564.

Racugno, W., Salvan, A., Ventura, L. (2010). Bayesian analysis in regression models using pseudo-likelihood. *Communications in Statistics-Theory and Methods*, **39**, 3444-3455.

Reid, N. (2003). Asymptotics and the theory of inference. *The Annual of Statistics*, **31**, 1695-1731.

Robert, C., Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York.

Ruli, E., Sartori, N., Ventura, L. (2013). Marginal posterior simulation via higher-order tail area approximations. *Bayesian Analysis*.

Severini, T.A. (2000). *Likelihood methods in statistics*. Oxford University Press, Oxford.

Severini, T.A. (2011). Frequency properties of inferences based on a integrated likelihood function. *Statistica Sinica*, **21**, 433-477.

Staicu, A.M., Reid, N. (2008). On Probability matching priors. *The Canadian Journal of Statistic*, **36**, 613–622.

Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika*, **76**, 604–608.

Tierney, L.J., Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association*, **81**, 82–86.

Tierney, L.J., Kass, R.E., Kadane, J.B. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions, *Journal of the American Statistical Association*, **84**, 710-716.

Ventura, L., Cabras, S., Racugno, W. (2009). Prior distribution from pseudo-likelihoods in the presence of nuisance parameters. *Journal of the American Statistical Association*, **486**, 768-773.

Ventura, L., Racugno, W. (2011). Recent advances on Bayesian inference for  $P(X<Y)$ . *Bayesian Analysis*, **6**, 411-428.

Ventura, L., Racugno, W. (2012). On interval and point estimators based on a penalization of the modified profile likelihood. *Statistics and Probability Letters*, **82**, 1285-1289.

Ventura, L., Sartori, N., Racugno, W. (2013). Objective Bayesian higher-order asymptotics in models with nuisance parameters. *Computational Statistics and Data Analysis*, **60**, 90-96.