



UNIVERSITÀ DEGLI STUDI DI PADOVA
FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA
IN SCIENZE STATISTICHE ECONOMICHE FINANZIARIE E
AZIENDALI

TESI DI LAUREA

**METODI DI VEROSIMIGLIANZA
MODIFICATA E MODELLI PER DATI
DI PANEL**

RELATORE: PROF. ALESSANDRA SALVAN

CORRELATORE: DOTT. NICOLA SARTORI

LAUREANDA: ANGELA CRISTOFORETTI

ANNO ACCADEMICO 2007/2008

Indice

Introduzione	III
1 L'inferenza di verosimiglianza	1
1.1 Introduzione	1
1.2 Specificazione del modello	2
1.3 Notazione ed assunzioni	4
1.4 Test e regioni di confidenza	6
1.5 Invarianza rispetto alla parametrizzazione	8
1.6 Parametrizzazione ortogonale	10
1.7 Principi per la riduzione dei dati	11
1.7.1 Statistiche sufficienti	12
1.7.2 Statistiche costanti in distribuzione	12
1.8 Classi di modelli parametrici	13
1.8.1 Famiglie esponenziali	14
1.8.2 Famiglie di gruppo	16
2 Parametri di disturbo e verosimiglianza	19
2.1 Introduzione	19
2.2 Inferenza in presenza di parametri di disturbo	20
2.3 Parametri di disturbo e riduzione dei dati	23
2.4 Pseudo-verosimiglianze	24
2.4.1 Verosimiglianza marginale e condizionata	25
2.4.2 Verosimiglianza profilo	28
2.5 Modificazioni della verosimiglianza profilo	32

2.5.1	Verosimiglianza condizionata approssimata	32
2.5.2	Verosimiglianza profilo modificata	33
3	Applicazioni a modelli per dati di panel	39
3.1	Introduzione	39
3.2	I parametri incidentali	41
3.3	Un modello per dati binari	43
3.3.1	Specificazione del modello	43
3.3.2	Simulazione	45
3.4	Modello autoregressivo con intercetta variabile	59
3.4.1	Specificazione del modello	59
3.4.2	Simulazione	62
	Conclusioni	81
A	Calcoli per il modello AR(1) con intercetta variabile	83
	Riferimenti bibliografici	89

Introduzione

In diversi ambiti applicativi è frequente dover analizzare insiemi di dati in cui è presente un raggruppamento naturale delle osservazioni. Ad esempio, un insieme di dati può contenere informazioni circa un individuo relativamente a più istanti temporali. In questo caso, si può pensare alle osservazioni relative al singolo individuo come ad un gruppo. Questo è il contesto usuale in cui si opera quando si ha a che fare con **dati di panel** e a cui si fa riferimento in questa tesi.

Altre situazioni in cui le osservazioni possono essere raggruppate si incontrano ad esempio quando, in un campione di individui, vi sono soggetti che appartengono alla stessa famiglia e che rappresentano quindi un gruppo. Il termine dati di panel è utilizzato per qualsiasi situazione in cui è evidente un raggruppamento naturale dei dati, poiché i metodi di analisi dei dati sono spesso adatti per diverse situazioni in cui si dispone di dati raggruppati.

Si pensi, per fissare le idee, al contesto, tipico in ambito econometrico, in cui si hanno q soggetti su cui vengono osservate alcune caratteristiche in T istanti temporali. Spesso, i dati ottenuti vengono analizzati sulla base di modelli statistici parametrici, in cui alcuni parametri, detti **parametri strutturali**, sono gli stessi per tutti i soggetti, mentre altri **parametri**, detti **incidentali**, variano da soggetto a soggetto.

Spesso i parametri strutturali rappresentano le quantità di interesse primario per l'inferenza, e sono detti **parametri di interesse**, mentre i parametri incidentali giocano il ruolo di **parametri di disturbo**.

In diversi contesti applicativi, in cui si desidera analizzare dati di panel, q è grande in relazione a T . Ciò pone dei problemi nell'impiego delle procedure inferenziali usuali basate sulla verosimiglianza, le cui ben note proprietà (consistenza degli stimatori, distribuzione asintotica normale o chi-quadrato delle quantità pivotali di verosimiglianza) valgono con riferimento alla prospettiva asintotica in cui T diverge mentre q rimane fissato.

In generale, se la dimensione del parametro di disturbo è fissata, al crescere della numerosità campionaria aumenta anche l'informazione contenuta nel campione, per cui si può fare riferimento alle procedure di verosimiglianza basate sulla teoria asintotica standard. In presenza di parametri incidentali, invece, la dimensione del parametro di disturbo aumenta al crescere del numero q di soggetti, così da risultare troppo elevata rispetto all'informazione disponibile.

In questo tipo di situazione, sorgono alcune complicazioni note come **problemi di Neyman e Scott**, in merito al contributo degli autori (Neyman e Scott, 1948) che per primi posero in evidenza gli aspetti problematici legati alla presenza di parametri incidentali. In questi casi, il ricorso alle usuali procedure di verosimiglianza, ed in particolar modo l'impiego della verosimiglianza profilo, può condurre a conclusioni inferenziali del tutto inattendibili. Infatti, l'elevata distorsione della funzione *score* profilo può portare ad ottenere una stima inconsistente del parametro d'interesse.

Il problema è dunque quello di individuare metodi alternativi, che permettano di eliminare, o perlomeno ridurre, l'effetto dei parametri incidentali nell'inferenza sulle quantità d'interesse. In questa tesi, si focalizza l'attenzione sulla verosimiglianza profilo modificata.

L'obiettivo principale della tesi è studiare la verosimiglianza profilo modificata e verificarne il comportamento in presenza di parametri incidentali. In particolare, tramite studi di simulazione, si mettono a confronto verosimiglianza profilo e profilo modificata al fine di mostrare come la seconda fornisca risultati migliori rispetto alla precedente. Il fine

ultimo è pertanto quello di confermare empiricamente le proprietà della verosimiglianza profilo modificata studiate.

Il Capitolo 1 è essenzialmente un capitolo di rassegna, volto a richiamare alcuni concetti base e a fornire delle definizioni utili nel seguito della tesi.

Nel Capitolo 2 si considerano le caratteristiche dell'inferenza in presenza di parametri di disturbo e si introduce il problema dei parametri incidentali. Successivamente, si richiamano alcune idee relative alla riduzione dei dati e vengono passati in rassegna alcuni tipi di pseudo-verosimiglianza. Nella parte finale del capitolo, si descrivono alcune modificazioni della verosimiglianza profilo e le rispettive proprietà. In particolare si introduce la versione della verosimiglianza profilo modificata introdotta da Severini (1998) che verrà impiegata nelle simulazioni.

Il Capitolo 3 vuole dimostrare empiricamente, tramite studi di simulazione, la validità delle proprietà della verosimiglianza profilo modificata presentate nel capitolo precedente. In particolare si verificano le proprietà di stimatori e regioni di confidenza basate su tale verosimiglianza in due esempi appartenenti alla categoria dei problemi di Neyman e Scott. Nello specifico, si considera un esempio di modello lineare generalizzato per dati binari e un modello autoregressivo del primo ordine con intercetta variabile.

Capitolo 1

L'inferenza di verosimiglianza

1.1 Introduzione

Il presente capitolo è essenzialmente un capitolo di rassegna, volto a richiamare alcuni concetti base dell'inferenza statistica, utili per contestualizzare la tesi.

Si richiamano la nozione di modello statistico e di funzione di verosimiglianza e le quantità ad essa associate, le definizioni di regioni di confidenza e test di verosimiglianza. Viene inoltre fissata la notazione utilizzata nella tesi.

Successivamente vengono date le definizioni di invarianza rispetto alla parametrizzazione e parametrizzazione ortogonale. Si richiamano inoltre i concetti di statistica sufficiente e costante in distribuzione. Infine si presentano sinteticamente le due principali classi di modelli parametrici: le famiglie esponenziali e di gruppo.

Per la stesura di questo capitolo si è fatto principalmente riferimento ai testi Azzalini (2001) e Pace e Salvan (1996, 2001).

1.2 Specificazione del modello

L'idealizzazione fondamentale, su cui poggia l'inferenza statistica, è che i dati osservati $y = (y_1, \dots, y_n)$ costituiscano una determinazione di un vettore casuale Y . Tale concetto è sintetizzabile con la scrittura y realizzazione di $Y \sim F_0(y)$, $y \in \mathcal{Y}$, ove F_0 rappresenta una funzione di ripartizione ignota e \mathcal{Y} è lo **spazio campionario**. L'obiettivo dell'inferenza è utilizzare y per trarre conclusioni sulla distribuzione $F_0(y)$ di Y , **modello probabilistico** generatore dei dati. Questo perché la legge di probabilità F_0 è almeno in parte ignota e occorre pertanto utilizzare l'informazione contenuta nei dati stessi per ottenerne una ricostruzione.

Le conclusioni su $F_0(y)$ sono naturalmente soggette a incertezza in quanto la natura casuale di Y , che determina y , si ripercuote sulle affermazioni fatte. Quello che è importante è fare in modo che il grado di incertezza sia il più piccolo possibile, compatibilmente con la natura casuale di Y , e che si sia in grado di valutare il livello di incertezza a cui si è sottoposti.

La prima fase dell'inferenza è quella di specificazione del modello, che consiste nell'individuare delle possibili forme per F_0 . In concreto, occorre specificare una famiglia di distribuzioni di probabilità almeno qualitativamente compatibili con la generazione dei dati y . Tale scelta va effettuata sulla base della natura del fenomeno che ha generato i dati y , dello schema di campionamento adottato e di altre eventuali informazioni disponibili sul fenomeno considerato. L'insieme delle distribuzioni così individuate viene indicato con \mathcal{F} e costituisce il **modello statistico**.

Dal punto di vista applicativo, un modello statistico è usualmente pensato come approssimazione, nel senso che non si presume che esso catturi esattamente il modello probabilistico, quanto piuttosto che lo rappresenti in un modo che possa ritenersi adeguato per gli scopi della ricerca. È intuitivo che l'inferenza sarà tanto più accurata quanto meglio è possibile delimitare la classe \mathcal{F} compatibilmente con la condizione $F_0(y) \in \mathcal{F}$. In particolare, se $F_0(y) \in \mathcal{F}$, si dice che il modello statistico è **correttamente**

specificato.

A seconda delle informazioni disponibili, si può ritenere opportuno dare al modello statistico \mathcal{F} un'estensione più o meno ampia. In ordine decrescente di estensione, e quindi crescente di informazione presupposta, si possono individuare i tre livelli di specificazione seguenti.

- **Specificazione non parametrica.** Il modello statistico \mathcal{F} è una restrizione dell'insieme di tutte le distribuzioni di probabilità sopra il supporto appropriato alla natura dei dati, definita tramite assunzioni semplificatrici globali, che non individuano espressamente un numero finito di parametri di interesse per l'inferenza.
- **Specificazione semiparametrica.** Gli elementi di \mathcal{F} sono individuati sia mediante una componente parametrica, sia mediante una componente non parametrica, ossia $\mathcal{F} = \{F(y; \theta), \theta \in \Theta\}$, dove $\theta = (\tau, h(\cdot))$, con $\tau \in T \subseteq \mathbb{R}^k$, mentre l'insieme di possibili specificazioni della funzione $h(\cdot)$ non è indicizzabile con un numero finito di parametri reali.
- **Specificazione parametrica.** Gli elementi che compongono \mathcal{F} sono indicizzabili con un numero finito p di parametri reali, ossia $\mathcal{F} = \{F(y; \theta), \theta \in \Theta \subseteq \mathbb{R}^p\}$, con Θ **spazio parametrico**. Se il modello è correttamente specificato si ha $F_0(y) = F(y; \theta_0)$ per un valore $\theta_0 \in \Theta$, detto **vero valore del parametro**.

La scelta del livello di specificazione dipende da quanta informazione ci si attende di poter ragionevolmente estrarre dai dati e dallo scopo per cui il modello è costruito. In seguito l'unico caso preso in considerazione sarà quello dei problemi di inferenza in ambito parametrico.

Occorre precisare che le funzioni di ripartizione $F(y; \theta)$ che descrivono un modello statistico parametrico, corrispondono tutte ad una variabile casuale discreta o continua. La classe \mathcal{F} può quindi essere specificata anche tramite le corrispondenti funzioni di probabilità o di densità, rispettivamente. Pertanto è possibile definire \mathcal{F} come insieme di funzioni di

densità invece che di funzioni di ripartizione e scrivere

$$\mathcal{F} = \{f(y; \theta), \theta \in \Theta \subseteq \mathbb{R}^p\} \quad (1.1)$$

per una qualche funzione di densità f . In seguito si farà riferimento a quest'ultima definizione di modello statistico.

1.3 Notazione ed assunzioni

Per un fissato modello statistico del tipo (1.1), una volta determinati i valori campionari $y = (y_1, \dots, y_n)$, la corrispondente funzione di densità si riduce ad una funzione unicamente del parametro $\theta = (\theta_1, \dots, \theta_p)$. Si definisce quindi **funzione di verosimiglianza** di θ basata sui dati y , la funzione da $\Theta \rightarrow \mathbb{R}^+$

$$L = L(\theta) = c(y)f(y; \theta),$$

con $c(y)$ costante positiva non dipendente da θ e per questo trascurabile dalla funzione stessa. Due funzioni di verosimiglianza che differiscono solamente per la costante moltiplicativa $c(y)$ si definiscono **equivalenti**. Quando si vorrà mettere in evidenza la dipendenza di $L(\theta)$ dai dati si userà la notazione $L(\theta; y)$.

Se si ha a che fare con osservazioni tratte da variabili casuali indipendenti e identicamente distribuite (i.i.d.) con funzione di densità marginale $f(y_i; \theta)$, la verosimiglianza complessiva è data dal prodotto delle funzioni di verosimiglianza ottenute nei singoli esperimenti, ovvero

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

Poiché $L(\theta)$ è non negativa, ha senso trattare con la **funzione di log-verosimiglianza**

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

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

Sotto campionamento da variabili casuali i.i.d., la funzione $l(\theta)$ assume la forma

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

L'utilizzo della funzione di log-verosimiglianza nelle procedure di inferenza comporta dei vantaggi in termini computazionali e rende più semplice la descrizione dei risultati teorici.

Nell'ambito di questa tesi, si assume che siano soddisfatte le condizioni di regolarità (cfr. Severini, 2000, § 3.4), sotto le quali valgono i risultati riportati in seguito. Per i modelli in cui la log-verosimiglianza soddisfa tali condizioni, si definiscono le seguenti **quantità di verosimiglianza**.

Funzione punteggio (funzione *score*) : vettore delle derivate parziali prime di $l(\theta)$, indicata con $l_* = l_*(\theta; y) = (l_1, \dots, l_p)$, con generico elemento $l_r = \partial l(\theta) / \partial \theta_r$, $r = 1, \dots, p$.

Informazione osservata: matrice di dimensione $p \times p$ delle derivate parziali seconde di $l(\theta)$ cambiate di segno, indicata con $j = j(\theta) = -l_{rs}$, con $l_{rs} = \partial^2 l(\theta) / (\partial \theta_r \partial \theta_s)$ e $r, s = 1, \dots, p$; l'informazione osservata rappresenta una misura della curvatura locale della log-verosimiglianza. Il generico elemento di posto (r, s) verrà indicato con la scrittura j_{rs} , mentre si indicheranno gli elementi della matrice inversa $j(\theta)^{-1}$ con la scrittura j^{rs} .

Informazione attesa: valore atteso dell'informazione osservata, indicata con $i = i(\theta) = E_\theta(j(\theta)) = E_\theta(-l_{rs}(\theta))$. Anche in questo caso il generico elemento di posto (r, s) verrà indicato con i_{rs} e gli elementi della matrice inversa $i(\theta)^{-1}$ con i^{rs} .

Inoltre, assumendo che siano soddisfatte le condizioni che permettono di scambiare le operazioni di derivazione e di integrazione, valgono i seguenti risultati di particolare interesse per l'inferenza.

- La funzione *score* ha valore atteso

$$E_\theta(l_*(\theta)) = 0, \quad \text{per ogni } \theta \in \Theta. \quad (1.2)$$

- Vale l'**identità dell'informazione**, ovvero

$$E_{\theta}(-l_{rs}(\theta)) = E_{\theta}(l_r(\theta) l_s(\theta)). \quad (1.3)$$

In altri termini, la matrice $i(\theta)$ è pari alla matrice di covarianza della *score* e come tale è una matrice definita non negativa.

Un valore $\theta \in \Theta$ che massimizza $L(\theta)$, o equivalentemente $l(\theta)$, è detto **stima di massima verosimiglianza** (s.m.v.) di θ ed è indicato con $\hat{\theta}$. In un modello con verosimiglianza regolare, $\hat{\theta}$ va cercato tra le soluzioni dell'**equazione di verosimiglianza** $l_*(\theta) = 0$. Nel seguito si assume che tale soluzione sia unica, almeno al divergere della numerosità campionaria. Lo stimatore di massima verosimiglianza $\hat{\theta}$ ha distribuzione approssimata sotto θ

$$\hat{\theta} \sim N_p(\theta, j(\hat{\theta})^{-1}). \quad (1.4)$$

Per indicare le quantità di verosimiglianza valutate in $\hat{\theta}$ si usano le notazioni $\hat{l} = l(\hat{\theta})$, $\hat{j} = j(\hat{\theta})$, $\hat{i} = i(\hat{\theta})$.

1.4 Test di verosimiglianza e regioni di confidenza

In generale, una stima $\hat{\theta}$ non coincide con il vero valore del parametro θ , per cui si ha l'esigenza di accompagnare la stima stessa con un insieme di valori plausibili per il parametro.

Data la funzione di verosimiglianza $L(\theta)$, è possibile ottenere, come stima intervallare per il parametro θ , una regione

$$\hat{\Theta}(y) = \{\theta \in \Theta : L(\theta) \geq c L(\hat{\theta})\},$$

con $c \in (0, 1)$, costituita dai valori del parametro che hanno elevata verosimiglianza.

Si può ricorrere alla verosimiglianza anche per verificare un'ipotesi nulla del tipo $H_0 : \theta \in \Theta_0 \subset \Theta$, confrontando $L(\hat{\theta}_0)$, massimo valore di

$L(\theta)$ per $\theta \in \Theta_0$, e $L(\hat{\theta})$, massimo valore di $L(\theta)$ per $\theta \in \Theta$. Ciò conduce a considerare la statistica test

$$\frac{L(\hat{\theta})}{L(\hat{\theta}_0)} \quad (1.5)$$

con regione di rifiuto unilaterale destra. Tale statistica test può essere espressa anche in termini di log-verosimiglianza, come $l(\hat{\theta}) - l(\hat{\theta}_0)$.

Per l'ipotesi semplice $H_0 : \theta = \theta_0$ contro l'alternativa $H_1 : \theta \neq \theta_0$, il logaritmo della statistica test (1.5) è proporzionale alla statistica

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

detta **log-rapporto di verosimiglianza**.

È possibile dimostrare che $W(\theta_0)$ ha distribuzione nulla approssimata chi-quadrato con p gradi di libertà, dove p è la dimensione del parametro θ . Si possono quindi costruire regioni di confidenza di livello nominale $1 - \alpha$, di forma

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

con $\chi_{p;1-\alpha}^2$ quantile- $(1 - \alpha)$ della distribuzione chi-quadrato con p gradi di libertà. La regione $\hat{\Theta}(y)$ può anche essere scritta come

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

Se $p = 1$ o 2 , la rappresentazione grafica di tali regioni è immediata sfruttando il grafico della log-verosimiglianza.

Statistiche asintoticamente equivalenti a W sono il **test score**

$$W_u = W_u(\theta_0) = l_*(\theta_0)^\top i(\theta_0)^{-1} l_*(\theta_0)$$

e il **test di Wald**

$$W_e = W_e(\theta_0) = (\hat{\theta} - \theta_0)^\top i(\theta_0) (\hat{\theta} - \theta_0).$$

Nonostante l'equivalenza asintotica fra le tre versioni del test del rapporto di verosimiglianza, è opportuno fare delle considerazioni circa gli aspetti che le distinguono. Innanzitutto solo i test W e W_u sono invarianti rispetto

alla parametrizzazione (cfr. § 1.5), inoltre solamente il test W non pone nelle regioni di confidenza dei valori per il parametro θ inammissibili. Un ulteriore aspetto che distingue W dalle altre due versioni è una maggiore adeguatezza qualitativa delle regioni di confidenza o di accettazione ad esso associate nei casi in cui il modello non ha verosimiglianza regolare.

Se il parametro θ è scalare ($p = 1$), può essere opportuno fare riferimento alle versioni unilaterali del test del rapporto di verosimiglianza, ossia

$$\begin{aligned} r &= r(\theta_0) = \text{sgn}(\hat{\theta} - \theta_0) \sqrt{W}, \\ r_u &= r_u(\theta_0) = l_*(\theta_0) (i(\theta_0))^{-1/2}, \\ r_e &= r_e(\theta_0) = (\hat{\theta} - \theta_0) (i(\theta_0))^{1/2}, \end{aligned}$$

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

È possibile costruire intervalli di confidenza di livello nominale $1 - \alpha$ per il parametro scalare θ del tipo

$$\hat{\Theta}(y) = \left\{ \theta \in \Theta : -z_{1-\alpha/2} < r(\theta) < z_{1-\alpha/2} \right\},$$

con $z_{1-\alpha/2}$ quantile- $(1 - \alpha/2)$ di una $N(0, 1)$.

Tali intervalli sono in genere preferibili rispetto a quelli basati sul test di Wald W_e calcolati come

$$\hat{\theta} \pm z_{1-\alpha/2} j(\hat{\theta})^{-1/2}. \quad (1.6)$$

1.5 Invarianza rispetto alla parametrizzazione

Nella specificazione operativa di un modello statistico \mathcal{F} , è possibile scegliere fra parametrizzazioni differenti ma equivalenti. Sia $\psi(\cdot)$ una funzione biunivoca e regolare (infinitamente derivabile assieme alla sua inversa) da Θ a Ψ , con $\theta(\psi)$ applicazione inversa di $\psi(\theta)$. La (1.1) può essere scritta come

$$\mathcal{F} = \{f^\Psi(y; \psi) = f(y; \theta(\psi)), \psi \in \Psi\},$$

con

$$\Psi = \{\psi \in \mathbb{R}^p : \psi = \psi(\theta), \theta \in \Theta\}.$$

Poichè θ e $\psi(\theta)$ individuano lo stesso elemento di \mathcal{F} , si può scrivere

$$L^\Psi(\psi) = L^\Theta(\theta(\psi))$$

e

$$l^\Psi(\psi) = l^\Theta(\theta(\psi)), \quad (1.7)$$

con $L^\Psi(\cdot)$ e $l^\Psi(\cdot)$ verosimiglianza e log-verosimiglianza nella parametrizzazione ψ e $L^\Theta(\cdot)$ e $l^\Theta(\cdot)$ le stesse funzioni nella parametrizzazione originaria θ .

Trattandosi di formulazioni equivalenti, la scelta della parametrizzazione da adottare è in parte di convenienza; una nuova parametrizzazione può permettere infatti di semplificare i calcoli nella fase di inferenza. In altre situazioni la riparametrizzazione del modello può essere giustificata dalla scelta di parametri che abbiano una determinata interpretazione nella distribuzione. Ciò che è auspicabile, è che le conclusioni inferenziali siano le stesse a prescindere dalla parametrizzazione adottata. Le procedure statistiche basate sul test W descritte nel § 1.4 godono proprio di questa proprietà.

Il principio di invarianza rispetto alla parametrizzazione, a cui si è appena fatto riferimento, è soddisfatto dalle procedure di stima secondo il metodo della massima verosimiglianza. Infatti le s.m.v. soddisfano la **proprietà di equivarianza**, secondo la quale se $\psi(\cdot)$ è una funzione biunivoca regolare da Θ a Ψ , allora la s.m.v. di $\psi(\theta)$ è $\psi(\hat{\theta})$, con $\hat{\theta}$ s.m.v. di θ relativa alla verosimiglianza $L(\theta)$.

Anche per quanto concerne la costruzione di regioni di confidenza, il soddisfacimento del principio di invarianza rispetto alla parametrizzazione si traduce in una richiesta di equivarianza. Se $\hat{\Theta}(y)$ costituisce una regione di confidenza per θ di livello $1 - \alpha$ con $\alpha \in (0, 1)$, ossia la **probabilità di copertura nulla** $P_\theta\{\theta \in \hat{\Theta}(Y)\}$ soddisfa la condizione

$$P_\theta\{\theta \in \hat{\Theta}(Y)\} = 1 - \alpha, \quad \theta \in \Theta,$$

si richiede che la medesima procedura di costruzione conduca, nella parametrizzazione ψ , a una regione

$$\hat{\Psi}(y) = \{\psi \in \Psi : \theta(\psi) \in \hat{\Theta}(y)\}, \quad y \in \mathcal{Y}.$$

Nei problemi di verifica di ipotesi, una riparametrizzazione comporta una traduzione del problema di verificare la conformità dei dati con l'ipotesi $H_0 : \theta \in \Theta_0 \subset \Theta$, nel problema di verificare l'ipotesi $H_0 = \psi \in \Psi_0 = \{\psi \in \Psi : \theta(\psi) \in \Theta_0\}$. Il test del rapporto di verosimiglianza riferito alla parametrizzazione ψ si basa su $l^\Psi(\hat{\psi}) - l^\Psi(\hat{\psi}_0)$, mentre nella parametrizzazione originale θ , il test è proporzionale alla quantità $l(\hat{\theta}) - l(\hat{\theta}_0)$ (cfr. § 1.4). Poiché $\hat{\psi} = \psi(\hat{\theta})$ e $\hat{\psi}_0 = \psi(\hat{\theta}_0)$, e tenuto conto della (1.7), il valore del test calcolato nelle due diverse parametrizzazioni è il medesimo. La statistica test W risulta pertanto invariante rispetto alla parametrizzazione.

La situazione a cui si fa riferimento in questa tesi, è quella in cui si considera una partizione del parametro θ del tipo $\theta = (\tau, \zeta)$ con τ parametro di interesse di dimensione k e ζ parametro di disturbo di dimensione $p - k$. In questi casi, si richiede che valga la proprietà di invarianza alle sole **riparametrizzazioni che non alterano l'interesse**, ossia a riparametrizzazioni del tipo $\psi = \psi(\theta) = \psi(\tau, \zeta)$, con $\psi = (\phi, \chi)$, tali che

$$\phi = \phi(\tau) \quad e \quad \chi = \chi(\tau, \zeta),$$

con $\phi(\cdot)$ funzione biunivoca. Questo perchè in generale una riparametrizzazione del tipo $\psi(\theta)$ non mantiene la distinzione tra componenti di interesse e di disturbo. Quando ci si troverà in questa situazione si farà pertanto riferimento al **principio di invarianza rispetto alle riparametrizzazioni che non alterano l'interesse**.

1.6 Parametrizzazione ortogonale

Un particolare tipo di riparametrizzazione è la **parametrizzazione ortogonale**. Partendo dalla partizione del parametro $\theta = (\tau, \zeta)$ considerata alla

fine del § 1.5, si dice che la riparametrizzazione (τ, χ) con $\chi = \chi(\tau, \zeta)$ è **ortogonale**, se i vettori *score* l_τ e l_χ sono incorrelati o, equivalentemente, se $i_{\tau\chi} = 0$. La conseguenza principale dell'ortogonalità è che le s.m.v. $\hat{\tau}$ e $\hat{\chi}$ risultano asintoticamente indipendenti. Inoltre, la matrice di covarianza di $\hat{\tau}$ può essere valutata come se χ fosse noto. Infatti, la matrice di covarianza asintotica di $\hat{\tau}$ è pari a

$$i^{\tau\tau} = (i_{\tau\tau} - i_{\tau\chi}(i_{\chi\chi})^{-1}i_{\chi\tau})^{-1} = (i_{\tau\tau})^{-1}, \quad \text{se } i_{\tau\chi} = 0.$$

Infine, si ha che

$$\hat{\chi}_\tau - \hat{\chi} = O_p(n^{-1}). \quad (1.8)$$

Assumendo per semplicità $p = 2$ e $k = 1$, è semplice descrivere come sia possibile reperire la nuova parametrizzazione (τ, χ) , tale che τ e χ risultino ortogonali. Deve valere $i_{\tau\chi} = 0$, ovvero $E((\partial^2/\partial\tau\partial\chi)l^\perp) = 0$, con $l^\perp = l^\perp(\tau, \chi) = l(\tau, \zeta(\tau, \chi))$, log-verosimiglianza nella nuova parametrizzazione. L'espressione della quantità di interesse $(\partial^2/\partial\tau\partial\chi)l^\perp$ è ricavata nel seguente modo

$$\begin{aligned} \frac{\partial l^\perp}{\partial \tau} &= \frac{\partial l}{\partial \tau} + \frac{\partial l}{\partial \zeta} \frac{\partial \zeta}{\partial \tau} \\ \frac{\partial^2 l^\perp}{\partial \tau \partial \chi} &= \frac{\partial^2 l}{\partial \tau \partial \zeta} \frac{\partial \zeta}{\partial \chi} + \frac{\partial^2 l}{\partial \zeta^2} \frac{\partial \zeta}{\partial \tau} \frac{\partial \zeta}{\partial \chi} + \frac{\partial l}{\partial \zeta} \frac{\partial^2 \zeta}{\partial \tau \partial \chi}. \end{aligned} \quad (1.9)$$

Calcolando il valore atteso del secondo membro della (1.9), il terzo addendo si annulla e si arriva quindi alla condizione

$$\frac{\partial \zeta}{\partial \chi} \left(-i_{\tau\zeta} - i_{\zeta\zeta} \frac{\partial \zeta}{\partial \tau} \right) = 0,$$

ossia

$$i_{\tau\zeta} + i_{\zeta\zeta} \frac{\partial \zeta(\tau, \chi)}{\partial \tau} = 0$$

a cui deve soddisfare la trasformazione $\chi(\tau, \zeta)$.

1.7 Principi per la riduzione dei dati

Una forma immediata di riduzione dei dati è data dall'utilizzo di una **statistica**, ossia di una funzione $t(\cdot) : \mathcal{Y} \rightarrow \mathbb{R}^r$, per un qualche r naturale, tale

che l'espressione di $t(y)$ non dipenda dal parametro θ . Il valore $t = t(y)$ corrispondente al valore osservato y è detto valore campionario della statistica. L'utilità di tale trasformazione per l'inferenza su θ dipende dalla **riduzione inferenziale** che essa opera, ossia dalla relazione fra il modello statistico dei dati trasformati e quello dei dati originari.

Vengono definite ora due tipologie di statistiche di particolare importanza per l'inferenza nei principali modelli statistici parametrici: le statistiche sufficienti e le statistiche costanti in distribuzione.

1.7.1 Statistiche sufficienti

In riferimento al modello statistico (1.1), una statistica s è detta **sufficiente** per l'inferenza su θ , se la densità di Y può essere fattorizzata nel seguente modo

$$f_Y(y; \theta) = f_S(s; \theta) f_{Y|S=s}(y; s), \quad (1.10)$$

dove la densità condizionata $f_{Y|S=s}(y; s)$ non dipende da θ .

Considerando la fattorizzazione (1.10), si può interpretare il processo di generazione dei dati y come un esperimento che avviene in due stadi. Nel primo stadio viene generato s secondo il modello probabilistico $f_S(s; \theta)$. Nel secondo stadio si genera y secondo il modello probabilistico $f_{Y|S=s}(y)$. Nell'ambito del modello \mathcal{F} , noto s , il secondo stadio è retto da una distribuzione non inferente su θ e pertanto può essere omesso, senza perdita d'informazione sul parametro.

Per un assegnato modello statistico parametrico \mathcal{F} , esiste usualmente una pluralità di statistiche sufficienti per θ . Una statistica è detta **sufficiente minimale** se è funzione di ogni altra statistica sufficiente.

1.7.2 Statistiche costanti in distribuzione

In riferimento al modello statistico \mathcal{F} , si dice che una statistica c è **costante in distribuzione** se la sua funzione di densità marginale non dipende da θ .

Una statistica costante in distribuzione c è detta **costante in distribuzione massimale** rispetto a \mathcal{F} se non esiste un'altra statistica costante in distribuzione di cui c è funzione non biunivoca. In generale, una statistica costante in distribuzione massimale non è unica.

Se c è costante in distribuzione e t è una statistica complementare a c , nel senso che (c, t) è una trasformazione biunivoca dei dati y , vale la fattorizzazione

$$f_{C,T}(c, t; \theta) = f_C(c) f_{T|C=c}(t; c, \theta)$$

e solo il modello condizionato con densità $f_{T|C=c}(t; c, \theta)$ è informativo su θ .

Il principio di condizionamento secondo Fisher (1934, 1935) fa riferimento ad una sottoclasse delle statistiche costanti in distribuzione, le statistiche ancillari. Data una statistica sufficiente minimale s e la stima di massima verosimiglianza $\hat{\theta}$, la statistica a è detta **ancillare** se è costante in distribuzione e se la relazione tra $(\hat{\theta}, a)$ ed s è biunivoca; a rappresenta quindi l'informazione mancante per ricostruire s a partire da $\hat{\theta}$.

Se a è ancillare, vale la fattorizzazione

$$f_{\hat{\theta}, A}(\hat{\theta}, a; \theta) = f_A(a) f_{\hat{\theta}|A=a}(\hat{\theta}; a, \theta),$$

per cui $\hat{\theta}$ è sufficiente minimale nel modello condizionato al valore osservato di a . Inoltre $l(\theta; y)$ è equivalente a $l(\theta; \hat{\theta}, a)$. Questa rappresentazione è utilizzata nel § 2.5.2 in cui si introduce la verosimiglianza profilo modificata.

1.8 Classi di modelli parametrici

Tra i modelli statistici parametrici, si possono individuare due ampie classi di particolare importanza, ossia le famiglie esponenziali e le famiglie di gruppo. La loro rilevanza nell'inferenza statistica è giustificata da due motivazioni: in primo luogo entrambe costituiscono un insieme ampio di modelli che ricorrono sovente nelle applicazioni; inoltre è possibile definire

procedure inferenziali generali, fundamentalmente semplici ed accurate, per i modelli che rientrano in tali classi.

In particolare, si possono ottenere riduzioni inferenziali non banali per sufficienza o per condizionamento e le quantità di verosimiglianza presentano un comportamento regolare che assicura esistenza e unicità delle stime, nonché altre utili proprietà. Inoltre i metodi per ottenere approssimazioni asintotiche di ordine elevato risultano di applicazione facilitata.

1.8.1 Famiglie esponenziali

L'importanza delle famiglie esponenziali con riferimento al problema della riduzione inferenziale è motivata dalla possibilità di ricorrere alla riduzione per sufficienza e al condizionamento per l'eliminazione dei parametri di disturbo.

La classe parametrica (1.1) costituisce una **famiglia esponenziale** se i suoi elementi sono del tipo

$$f(y; \theta) = h(y) \exp \left\{ \psi(\theta) \cdot t(y)^\top - K(\theta) \right\}, \quad (1.11)$$

dove $h(\cdot) \geq 0$ è funzione di y non dipendente da θ , così come la statistica $t(y) = (t_1(y), \dots, t_r(y))$; $\psi(\theta) = (\psi_1(\theta), \dots, \psi_r(\theta))$, funzione con dominio $\Theta \subseteq \mathbb{R}^p$ e codominio $\Psi = \psi(\Theta) \subseteq \mathbb{R}^r$ e $K(\theta)$ sono invece funzioni di θ che non dipendono da y .

Affinché $\theta = (\theta_1, \dots, \theta_p)$ sia identificabile, si richiede che $\psi(\theta)$ sia biunivoca. Inoltre se si vuole che la (1.11) sia una **rappresentazione minimale** (o **in forma ridotta**), ossia coinvolga il minimo numero di funzioni $\psi_i(\theta)$ e di associate statistiche $t_i(\theta)$ per $i = 1, \dots, r$, devono essere soddisfatte le seguenti tre condizioni:

1. lo spazio parametrico Θ deve contenere almeno $r + 1$ elementi;
2. le $r + 1$ funzioni reali $1, \psi_1(\theta), \dots, \psi_r(\theta)$ devono essere linearmente indipendenti in Θ , cioè non deve esserci alcuna combinazione lineare

di queste che si annulli identicamente per $\theta \in \Theta$; ciò implica che nessuna di queste è esprimibile come funzione lineare delle altre;

3. analogamente, anche le $r + 1$ funzioni reali $1, t_1(y), \dots, t_r(y)$ devono essere linearmente indipendenti.

Quando una famiglia esponenziale è scritta in forma ridotta, si dice che r è l'**ordine della famiglia** e $t(y)$ è una **statistica canonica** di \mathcal{F} , sufficiente minimale per θ . Nel caso in cui l'ordine della famiglia coincida con la dimensione di Θ , ossia $r = p$, e $\psi(\theta)$ sia una riparametrizzazione del modello, $\psi = \psi(\theta)$ è detto **parametro canonico** (o **naturale**).

Momenti della statistica canonica

Per una famiglia esponenziale regolare (ossia con spazio parametrico Ψ il più ampio possibile e aperto) di ordine p , sotto campionamento casuale semplice di numerosità n , la funzione del modello per la statistica sufficiente minimale $t = t(y)$ ha la forma

$$p(t; \theta) = \tilde{h}(t) \exp \{ \psi(\theta) \cdot t - nK(\theta) \},$$

con $\psi(\theta) = (\psi_1(\theta), \dots, \psi_p(\theta))$ e $t = (t_1, \dots, t_p)$. La log-verosimiglianza per θ risulta dunque pari a

$$l(\theta; t) = \psi(\theta) \cdot t - nK(\theta),$$

e quindi nella parametrizzazione canonica, si ha

$$l(\psi; t) = \psi \cdot t - nG(\psi),$$

con $G(\psi) = K(\theta(\psi))$.

La funzione punteggio è allora

$$l_r(\psi; t) = \frac{\partial l(\psi)}{\partial \psi_r} = t_r - n \frac{\partial G(\psi)}{\partial \psi_r}, \quad r = 1, \dots, p,$$

e l'informazione osservata è

$$j(\psi) = n \frac{\partial^2 G(\psi)}{\partial \psi_r \partial \psi_s}, \quad r, s = 1, \dots, p.$$

Poiché $j(\psi)$ è non stocastica, essa coincide con l'informazione attesa $i(\psi)$.

L'identità (1.2), applicata al caso in esame, diviene $E_\psi(l_*(\psi; t)) = 0$, per cui

$$E_\psi(T_r) = n \frac{\partial K(\psi)}{\partial \psi_r}, \quad r = 1, \dots, p,$$

ossia la statistica canonica ha valore atteso pari al gradiente di $nK(\psi)$.

Dall'identità dell'informazione (1.3), segue inoltre che

$$\text{Cov}(T_r, T_s) = E_\psi(l_r(\psi; T) l_s(\psi; T)) = n \frac{\partial^2 K(\psi)}{\partial \psi_r \partial \psi_s} = i(\psi).$$

1.8.2 Famiglie di gruppo

Le famiglie di gruppo costituiscono la seconda ampia classe di modelli statistici di rilievo, in quanto rappresentano la situazione ideale in cui trovano applicazione sia il principio del condizionamento rispetto a statistiche costanti in distribuzione, sia il ricorso alla verosimiglianza marginale per l'eliminazione dei parametri di disturbo.

Prima di introdurre il concetto di famiglia di gruppo, è opportuno dare alcune definizioni. Sia \mathcal{G} una classe di trasformazioni biunivoche che agiscono su uno spazio \mathcal{Y} , l'applicazione di una trasformazione $g_1 \in \mathcal{G}$ seguita dall'applicazione di una trasformazione $g_2 \in \mathcal{G}$ è detta composizione di g_1 e g_2 e viene indicata con $g_2 \circ g_1$. La classe \mathcal{G} è detta **chiusa rispetto alla composizione** se

$$g_1, g_2 \in \mathcal{G} \quad \Rightarrow \quad g_2 \circ g_1 \in \mathcal{G}.$$

Sia ora $e = e(y) = y$ la trasformazione identità e si indichi con g^{-1} la funzione inversa per $g \in \mathcal{G}$, tale che $g^{-1} \circ g(y) = e(y)$. La classe \mathcal{G} è detta **chiusa rispetto all'inversione** se

$$g \in \mathcal{G} \quad \Rightarrow \quad g^{-1} \in \mathcal{G}.$$

Una classe \mathcal{G} di trasformazioni biunivoche che agiscono su uno spazio \mathcal{Y} è detta **gruppo di trasformazioni** se è chiusa sia rispetto alla composizione sia rispetto all'inversione.

Data una funzione $u(\cdot)$ da \mathcal{Y} in un codominio \mathcal{U} , essa è detta **invariante rispetto a \mathcal{G}** se per ogni $g \in \mathcal{G}$ vale $u(g(y)) = u(y)$.

Si supponga dunque assegnata una famiglia parametrica

$$\mathcal{F}_0 = \{f_0(y; \tau), \tau \in T \subseteq \mathbb{R}^k, y \in \mathcal{Y} \subseteq \mathbb{R}^d\}$$

e sia \mathcal{G} un gruppo di trasformazioni $g : \mathcal{Y} \rightarrow \mathcal{Y}$ tale che \mathcal{F}_0 non sia invariante rispetto a \mathcal{G} . Sia $gf_0(y; \tau)$ la densità di $g(Y)$ quando Y ha densità $f_0(y; \tau)$, si definisce una **famiglia di gruppo composta** generata da \mathcal{F}_0 sotto l'azione di \mathcal{G} , la classe di densità

$$\mathcal{F} = \{gf_0(y; \tau), \tau \in T, g \in \mathcal{G}, y \in \mathcal{Y}\}.$$

Si osservi che \mathcal{F} rappresenta in generale un modello semiparametrico; si ha un modello parametrico se gli elementi di \mathcal{G} sono identificabili tramite un numero finito di parametri reali.

Un esempio di famiglie di gruppo è dato dalle **famiglie di posizione e scala**. Si definisce tale famiglia come un modello statistico parametrico con parametro $\theta = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$ per una variabile casuale univariata Y , con funzione di densità

$$f(y; \mu, \sigma) = \frac{1}{\sigma} f_0\left(\frac{y - \mu}{\sigma}\right), \quad y \in \mathbb{R},$$

dove $f_0(\cdot)$ è un'assegnata funzione di densità. In questo caso lo spazio parametrico T è degenere, per cui si ha una **famiglia di gruppo semplice**.

Poiché vale la rappresentazione $Y = \mu + \sigma Y_0$, dove Y_0 ha densità $f_0(\cdot)$, si specifica una famiglia di posizione e scala quando la legge del modello probabilistico generatore di un'osservazione è nota a meno di una traslazione e di un cambiamento dell'unità di misura. In una famiglia di posizione e scala, il parametro μ è detto **parametro di posizione**, mentre σ è detto **parametro di scala**.

La legge di Y_0 è detta legge tipo della famiglia di posizione e scala ed è espressa dalla densità $f_0(y) = f(y; 0, 1)$.

Se $y = (y_1, \dots, y_n)$ è un campione casuale semplice tratto da una distribuzione univariata in una famiglia di posizione e scala, la funzione di densità del modello statistico risulta pari a

$$f(y; \mu, \sigma) = \frac{1}{\sigma^n} \prod_{i=1}^n f_0\left(\frac{y_i - \mu}{\sigma}\right).$$

Una proprietà verificata in generale è che il modello statistico di una statistica univariata $t = t(y_1, \dots, y_n)$ tale che $t(a + by_1, \dots, a + by_n) = bt(y_1, \dots, y_n) + a$ per ogni $b \in \mathbb{R}^+$ e per ogni $a \in \mathbb{R}$, risulta ancora una famiglia di posizione e scala.

Capitolo 2

Parametri di disturbo e verosimiglianza

2.1 Introduzione

Introdotta le nozioni base dell'inferenza statistica, si vuole in questo capitolo incentrare l'attenzione sulle tecniche inferenziali utilizzate in presenza di parametri di disturbo. Si introduce inoltre il problema dei parametri incidentali tramite un esempio e si definiscono le verosimiglianze che permettono di risolvere, o perlomeno ridurre, tale problema.

Il paragrafo 2.2 introduce i principi dell'inferenza condotta in presenza di parametri di disturbo e i problemi legati alla presenza di parametri incidentali.

Nel paragrafo 2.3 si definisce il concetto di riduzione inferenziale nel caso in cui l'interesse è parziale, ossia quando solo una parte del parametro θ è di interesse per l'inferenza. In questi casi l'obiettivo è quello di semplificare il problema individuando un fattore di verosimiglianza dipendente dal solo parametro di interesse, sul quale si possa basare l'inferenza. L'utilità di tale riduzione è tanto maggiore quanto più grande è la dimensione del parametro di disturbo.

Nel paragrafo 2.4 si descrivono alcune caratteristiche legate alla di-

stinzione tra verosimiglianza propria e non propria, che introducono dei sottoparagrafi dedicati alla descrizione delle verosimiglianze marginale e condizionata e della verosimiglianza profilo. In presenza di parametri incidentali, la verosimiglianza profilo presenta degli inconvenienti, per cui le relative procedure inferenziali per la stima del parametro di interesse non risultano soddisfacenti. Di qui l'esigenza di trovare soluzioni alternative che permettano di ottenere risultati attendibili anche quando l'informazione contenuta nel campione è ridotta.

Il paragrafo 2.5 è quindi interamente dedicato alle modificazioni della verosimiglianza profilo. In particolare, sono presentate la verosimiglianza profilo condizionata approssimata di Cox e Reid (1987), utilizzabile solamente qualora si disponga di una parametrizzazione ortogonale, e la verosimiglianza profilo modificata di Barndorff-Nielsen (1980, 1983), applicabile invece a modelli più generali.

2.2 Inferenza in presenza di parametri di disturbo

Per ottenere una corretta specificazione del modello statistico \mathcal{F} , è necessario tener conto di tutti gli aspetti connaturati alla struttura probabilistica in esame, sia quelli di effettivo interesse, sia quelli accessori che, pur essendo trascurabili ai fini dell'analisi, permettono di catturare al meglio la variabilità di popolazione.

Nell'ambito dell'inferenza statistica parametrica gli aspetti di interesse sono esprimibili tramite un parametro τ , k -dimensionale, che prende il nome di **parametro di interesse**. Gli aspetti accessori di variabilità sono indicati dal parametro ζ , di dimensione $p - k$, che prende il nome di **parametro di disturbo**. È di disturbo nel senso convenzionale che, se fosse noto il suo vero valore ζ_0 , sarebbe ovvio come procedere all'inferenza su τ tramite $L(\tau) = L(\tau, \zeta_0)$, verosimiglianza del modello ridotto. Se ζ non

è noto, è necessario ricorrere a procedure inferenziali su τ che godano di proprietà soddisfacenti per tutti i possibili valori di ζ .

Si considerino i dati y costituiti da n osservazioni (y_1, \dots, y_n) indipendenti, con densità congiunta

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

si possono verificare due situazioni che fanno riferimento a partizioni differenti del parametro $\theta = (\theta_1, \dots, \theta_n)$, ovvero:

1. $\theta_i = (\tau, \zeta), i = 1, \dots, n,$
2. $\theta_i = (\tau, \zeta_i), i = 1, \dots, n.$

Nel primo caso il parametro di disturbo ζ , così come quello di interesse τ , è comune a tutte le osservazioni, che sono pertanto identicamente distribuite; la sua dimensione è quindi indipendente dalla numerosità campionaria n .

Nel secondo caso invece, la dimensione del parametro di disturbo $\zeta = (\zeta_1, \dots, \zeta_n)$ dipende dalla numerosità campionaria n , mentre il parametro di interesse τ rimane comune a tutte le osservazioni. In particolare, ad ogni osservazione y_i è associata una componente di disturbo ζ_i che modella l'eterogeneità delle osservazioni. In questo contesto, τ è il **parametro strutturale**, mentre ζ_1, \dots, ζ_n sono **parametri incidentali**.

I problemi che coinvolgono parametri incidentali, di cui si sono occupati, fra gli altri, Arellano e Hahn (2006), Honorè (2002) e Lancaster (2000), vanno sotto il nome di **problemi di Neyman e Scott** e richiedono soluzioni diverse da quelle standard.

L'esempio seguente chiarisce il concetto di parametri strutturali e incidentali e mette in evidenza le debolezze dello stimatore di massima verosimiglianza in presenza di parametri incidentali.

Esempio 1.1 Il problema di Neyman e Scott

Siano y_{ij} realizzazioni di variabili casuali indipendenti con

funzione di densità

$$p(y_{ij}; \mu_i, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y_{ij} - \mu_i)^2}{2\sigma^2}\right\},$$

$$i = 1, \dots, q, \quad j = 1, \dots, m_i.$$

Il parametro μ_i è incidentale poiché appare nella legge di un numero finito m_i di variabili casuali; σ^2 è un parametro strutturale poiché appare nella legge di ciascuna variabile casuale.

Se si considerano le stime di massima verosimiglianza per μ_i e σ^2 , ponendo per semplicità $m_i = m$ costante, si ha che naturalmente $\hat{\mu}_i = \bar{y}_i$ e

$$\hat{\sigma}^2 = \sum_{i=1}^q \sum_{j=1}^m \frac{(y_{ij} - \hat{\mu}_i)^2}{qm}.$$

Dato che la somma di quadrati $\sum_{j=1}^m \frac{(Y_{ij} - \hat{\mu}_i)^2}{\sigma^2}$ si distribuisce come un chi-quadrato con $m - 1$ gradi di libertà, si ha che $\hat{\sigma}^2 \sim \frac{\sigma^2 \chi_{q(m-1)}^2}{qm}$.

Si ha quindi che

$$E(\hat{\sigma}^2) = \frac{\sigma^2(m-1)}{m} \quad \text{e} \quad V(\hat{\sigma}^2) = \frac{2(m-1)\sigma^4}{m^2q}.$$

Si mostra così che lo stimatore di massima verosimiglianza di σ^2 , $\hat{\sigma}^2$, non è consistente se q diverge con m fissato.

Questo è un classico esempio utilizzato per mostrare l'inconsistenza dello stimatore di massima verosimiglianza in presenza di parametri incidentali. Per meglio comprendere la relazione tra eliminazione dei parametri di disturbo e inconsistenza dello stimatore di massima verosimiglianza, si consideri la stima di massima verosimiglianza ottenuta in due stadi: nel primo stadio si eliminano i parametri di disturbo dalla verosimiglianza sostituendoli con le rispettive s.m.v. per valori fissati del parametro di interesse, ottenendo quella che viene comunemente chiamata **verosimiglianza profilo**; nel secondo stadio si massimizza la verosimiglianza ottenuta al primo stadio rispetto ai parametri di interesse.

La sostituzione con le stime di massima verosimiglianza può essere interpretata quindi come uno dei metodi possibili per eliminare i parametri di disturbo. Tuttavia in presenza di un numero elevato di parametri di disturbo, tale approccio può portare a stimatori dei parametri di interesse con proprietà insoddisfacenti, come evidenziato nell'Esempio 1.1.

2.3 Parametri di disturbo e riduzione dei dati

L'obiettivo della riduzione inferenziale è quello di evidenziare situazioni di assenza di informazione sul parametro θ , o su una parte di esso, in un fattore di verosimiglianza, in modo tale che questo risulti trascurabile ai fini dell'inferenza.

Si consideri in particolare il caso in cui l'interesse è parziale, ossia quando $\theta = (\tau, \zeta)$ e solo τ è d'interesse per l'inferenza. Per comprendere come è possibile operare la riduzione in presenza di parametri di disturbo, è necessario introdurre i concetti di assenza di informazione sul parametro di interesse (o non informatività parziale), statistica parzialmente sufficiente e statistica parzialmente costante in distribuzione.

Date due statistiche u e v , si dice che u non è informativa su τ se la distribuzione di U condizionata a $V = v$ non dipende da τ . Tale nozione è detta di **non informatività parziale**, intendendo non informatività sulla parte di interesse del parametro.

Si consideri la fattorizzazione

$$f_Y(y; \tau, \zeta) = f_V(v; \tau, \zeta) f_{U|V=v}(u; v, \tau, \zeta) f_{Y|U=u, V=v}(y; u, v, \tau, \zeta). \quad (2.1)$$

Sia y in corrispondenza biunivoca con (u, v) , si dice che v è **parzialmente sufficiente** per τ se $f_{U|V=v}(u; v, \tau, \zeta) = f_{U|V=v}(u; v, \zeta)$.

Sia V degenera, si dice che u è **parzialmente costante in distribuzione** per τ se $f_{U|V=v}(u; v, \tau, \zeta) = f_U(u; \zeta)$.

In riferimento alla fattorizzazione (2.1) se v è parzialmente sufficiente per τ , il contributo alla verosimiglianza del fattore $f_{U|V=v}(u; v, \zeta)$ è trascu-

rabile per l'inferenza su τ . Se u è parzialmente costante in distribuzione per τ , è trascurabile il contributo del fattore $f_U(u; \zeta)$.

Le nozioni di sufficienza parziale e costanza in distribuzione parziale si accompagnano ai corrispondenti principi di riduzione dei dati, ossia il **principio di sufficienza parziale** e il **principio di condizionamento rispetto a statistiche parzialmente costanti in distribuzione** che verranno ripresi nel § 2.4.1.

2.4 Pseudo-verosimiglianze

In presenza di parametri di disturbo si è già visto che sarebbe conveniente poter basare l'inferenza su un fattore di verosimiglianza che dipenda solo dal parametro di interesse τ . Tale riduzione di complessità del problema inferenziale risulta tanto più vantaggiosa quanto maggiore è la dimensione del parametro di disturbo ζ , soprattutto se la perdita di informazione su τ ad essa associata è nulla o trascurabile.

Una qualunque funzione dipendente soltanto dai dati e dal parametro di interesse che si comporti sotto determinati aspetti come una verosimiglianza in senso proprio si definisce **pseudo-verosimiglianza**. Gli aspetti a cui si fa riferimento sono le usuali proprietà della verosimiglianza in condizioni di regolarità, ossia *score* con media nulla, s.m.v. con distribuzione asintotica normale, test del rapporto di verosimiglianza con distribuzione asintotica nulla di tipo χ^2 e così via.

È possibile distinguere due situazioni.

1. **Verosimiglianza propria.** La pseudo-verosimiglianza è ottenuta a partire da un modello statistico definito come riduzione del modello \mathcal{F} originario, tipicamente per marginalizzazione o condizionamento.
2. **Verosimiglianza non propria.** Non è definito un modello ridotto da cui la pseudo-verosimiglianza possa pensarsi dedotta in via diretta, ossia la pseudo-verosimiglianza non è ottenuta da una fattorizzazio-

ne del tipo (2.1). Tutte le proprietà vanno quindi indagate caso per caso.

Seguono alcuni esempi relativi alla precedente classificazione.

2.4.1 Verosimiglianza marginale e condizionata

Il primo passo da effettuare nella ricerca di una pseudo-verosimiglianza propria è quello di identificare delle fattorizzazioni della forma (2.1) in cui almeno un fattore dipenda solo da τ .

Due casi particolari di questa fattorizzazione si hanno quando la statistica v è sia parzialmente sufficiente per τ , sia parzialmente costante in distribuzione per ζ , ossia vale la fattorizzazione

$$f_{U,V}(u, v; \tau, \zeta) = f_V(v; \tau) f_{U|V=v}(u; v, \zeta), \quad (2.2)$$

oppure quando la statistica u è sia parzialmente costante in distribuzione per τ , sia parzialmente sufficiente per ζ , ossia vale la fattorizzazione

$$f_Y(y; \tau, \zeta) = f_U(u; \zeta) f_{Y|U=u}(y; u, \tau). \quad (2.3)$$

Se vale la (2.2) o la (2.3), la funzione di verosimiglianza $L(\theta)$ mostra una fattorizzazione del tipo

$$L(\theta) = L_1(\tau) L_2(\zeta). \quad (2.4)$$

Una funzione di verosimiglianza $L(\theta)$ per cui vale la (2.4) è detta **con parametri separabili**. L'informazione su τ è catturata interamente dal fattore $L_1(\tau)$ e l'inferenza su τ può procedere separatamente da quella su ζ .

Nella maggior parte delle applicazioni tuttavia non si dispone di funzioni di verosimiglianza con parametri separabili, per cui si cerca perlomeno di basare l'inferenza su un fattore di verosimiglianza dipendente solo da τ , trascurando altri fattori dipendenti sia da τ che da ζ per i quali valga però la proprietà di **assenza di informazione sul parametro di interesse utilmente estraibile**. Tale proprietà viene indicata anche come **non informatività parziale in senso esteso** (cfr. Pace e Salvan, 1996, § 4.2.3).

In quest'ambito risultano particolarmente interessanti le seguenti fattorizzazioni che conducono alle definizioni di verosimiglianza marginale e condizionata.

Sia y in corrispondenza biunivoca con (u, v) , o, più in generale, sia (u, v) statistica sufficiente per θ , allora l'ultimo fattore della (2.1) può essere trascurato, in quanto privo di informazione su θ . Pertanto la (2.1) può essere riscritta come

$$f_{U,V}(u, v; \tau, \zeta) = f_V(v; \tau, \zeta) f_{U|V=v}(u; v, \tau, \zeta).$$

Nel caso in cui v sia parzialmente costante in distribuzione per ζ , la densità congiunta di (U, V) può essere scomposta come

$$f_{U,V}(u, v; \tau, \zeta) = f_V(v; \tau) f_{U|V=v}(u; v, \tau, \zeta); \quad (2.5)$$

se il fattore di verosimiglianza corrispondente a $f_{U|V}(\cdot)$ è trascurabile, si può basare l'inferenza su τ sul modello $f_V(v; \tau)$ ottenuto per marginalizzazione a V . La funzione di verosimiglianza corrispondente

$$L_M(\tau) = L_M(\tau; v) = f_V(v; \tau)$$

è detta **verosimiglianza marginale** basata su v .

Sia invece u una statistica parzialmente sufficiente per ζ , per cui si ha la fattorizzazione

$$f_Y(y; \tau, \zeta) = f_U(u; \tau, \zeta) f_{Y|U=u}(y; u, \tau), \quad (2.6)$$

allora, purchè il fattore di verosimiglianza corrispondente a $f_U(\cdot)$ sia trascurabile, l'inferenza su τ può essere basata sul modello ridotto ottenuto per condizionamento ad U . La funzione di verosimiglianza corrispondente

$$L_C(\tau) = L_C(\tau; y) = f_{Y|U=u}(y; u, \tau)$$

è detta **verosimiglianza condizionata** ad u .

La possibilità di ottenere fattorizzazioni della forma (2.5) oppure (2.6) è essenzialmente limitata alle famiglie di gruppo e alle famiglie

esponenziali. Quando non è possibile ottenere tale separazione inferenziale, rimane la possibilità di identificare opportune funzioni di pseudo-verosimiglianza non corrispondenti ad una riduzione esatta del modello \mathcal{F} . Un esempio di verosimiglianza non propria è dato dalla verosimiglianza profilo presentata nel § 2.4.2.

Riduzione inferenziale nelle famiglie esponenziali e di posizione e scala

In riferimento alla definizione di famiglia esponenziale data nel § 1.8.1, qualora si consideri la partizione del vettore delle statistiche canoniche nella forma $t(y) = (v(y), u(y))$, con $v(\cdot)$ di dimensione k e $u(\cdot)$ di dimensione $p - k$, valgono importanti risultati sulle distribuzioni marginali e condizionate. Si consideri la corrispondente partizione del parametro naturale $\psi = (\tau, \zeta)$. La densità di un generico elemento della famiglia è

$$f(y; \tau, \zeta) = h(y) \exp\{\tau \cdot v(y) + \zeta \cdot u(y) - K(\tau, \zeta)\},$$

con $(\tau, \zeta) \in \Psi$.

L'inferenza su componenti d'interesse del parametro naturale è assai agevolata dal fatto che la distribuzione marginale di $u(Y)$ è del tipo

$$f_U(u; \tau, \zeta) = h_\tau(u) \exp\{\zeta \cdot u - K_\tau(\zeta)\}$$

per ogni fissato valore di τ , e la distribuzione condizionata di $v(Y)$ dato $u(Y) = u$ è del tipo

$$f_{V|U=u}(v; u, \tau) = h_u(v) \exp\{\tau \cdot v - K_u(\tau)\} \quad (2.7)$$

indipendente da ζ .

Questi risultati hanno implicazioni importanti dal punto di vista della riduzione inferenziale. Si ha infatti che u è parzialmente sufficiente per ζ , poiché $f_{V|U=u}(v; u, \tau, \zeta) = f_{V|U=u}(v; u, \tau)$. Per cui se τ è il parametro di interesse, risulta definita in generale una verosimiglianza condizionata basata sulla (2.7)

$$L_C(\tau) = \exp\{\tau \cdot v - K_u(\tau)\},$$

mediante la quale è possibile fare inferenza sul parametro di interesse.

Anche per la classe delle famiglie di posizione e scala valgono alcune proprietà interessanti dal punto di vista della riduzione inferenziale. Siano y_1, \dots, y_n realizzazioni indipendenti di una variabile casuale con densità appartenente a tale classe, e siano $\bar{y}_n = n^{-1} \sum y_i$, $d_n^2 = \sum (y_i - \bar{y}_n)^2$ la media e la devianza campionarie. Si ha che la statistica $c = \left(\frac{y_1 - \bar{y}_n}{\sqrt{d_n^2}}, \dots, \frac{y_n - \bar{y}_n}{\sqrt{d_n^2}} \right)$ è costante in distribuzione e il condizionamento a tale statistica è opportuno per l'inferenza su (μ, σ) . Inoltre la statistica $c' = (y_1 - \bar{y}_n, \dots, y_n - \bar{y}_n)$ ha distribuzione dipendente solo da σ e pertanto il modello marginale indotto da c' può essere utile per l'inferenza su σ quando μ è trattato come parametro di disturbo.

2.4.2 Verosimiglianza profilo

Una via semplice e di vasta applicabilità per reperire una pseudo-verosimiglianza per il parametro di interesse τ è quella di sostituire, nella verosimiglianza complessiva $L(\tau, \zeta)$, il parametro di disturbo ζ con la sua stima di massima verosimiglianza per τ fissato.

In un modello statistico parametrico con funzione di verosimiglianza $L(\theta)$, in cui $\theta = (\tau, \zeta)$ e τ è d'interesse, si definisce **verosimiglianza profilo** per τ la funzione

$$L_P(\tau) = L(\tau, \hat{\zeta}_\tau),$$

dove $\hat{\zeta}_\tau$ è la s.m.v. di ζ nel sottomodello con τ fissato. Si indica inoltre con $l_P(\tau) = \log L_P(\tau)$ la funzione di **log-verosimiglianza profilo**.

Il valore di $\hat{\zeta}_\tau$ è determinato come soluzione in ζ dell'equazione di verosimiglianza parziale, del sottomodello con τ fissato, $l_\zeta(\tau, \zeta) = 0$, con $l_\zeta(\tau, \zeta) = \frac{\partial l(\tau, \zeta)}{\partial \zeta}$. La soluzione di tale equazione può essere ottenuta per via algebrica o più frequentemente attraverso l'impiego di metodi numerici.

La verosimiglianza profilo $L_P(\tau)$ non è una verosimiglianza propria, tuttavia gode di alcune proprietà interessanti che l'assimilano ad una verosimiglianza propria.

- i. La stima di massima verosimiglianza profilo coincide con la s.m.v. di τ basata su $L(\tau, \zeta)$, $\hat{\tau}$, ossia

$$\sup_{\tau} L_P(\tau) = L_P(\hat{\tau}).$$

Il fatto è evidente dalla definizione di $L_P(\tau)$, poiché $\hat{\zeta}_{\hat{\tau}} = \hat{\zeta}$.

- ii. Il log-rapporto di verosimiglianza profilo coincide con l'usuale log-rapporto di verosimiglianza basato su $L(\tau, \zeta)$ considerato per la verifica d'ipotesi su τ , con ζ ignoto. Si ha infatti

$$\begin{aligned} W_P = W_P(\tau) &= 2 \{l_P(\hat{\tau}) - l_P(\tau)\} \\ &= 2 \{l(\hat{\tau}, \hat{\zeta}) - l(\tau, \hat{\zeta}_{\tau})\}. \end{aligned}$$

Sotto condizioni di regolarità, $W_P(\tau)$ ha distribuzione asintotica nulla χ_k^2 , dove k è la dimensione del parametro d'interesse.

Si possono considerare anche delle versioni asintoticamente equivalenti a W_P , analoghe al caso senza parametri di disturbo, ossia

$$W_{u_P} = W_{u_P}(\tau) = (l_{\tau}(\tau, \hat{\zeta}_{\tau}))^{\top} i^{\tau\tau}(\tau, \hat{\zeta}_{\tau}) l_{\tau}(\tau, \hat{\zeta}_{\tau})$$

e

$$W_{e_P} = W_{e_P}(\tau) = (\hat{\tau} - \tau)^{\top} (i^{\tau\tau}(\tau, \hat{\zeta}_{\tau}))^{-1} (\hat{\tau} - \tau).$$

In tali statistiche la matrice di informazione attesa i può essere sostituita con la matrice di informazione osservata j o \hat{j} , senza alterare il risultato di distribuzione asintotica nulla. Le regioni di accettazione dei tre test riportati sono pertanto asintoticamente simili, ossia con livello asintotico indipendente dal valore di ζ . Un aspetto importante da tenere in considerazione è che solamente le procedure inferenziali basate su W_P e W_{u_P} sono invarianti rispetto alle riparametrizzazioni che non alterano l'interesse.

Se τ è scalare, si possono considerare anche le rispettive versioni unilaterali

$$r_P = r_P(\tau) = \text{sgn}(\hat{\tau} - \tau) \sqrt{W_P},$$

$$\begin{aligned} r_{u_P} &= r_{u_P}(\tau) = l_\tau(\tau, \hat{\zeta}_\tau) (i^{\tau\tau}(\tau, \hat{\zeta}_\tau))^{1/2}, \\ r_{e_P} &= r_{e_P}(\tau) = (\hat{\tau} - \tau) (i^{\tau\tau}(\tau, \hat{\zeta}_\tau))^{-1/2}, \end{aligned}$$

che, sotto τ , si distribuiscono asintoticamente come una normale standard.

iii Dalle considerazioni del punto precedente segue che

$$\hat{T}(y) = \left\{ \tau \in T : W_P(\tau) < \chi_{k;1-\alpha}^2 \right\},$$

costituisce una regione di confidenza per il parametro di interesse τ di livello $1 - \alpha$ per ogni valore di ζ . La regione di confidenza si può basare anche sulle versioni asintoticamente equivalenti di W_P .

Analogamente, quando il parametro τ è scalare, si possono costruire regioni di confidenza per il parametro di interesse di livello $1 - \alpha$ del tipo

$$\hat{T}(y) = \left\{ \tau \in T : -z_{1-\alpha/2} < r_P(\tau) < z_{1-\alpha/2} \right\}.$$

iv. **L'informazione osservata profilo** è definita come

$$j_P(\tau) = -\frac{\partial^2}{\partial\tau\partial\tau'} l_P(\tau) = -\frac{\partial^2}{\partial\tau\partial\tau'} l(\tau, \hat{\zeta}_\tau).$$

È possibile dimostrare che l'inversa dell'informazione osservata profilo è uguale al blocco (τ, τ) dell'inversa della matrice d'informazione osservata complessiva, valutato in $(\tau, \hat{\zeta}_\tau)$. Infatti

$$\frac{\partial}{\partial\tau} l_P(\tau) = l_\tau(\tau, \hat{\zeta}_\tau) + l_\zeta(\tau, \hat{\zeta}_\tau) \frac{\partial}{\partial\tau} \hat{\zeta}_\tau.$$

Dato che $l_\zeta(\tau, \hat{\zeta}_\tau) = 0$, risulta

$$\frac{\partial}{\partial\tau} l_P(\tau) = l_\tau(\tau, \hat{\zeta}_\tau).$$

La matrice delle derivate seconde è data da

$$\frac{\partial^2}{\partial\tau\partial\tau'} l_P(\tau) = l_{\tau\tau}(\tau, \hat{\zeta}_\tau) + l_{\tau\zeta}(\tau, \hat{\zeta}_\tau) \frac{\partial}{\partial\tau} \hat{\zeta}_\tau.$$

Dato che $l_{\zeta}(\tau, \hat{\zeta}_{\tau}) = 0$, anche la sua derivata prima calcolata rispetto a τ risulterà nulla, ossia

$$l_{\zeta\tau}(\tau, \hat{\zeta}_{\tau}) + l_{\zeta\zeta}(\tau, \hat{\zeta}_{\tau}) \frac{\partial}{\partial \tau} \hat{\zeta}_{\tau} = 0$$

da cui

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

ne segue che

$$j_P(\tau) = -(l_{\tau\tau} - l_{\tau\zeta} (l_{\zeta\zeta})^{-1} l_{\zeta\tau}),$$

dove tutte le derivate sono valutate in $(\tau, \hat{\zeta}_{\tau})$. Pertanto

$$(j_P(\tau))^{-1} = j^{\tau\tau}(\tau, \hat{\zeta}_{\tau}).$$

Le proprietà appena elencate rendono la verosimiglianza profilo estremamente interessante anche in considerazione del fatto che, qualora esista una verosimiglianza marginale o condizionata, $L_P(\tau)$ ne rappresenta un'approssimazione del primo ordine. Tuttavia, come già accennato, non si tratta di una verosimiglianza in senso proprio. Ciò si riflette principalmente sul valore atteso della funzione *score* profilo, $\frac{\partial L_P(\tau)}{\partial \tau}$, che non è pari a zero. In situazioni non regolari, come il caso di problemi con parametri incidentali, la non consistenza della componente dello stimatore di massima verosimiglianza relativa a τ è riconducibile al fatto che la funzione *score* profilo ha valore atteso legato direttamente alla numerosità campionaria. Ricorrere a $L_P(\tau)$ equivale infatti a comportarsi come se ζ fosse noto e pari a $\hat{\zeta}_{\tau}$. Ciò può non essere appropriato se i dati sono carenti di informazione su ζ , cosa che si verifica tipicamente se la dimensione di ζ è elevata.

È conveniente quindi compensare tale mancanza di informazione su ζ attraverso dei fattori di correzione della verosimiglianza profilo. Alcuni esempi di tali modificazioni verranno illustrati in seguito.

2.5 Modificazioni della verosimiglianza profilo

In anni recenti sono state studiate varie proposte di modificazione di $l_P(\tau)$, con l'obiettivo di correggere la distorsione della *score* profilo. Si vedano per esempio Barndorff-Nielsen (1983, 1994, 1995), Cox e Reid (1987), Fraser e Reid (1988), McCullagh e Tibshirani (1990). Occorre tener presente che, nonostante le varie modificazioni permettano tutte di ridurre la distorsione della funzione *score* profilo, tale riduzione non riflette in generale un miglioramento nelle proprietà asintotiche usuali.

Nei paragrafi successivi vengono descritte due tipologie di modificazione della verosimiglianza profilo. Nel paragrafo 2.5.1 si presenta la verosimiglianza condizionata approssimata di Cox e Reid (1987), basata sulla scelta di una parametrizzazione ortogonale. Nel paragrafo 2.5.2 è descritta un'ulteriore modificazione della verosimiglianza profilo applicabile a situazioni più generali e a cui si farà ricorso in questa tesi per risolvere il problema dei parametri incidentali. Tale modificazione è stata introdotta da Barndorff-Nielsen (1980, 1983) con l'obiettivo di approssimare opportune verosimiglianze marginali o condizionate qualora esse esistano.

2.5.1 Verosimiglianza condizionata approssimata

Cox e Reid (1987) definiscono una versione modificata della verosimiglianza profilo sfruttando le conseguenze dell'ortogonalità. Se $k = 1$ e τ e ζ sono ortogonali, è possibile dimostrare che $\hat{\zeta}$ è asintoticamente una statistica parzialmente sufficiente per ζ . Si può allora basare l'inferenza su τ sul modello condizionato a $\hat{\zeta}$, che costituisce una pseudo-verosimiglianza più accurata della verosimiglianza profilo. La distribuzione asintotica di $\hat{\zeta}$ è in generale approssimabile con una distribuzione $N(\zeta, j^{\zeta\zeta})$ e, per l'ortogonalità, $j^{\zeta\zeta} \doteq (j_{\zeta\zeta})^{-1}$. Si dimostra che la log-verosimiglianza condizionata a $\hat{\zeta}$ è approssimativamente pari a

$$\log f_{y|\hat{\zeta}}(y; \hat{\zeta}, \tau, \zeta) = c(y) + l(\tau, \zeta) - \frac{1}{2} \log |j_{\zeta\zeta}(\tau, \zeta)| + \frac{1}{2} (\hat{\zeta} - \zeta)^\top (j^{\zeta\zeta})^{-1} (\hat{\zeta} - \zeta). \quad (2.8)$$

Tuttavia la log-verosimiglianza condizionata non risulta indipendente da ζ in quanto $\hat{\zeta}$ è solo asintoticamente parzialmente sufficiente per ζ ed inoltre si è fatto ricorso alla distribuzione approssimata di $\hat{\zeta}$. Una soluzione naturale per eliminare il parametro di disturbo dalla verosimiglianza condizionata è quello di sostituire ζ con $\hat{\zeta}_\tau$. Così facendo, in virtù della (1.8), il contributo dell'ultimo termine della (2.8) risulta trascurabile e si definisce quindi la **log-verosimiglianza condizionata approssimata** come

$$\begin{aligned} l_{CA}(\tau) &= l(\tau, \hat{\zeta}_\tau) - \frac{1}{2} \log |j_{\zeta\zeta}(\tau, \hat{\zeta}_\tau)| \\ &= l_P(\tau) - \frac{1}{2} \log |j_{\zeta\zeta}(\tau, \hat{\zeta}_\tau)|. \end{aligned} \quad (2.9)$$

La (2.9) costituisce una versione penalizzata della log-verosimiglianza profilo, in cui il termine di penalizzazione tiene conto dell'informazione sul parametro di disturbo al variare di τ .

La log-verosimiglianza condizionata approssimata presenta tuttavia degli inconvenienti non indifferenti. Innanzitutto, è necessario disporre di una parametrizzazione ortogonale per esplicitare la (2.9) ed è pertanto limitata usualmente al caso in cui τ è scalare. Inoltre $l_{CA}(\tau)$ non gode della proprietà di invarianza rispetto a riparametrazioni che non alterano l'interesse.

Un possibile modo per ovviare a queste difficoltà è quello di utilizzare la verosimiglianza profilo modificata proposta da Barndorff-Nielsen (1980, 1983).

2.5.2 Verosimiglianza profilo modificata

La **verosimiglianza profilo modificata** è definita come

$$L_M(\tau) = L_P(\tau) M(\tau), \quad (2.10)$$

con $M(\tau)$ fattore di aggiustamento pari a

$$M(\tau) = |l_{\zeta; \hat{\zeta}}(\tau, \hat{\zeta}_\tau; \hat{\tau}, \hat{\zeta}, a)|^{-1} |j_{\zeta\zeta}(\tau, \hat{\zeta}_\tau; \hat{\tau}, \hat{\zeta}, a)|^{1/2}. \quad (2.11)$$

Si osservi che nel primo determinante della (2.11) compare la derivata mista $l_{\zeta;\hat{\zeta}}(\tau, \hat{\zeta}_\tau; \hat{\tau}, \hat{\zeta}, a) = \frac{\partial^2}{\partial \zeta \partial \hat{\zeta}} l(\tau, \hat{\zeta}_\tau; \hat{\tau}, \hat{\zeta}, a)$, con a statistica ancillare e utilizzando la versione $l(\theta; \hat{\theta}, a)$ della log-verosimiglianza. Quantità come $l_{\zeta;\hat{\zeta}}$, che coinvolgono derivate rispetto a componenti di $\hat{\theta}$, sono dette **derivate nello spazio campionario**.

Per il calcolo di $L_M(\tau)$ è necessaria quindi l'esplicitazione di una statistica ancillare. La (2.10) può essere riscritta nella forma

$$L_M(\tau) = L_P(\tau) D(\tau) \left| j_{\zeta\zeta}(\tau, \hat{\zeta}_\tau; \hat{\tau}, \hat{\zeta}, a) \right|^{-1/2}, \quad (2.12)$$

con

$$D(\tau) = \frac{\left| j_{\zeta\zeta}(\tau, \hat{\zeta}_\tau; \hat{\tau}, \hat{\zeta}, a) \right|}{\left| l_{\zeta;\hat{\zeta}}(\tau, \hat{\zeta}_\tau; \hat{\tau}, \hat{\zeta}, a) \right|}. \quad (2.13)$$

Derivando $l_{\zeta}(\tau, \hat{\zeta}_\tau; \hat{\tau}, \hat{\zeta}, a) = 0$ rispetto $\hat{\zeta}$, si mostra immediatamente che

$$D(\tau) = \left| \frac{\partial \hat{\zeta}_\tau}{\partial \hat{\zeta}} \right|^{-1}.$$

Data l'espressione della verosimiglianza profilo modificata (2.10), si ottiene immediatamente la log-verosimiglianza profilo modificata, definita come

$$l_M(\tau) = \log(L_M(\tau)) = l_P(\tau) + \log \frac{\left| j_{\zeta\zeta}(\tau, \hat{\zeta}_\tau; \hat{\tau}, \hat{\zeta}, a) \right|^{1/2}}{\left| l_{\zeta;\hat{\zeta}}(\tau, \hat{\zeta}_\tau; \hat{\tau}, \hat{\zeta}, a) \right|}.$$

Per il calcolo della log-verosimiglianza profilo modificata occorre conoscere le stime di massima verosimiglianza globali $(\hat{\tau}, \hat{\zeta})$ e vincolate $(\tau, \hat{\zeta}_\tau)$, il blocco (ζ, ζ) della matrice di informazione osservata, ossia $j_{\zeta\zeta}$, e la derivata della log-verosimiglianza rispetto ζ e $\hat{\zeta}$, ossia $l_{\zeta;\hat{\zeta}}$. Il calcolo diretto di $l_{\zeta;\hat{\zeta}}$ è possibile soltanto per particolari classi di modelli; nei casi in cui questo non è possibile, si ricorre a delle semplici approssimazioni, alcune delle quali verranno descritte in seguito.

Per una famiglia esponenziale multiparametrica con densità

$$f_{V,U}(v, u; \tau, \zeta) = h(v, u) \exp\{\tau v + \zeta u - nK(\tau, \zeta)\},$$

in cui vale una fattorizzazione del tipo

$$f_{\hat{\tau}, \hat{\zeta}}(\hat{\tau}, \hat{\zeta}; \tau, \zeta) = f_{\hat{\zeta}}(\hat{\zeta}; \tau, \zeta) f_{\hat{\tau}|\hat{\zeta}}(\hat{\tau}; \tau, \hat{\zeta}),$$

si ha che

$$l(\tau, \zeta) = \tau v + \zeta u - nK(\tau, \zeta),$$

e quindi

$$\begin{aligned} l_{\zeta} &= u - nK_{\zeta}(\tau, \zeta) \\ &= nK_{\zeta}(\hat{\tau}, \hat{\zeta}) - nK_{\zeta}(\tau, \zeta). \end{aligned}$$

La derivata nello spazio campionario $l_{\zeta; \hat{\zeta}}$ è quindi pari a

$$l_{\zeta; \hat{\zeta}} = nK_{\zeta\zeta}(\hat{\tau}, \hat{\zeta}) = \hat{J}_{\zeta\zeta}.$$

Poiché la quantità $l_{\zeta; \hat{\zeta}}$ è funzione solamente dei dati, essa risulta trascurabile nella modificazione della verosimiglianza profilo. Per cui si ha che la verosimiglianza profilo modificata è pari a

$$L_M(\tau) = L_P(\tau) \left| K_{\zeta\zeta}(\tau, \hat{\zeta}_{\tau}) \right|^{1/2}.$$

Si osservi inoltre che, in generale, esplicitare la (2.10) richiede la specificazione di una statistica ancillare a . Tuttavia se $\hat{\zeta}_{\tau} = \hat{\zeta}$, la matrice $D(\tau)$ è la matrice identità e pertanto, utilizzando la rappresentazione (2.12) della verosimiglianza profilo modificata, si osserva che

$$L_M(\tau) = L_{CA}(\tau), \quad (2.14)$$

con $L_{CA}(\tau)$ funzione di verosimiglianza condizionata approssimata di Cox e Reid. Se i parametri τ e ζ sono ortogonali, la relazione (2.14) vale in via approssimata.

La verosimiglianza profilo modificata possiede delle proprietà che la rendono interessante. Innanzitutto, a differenza della verosimiglianza condizionata approssimata, è invariante rispetto a riparametrazioni che non alterano l'interesse e non richiede la specificazione di una parametrizzazione ortogonale. Inoltre, il fattore di modificazione $M(\tau)$ è di ordine

$O_p(1)$, per cui $L_P(\tau)$ e $L_M(\tau)$ sono asintoticamente equivalenti al primo ordine. Infine la distorsione della funzione *score* associata a $L_M(\tau)$ è di ordine inferiore rispetto a quella associata a $L_P(\tau)$ di ordine $O(1)$. Un ulteriore aspetto di particolare rilievo è che la verosimiglianza profilo modificata costituisce un'approssimazione sufficientemente accurata per le verosimiglianze proprie per τ , quali la verosimiglianza marginale e condizionata per τ , qualora esse esistano. Per una trattazione più approfondita si veda ad esempio il § 11.6 di Pace e Salvan, 1996.

L'impiego della verosimiglianza profilo modificata non comporta in generale un miglioramento in termini di distribuzione asintotica delle statistiche W e Z , che mantengono un'errore del primo ordine, ma solo in termini inferenziali, riducendo la distorsione della funzione *score*. Tuttavia nei modelli per dati stratificati si possono ottenere dei miglioramenti anche in termini di proprietà asintotiche.

Sia q il numero di strati ed m la numerosità di ciascuno strato, si ha che per q ed m che divergono, la distorsione della funzione *score* associata alla verosimiglianza profilo si accumula tra gli strati. Essa risulta pertanto di ordine $O(q)$, e condizione sufficiente per ottenere gli usuali risultati asintotici è che m cresca più rapidamente di q . La capacità della verosimiglianza profilo modificata di ridurre la distorsione della funzione *score* in ciascuno strato permette di ottenere delle proprietà asintotiche migliori. In particolare, una condizione sufficiente per ottenere gli usuali risultati asintotici è che m cresca più rapidamente di $q^{1/3}$. Inoltre, a prescindere da q ed m , la distorsione di $\hat{\tau}$ è di ordine $O(m^{-1})$, mentre la distorsione di $\hat{\tau}_M$ è di ordine $O(m^{-2})$ (Sartori, 2003).

Approssimazioni per la verosimiglianza profilo modificata

Si è visto in precedenza che la verosimiglianza profilo modificata possiede diverse proprietà che la rendono particolarmente interessante; tuttavia non sempre è possibile calcolarla direttamente. In particolare la quantità $l_{\zeta; \hat{\zeta}}$ che compare nel termine $D(\tau)$ nella (2.13) richiede che il contributo

dei dati alla log-verosimiglianza sia espresso tramite $(\hat{\tau}, \hat{\zeta}, a)$, con a statistica ancillare. Il calcolo è immediato nei modelli appartenenti alle famiglie esponenziale, in cui non è richiesta l'esplicitazione di una statistica ancillare. In altri casi non è possibile calcolare la verosimiglianza profilo modificata e si utilizzano quindi delle approssimazioni.

In presenza di parametri ortogonali, una prima approssimazione per la verosimiglianza profilo modificata è data dalla verosimiglianza approssimata condizionata di Cox e Reid già presentata nel § 2.5.1. Si è già visto che in questo caso la matrice $D(\tau)$, matrice identità, è trascurabile, per cui il problema di calcolare la quantità $l_{\zeta;\hat{\zeta}}$ non si pone. Se non si dispone invece di una parametrizzazione ortogonale, è necessario trovare un'approssimazione per la derivata nello spazio campionario $l_{\zeta;\hat{\zeta}}$ presente nella (2.13).

Un primo approccio a tale problema, proposto da Fraser e Reid (1989) e riportato in Severini (2000, § 9.5.3), consiste nell'approssimare $l_{\zeta;\hat{\zeta}}(\hat{\theta}_\tau)$ con la quantità

$$\tilde{l}_{\zeta;\hat{\zeta}}(\hat{\theta}_\tau) = l_{\zeta;y}(\hat{\theta}_\tau) \hat{V}_\zeta \left(\frac{\partial \hat{\zeta}}{\partial y} \hat{V}_\zeta \right)^{-1},$$

dove $l_{\zeta;y}$ è una matrice contenente le derivate della log-verosimiglianza rispetto a ζ e ai dati $y = (y_1, \dots, y_n)$ e la matrice \hat{V}_ζ è data da

$$\hat{V}_\zeta = \begin{pmatrix} -\frac{\partial F(y_1; \hat{\theta}) / \partial \hat{\zeta}}{f(y_1; \hat{\theta})} \\ \vdots \\ -\frac{\partial F(y_n; \hat{\theta}) / \partial \hat{\zeta}}{f(y_n; \hat{\theta})} \end{pmatrix},$$

con F_i e f_i rispettivamente funzione di ripartizione e di densità della generica variabile casuale $Y_i, i = 1, \dots, n$.

Pertanto, trascurando i termini che non dipendono da τ , un'approssimazione per $l_M(\tau)$ è data da

$$\tilde{l}_M(\tau) = l_P(\tau) + \log \frac{|j_{\zeta\zeta}(\hat{\theta}_\tau)|^{1/2}}{|l_{\zeta;y}(\hat{\theta}_\tau) \hat{V}_\zeta|}.$$

Un approccio alternativo, proposto da Severini (1998) e riportato in Severini (2000, §9.5.4), consiste invece nell'approssimare $l_{\zeta;\hat{\zeta}}(\hat{\theta}_\tau)$ con la quantità

$$\bar{l}_{\zeta;\hat{\zeta}}(\hat{\theta}_\tau) = I_{\zeta;\zeta}(\hat{\theta}_\tau; \hat{\theta}),$$

dove

$$I_{\zeta;\zeta}(\theta; \theta_0) = E_{\theta_0} \{ l_\zeta(\theta) l_\zeta(\theta_0)^\top \} \quad (2.15)$$

è una particolare covarianza tra componenti della funzione *score*.

Pertanto si può approssimare la funzione di log-verosimiglianza profilo modificata $l_M(\tau)$ con

$$\bar{l}_M(\tau) = l_P(\tau) + \log \frac{|j_{\zeta\zeta}(\hat{\theta}_\tau)|^{1/2}}{|I_{\zeta;\zeta}(\hat{\theta}_\tau; \hat{\theta})|}.$$

Per un'analisi maggiormente approfondita circa le approssimazioni della verosimiglianza profilo modificata, si veda il § 9.5 di Severini, 2000.

Capitolo 3

Applicazioni a modelli per dati di panel

3.1 Introduzione

Il presente capitolo ha l'obiettivo di sostenere le giustificazioni teoriche all'uso della verosimiglianza profilo modificata, discusse nel capitolo precedente.

Si cercherà di dimostrare empiricamente, mediante alcuni studi di simulazione relativi a problemi di Neyman e Scott, che l'uso della verosimiglianza profilo modificata può fornire un sostanziale miglioramento rispetto a quello della profilo.

Nel paragrafo 3.2 si riprendono la definizione di parametri incidentali e i problemi ad essi associati. Successivamente si fissa la notazione per le quantità di verosimiglianza relative a modelli per dati stratificati e si illustrano le relative proprietà; in particolare la separabilità della log-verosimiglianza rispetto ai parametri di disturbo e la possibilità di calcolare la modificazione della log-verosimiglianza profilo come somma delle modificazioni dei singoli strati. Tali proprietà sono entrambe giustificate dal fatto che i parametri incidentali sono specifici di ciascuno strato e gli strati sono assunti indipendenti tra di loro.

Nei paragrafi successivi vengono riportati due esempi, appartenenti alla categoria dei problemi di Neyman e Scott, in cui la presenza di parametri incidentali rende la verosimiglianza profilo uno strumento inadatto all'inferenza sul parametro di interesse.

Per ciascuno degli esempi riportati, ad un'analisi condotta mediante la verosimiglianza profilo, ne verrà affiancata una basata sull'uso della verosimiglianza profilo modificata.

Il primo esempio, affrontato nel § 3.3, parte da un modello per dati discreti riportato da Sartori (2001, Esempio 2.7). Tale modello è stato successivamente utilizzato in un esempio proposto da Greene (2004) e ripreso da Sartori (2005). In questa tesi si vuole sviluppare ed approfondire lo studio considerato da Sartori (2005).

Il modello considerato è un modello per dati binari appartenente alla categoria dei modelli lineari generalizzati. L'oggetto dell'inferenza è il parametro di interesse bivariato $\tau = (\beta, \delta)$, dove β e δ rappresentano i coefficienti rispettivamente di una variabile continua ed una discreta presenti nel predittore lineare associato al modello.

Il secondo esempio, affrontato nel § 3.4, fa riferimento ad un modello autoregressivo del primo ordine con intercetta variabile. In questo caso si ha che le osservazioni y_{i1}, \dots, y_{iT} si susseguono secondo un ordine temporale e la distribuzione di ciascuna osservazione dipende solo dal valore di quella immediatamente precedente. Il modello è caratterizzato da osservazioni dipendenti all'interno di ciascuno strato, ma assunte indipendenti da uno strato all'altro.

In questo esempio il parametro di interesse è $\tau = (\rho, \sigma^2)$, dove σ^2 rappresenta la varianza condizionata del processo autoregressivo, mentre ρ costituisce il coefficiente che lega una generica osservazione y_{it} a quella immediatamente precedente, ossia y_{it-1} , e misura pertanto il livello di dipendenza tra le due osservazioni. Lo studio vuole analizzare come si comportano la verosimiglianza profilo e profilo modificata anche al variare del valore assunto dal parametro ρ .

3.2 I parametri incidentali

I problemi che coinvolgono parametri incidentali risultano particolarmente interessanti essenzialmente per due motivi. Innanzitutto, un numero elevato di parametri, permette in generale di ottenere una modellazione più flessibile dei vari aspetti del fenomeno che si deve rappresentare; ciò comporta una maggiore coerenza, qualora la scelta della classe parametrica effettuata risulti adeguata. Inoltre, situazioni caratterizzate da un parametro di disturbo di dimensione elevata, possono aiutare maggiormente ad evidenziare il miglioramento raggiungibile attraverso l'impiego della verosimiglianza profilo modificata al posto della verosimiglianza profilo. Nell'ambito di questa tesi, si vuole prestare maggiore attenzione proprio a quest'ultimo aspetto.

Nel seguito si considerano modelli per osservazioni raggruppate indipendenti del tipo

$$Y_{ij} \sim f_{ij}(y_{ij}; \tau, \zeta_i),$$

dove $i = 1, \dots, q$, con q numero di strati e $j = 1, \dots, m_i$, con m_i numero di osservazioni dello strato i -mo. Per semplicità si assume che tutti gli strati abbiano la stessa dimensione, ossia $m_i = m$. I parametri coinvolti sono τ , parametro di interesse e ζ_1, \dots, ζ_q , parametri di disturbo incidentali. Si osservi che la notazione f_{ij} è volta a includere anche i modelli di regressione in cui $f_{ij}(y_{ij}; \tau, \zeta_i) = f_{ij}(y_{ij}; \tau, \zeta_i, x_{ij})$, con x_{ij} variabili esplicative note.

La funzione di log-verosimiglianza può essere scritta come

$$l(\theta) = \sum_{i=1}^q l^i(\tau, \zeta_i), \quad (3.1)$$

dove

$$l^i(\tau, \zeta_i) = \sum_{j=1}^m \log f_{ij}(y_{ij}; \tau, \zeta_i)$$

è la funzione di log-verosimiglianza relativa all' i -mo strato. Come si può osservare dalla (3.1), la funzione di log-verosimiglianza è separabile rispetto ai parametri di disturbo. Questo perché gli strati sono indipendenti

e il parametro ζ_i è specifico dello strato i -mo. La separabilità dei parametri di disturbo fa sì che la stima di massima verosimiglianza vincolata di ζ per τ fissato, $\hat{\zeta}_\tau$, sia soluzione delle q equazioni di verosimiglianza indipendenti associate agli strati. Pertanto la log-verosimiglianza profilo può essere scritta come somma delle q funzioni di log-verosimiglianza profilo relative agli strati, ossia

$$l_P(\tau) = l(\tau, \hat{\zeta}_\tau) = \sum_{i=1}^q l_P^i(\tau),$$

con $l_P^i(\tau) = l^i(\tau, \hat{\zeta}_{i\tau})$ log-verosimiglianza profilo relativa all' i -mo strato.

Per quanto riguarda invece gli aggiustamenti della log-verosimiglianza profilo, in generale si considerano modificazioni della forma

$$l_M(\tau) = l_P(\tau) + M'(\tau),$$

in cui il termine di modificazione $M'(\tau)$ soddisfa la condizione

$$M'(\tau) = \sum_{i=1}^q M'_i(\tau), \quad (3.2)$$

con $M'_i(\tau)$ modificazione relativa all' i -mo strato.

La proprietà (3.2) è soddisfatta dalla maggior parte delle modificazioni proposte in letteratura, in particolare da quella fornita da Barndorff-Nielsen (1994, 1995) introdotta nel § 2.5.2 e a cui si farà ricorso nelle simulazioni. Per modelli appartenenti alla famiglia esponenziale, la verosimiglianza profilo modificata soddisfa la (3.2), mentre per modelli più generali vengono utilizzate delle approssimazioni per $M'(\tau)$, come quelle introdotte sempre nel § 2.5.2. È immediato verificare che le verosimiglianze profilo modificate approssimate soddisfano anch'esse la (3.2) (Sartori, 2001).

In seguito si indica con $\hat{\tau}_M$ il valore che massimizza la quantità $l_M(\tau)$, mentre il test del log-rapporto di verosimiglianza viene indicato con W_M e, analogamente a quello relativo alla verosimiglianza profilo riportato nel § 2.4.2, è dato da

$$W_M(\tau) = 2 \{l_M(\hat{\tau}_M) - l_M(\tau)\}.$$

Il test r per la log-verosimiglianza profilo modificata viene invece indicato con r_M ed è pari a

$$r_M(\tau) = \text{sgn}(\hat{\tau}_M - \hat{\tau})\sqrt{W_M}.$$

3.3 Un modello per dati binari

3.3.1 Specificazione del modello

Sia $Y \in \mathcal{Y}$ la variabile risposta, descritta da un modello lineare generalizzato, con funzione di densità pari a

$$f(y; \eta, \phi) = \exp[\{y \theta(\eta) - K(\eta)\}/a(\phi) + c(y, \phi)], \quad (3.3)$$

dove a, c, θ e K sono funzioni date, $\phi > 0$ è il parametro di dispersione e η è il predittore lineare. Si supponga di osservare un campione stratificato y_{ij} , $i = 1, \dots, q$, $j = 1, \dots, m_i$, in cui ciascuna osservazione y_{ij} è realizzazione di una variabile casuale con densità (3.3) con ϕ noto e predittore lineare

$$\eta_{ij} = \alpha_i + x_{ij}^T \beta,$$

con $\alpha_i, \beta \in \mathbb{R}$.

In questo contesto, il parametro di interesse è $\tau = \beta$, mentre $\alpha_1, \dots, \alpha_q$ sono i parametri di disturbo incidentali.

Per eliminare i parametri di disturbo $\alpha_1, \dots, \alpha_q$ dalla funzione di verosimiglianza, si può ricorrere alla verosimiglianza condizionata, ma solamente nel caso in cui la funzione di legame utilizzata sia quella canonica. Si può dimostrare che lo stimatore di β così ottenuto ha proprietà migliori rispetto allo stimatore non condizionato $\hat{\beta}$, soprattutto quando il numero di strati q è grande rispetto alla dimensione m_i di ciascuno strato. Qualora la funzione di legame non sia quella canonica, una possibile soluzione al problema dei parametri incidentali è data dalla verosimiglianza profilo modificata, che approssima accuratamente la verosimiglianza condizionata, con il vantaggio di poter essere usata in modelli più generali.

Si consideri ora un esempio particolare della (3.3), in cui Y è una variabile binaria con probabilità di successo $G(\eta)$ e G è una funzione di ripartizione assegnata, con corrispondente funzione di densità g .

La log-verosimiglianza profilo per β è data da

$$l_P(\beta) = \sum_{i=1}^q l_P^i(\beta), \quad (3.4)$$

con

$$l_P^i(\beta) = l^i(\beta, \hat{\alpha}_{i\tau}) = \sum_{j=1}^{m_i} [y_{ij} \log G(\tilde{\eta}_{ij}) + (1 - y_{ij}) \log \{1 - G(\tilde{\eta}_{ij})\}],$$

e $\tilde{\eta}_{ij} = \eta_{ij}(\beta, \hat{\alpha}_{i\tau})$. La quantità $l_P^i(\beta)$ rappresenta la log-verosimiglianza profilo dello strato i -mo, mentre $\tilde{\eta}_{ij}$ è il predittore lineare calcolato in $\hat{\alpha}_{i\tau}$ e quindi dipendente dal solo parametro di interesse β .

La log-verosimiglianza profilo modificata, calcolata utilizzando l'approssimazione (2.15), è ottenuta partendo dalla log-verosimiglianza profilo indicata nella (3.4) a cui si somma il termine di correzione $M'(\tau) = \sum_{i=1}^q M'_i(\tau)$, con

$$M'_i(\tau) = \frac{1}{2} \log | j_{\alpha_i \alpha_i}(\beta, \hat{\alpha}_{i\tau}) | - \log | I_{\alpha_i \alpha_i}^i(\beta, \hat{\alpha}_{i\tau}; \hat{\beta}, \hat{\alpha}_i) | \quad (3.5)$$

correzione per ciascuno strato. Nell'esempio considerato, le quantità nella (3.5) sono pari a

$$I_{\alpha_i \alpha_i}^i(\beta, \hat{\alpha}_{i\tau}; \hat{\beta}, \hat{\alpha}_i) = \sum_{j=1}^{m_i} \frac{g(\hat{\eta}_{ij}) g(\tilde{\eta}_{ij})}{\{1 - G(\tilde{\eta}_{ij})\} G(\tilde{\eta}_{ij})}, \quad (3.6)$$

$$j_{\alpha_i \alpha_i}(\beta, \hat{\alpha}_{i\tau}) = I_{\alpha_i \alpha_i}^i(\beta, \hat{\alpha}_{i\tau}; \beta, \hat{\alpha}_{i\tau}) - \sum_{j=1}^{m_i} \{y_{ij} - G(\tilde{\eta}_{ij})\} c_{ij}, \quad (3.7)$$

$$c_{ij} = \left[\frac{g'(\tilde{\eta}_{ij})}{\{1 - G(\tilde{\eta}_{ij})\} G(\tilde{\eta}_{ij})} - \frac{g^2(\tilde{\eta}_{ij}) \{1 - 2G(\tilde{\eta}_{ij})\}}{\{1 - G(\tilde{\eta}_{ij})\}^2 G(\tilde{\eta}_{ij})^2} \right], \quad (3.8)$$

con $\hat{\eta}_{ij} = \eta_{ij}(\hat{\beta}, \hat{\alpha}_i)$, predittore lineare calcolato nelle stime di massima verosimiglianza.

L'espressione risultante della log-verosimiglianza profilo modificata $l_M(\beta)$ può essere massimizzata numericamente ottenendo così uno stimatore di massima verosimiglianza modificato per il parametro di interesse.

Funzione di legame canonica e semplificazioni

Qualora si consideri il caso con funzione di legame canonica, nell'esempio specifico la funzione logistica, si ottengono rilevanti semplificazioni nella specificazione del modello e in particolare nella modificazione della verosimiglianza profilo, in cui alcune componenti risultano nulle o trascurabili.

Si consideri quindi il modello per dati binari definito precedentemente con funzione di legame logistica. La probabilità di successo $G(\eta)$ associata alla variabile Y assume la forma

$$G(\eta_{ij}) = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}},$$

mentre la relativa funzione di densità $g(\eta)$ ha la seguente espressione

$$g(\eta_{ij}) = G'(\eta_{ij}) = \frac{e^{\eta_{ij}}}{(1 + e^{\eta_{ij}})^2},$$

e la derivata prima della funzione di densità, $g'(\eta)$ è pari a

$$g'(\eta_{ij}) = \frac{e^{\eta_{ij}}(1 - e^{\eta_{ij}})}{(1 + e^{\eta_{ij}})^3} = \frac{1 - e^{\eta_{ij}}}{1 + e^{\eta_{ij}}} g(\eta_{ij}).$$

Impiegando queste quantità nelle formule (3.6), (3.7) e (3.8), si ottiene che il termine $I_{\alpha;\alpha}^i(\beta, \hat{\alpha}_{i\tau}; \hat{\beta}, \hat{\alpha}_i)$ non dipende da $\tilde{\eta}_{ij}$ e risulta quindi trascurabile ai fini dell'inferenza su β , mentre la costante c_{ij} si annulla. Pertanto la funzione di log-verosimiglianza profilo modificata è data da

$$l_M(\beta) = l_P(\beta) + \frac{1}{2} \sum_{i=1}^q \log |j_{\alpha_i, \alpha_i}(\beta, \hat{\alpha}_{i\tau})|,$$

con

$$j_{\alpha_i, \alpha_i}(\beta, \hat{\alpha}_{i\tau}) = I_{\alpha;\alpha}^i(\beta, \hat{\alpha}_{i\tau}; \beta, \hat{\alpha}_{i\tau}).$$

3.3.2 Simulazione

Funzione di legame logit

Per valutare l'accuratezza delle procedure inferenziali basate sulla verosimiglianza profilo modificata, soprattutto quando l'informazione disponibile sul parametro di disturbo è ridotta, è stato eseguito uno studio di

simulazione, volto ad approfondire quello considerato da Sartori (2005). In quest'ultimo, l'autore si limita a confrontare le distorsioni degli stimatori $\hat{\tau}$ e $\hat{\tau}_M$ per diversi valori di m . In questa tesi si vuole invece verificare come si comportano la verosimiglianza profilo e profilo modificata al variare della dimensione di ciascuno strato m , ma anche del numero di strati q . L'analisi include il confronto fra gli stimatori di massima verosimiglianza per τ , ottenuti tramite l'impiego della verosimiglianza profilo e profilo modificata, in termini di distorsione, nonché il confronto dei livelli di copertura stimati di regioni e intervalli di confidenza.

Il modello considerato appartiene alla classe dei modelli lineari generalizzati; la variabile risposta binaria ha distribuzione

$$Y_{ij} \sim Bi(1, \pi_{ij}),$$

con probabilità di successo

$$\pi_{ij} = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}}.$$

Il predittore lineare η_{ij} include due variabili esplicative x e d , rispettivamente continua e dicotomica. L'intercetta α_i è specifica di ciascuno strato e rappresenta il parametro di disturbo incidentale. Il parametro d'interesse è $\tau = (\beta, \delta)$ bivariato, con β coefficiente della variabile continua x e δ coefficiente della variabile discreta d . Il predittore lineare assume dunque la forma

$$\eta_{ij} = \alpha_i + \beta x_{ij} + \delta d_{ij}.$$

Nella simulazione, si è posto $\beta = \delta = 1$.

Le variabili coinvolte sono state generate come segue.

$$\begin{aligned} x_{ij} &\sim N(0, 1), \\ d_{ij} &= I[x_{ij} + h_{ij} > 0], \end{aligned}$$

con $h_{ij} \sim N(0, 1)$,

$$\alpha_i \sim N(\sqrt{m} \bar{x}_i, 1),$$

con \bar{x}_i media relativa alla variabile x dello strato i -mo (Greene, 2004).

Nella simulazione, le variabili x_{ij} e d_{ij} e i valori dei parametri di disturbo α_i sono generati e poi mantenuti fissati, mentre le y_{ij} variano da campione a campione. Sono stati considerati i valori $m = (3, 5, 7, 10)$ e $q = (20, 50, 100, 250)$.

I risultati sono riportati in più tabelle. La Tabella 3.1 contiene le stime basate sulla simulazione delle distorsioni degli stimatori $\hat{\beta}$ e $\hat{\delta}$ ottenuti massimizzando la funzione di log-verosimiglianza profilo e le distorsioni degli stimatori $\hat{\beta}_M$ e $\hat{\delta}_M$ ottenuti dalla massimizzazione della log-verosimiglianza profilo modificata. Nella Tabella 3.2 sono invece riportati i livelli di copertura stimati delle regioni di confidenza basate sulle statistiche log-rapporto di verosimiglianza W_P e W_M , con livelli di copertura nominali pari a 0.90, 0.95 e 0.99. Infine, nelle Tabelle 3.3 e 3.4 vengono presentati i valori stimati delle coperture degli intervalli di confidenza alla Wald, calcolati secondo la (1.6), per le singole componenti del parametro di interesse $\tau = (\beta, \delta)$, sempre con livelli di copertura nominali pari a 0.90, 0.95 e 0.99. Gli intervalli, così calcolati, sono centrati nelle stime di massima verosimiglianza ottenute tramite la verosimiglianza profilo e profilo modificata, mentre gli errori standard degli stimatori di massima verosimiglianza, necessari nel determinare i limiti degli intervalli, sono stati calcolati facendo riferimento alla distribuzione approssimata dello s.m.v. riportata nella (1.4). Pertanto gli errori standard per $\hat{\beta}$ e $\hat{\delta}$ sono dati dalla radice degli elementi presenti sulla diagonale principale dell'inversa della matrice di informazione osservata \hat{j}^{-1} . Gli errori standard per $\hat{\beta}_M$ e $\hat{\delta}_M$ sono calcolati nello stesso modo, ma utilizzando l'inversa della matrice di informazione osservata \hat{j}_M^{-1} relativa alla verosimiglianza profilo modificata.

Nelle Figure 3.1 e 3.2 sono rappresentate le curve di livello associate alla log-verosimiglianza profilo relativa e alla log-verosimiglianza profilo modificata relativa di un singolo campione. Sono stati considerati due casi estremi tra quelli presentati nella simulazione. In particolare, nella Figura

3.1 il campione comprende 250 strati di numerosità 3 ed è il caso in cui il rapporto tra numerosità degli strati m e numero di strati q è minimo. Nella Figura 3.2, invece, il campione è composto da soli 20 strati di numerosità 10, caso in cui il rapporto tra m ed q è massimo.

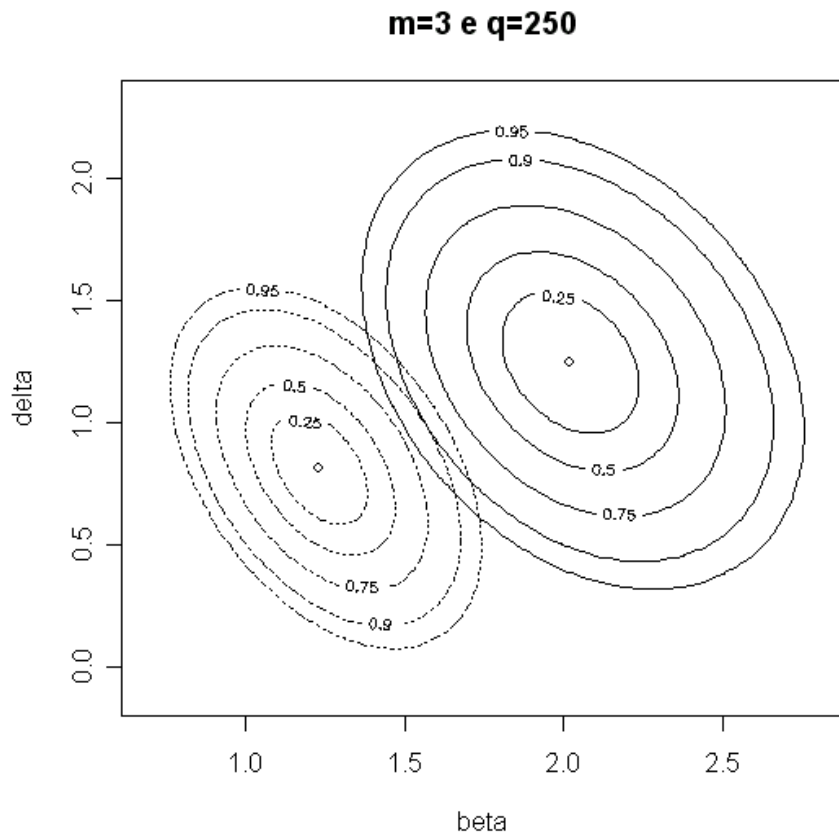


Figura 3.1: Modello per dati binari e funzione di legame logit. Curve di livello riferite alla log-verosimiglianza profilo relativa (linea continua) e alla log-verosimiglianza profilo modificata relativa (linea tratteggiata), calcolate su un campione di $q = 250$ strati di dimensione $m = 3$.

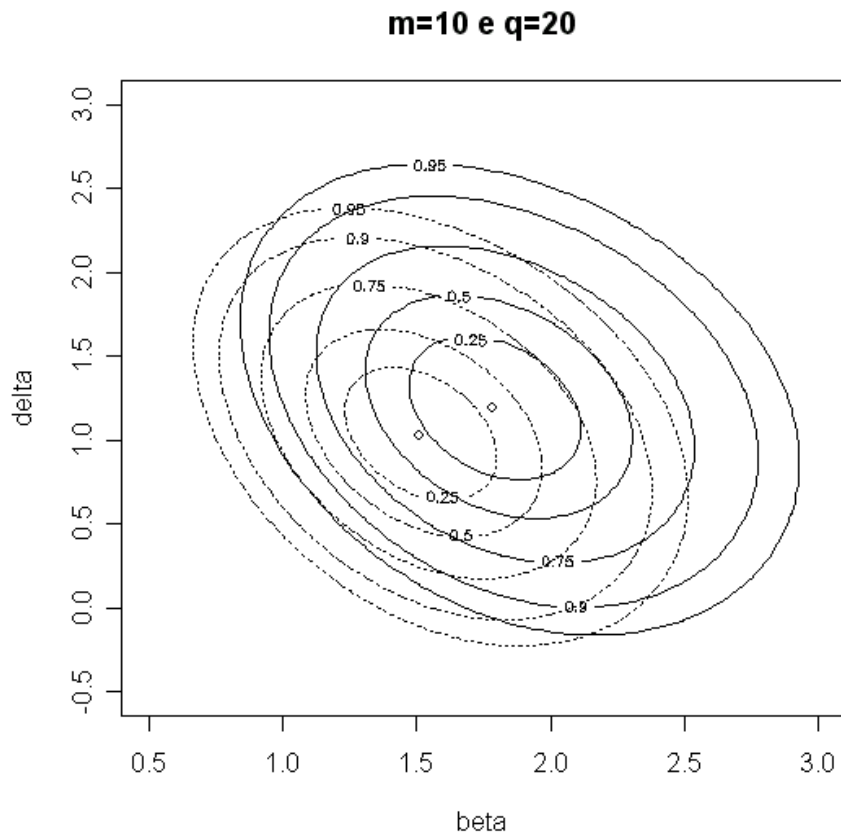


Figura 3.2: Modello per dati binari e funzione di legame logit. Curve di livello riferite alla log-verosimiglianza profilo relativa (linea continua) e alla log-verosimiglianza profilo modificata relativa (linea tratteggiata), calcolate su un campione di $q = 20$ strati di dimensione $m = 10$.

		$m = 3$		$m = 5$		$m = 7$		$m = 10$	
		P	M	P	M	P	M	P	M
$q = 20$	$\hat{\beta}$	0.779	0.109	0.461	0.074	0.336	0.097	0.226	0.058
	$\hat{\delta}$	0.710	0.152	0.682	0.152	0.205	0.002	0.090	-0.033
$q = 50$	$\hat{\beta}$	0.827	-0.008	0.456	0.085	0.256	0.042	0.182	0.030
	$\hat{\delta}$	0.921	0.153	0.436	0.090	0.246	0.041	0.140	0.012
$q = 100$	$\hat{\beta}$	0.825	0.043	0.369	0.029	0.216	0.006	0.156	0.015
	$\hat{\delta}$	0.671	0.066	0.358	0.060	0.247	0.052	0.126	-0.001
$q = 250$	$\hat{\beta}$	0.734	0.045	0.330	0.020	0.224	0.016	0.151	0.012
	$\hat{\delta}$	0.682	0.086	0.334	0.045	0.221	0.030	0.124	-0.002

Tabella 3.1: *Modello per dati binari. Risultati della simulazione con $N = 200$ repliche e funzione di legame logit. Distorsioni stimate degli stimatori $\hat{\beta}$ e $\hat{\delta}$ ottenuti tramite la verosimiglianza profilo (P) e la verosimiglianza profilo modificata (M).*

		$m = 3$		$m = 5$		$m = 7$		$m = 10$	
	$1 - \alpha$	P	M	P	M	P	M	P	M
$q = 20$	0.90	0.550	0.930	0.660	0.880	0.780	0.865	0.805	0.895
	0.95	0.645	0.960	0.755	0.925	0.835	0.925	0.875	0.950
	0.99	0.795	1.000	0.885	0.980	0.950	0.985	0.975	0.975
$q = 50$	0.90	0.450	0.925	0.565	0.895	0.680	0.905	0.685	0.900
	0.95	0.590	0.980	0.645	0.950	0.755	0.950	0.770	0.960
	0.99	0.785	0.995	0.825	0.975	0.900	0.990	0.925	0.995
$q = 100$	0.90	0.225	0.895	0.315	0.890	0.460	0.895	0.585	0.925
	0.95	0.310	0.945	0.425	0.960	0.625	0.945	0.710	0.965
	0.99	0.495	0.980	0.710	0.995	0.795	1.000	0.870	0.995
$q = 250$	0.90	0.015	0.940	0.105	0.880	0.180	0.860	0.320	0.920
	0.95	0.035	0.965	0.185	0.950	0.220	0.915	0.415	0.955
	0.99	0.110	0.990	0.350	0.980	0.395	0.995	0.635	0.980

Tabella 3.2: *Modello per dati binari. Risultati della simulazione con $N = 200$ replicazioni e funzione di legame logit. Stime delle probabilità di copertura delle regioni di confidenza per β e δ congiuntamente con livello nominale $1 - \alpha$, basate su W ottenuto tramite la verosimiglianza profilo (P) e la verosimiglianza profilo modificata (M).*

		$m = 3$		$m = 5$		$m = 7$		$m = 10$	
	$1 - \alpha$	P	M	P	M	P	M	P	M
$q = 20$	0.90	0.750	0.910	0.810	0.885	0.810	0.890	0.845	0.895
	0.95	0.800	0.970	0.905	0.935	0.905	0.940	0.920	0.960
	0.99	0.850	0.995	0.980	0.990	0.970	0.990	0.990	0.995
$q = 50$	0.90	0.665	0.930	0.680	0.940	0.770	0.900	0.775	0.925
	0.95	0.770	0.960	0.805	0.975	0.850	0.945	0.890	0.965
	0.99	0.925	1.000	0.955	0.995	0.955	0.995	0.970	0.995
$q = 100$	0.90	0.475	0.895	0.575	0.930	0.670	0.920	0.705	0.925
	0.95	0.590	0.945	0.685	0.955	0.815	0.950	0.795	0.955
	0.99	0.785	0.990	0.890	0.995	0.930	0.990	0.930	1.000
$q = 250$	0.90	0.160	0.945	0.290	0.895	0.380	0.890	0.480	0.875
	0.95	0.255	0.975	0.405	0.935	0.460	0.950	0.575	0.940
	0.99	0.455	0.995	0.675	0.995	0.645	0.985	0.770	0.985

Tabella 3.3: *Modello per dati binari. Risultati della simulazione con $N = 200$ repliche e funzione di legame logit. Stime delle probabilità di copertura dell'intervallo per β con livello nominale $1 - \alpha$, ottenute tramite la verosimiglianza profilo (P) e la verosimiglianza profilo modificata (M).*

		$m = 3$		$m = 5$		$m = 7$		$m = 10$	
	$1 - \alpha$	P	M	P	M	P	M	P	M
$q = 20$	0.90	0.770	0.915	0.810	0.860	0.875	0.910	0.840	0.900
	0.95	0.83	0.965	0.890	0.965	0.910	0.960	0.920	0.940
	0.99	0.960	1.000	0.985	0.990	1.000	1.000	0.980	0.985
$q = 50$	0.90	0.730	0.915	0.790	0.840	0.815	0.890	0.805	0.855
	0.95	0.815	0.960	0.845	0.930	0.895	0.960	0.895	0.905
	0.99	0.965	0.995	0.940	0.995	0.985	0.985	0.975	0.995
$q = 100$	0.90	0.620	0.895	0.730	0.885	0.720	0.915	0.840	0.925
	0.95	0.730	0.935	0.830	0.945	0.840	0.955	0.920	0.965
	0.99	0.930	0.985	0.930	0.995	0.960	0.995	0.980	0.995
$q = 250$	0.90	0.505	0.900	0.580	0.890	0.605	0.880	0.800	0.920
	0.95	0.645	0.955	0.670	0.930	0.700	0.915	0.875	0.970
	0.99	0.820	0.990	0.860	0.980	0.875	0.985	0.960	0.990

Tabella 3.4: *Modello per dati binari. Risultati della simulazione con $N = 200$ replicazioni e funzione di legame logit. Stime delle probabilità di copertura dell'intervallo per δ con livello nominale $1 - \alpha$, ottenute tramite la verosimiglianza profilo (P) e la verosimiglianza profilo modificata (M).*

Osservazioni

Nella Figura 3.1 si può osservare, per un singolo campione composto da 250 strati di numerosità 3, che le linee di livello associate alla verosimiglianza profilo e profilo modificata sono distanti tra loro. In particolare, per la profilo modificata esse sono più vicine al punto $(1, 1)$, che rappresenta il vero valore del parametro di interesse $\tau = (\beta, \delta)$. Le curve di livello della profilo modificata risultano inoltre più concentrate attorno alla stima di massima verosimiglianza $\hat{\tau}_M$.

Nella Figura 3.2 le curve di livello si riferiscono invece ad un campione di soli 20 strati di numerosità 10. Il rapporto tra numerosità degli strati m e numero di strati q è in questo caso elevato e si può osservare come in una situazione simile sia possibile ottenere dei risultati accettabili anche mediante l'impiego della verosimiglianza profilo. Le curve di livello delle due log-verosimiglianze sono infatti entrambe vicine al punto $(1, 1)$ e tendono a sovrapporsi.

Analizzando ora i risultati della simulazione, ed in particolare le distorsioni stimate riportate nella Tabella 3.1, si può notare come, in generale, per ogni combinazione di valori di q ed m considerati, la log-verosimiglianza profilo modificata fornisca stime per β e δ più precise rispetto alla log-verosimiglianza profilo, ossia stime mediamente più vicine al vero valore assunto dai due parametri.

Per entrambe le log-verosimiglianze, si osserva invece che, a parità di q , per valori grandi di m ($m = 7, 10$) le distorsioni sono più basse e quindi le stime dei due parametri sono migliori rispetto ai casi in cui la dimensione di ciascuno strato è ridotta ($m = 3, 5$). Questo miglioramento delle stime associato a valori di m elevati è essenzialmente dovuto al fatto che, essendo la dimensione di ciascuno strato abbastanza grande, l'informazione disponibile utile a stimare i parametri incidentali α_i specifici di ciascuno strato è maggiore rispetto ai casi in cui il valore di m è basso.

Un aspetto importante da prendere in considerazione, sempre in riferimento alla Tabella 3.1, è che la log-verosimiglianza profilo presenta

delle distorsioni elevate soprattutto per $m = 3$, situazione in cui la log-verosimiglianza profilo modificata produce invece delle stime soddisfacenti. Una considerazione analoga può essere fatta anche per le tabelle seguenti, in cui la distinzione tra i risultati ottenuti tramite la verosimiglianza profilo e la profilo modificata per valori di m bassi è ancora più evidente.

Osservando la Tabella 3.2, si può confermare quanto è stato detto per le distorsioni. Le probabilità di copertura stimate relative alla verosimiglianza profilo modificata sono più vicine ai livelli nominali fissati rispetto a quelle della verosimiglianza profilo.

Le probabilità di copertura stimate associate alla log-verosimiglianza profilo sono insoddisfacenti soprattutto quando il numero di strati q è elevato ($q = 100, 250$) rispetto alla numerosità di ciascuno strato m . Esse infatti si discostano di molto dai valori teorici 0.90, 0.95 e 0.99. La log-verosimiglianza profilo sembra invece dare dei risultati accettabili quando $q = 20$ mentre m assume valori elevati, ossia quando numero di strati e numerosità di ciascuno strato non sono troppo distanti tra loro.

Le Tabelle 3.3 e 3.4 mostrano i valori stimati delle coperture degli intervalli di confidenza alla Wald, calcolati secondo la (1.6), ottenuti relativamente alle componenti del parametro di interesse β e δ considerate singolarmente. Anche in questo caso, si può confermare quanto è stato detto in precedenza. La log-verosimiglianza profilo modificata è più precisa nello stimare i singoli parametri in tutte le situazioni considerate, ma soprattutto quando la numerosità degli strati è ridotta, caso in cui la log-verosimiglianza profilo presenta invece dei problemi evidenti.

Dal confronto delle Tabelle 3.3 e 3.4 si può notare, relativamente alla log-verosimiglianza profilo, che le probabilità di copertura stimate per il parametro β associato alla variabile x continua, quando il numero di strati q è elevato ($q = 100, 250$), risultano peggiori rispetto a quelle per il parametro δ associato invece alla variabile d discreta. In entrambi i casi le probabilità di copertura effettive, stimate tramite la simulazione, sono

comunque lontane da quelle nominali.

Funzione di legame probit

Si desidera ora analizzare come si comportano verosimiglianza profilo e profilo modificata quando nel modello considerato si utilizza la funzione di legame probit anziché quella logistica. In questo caso si ha quindi che la variabile risposta binaria ha distribuzione

$$Y_{ij} \sim Bi(1, \pi_{ij}),$$

con probabilità di successo

$$\pi_{ij} = \Phi(\eta_{ij}),$$

dove $\Phi(\cdot)$ rappresenta la funzione di ripartizione di una normale standard.

Il contesto da cui si parte è quello della simulazione precedente, per cui l'intercetta del predittore lineare α_i e le variabili x_{ij} e d_{ij} sono state generate nello stesso modo e i parametri β e δ sono posti entrambi pari a 1. L'unica variante è la funzione di legame che caratterizza il modello.

Nella simulazione sono state stimate le distorsioni per gli stimatori $\hat{\beta}$, $\hat{\beta}_M$ e $\hat{\delta}$, $\hat{\delta}_M$ (Tabella 3.5) e le probabilità di copertura delle regioni di confidenza basate sulle statistiche test W_P e W_M (Tabella 3.6) per $m = (3, 5, 7, 19)$ e $q = 100$ strati.

Osservazioni

Osservando entrambe le tabelle, è evidente come, anche in questo caso, la verosimiglianza profilo modificata fornisca, per ogni valore di m , risultati migliori rispetto alla profilo. In particolare, le distorsioni per gli stimatori $\hat{\beta}_M$ e $\hat{\delta}_M$ sono sempre inferiori rispetto a quelle degli stimatori $\hat{\beta}$ e $\hat{\delta}$ e le probabilità di copertura stimate delle regioni di confidenza basate su W_P sono molto più basse rispetto ai livelli di confidenza nominali 0.90, 0.95 e 0.99, mentre le probabilità di copertura delle regioni basate su W_M risultano prossime a tali valori.

	$m = 3$		$m = 5$		$m = 7$		$m = 10$	
	P	M	P	M	P	M	P	M
$\hat{\beta}$	1.072	-0.168	0.548	-0.058	0.358	0.003	0.216	0.005
$\hat{\delta}$	1.317	0.007	0.493	0.000	0.312	0.025	0.204	0.017

Tabella 3.5: *Modello per dati binari. Risultati della simulazione con $N = 200$ replicazioni, $q = 100$ strati e funzione di legame probit. Distorsioni stimate degli stimatori $\hat{\beta}$ e $\hat{\delta}$ ottenuti tramite la verosimiglianza profilo (P) e la verosimiglianza profilo modificata (M).*

	$m = 3$		$m = 5$		$m = 7$		$m = 10$	
$1 - \alpha$	P	M	P	M	P	M	P	M
0.90	0.135	0.910	0.135	0.930	0.180	0.950	0.200	0.950
0.95	0.185	0.970	0.175	0.975	0.290	0.970	0.400	0.980
0.99	0.310	0.995	0.315	0.995	0.510	0.985	0.660	1.000

Tabella 3.6: *Modello per dati binari. Risultati della simulazione con $N = 200$ replicazioni, $q = 100$ strati e funzione di legame probit. Stime delle probabilità di copertura delle regioni di confidenza per β e δ congiuntamente con livello nominale $1 - \alpha$, basate su W ottenuto tramite la verosimiglianza profilo (P) e la verosimiglianza profilo modificata (M).*

Dalla Tabella 3.5 si può osservare che, per entrambe le verosimiglianze, all'aumentare di m le distorsioni degli stimatori diminuiscono; questo perché, per strati numerosi, l'informazione disponibile sui parametri di disturbo è maggiore, per cui è possibile stimare le quantità α_i con più precisione.

Dal confronto con i risultati ottenuti nella simulazione precedente, si osserva che le distorsioni per gli stimatori $\hat{\beta}$ e $\hat{\delta}$ risultano maggiori nel caso di funzione di legame probit (cfr. Tabella 3.1). Tale aspetto si riflette anche sulle probabilità di copertura stimate che per la verosimiglianza profilo risultano ancora più basse rispetto al caso precedente in cui si utilizzava

la funzione logit (cfr. Tabella 3.2).

3.4 Modello autoregressivo con intercetta variabile

3.4.1 Specificazione del modello

In questo esempio si utilizza un modello autoregressivo del primo ordine, AR(1), con l'obiettivo di confrontare le distorsioni degli stimatori del parametro di interesse $\hat{\tau}$ e $\hat{\tau}_M$ e le procedure inferenziali basate sulla verosimiglianza profilo e profilo modificata. Il contesto di riferimento è nuovamente quello dei parametri incidentali.

Tipicamente, i modelli di tipo autoregressivo sono utilizzati per serie storiche, ossia osservazioni che si susseguono secondo un ordine temporale. Per questo motivo, sembra opportuno ricorrere ad una notazione in parte differente rispetto a quella utilizzata nell'esempio precedente, anche se in quest'ambito i dati sono comunque simulati. Le osservazioni disponibili sono quindi del tipo y_{it} , con $i = 1, \dots, q$, numero di strati e $t = 1, \dots, T_i$, numerosità di ciascuno strato. L'indice t sta appunto ad indicare la dimensione temporale delle osservazioni. Anche in questo caso si suppone per semplicità $T_i = T$ costante.

Si assume che il processo autoregressivo sia debolmente stazionario, ossia che, per ogni t , media e varianza non dipendano da t , mentre la covarianza $\text{cov}(Y_{it}, Y_{it'})$ dipenda solo dalla distanza tra le due osservazioni $|t - t'|$.

L'espressione del modello considerato è la seguente

$$Y_{i1} \sim N\left(\mu_i, \frac{\sigma^2}{1 - \rho^2}\right),$$

$$Y_{it} - \mu_i = \rho(Y_{it-1} - \mu_i) + \sigma\epsilon_{it}, \quad t = 2, \dots, T, \quad (3.9)$$

con $\epsilon_{it} \sim i.i.d.N(0, 1)$.

Segue che

$$\begin{aligned} Y_{it} &= \mu_i + \rho Y_{it-1} - \rho \mu_i + \sigma \epsilon_{it} \\ &= \mu_i(1 - \rho) + \rho Y_{it-1} + \sigma \epsilon_{it}, \end{aligned}$$

con $\mu_i(1 - \rho)$ intercetta del modello specifica dello strato i -mo.

In questo contesto, il parametro di interesse è $\tau = (\rho, \sigma^2)$ bivariato, con $\rho \in (-1, 1)$ e $\sigma^2 \in \mathbb{R}^+$, mentre $\mu_1, \dots, \mu_q \in \mathbb{R}^q$, sono parametri di disturbo incidentali.

Per il modello considerato, il valore atteso di Y_{it} , per $t = 2, \dots, T$, condizionato alla storia passata Y_{it-1}, \dots, Y_{i1} , è dato da

$$E(Y_{it} | I_t) = E(Y_{it} | Y_{it-1}) = \mu_i(1 - \rho) + \rho y_{it-1},$$

mentre la varianza condizionata è pari a

$$V(Y_{it} | I_t) = V(Y_{it} | Y_{it-1}) = \sigma^2,$$

con I_t informazione disponibile fino al tempo t .

Pertanto si ha che

$$Y_{it} | I_t \sim N(\mu_i(1 - \rho) + \rho y_{it-1}, \sigma^2).$$

Si ricava quindi la funzione di verosimiglianza associata al modello

$$\begin{aligned} L(\mu_i, \rho, \sigma^2) &= \prod_{i=1}^q \left(f(y_{i1}; \mu_i, \rho, \sigma^2) \prod_{t=2}^T f(y_{it} | y_{it-1}; \mu_i, \rho, \sigma^2) \right) \\ &= \prod_{i=1}^q \left(\frac{\sqrt{1 - \rho^2}}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(1 - \rho^2)}{2\sigma^2} (y_{i1} - \mu_i)^2 \right\} \right. \\ &\quad \left. \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_{it} - \mu_i(1 - \rho) - \rho y_{it-1})^2 \right\} \right), \end{aligned}$$

da cui si ottiene la funzione di log-verosimiglianza

$$\begin{aligned} l(\mu_i, \rho, \sigma^2) &= -\frac{qT}{2} \log \sigma^2 + \frac{q}{2} \log(1 - \rho^2) - \frac{1}{2\sigma^2} \sum_{i=1}^q \left\{ \sum_{t=1}^T (y_{it} - \mu_i)^2 + \right. \\ &\quad \left. + \rho^2 \sum_{t=2}^{T-1} (y_{it} - \mu_i)^2 - 2\rho \sum_{t=2}^T (y_{it} - \mu_i)(y_{it-1} - \mu_i) \right\}. \quad (3.10) \end{aligned}$$

Partendo dalla (3.10) si ricava facilmente l'espressione della stima di massima verosimiglianza vincolata del parametro di disturbo, $\hat{\mu}_{i\rho\sigma^2} = \hat{\mu}_{i\rho}$, che è pari a

$$\hat{\mu}_{i\rho} = \frac{\sum_{t=1}^T y_{it} - \rho(\sum_{t=2}^T y_{it} + \sum_{t=1}^{T-1} y_{it}) + \rho^2 \sum_{t=2}^{T-1} y_{it}}{T - 2(T-1)\rho + (T-2)\rho^2}.$$

La log-verosimiglianza profilo per il modello considerato non è altro che la quantità $l_P(\rho, \sigma^2) = l(\hat{\mu}_{i\rho}, \rho, \sigma^2)$, log-verosimiglianza calcolata nella s.m.v. vincolata $\hat{\mu}_{i\rho}$.

La log-verosimiglianza profilo modificata, calcolata utilizzando l'approssimazione (2.15), è ottenuta sempre partendo dalla log-verosimiglianza profilo indicata nella (3.10), a cui si somma il termine di correzione $M'(\tau) = \sum_{i=1}^q M'_i(\tau)$, con

$$M'_i(\tau) = \frac{1}{2} \log | j_{\mu_i\mu_i}(\tau, \hat{\mu}_{i\rho}) | - \log | I_{\mu;\mu}^i(\tau, \hat{\mu}_{i\rho}; \hat{\tau}, \hat{\mu}_i) | \quad (3.11)$$

modificazione relativa allo strato i -mo. Per il modello considerato, le quantità presenti nella (3.11) risultano pari a

$$\begin{aligned} j_{\mu_i\mu_i}(\tau, \hat{\mu}_{i\rho}) &= \frac{1}{\sigma^2} \{T - 2(T-1)\rho + (T-2)\rho^2\}, \\ I_{\mu;\mu}^i(\tau, \hat{\mu}_{i\rho}; \hat{\tau}, \hat{\mu}_i) &= \frac{1}{\sigma^2(1-\hat{\rho}^2)} \{A + (\hat{\rho} + \rho)[(\hat{\rho} + \rho)B - 2C - 2\hat{\rho}\rho E] + \\ &\quad + 2\hat{\rho}\rho(B + 1 + \hat{\rho}^{T-1}) + \hat{\rho}^2\rho^2 D\}, \end{aligned} \quad (3.12)$$

con

$$\begin{aligned} A &= \sum_{t=1}^T \sum_{t'=1}^T \hat{\rho}^{|t-t'|}, \\ B &= \sum_{t=1}^T \sum_{t'=2}^{T-1} \hat{\rho}^{|t-t'|} = \sum_{t=2}^{T-1} \sum_{t'=1}^T \hat{\rho}^{|t-t'|} = \sum_{t=2}^T \sum_{t'=1}^{T-1} \hat{\rho}^{|t-t'|} - \hat{\rho}^{T-1} \\ &= \sum_{t=1}^{T-1} \sum_{t'=2}^T \hat{\rho}^{|t-t'|} - \hat{\rho}^{T-1} = \sum_{t=2}^T \sum_{t'=2}^T \hat{\rho}^{|t-t'|} - 1 = \sum_{t=1}^{T-1} \sum_{t'=1}^{T-1} \hat{\rho}^{|t-t'|} - 1, \\ C &= \sum_{t=1}^T \sum_{t'=2}^T \hat{\rho}^{|t-t'|} = \sum_{t=2}^T \sum_{t'=1}^T \hat{\rho}^{|t-t'|} = \sum_{t=1}^T \sum_{t'=1}^{T-1} \hat{\rho}^{|t-t'|} = \sum_{t=1}^{T-1} \sum_{t'=1}^T \hat{\rho}^{|t-t'|}, \end{aligned}$$

$$D = \sum_{t=2}^{T-1} \sum_{t'=2}^{T-1} \hat{\rho}^{|t-t'|},$$

$$E = \sum_{t=2}^{T-1} \sum_{t'=2}^T \hat{\rho}^{|t-t'|} = \sum_{t=2}^T \sum_{t'=2}^{T-1} \hat{\rho}^{|t-t'|} = \sum_{t=2}^{T-1} \sum_{t'=1}^{T-1} \hat{\rho}^{|t-t'|} = \sum_{t=1}^{T-1} \sum_{t'=2}^{T-1} \hat{\rho}^{|t-t'|},$$

per cui la modificazione $M'(\tau)$ è pari a

$$M'(\rho, \sigma^2) = \frac{q}{2} \log \sigma^2 + \frac{q}{2} \log \{T - 2(T-1)\rho + (T-2)\rho^2\} - q \log K(\rho),$$

con

$$K(\rho) = A + (\hat{\rho} + \rho)[(\hat{\rho} + \rho)B - 2C - 2\hat{\rho}\rho] + 2\hat{\rho}\rho(B + 1 + \hat{\rho}^{T-1}) + \hat{\rho}^2\rho^2D.$$

3.4.2 Simulazione

Anche in questo esempio, come nel precedente, si vuole valutare l'accuratezza delle procedure inferenziali basate sulla verosimiglianza profilo modificata, con particolare riferimento al caso in cui la dimensione del parametro di disturbo è elevata e l'informazione disponibile su di esso è ridotta. È proprio in queste situazioni infatti che le usali procedure di verosimiglianza riscontrano dei problemi nella stima delle quantità di interesse, ed è quindi necessario ricorrere a metodi di stima differenti.

In questo studio si vuole in particolare verificare come si comportano la verosimiglianza profilo e profilo modificata al variare del numero di strati q e della numerosità di ciascuno strato T . Si vuole inoltre osservare il livello di accuratezza dei due metodi al variare del valore assunto dalla componente ρ del parametro di interesse.

Nella simulazione si è deciso di utilizzare la parametrizzazione $(\rho, \psi) = (\rho, \log \sigma^2)$ con $\rho \in (-1, 1)$ e $\psi \in \mathbb{R}$, per evitare di incorrere in problemi di tipo numerico. L'analisi include il confronto fra gli stimatori di massima verosimiglianza per ρ e ψ , ottenuti tramite l'impiego della verosimiglianza profilo e profilo modificata, in termini di distorsione, nonché il confronto dei livelli di copertura stimati di regioni e intervalli di confidenza.

I dati sono stati generati da un modello AR(1) del tipo (3.9), con media $\mu_i \sim N(0, 1)$, $\sigma^2 = 1$ e quindi $\psi = \exp \sigma^2 = 0$ e per $\rho = 0, \pm 0.5, \pm 0.9$. Nella simulazione i valori dei parametri di disturbo μ_i sono generati e poi mantenuti fissati. Sono stati considerati i valori $T = (3, 5, 7, 10)$ e $q = (20, 50, 100, 250)$.

Le Tabelle 3.7 e 3.8 riportano le stime delle distorsioni degli stimatori per ρ e $\log \sigma^2$ ottenuti tramite la log-verosimiglianza profilo e profilo modificata. Le Tabelle 3.9 e 3.10 contengono le probabilità di copertura stimate delle regioni di confidenza basate sulle statistiche W_P e W_M , con livelli di copertura nominali pari a 0.90, 0.95 e 0.99. Nelle Tabelle 3.11 e 3.13 sono presentati i valori stimati delle coperture degli intervalli alla Wald, calcolati secondo la (1.6), per i parametri ρ e $\log \sigma^2$ ottenuti tramite la verosimiglianza profilo, mentre nelle Tabelle 3.12 e 3.14 sono riportate le stesse quantità ottenute però attraverso l'impiego della verosimiglianza profilo modificata.

Gli intervalli di confidenza alla Wald per ρ possono tuttavia risultare inadeguati, soprattutto quando il valore del parametro è vicino agli estremi del relativo spazio parametrico, ossia quando ρ è prossimo a 1 o -1 . In queste situazioni c'è infatti il rischio di ottenere intervalli di confidenza per ρ che includono valori inammissibili per il parametro stesso. Sembra quindi opportuno calcolare degli intervalli che non presentino tale inconveniente. A tal fine è stata calcolata la log-verosimiglianza profilo per ρ di $l_P(\rho, \log \sigma^2)$ e $l_M(\rho, \log \sigma^2)$ e sono stati calcolati successivamente gli intervalli per ρ basati sulle statistiche r_P e r_M , che includono solo valori ammissibili per il parametro ρ . Le coperture stimate relative a tali intervalli sono riportate nelle Tabelle 3.15 e 3.16.

Nelle Figure 3.3, 3.4, 3.5, 3.6 e 3.7 sono rappresentate le curve di livello associate alla log-verosimiglianza profilo relativa e alla log-verosimiglianza profilo modificata relativa di un singolo campione con $q = 20$ strati di dimensione $T = 10$. Nei campioni considerati $\log \sigma^2 = 0$ mentre $\rho = -0.9, -0.5, 0, 0.5$ e 0.9 rispettivamente.

		$T = 3$		$T = 5$		$T = 7$		$T = 10$	
	ρ	$\hat{\rho}$	$\log \hat{\sigma}^2$	$\hat{\rho}$	$\log \hat{\sigma}^2$	$\hat{\rho}$	$\log \hat{\sigma}^2$	$\hat{\rho}$	$\log \hat{\sigma}^2$
$q = 20$	-0.9	-0.050	-0.784	-0.019	-0.272	-0.011	-0.188	-0.007	-0.127
	-0.5	-0.281	-0.811	-0.134	-0.311	-0.077	-0.201	-0.055	-0.124
	0.0	-0.623	-0.902	-0.271	-0.341	-0.175	-0.206	-0.116	-0.130
	0.5	-0.985	-0.918	-0.450	-0.327	-0.279	-0.200	-0.177	-0.120
	0.9	-1.308	-1.006	-0.638	-0.338	-0.394	-0.210	-0.263	-0.135
$q = 50$	-0.9	-0.050	-0.729	-0.024	-0.286	-0.017	-0.183	-0.011	-0.119
	-0.5	-0.287	-0.801	-0.134	-0.305	-0.078	-0.190	-0.052	-0.131
	0.0	-0.637	-0.875	-0.269	-0.306	-0.175	-0.190	-0.109	-0.124
	0.5	-0.996	-0.905	-0.450	-0.318	-0.273	-0.191	-0.175	-0.133
	0.9	-1.304	-0.940	-0.619	-0.328	-0.397	-0.202	-0.256	-0.131
$q = 100$	-0.9	-0.053	-0.700	-0.024	-0.284	-0.016	-0.180	-0.011	-0.114
	-0.5	-0.291	-0.796	-0.127	-0.297	-0.080	-0.184	-0.055	-0.114
	0.0	-0.629	-0.843	-0.274	-0.304	-0.174	-0.195	-0.113	-0.126
	0.5	-1.001	-0.910	-0.445	-0.310	-0.275	-0.188	-0.177	-0.114
	0.9	-1.305	-0.935	-0.612	-0.321	-0.384	-0.185	-0.249	-0.119
$q = 250$	-0.9	-0.051	-0.695	-0.025	-0.288	-0.018	-0.183	-0.011	-0.117
	-0.5	-0.290	-0.775	-0.130	-0.296	-0.085	-0.183	-0.055	-0.117
	0.0	-0.633	-0.847	-0.274	-0.297	-0.173	-0.181	-0.114	-0.116
	0.5	-1.001	-0.891	-0.443	-0.308	-0.278	-0.189	-0.176	-0.118
	0.9	-1.305	-0.926	-0.609	-0.317	-0.388	-0.195	-0.251	-0.122

Tabella 3.7: Modello autoregressivo con intercetta variabile. Risultati della simulazione con $N = 200$ replicazioni. Distorsioni stimate degli stimatori $\hat{\rho}$ e $\log \hat{\sigma}^2$ per valori di $\rho = 0, \pm 0.5, \pm 0.9$ e $\sigma^2 = 1$.

		$T = 3$		$T = 5$		$T = 7$		$T = 10$	
	ρ	$\hat{\rho}_M$	$\log \hat{\sigma}_M^2$	$\hat{\rho}_M$	$\log \hat{\sigma}_M^2$	$\hat{\rho}_M$	$\log \hat{\sigma}_M^2$	$\hat{\rho}_M$	$\log \hat{\sigma}_M^2$
$q = 20$	-0.9	0.005	-0.060	0.010	0.014	0.006	-0.002	0.005	-0.009
	-0.5	0.044	-0.034	0.001	-0.020	0.013	-0.017	0.002	-0.006
	0.0	0.042	-0.068	0.007	-0.045	-0.001	-0.022	-0.003	-0.011
	0.5	0.028	-0.048	0.000	-0.020	0.001	-0.013	0.001	-0.001
	0.9	-0.134	-0.145	-0.066	-0.032	-0.027	-0.022	-0.020	-0.015
$q = 50$	-0.9	0.003	0.002	0.006	0.014	-0.001	0.017	0.000	-0.001
	-0.5	0.011	-0.031	-0.005	-0.014	0.008	-0.007	0.005	-0.013
	0.0	-0.012	-0.045	0.008	-0.010	-0.001	-0.006	0.005	-0.006
	0.5	0.007	-0.023	-0.009	-0.014	0.003	-0.005	0.001	-0.014
	0.9	-0.074	-0.054	-0.022	-0.014	-0.015	-0.010	-0.010	-0.010
$q = 100$	-0.9	0.000	-0.004	0.000	0.036	0.001	0.008	0.000	0.004
	-0.5	0.000	-0.026	0.006	-0.006	0.006	0.000	0.001	0.003
	0.0	0.007	-0.010	-0.002	-0.010	0.000	-0.011	0.000	-0.008
	0.5	-0.013	-0.028	-0.003	-0.005	0.002	-0.001	-0.001	0.004
	0.9	-0.048	-0.034	-0.013	-0.006	-0.001	0.006	-0.001	0.003
$q = 250$	-0.9	0.002	0.009	-0.002	0.022	-0.001	0.000	0.000	0.001
	-0.5	0.001	-0.003	0.001	-0.005	0.000	0.000	0.001	0.001
	0.0	-0.008	-0.016	-0.002	-0.001	0.001	0.003	-0.001	0.002
	0.5	-0.013	-0.009	0.000	-0.004	-0.001	-0.002	0.001	0.000
	0.9	-0.023	-0.012	-0.002	0.000	-0.002	-0.001	0.000	0.000

Tabella 3.8: Modello autoregressivo con intercetta variabile. Risultati della simulazione con $N = 200$ replicazioni. Distorsioni stimate degli stimatori $\hat{\rho}_M$ e $\log \hat{\sigma}_M^2$ per valori di $\rho = 0, \pm 0.5, \pm 0.9$ e $\sigma^2 = 1$.

		$T = 3$			$T = 5$			$T = 7$			$T = 10$		
		$1 - \alpha$											
	ρ	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$q = 20$	-0.9	0.15	0.23	0.43	0.60	0.70	0.85	0.68	0.81	0.91	0.77	0.86	0.94
	-0.5	0.08	0.14	0.30	0.36	0.49	0.73	0.57	0.70	0.90	0.67	0.81	0.94
	0.0	0.00	0.01	0.03	0.15	0.23	0.50	0.30	0.44	0.71	0.51	0.65	0.81
	0.5	0.00	0.00	0.00	0.00	0.01	0.09	0.07	0.15	0.33	0.22	0.30	0.53
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00
$q = 50$	-0.9	0.01	0.03	0.06	0.20	0.29	0.51	0.35	0.49	0.74	0.50	0.63	0.85
	-0.5	0.00	0.00	0.00	0.06	0.10	0.22	0.22	0.32	0.57	0.36	0.46	0.73
	0.0	0.00	0.00	0.00	0.01	0.02	0.04	0.03	0.08	0.22	0.17	0.28	0.48
	0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.10
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$q = 100$	-0.9	0.01	0.01	0.01	0.02	0.05	0.15	0.11	0.20	0.39	0.27	0.40	0.66
	-0.5	0.00	0.00	0.00	0.00	0.00	0.03	0.04	0.08	0.19	0.15	0.22	0.45
	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.03	0.11
	0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$q = 250$	-0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.05	0.15
	-0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Tabella 3.9: Modello autoregressivo con intercetta variabile. Risultati della simulazione con $N = 200$ repliche. Probabilità di copertura stimate delle regioni di confidenza per ρ e $\log \sigma^2$ congiuntamente con livello nominale $1 - \alpha$, basate su W_P per valori di $\rho = 0, \pm 0.5, \pm 0.9$ e $\sigma^2 = 1$.

		$T = 3$			$T = 5$			$T = 7$			$T = 10$		
		$1 - \alpha$											
	ρ	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$q = 20$	-0.9	0.92	0.95	0.99	0.92	0.96	0.98	0.90	0.96	0.99	0.91	0.95	0.99
	-0.5	0.87	0.93	0.99	0.91	0.94	0.98	0.89	0.93	0.98	0.91	0.95	1.00
	0.0	0.91	0.94	0.99	0.88	0.94	0.98	0.92	0.98	0.99	0.91	0.94	0.97
	0.5	0.91	0.98	0.99	0.91	0.95	1.00	0.91	0.95	1.00	0.91	0.97	0.99
	0.9	0.92	0.96	0.98	0.92	0.95	0.98	0.93	0.97	1.00	0.90	0.96	1.00
$q = 50$	-0.9	0.87	0.93	1.00	0.92	0.97	1.00	0.90	0.94	0.98	0.93	0.96	0.98
	-0.5	0.87	0.95	0.98	0.91	0.96	0.98	0.90	0.97	0.99	0.90	0.94	0.99
	0.0	0.90	0.95	0.99	0.87	0.95	0.99	0.92	0.96	1.00	0.92	0.96	0.99
	0.5	0.89	0.92	0.98	0.92	0.99	1.00	0.89	0.94	0.99	0.88	0.92	0.98
	0.9	0.96	0.98	1.00	0.95	1.00	1.00	0.91	0.97	1.00	0.90	0.95	0.99
$q = 100$	-0.9	0.90	0.94	0.98	0.92	0.95	0.99	0.88	0.93	0.99	0.88	0.94	0.98
	-0.5	0.85	0.96	1.00	0.89	0.95	0.97	0.89	0.93	0.99	0.88	0.95	0.99
	0.0	0.95	0.98	0.99	0.92	0.97	1.00	0.91	0.98	0.99	0.90	0.95	0.99
	0.5	0.89	0.93	1.00	0.88	0.94	0.99	0.92	0.94	0.99	0.89	0.93	0.98
	0.9	0.93	0.95	0.99	0.94	0.97	0.99	0.88	0.94	0.98	0.92	0.95	0.97
$q = 250$	-0.9	0.89	0.93	0.98	0.89	0.95	0.99	0.89	0.95	0.98	0.90	0.94	0.99
	-0.5	0.93	0.97	0.98	0.88	0.91	0.98	0.92	0.95	1.00	0.94	0.96	0.99
	0.0	0.90	0.95	0.99	0.89	0.92	0.98	0.88	0.94	1.00	0.88	0.95	0.99
	0.5	0.89	0.95	0.99	0.87	0.92	0.98	0.92	0.95	0.99	0.92	0.97	0.98
	0.9	0.91	0.95	0.99	0.88	0.95	0.97	0.93	0.97	0.98	0.90	0.94	0.98

Tabella 3.10: *Modello autoregressivo con intercetta variabile. Risultati della simulazione con $N = 200$ repliche. Probabilità di copertura stimate delle regioni di confidenza per ρ e $\log \sigma^2$ congiuntamente con livello nominale $1 - \alpha$, basate sul test W_M , per valori di $\rho = 0, \pm 0.5, \pm 0.9$ e $\sigma^2 = 1$.*

		$T = 3$			$T = 5$			$T = 7$			$T = 10$		
		$1 - \alpha$											
	ρ	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$q = 20$	-0.9	0.24	0.30	0.37	0.71	0.79	0.89	0.84	0.91	0.97	0.85	0.91	0.95
	-0.5	0.09	0.12	0.26	0.49	0.59	0.75	0.65	0.77	0.92	0.74	0.81	0.94
	0.0	0.00	0.01	0.01	0.21	0.30	0.58	0.37	0.47	0.77	0.54	0.69	0.84
	0.5	0.00	0.00	0.00	0.00	0.01	0.11	0.07	0.15	0.34	0.18	0.27	0.54
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$q = 50$	-0.9	0.05	0.06	0.14	0.52	0.61	0.75	0.65	0.73	0.85	0.76	0.83	0.94
	-0.5	0.01	0.01	0.01	0.16	0.24	0.44	0.43	0.54	0.78	0.60	0.73	0.87
	0.0	0.00	0.00	0.00	0.01	0.03	0.11	0.07	0.12	0.33	0.25	0.34	0.60
	0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.05	0.12
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$q = 100$	-0.9	0.00	0.01	0.01	0.31	0.40	0.57	0.49	0.57	0.78	0.68	0.76	0.90
	-0.5	0.00	0.00	0.00	0.06	0.08	0.21	0.16	0.25	0.51	0.32	0.45	0.64
	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.05	0.10	0.23
	0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$q = 250$	-0.9	0.00	0.00	0.00	0.02	0.07	0.18	0.14	0.25	0.45	0.46	0.55	0.73
	-0.5	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.04	0.05	0.08	0.25
	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Tabella 3.11: Modello autoregressivo con intercetta variabile. Risultati della simulazione con $N = 200$ replicazioni. Probabilità di copertura stimate dell'intervallo per ρ con livello nominale $1 - \alpha$, ottenute tramite la verosimiglianza profilo per valori di $\rho = 0, \pm 0.5, \pm 0.9$ e $\sigma^2 = 1$.

		$T = 3$			$T = 5$			$T = 7$			$T = 10$		
		$1 - \alpha$											
	ρ	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$q = 20$	-0.9	0.84	0.88	0.94	0.92	0.96	0.98	0.91	0.97	0.98	0.91	0.96	0.99
	-0.5	0.89	0.90	0.97	0.92	0.96	0.97	0.91	0.96	0.99	0.89	0.95	1.00
	0.0	0.90	0.92	0.97	0.88	0.95	0.99	0.91	0.97	0.99	0.91	0.94	0.99
	0.5	0.94	0.96	0.99	0.88	0.96	0.98	0.90	0.96	0.99	0.90	0.95	0.99
	0.9	0.90	0.93	0.97	0.92	0.95	0.99	0.94	0.96	0.99	0.91	0.96	1.00
$q = 50$	-0.9	0.86	0.91	0.95	0.90	0.96	0.99	0.89	0.92	0.97	0.92	0.96	0.98
	-0.5	0.90	0.93	0.95	0.89	0.94	0.98	0.92	0.96	1.00	0.88	0.95	0.99
	0.0	0.81	0.89	0.95	0.88	0.94	0.99	0.89	0.95	1.00	0.86	0.94	0.98
	0.5	0.89	0.95	0.97	0.91	0.98	1.00	0.89	0.95	0.99	0.88	0.92	0.98
	0.9	0.93	0.98	0.99	0.95	0.98	1.00	0.92	0.96	0.99	0.91	0.95	0.99
$q = 100$	-0.9	0.91	0.94	0.99	0.87	0.93	0.95	0.92	0.94	0.99	0.92	0.96	1.00
	-0.5	0.88	0.93	0.96	0.87	0.93	0.98	0.90	0.96	0.99	0.88	0.94	0.98
	0.0	0.92	0.98	0.99	0.91	0.96	0.99	0.91	0.95	1.00	0.87	0.95	0.99
	0.5	0.88	0.92	0.98	0.86	0.92	1.00	0.90	0.96	0.98	0.87	0.93	0.99
	0.9	0.93	0.96	0.98	0.95	0.98	0.99	0.92	0.97	0.98	0.89	0.94	0.99
$q = 250$	-0.9	0.90	0.94	0.97	0.86	0.93	0.97	0.91	0.95	1.00	0.88	0.94	0.99
	-0.5	0.90	0.95	0.98	0.90	0.95	0.98	0.91	0.96	0.99	0.89	0.96	0.99
	0.0	0.88	0.94	0.98	0.90	0.93	0.97	0.90	0.94	0.99	0.86	0.93	0.99
	0.5	0.90	0.93	0.98	0.89	0.94	0.98	0.92	0.94	0.99	0.89	0.95	0.99
	0.9	0.95	0.98	0.99	0.87	0.97	0.99	0.92	0.96	0.99	0.89	0.94	0.98

Tabella 3.12: *Modello autoregressivo con intercetta variabile. Risultati della simulazione con $N = 200$ replicazioni. Probabilità di copertura stimate dell'intervallo per ρ con livello nominale $1 - \alpha$, ottenute tramite la verosimiglianza profilo modificata per valori di $\rho = 0, \pm 0.5, \pm 0.9$ e $\sigma^2 = 1$.*

		$T = 3$			$T = 5$			$T = 7$			$T = 10$		
		$1 - \alpha$											
	ρ	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$q = 20$	-0.9	0.09	0.15	0.26	0.50	0.60	0.73	0.58	0.69	0.84	0.65	0.76	0.91
	-0.5	0.05	0.08	0.20	0.35	0.45	0.68	0.49	0.65	0.84	0.64	0.78	0.91
	0.0	0.02	0.04	0.09	0.26	0.38	0.58	0.52	0.62	0.78	0.64	0.77	0.89
	0.5	0.01	0.01	0.06	0.29	0.39	0.60	0.52	0.61	0.79	0.65	0.76	0.90
	0.9	0.00	0.00	0.02	0.27	0.40	0.58	0.47	0.57	0.81	0.61	0.74	0.89
$q = 50$	-0.9	0.00	0.00	0.04	0.11	0.17	0.36	0.23	0.34	0.63	0.41	0.55	0.75
	-0.5	0.00	0.00	0.00	0.07	0.11	0.26	0.20	0.33	0.53	0.36	0.47	0.69
	0.0	0.00	0.00	0.00	0.06	0.11	0.27	0.19	0.30	0.52	0.38	0.50	0.73
	0.5	0.00	0.00	0.00	0.04	0.06	0.19	0.25	0.32	0.51	0.35	0.43	0.68
	0.9	0.00	0.00	0.00	0.05	0.07	0.18	0.14	0.24	0.45	0.36	0.48	0.65
$q = 100$	-0.9	0.00	0.00	0.00	0.01	0.02	0.08	0.08	0.13	0.29	0.16	0.27	0.58
	-0.5	0.00	0.00	0.00	0.00	0.00	0.02	0.07	0.10	0.22	0.21	0.31	0.52
	0.0	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.05	0.16	0.10	0.19	0.44
	0.5	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.08	0.21	0.21	0.31	0.50
	0.9	0.00	0.00	0.00	0.00	0.00	0.01	0.06	0.10	0.23	0.15	0.24	0.48
$q = 250$	-0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.09
	-0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.04
	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.07
	0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.06

Tabella 3.13: *Modello autoregressivo con intercetta variabile. Risultati della simulazione con $N = 200$ replicazioni. Probabilità di copertura stimate dell'intervallo per $\log \sigma^2$ con livello nominale $1 - \alpha$, ottenute tramite la verosimiglianza profilo per valori di $\rho = 0, \pm 0.5, \pm 0.9$ e $\sigma^2 = 1$.*

		$T = 3$			$T = 5$			$T = 7$			$T = 10$		
		$1 - \alpha$											
	ρ	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$q = 20$	-0.9	0.86	0.93	0.97	0.90	0.95	0.99	0.90	0.94	0.98	0.90	0.95	1.00
	-0.5	0.86	0.91	0.99	0.90	0.93	0.98	0.92	0.94	0.99	0.92	0.96	1.00
	0.0	0.88	0.96	0.98	0.85	0.91	0.99	0.91	0.96	1.00	0.91	0.94	0.98
	0.5	0.90	0.96	0.99	0.93	0.98	0.99	0.91	0.96	0.99	0.92	0.95	0.99
	0.9	0.90	0.95	0.98	0.88	0.93	0.98	0.93	0.96	1.00	0.89	0.93	0.98
$q = 50$	-0.9	0.84	0.89	0.96	0.92	0.96	0.99	0.89	0.93	0.98	0.92	0.95	0.98
	-0.5	0.85	0.95	0.97	0.91	0.93	0.98	0.89	0.95	0.98	0.88	0.94	0.99
	0.0	0.87	0.93	0.99	0.89	0.95	0.98	0.91	0.96	1.00	0.89	0.95	0.99
	0.5	0.90	0.95	0.99	0.92	0.95	0.99	0.91	0.95	0.98	0.89	0.94	0.98
	0.9	0.93	0.97	0.99	0.92	0.97	1.00	0.91	0.98	0.99	0.91	0.94	0.99
$q = 100$	-0.9	0.90	0.95	0.98	0.88	0.93	0.96	0.85	0.90	0.97	0.90	0.95	0.97
	-0.5	0.86	0.92	0.98	0.93	0.97	0.99	0.87	0.91	0.97	0.90	0.95	0.99
	0.0	0.91	0.96	0.99	0.92	0.96	0.99	0.92	0.95	0.99	0.90	0.95	0.98
	0.5	0.87	0.91	0.98	0.90	0.95	0.99	0.90	0.95	0.99	0.91	0.95	0.98
	0.9	0.93	0.96	0.98	0.90	0.95	0.98	0.88	0.93	0.98	0.92	0.96	0.97
$q = 250$	-0.9	0.89	0.94	0.98	0.87	0.94	0.96	0.92	0.94	0.98	0.90	0.95	0.98
	-0.5	0.89	0.96	1.00	0.89	0.92	0.97	0.95	0.97	0.99	0.94	0.97	0.99
	0.0	0.91	0.94	0.99	0.89	0.96	0.99	0.88	0.95	1.00	0.90	0.96	1.00
	0.5	0.91	0.95	0.99	0.86	0.92	0.98	0.93	0.95	0.99	0.92	0.96	0.99
	0.9	0.91	0.95	0.98	0.90	0.94	0.98	0.88	0.94	0.99	0.90	0.95	0.98

Tabella 3.14: *Modello autoregressivo con intercetta variabile. Risultati della simulazione con $N = 200$ replicazioni. Probabilità di copertura stimate dell'intervallo per $\log \sigma^2$ con livello nominale $1 - \alpha$, ottenute tramite la verosimiglianza profilo modificata per valori di $\rho = 0, \pm 0.5, \pm 0.9$ e $\sigma^2 = 1$.*

		$T = 3$			$T = 5$			$T = 7$			$T = 10$		
		$1 - \alpha$											
	ρ	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$q = 20$	-0.9	0.32	0.40	0.69	0.78	0.87	0.95	0.90	0.94	0.99	0.88	0.93	0.98
	-0.5	0.12	0.24	0.42	0.52	0.66	0.85	0.68	0.80	0.96	0.74	0.84	0.96
	0.0	0.00	0.01	0.05	0.21	0.31	0.60	0.37	0.48	0.68	0.55	0.70	0.84
	0.5	0.00	0.00	0.00	0.00	0.00	0.10	0.07	0.13	0.33	0.18	0.26	0.54
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$q = 50$	-0.9	0.05	0.12	0.28	0.58	0.70	0.83	0.69	0.79	0.90	0.78	0.85	0.96
	-0.5	0.01	0.01	0.03	0.18	0.25	0.46	0.44	0.56	0.80	0.60	0.73	0.88
	0.0	0.00	0.00	0.00	0.01	0.03	0.12	0.07	0.13	0.35	0.26	0.35	0.62
	0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.12
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$q = 100$	-0.9	0.00	0.01	0.02	0.35	0.46	0.65	0.52	0.65	0.86	0.70	0.78	0.94
	-0.5	0.00	0.00	0.00	0.06	0.10	0.25	0.17	0.26	0.52	0.34	0.46	0.66
	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.05	0.10	0.24
	0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$q = 250$	-0.9	0.00	0.00	0.00	0.03	0.09	0.21	0.15	0.26	0.48	0.49	0.57	0.77
	-0.5	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.05	0.05	0.09	0.27
	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Tabella 3.15: *Modello autoregressivo con intercetta variabile. Risultati della simulazione con $N = 200$ repliche. Probabilità di copertura stimate dell'intervallo per ρ con livello nominale $1 - \alpha$, basato sul test r_P per valori di $\rho = 0, \pm 0.5, \pm 0.9$ e $\sigma^2 = 1$.*

		$T = 3$			$T = 5$			$T = 7$			$T = 10$		
		$1 - \alpha$											
	ρ	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$q = 20$	-0.9	0.92	0.93	0.98	0.91	0.98	0.99	0.93	0.95	0.99	0.91	0.95	0.99
	-0.5	0.92	0.96	0.98	0.93	0.97	0.98	0.92	0.96	0.98	0.90	0.95	0.99
	0.0	0.91	0.96	0.99	0.91	0.95	0.99	0.91	0.96	1.00	0.92	0.94	0.98
	0.5	0.91	0.97	1.00	0.90	0.96	1.00	0.92	0.96	0.99	0.90	0.96	0.99
	0.9	0.93	0.97	0.99	0.94	0.97	0.99	0.93	0.97	1.00	0.87	0.95	1.00
$q = 50$	-0.9	0.88	0.93	0.97	0.92	0.97	1.00	0.88	0.96	0.99	0.91	0.97	0.98
	-0.5	0.89	0.93	0.98	0.90	0.95	0.98	0.92	0.96	1.00	0.89	0.95	0.99
	0.0	0.85	0.90	0.98	0.89	0.94	0.99	0.90	0.95	1.00	0.85	0.94	0.98
	0.5	0.91	0.94	0.99	0.92	0.97	1.00	0.89	0.95	0.99	0.88	0.92	0.98
	0.9	0.95	0.99	1.00	0.96	0.98	1.00	0.89	0.95	1.00	0.91	0.96	0.99
$q = 100$	-0.9	0.93	0.96	0.98	0.90	0.94	0.99	0.92	0.95	0.99	0.92	0.96	1.00
	-0.5	0.89	0.94	0.98	0.88	0.94	0.98	0.90	0.95	0.99	0.88	0.94	0.98
	0.0	0.93	0.98	0.99	0.91	0.95	0.99	0.89	0.96	1.00	0.87	0.95	0.99
	0.5	0.88	0.94	0.99	0.86	0.93	0.99	0.91	0.96	0.98	0.87	0.93	0.99
	0.9	0.93	0.97	0.98	0.94	0.98	0.99	0.91	0.94	0.99	0.89	0.95	0.99
$q = 250$	-0.9	0.89	0.94	0.98	0.88	0.96	1.00	0.91	0.96	1.00	0.89	0.94	0.99
	-0.5	0.91	0.95	0.98	0.90	0.95	0.98	0.91	0.97	0.99	0.89	0.96	0.99
	0.0	0.87	0.94	0.99	0.88	0.93	0.98	0.90	0.94	0.99	0.86	0.93	0.99
	0.5	0.89	0.94	0.98	0.88	0.93	0.98	0.92	0.94	0.99	0.88	0.95	0.99
	0.9	0.95	0.98	0.99	0.88	0.93	0.99	0.92	0.96	0.99	0.89	0.94	0.98

Tabella 3.16: Modello autoregressivo con intercetta variabile. Risultati della simulazione con $N = 200$ replicazioni. Coperture dell'intervallo per ρ con livello nominale $1 - \alpha$, basato sul test r_M per valori di $\rho = 0, \pm 0.5, \pm 0.9$ e $\sigma^2 = 1$.

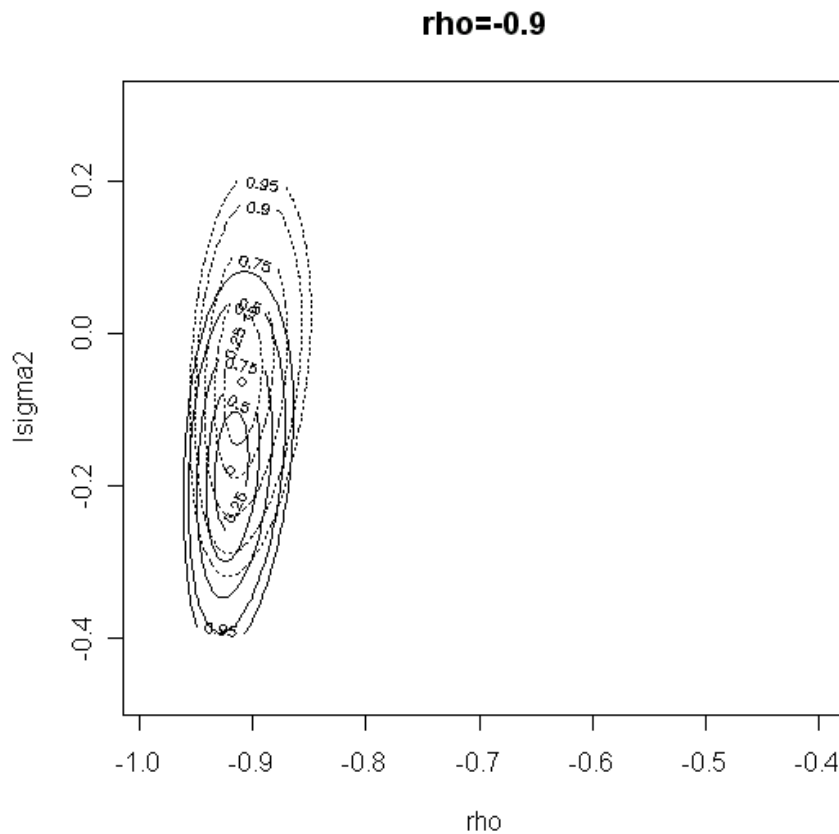


Figura 3.3: Modello autoregressivo con intercetta variabile. Curve di livello riferite alla log-verosimiglianza profilo relativa (linea continua) e alla log-verosimiglianza profilo modificata relativa (linea tratteggiata), calcolate su un campione di $q = 20$ strati di dimensione $T = 10$, con $\log \sigma^2 = 0$ e $\rho = -0.9$.

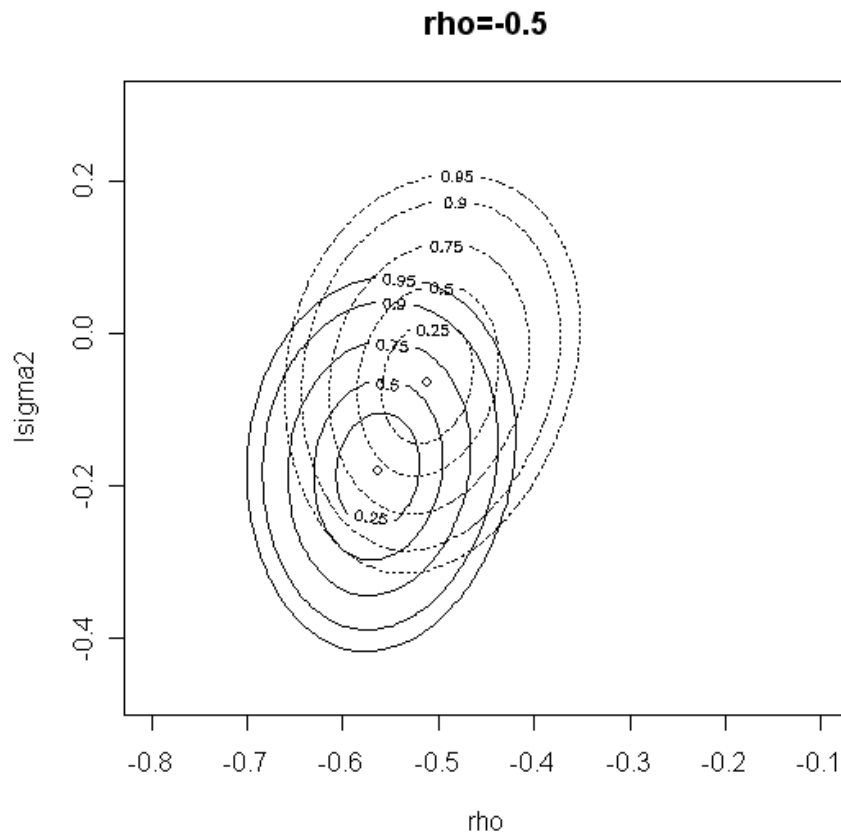


Figura 3.4: Modello autoregressivo con intercetta variabile. Curve di livello riferite alla log-verosimiglianza profilo relativa (linea continua) e alla log-verosimiglianza profilo modificata relativa (linea tratteggiata), calcolate su un campione di $q = 20$ strati di dimensione $T = 10$, con $\log \sigma^2 = 0$ e $\rho = -0.5$.

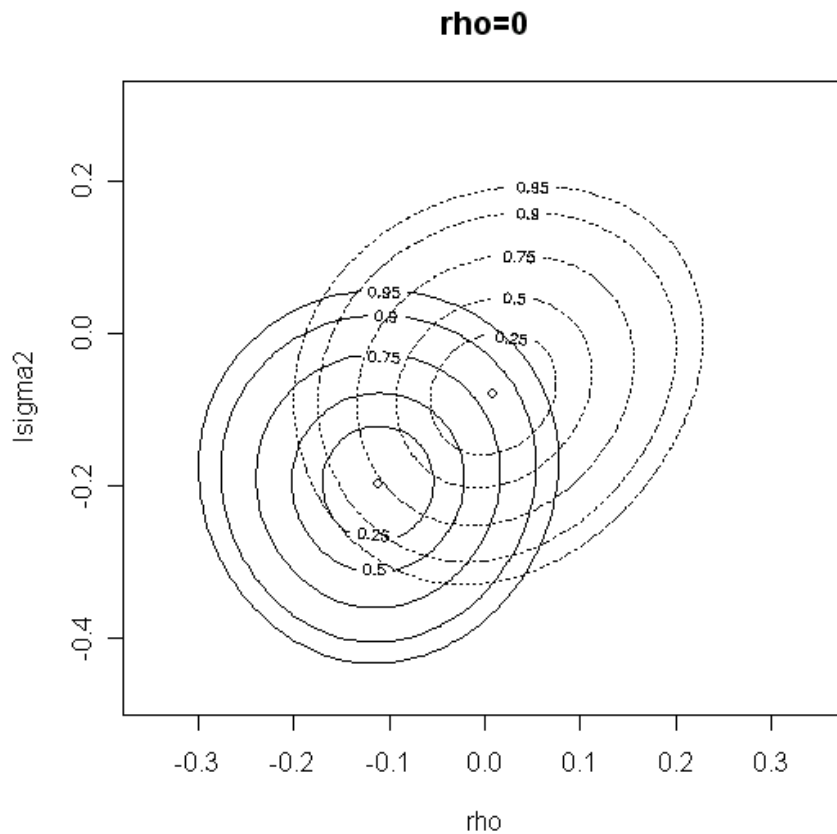


Figura 3.5: Modello autoregressivo con intercetta variabile. Curve di livello riferite alla log-verosimiglianza profilo relativa (linea continua) e alla log-verosimiglianza profilo modificata relativa (linea tratteggiata), calcolate su un campione di $q = 20$ strati di dimensione $T = 10$, con $\log \sigma^2 = 0$ e $\rho = 0$.

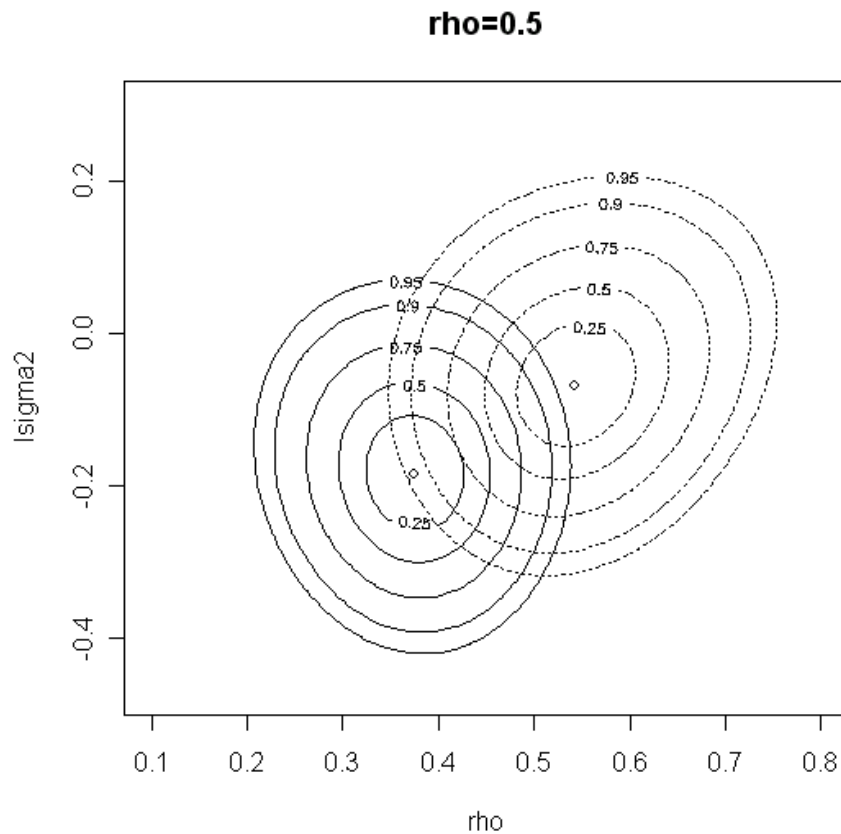


Figura 3.6: Modello autoregressivo con intercetta variabile. Curve di livello riferite alla log-verosimiglianza profilo relativa (linea continua) e alla log-verosimiglianza profilo modificata relativa (linea tratteggiata), calcolate su un campione di $q = 20$ strati di dimensione $T = 10$, con $\log \sigma^2 = 0$ e $\rho = 0.5$.

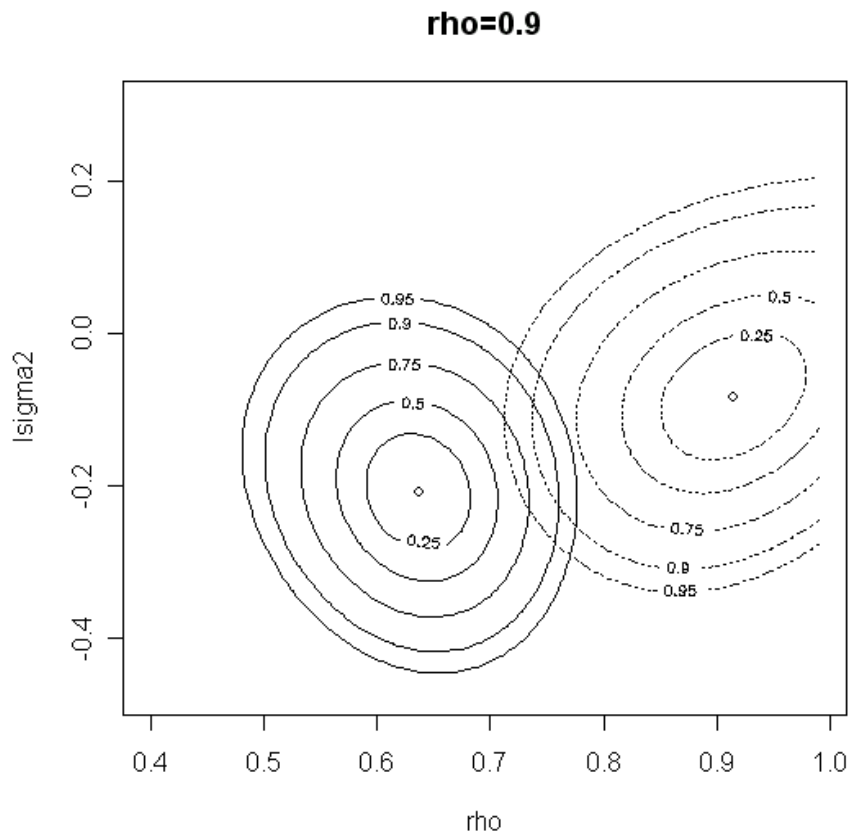


Figura 3.7: Modello autoregressivo con intercetta variabile. Curve di livello riferite alla log-verosimiglianza profilo relativa (linea continua) e alla log-verosimiglianza profilo modificata relativa (linea tratteggiata), calcolate su un campione di $q = 20$ strati di dimensione $T = 10$, con $\log \sigma^2 = 0$ e $\rho = 0.9$.

Osservazioni

In questo esempio, più che nel precedente, è evidente come le procedure inferenziali basate sulla verosimiglianza profilo modificata risultino decisamente più adeguate ed accurate rispetto a quelle basate sulla verosimiglianza profilo. In particolare, le probabilità di copertura stimate delle regioni di confidenza per il parametro di interesse $\tau = (\rho, \log \sigma^2)$ e degli intervalli per ρ e $\log \sigma^2$ relative alla verosimiglianza profilo sono, in quasi la metà dei casi considerati, addirittura nulle. In tutte le situazioni considerate, le probabilità di copertura stimate relative alla verosimiglianza profilo modificata non si scostano invece di molto dai livelli di confidenza nominali pari al 90%, 95% e 99%.

Osservando le distorsioni stimate per $\hat{\rho}$ e $\log \hat{\sigma}^2$ riportate nella Tabella 3.7, si può notare come esse tendano a diminuire all'aumentare di T per q fissato. Questo perché, all'aumentare della numerosità degli strati T , aumenta anche l'informazione disponibile utile a stimare il parametro μ_i , permettendo di conseguenza di ottenere stime più precise anche per il parametro di interesse $\tau = (\rho, \log \sigma^2)$.

Analizzando ora le distorsioni stimate per $\hat{\rho}$ e $\log \hat{\sigma}^2$ per T e q fissati, si può innanzitutto notare come sia presente una distorsione verso il basso per entrambi gli stimatori. Inoltre, si osserva che le distorsioni di $\hat{\rho}$ tendono ad aumentare in valore assoluto al crescere del valore assunto da ρ . Ciò sta a significare che l'impiego della verosimiglianza profilo permette di ottenere stime più precise per valori piccoli di ρ (in questo caso valori negativi), mentre incontra delle difficoltà per valori di ρ elevati e positivi, anche quando la numerosità degli strati è elevata. Per quanto riguarda invece lo stimatore $\log \hat{\sigma}^2$, le distorsioni non variano di molto al variare del valore assunto da ρ .

Tali aspetti possono essere riscontrati anche nelle probabilità di copertura stimate degli intervalli di confidenza per ρ (cfr. Tabelle 3.11 e 3.15), le quali, all'aumentare del valore assunto da ρ , tendono ad allontanarsi dai livelli di confidenza nominali e nelle coperture stimate degli intervalli per

$\log \sigma^2$ (cfr. Tabella 3.13), le quali invece non variano di molto al variare di ρ .

Sempre in riferimento alle distorsioni stimate riportate nella Tabella 3.7, si osserva che le distorsioni per $\hat{\rho}$ più elevate si hanno per $T = 3$ e $\rho = 0.5, 0.9$, mentre le distorsioni più basse si hanno per $T = 10$ e $\rho = -0.5, -0.9$.

Per quanto riguarda invece le distorsioni stimate per $\hat{\rho}_M$ e $\log \hat{\sigma}_M^2$ (cfr. Tabella 3.8), queste assumono valori molto bassi, prossimi allo zero a prescindere dai valori assunti da T , q e ρ .

Osservando le probabilità di copertura stimate delle regioni di confidenza basate su W_P riportate nella Tabella 3.9 e le probabilità di copertura stimate degli intervalli per ρ e $\log \sigma^2$ considerati singolarmente, si può notare come forse l'unico caso in cui l'impiego della verosimiglianza profilo possa fornire dei risultati soddisfacenti è quando $q = 20$, $T = 10$ e $\rho = -0.9$.

Nelle Figure 3.3, 3.4, 3.5, 3.6 e 3.7 si osserva, per un singolo campione di 20 strati di numerosità 10, che le curve relative alla log-verosimiglianza profilo modificata sono, in tutti i casi considerati graficamente, più vicine al vero valore del parametro di interesse $\tau = (\rho, \log \sigma^2)$. Inoltre, le curve di livello relative alla log-verosimiglianza profilo si allontanano progressivamente, all'aumentare del valore assunto da ρ , da quelle relative alla log-verosimiglianza profilo modificata. In particolare, le curve per la verosimiglianza profilo sono posizionate sulla sinistra rispetto alle curve per la verosimiglianza profilo modificata. Ciò è in accordo con le distorsioni di segno negativo per $\hat{\rho}$ stimate nella simulazione. Per i campioni considerati, si ha quindi che la verosimiglianza profilo modificata sembra preferibile rispetto alla profilo, indipendentemente dal valore assunto da ρ , e che i risultati ottenuti tramite l'impiego della verosimiglianza profilo peggiorano all'aumentare del valore assunto da ρ per q e T fissati.

Conclusioni

L'obiettivo principale della tesi era quello di studiare le proprietà della verosimiglianza profilo modificata e di verificarle empiricamente. Si è cercato di mettere in evidenza gli aspetti che caratterizzano la verosimiglianza profilo modificata e che la distinguono dalla verosimiglianza profilo, rendendola in determinate situazioni più appropriata. Si fa in particolare riferimento a situazioni in cui vi è la presenza di parametri incidentali.

I risultati ottenuti nelle simulazioni confermano le proprietà delle verosimiglianza profilo modificata espone alla fine del Capitolo 2, evidenziando che, in presenza di parametri incidentali, l'impiego della verosimiglianza profilo modificata consente di ottenere risultati migliori rispetto alle usuali procedure inferenziali.

In particolare, quello che si è appreso dagli esempi riportati nel Capitolo 3 è che la verosimiglianza profilo modificata permette di ottenere risultati soddisfacenti a prescindere dalla quantità di informazione disponibile. L'impiego risulta tuttavia conveniente nei casi in cui la dimensione del parametro di disturbo è talmente grande, da risultare troppo elevata rispetto all'informazione disponibile. È infatti in queste situazioni che l'impiego della verosimiglianza profilo tende a condurre a conclusioni inferenziali del tutto inattendibili.

Quando invece la dimensione del parametro di disturbo è ridotta e l'informazione disponibile per stimare i parametri di disturbo incidentali è sufficientemente elevata, le due procedure tendono a coincidere, per cui può risultare più conveniente utilizzare la verosimiglianza profilo che è

analiticamente più semplice.

Si è verificato inoltre che, per entrambe le procedure, la distorsione dello stimatore del parametro di interesse e le quantità di verosimiglianza ad esso associate tendono a migliorare all'aumentare della numerosità degli strati, poiché, maggiore è la numerosità degli strati, maggiore è l'informazione disponibile utile a stimare i parametri di disturbo vincolati.

Le conclusioni riportate sono valide sia per il modello per dati discreti (§ 3.3), che per il modello per dati continui (§ 3.4).

Nel secondo esempio, in cui le osservazioni sono dipendenti all'interno di ciascuno strato, l'impiego della verosimiglianza profilo conduce a risultati peggiori rispetto al modello precedente. Tale peggioramento va probabilmente attribuito alla dipendenza tra le osservazioni, a causa della quale l'informazione complessiva disponibile risulta essere inferiore. Infatti il peggioramento si osserva, in particolare, per valori di ρ elevati, dove ρ è il parametro che misura la dipendenza tra le osservazioni all'interno del singolo strato.

Appendice A

Calcoli per il modello AR(1) con intercetta variabile

In questa appendice si riportano i calcoli omissi precedentemente, che sono stati eseguiti per ottenere le espressioni delle stime di massima verosimiglianza vincolate dei parametri di disturbo e i termini della modificazione della verosimiglianza profilo per il modello AR(1) con intercetta variabile.

Si consideri il modello autoregressivo del primo ordine definito nel paragrafo 3.4.1 ed in particolare la funzione di log-verosimiglianza

$$l(\mu, \rho, \sigma^2) = -\frac{qT}{2} \log \sigma^2 + \frac{q}{2} \log(1 - \rho^2) - \frac{1}{2\sigma^2} \sum_{i=1}^q \left\{ \sum_{t=1}^T (y_{it} - \mu_i)^2 + \rho^2 \sum_{t=2}^{T-1} (y_{it} - \mu_i)^2 - 2\rho \sum_{t=2}^T (y_{it} - \mu_i)(y_{it-1} - \mu_i) \right\},$$

con $\mu = (\mu_1, \dots, \mu_q)$.

Per ottenere le stime di massima verosimiglianza vincolate $\hat{\mu}_{i\rho}$ occorre derivare la funzione di log-verosimiglianza rispetto al parametro μ_i , per i fissato. Tale derivata è pari a

$$l_{\mu_i} = -\frac{1}{2\sigma^2} \left\{ -2 \sum_{t=1}^T (y_{it} - \mu_i) - 2\rho^2 \sum_{t=2}^{T-1} (y_{it} - \mu_i) + \right.$$

$$\begin{aligned}
& +2\rho \sum_{t=2}^T (y_{it} + y_{it-1} - 2\mu_i) \Big\} \\
= & \frac{1}{\sigma^2} \left\{ \sum_{t=1}^T y_{it} - T\mu_i + \rho^2 \sum_{t=2}^{T-1} y_{it} - (T-2)\rho^2\mu_i - \rho \sum_{t=2}^T y_{it} + \right. \\
& \left. -\rho \sum_{t=2}^T y_{it-1} + 2(T-1)\rho\mu_i \right\}.
\end{aligned}$$

Ponendo $l_{\mu_i} = 0$ e risolvendo rispetto a μ_i si ottiene

$$\hat{\mu}_{i\rho} = \frac{\sum_{t=1}^T y_{it} - \rho(\sum_{t=2}^T y_{it} + \sum_{t=1}^{T-1} y_{it}) + \rho^2 \sum_{t=2}^{T-1} y_{it}}{T - 2(T-1)\rho + (T-2)\rho^2}.$$

La log-verosimiglianza profilo per il modello considerato è data dalla quantità $l_P(\rho, \sigma^2) = l(\hat{\mu}_{i\rho}, \rho, \sigma^2)$, log-verosimiglianza calcolata nelle s.m.v. vincolate $\hat{\mu}_{i\rho}$.

Per quanto riguarda invece i termini della modificazione della verosimiglianza profilo, occorre calcolare le quantità $j_{\mu_i\mu_i}(\tau, \hat{\mu}_{i\rho})$ e $I_{\mu_i\mu}^i(\tau, \hat{\mu}_{i\rho}; \hat{\tau}, \hat{\mu}_i)$.

Per il termine $j_{\mu_i\mu_i}(\tau, \hat{\mu}_{i\rho})$, pari alla derivata seconda $l_{\mu_i\mu_i}(\tau, \hat{\mu}_{i\rho})$ cambiata di segno, è immediato verificare che esso è pari a

$$\frac{1}{\sigma^2} \{T - 2(T-1)\rho + (T-2)\rho^2\}.$$

Il calcolo del termine $I_{\mu_i\mu}^i(\tau, \hat{\mu}_{i\rho}; \hat{\tau}, \hat{\mu}_i)$ risulta invece più laborioso e consiste nel calcolare il valore atteso $E_{\theta_0} \{l_{\mu_i}(\theta_0) l_{\mu_i}(\theta_1)\}$, con $\theta_0 = (\hat{\tau}, \hat{\mu}_i)$ e $\theta_1 = (\tau, \hat{\mu}_{i\rho})$. Considerando che la derivata prima l_{μ_i} può essere anche scritta come

$$\begin{aligned}
l_{\mu_i} = & \frac{1}{\sigma^2} \left\{ \sum_{t=1}^T (y_{it} - \mu_i) + \rho^2 \sum_{t=2}^{T-1} (y_{it} - \mu_i) - \rho \sum_{t=2}^T (y_{it} - \mu_i) + \right. \\
& \left. -\rho \sum_{t=1}^{T-1} (y_{it} - \mu_i) \right\},
\end{aligned}$$

si ha che

$$I_{\mu_i\mu}^i(\theta_0, \theta_1) = \frac{1}{\sigma_0^2 \sigma_1^2} E_{\theta_0} \left\{ \left[\sum_{t=1}^T (y_{it} - \mu_{i0}) + \rho_0^2 \sum_{t=2}^{T-1} (y_{it} - \mu_{i0}) + \right. \right.$$

$$\begin{aligned}
& -\rho_0 \sum_{t=2}^T (y_{it} - \mu_{i0}) - \rho_0 \sum_{t=1}^{T-1} (y_{it} - \mu_{i0}) \left[\sum_{t'=1}^T (y_{it'} - \mu_{i0}) + \right. \\
& + T(\mu_0 - \mu_1) + \rho_1^2 \sum_{t'=2}^{T-1} (y_{it'} - \mu_{i0}) + \rho_1^2 (T-2)(\mu_0 - \mu_1) + \\
& -\rho_1 \sum_{t'=2}^T (y_{it'} - \mu_{i0}) - \rho_1 (T-1)(\mu_0 - \mu_1) + \\
& \left. -\rho_1 \sum_{t'=1}^{T-1} (y_{it'} - \mu_{i0}) - \rho_1 (T-1)(\mu_0 - \mu_1) \right] \Big\}.
\end{aligned}$$

Dato che $E_\theta(Y_{it}) = \mu_i$, $V(Y_{it}) = \frac{\sigma^2}{1-\rho^2}$ e $cov(Y_{it}, Y_{it'}) = E_\theta\{(Y_{it} - \mu_i)(Y_{it'} - \mu_i)\} = \frac{\sigma^2}{1-\rho^2} \rho^{|t-t'|}$ e tenendo presente che $E_{\theta_0}(Y_{it} - \mu_{i0}) = 0$, si ha che

$$\begin{aligned}
I_{\mu;\mu}^i(\theta_0, \theta_1) &= \frac{1}{(1-\rho_0^2)\sigma_1^2} \left\{ \sum_{t=1}^T \sum_{t'=1}^T \hat{\rho}^{|t-t'|} + \rho_1^2 \sum_{t=1}^T \sum_{t'=2}^{T-1} \hat{\rho}^{|t-t'|} + \right. \\
& -\rho_1 \sum_{t=1}^T \sum_{t'=2}^T \hat{\rho}^{|t-t'|} - \rho_1 \sum_{t=1}^T \sum_{t'=1}^{T-1} \hat{\rho}^{|t-t'|} + \rho_0^2 \sum_{t=2}^{T-1} \sum_{t'=1}^T \hat{\rho}^{|t-t'|} + \\
& + \rho_0^2 \rho_1^2 \sum_{t=2}^{T-1} \sum_{t'=2}^{T-1} \hat{\rho}^{|t-t'|} - \rho_1 \rho_0^2 \sum_{t=2}^{T-1} \sum_{t'=2}^T \hat{\rho}^{|t-t'|} + \\
& -\rho_1 \rho_0^2 \sum_{t=2}^{T-1} \sum_{t'=1}^{T-1} \hat{\rho}^{|t-t'|} - \rho_0 \sum_{t=2}^T \sum_{t'=1}^T \hat{\rho}^{|t-t'|} - \rho_0 \rho_1^2 \sum_{t=2}^T \sum_{t'=2}^{T-1} \hat{\rho}^{|t-t'|} + \\
& + \rho_1 \rho_0 \sum_{t=2}^T \sum_{t'=2}^T \hat{\rho}^{|t-t'|} + \rho_0 \rho_1 \sum_{t=2}^T \sum_{t'=1}^{T-1} \hat{\rho}^{|t-t'|} - \rho_0 \sum_{t=1}^{T-1} \sum_{t'=1}^T \hat{\rho}^{|t-t'|} + \\
& -\rho_0 \rho_1^2 \sum_{t=1}^{T-1} \sum_{t'=2}^{T-1} \hat{\rho}^{|t-t'|} + \rho_0 \rho_1 \sum_{t=1}^{T-1} \sum_{t'=2}^T \hat{\rho}^{|t-t'|} + \\
& \left. + \rho_0 \rho_1 \sum_{t=1}^{T-1} \sum_{t'=1}^{T-1} \hat{\rho}^{|t-t'|} \right\}.
\end{aligned}$$

Ora sostituendo a (ρ_0, σ_0^2) le s.m.v. $(\hat{\rho}, \hat{\sigma}^2)$ e a (ρ_1, σ_1^2) i valori (ρ, σ^2) e raccogliendo opportunamente, secondo le uguaglianze riportate nel paragrafo 3.4.1, si ottiene l'espressione nella (3.12).

Nella simulazione si sono calcolate anche la log-verosimiglianza profilo per ρ di $l_P(\rho, \sigma^2)$ e di $l_M(\rho, \sigma^2)$ con l'obiettivo di individuare degli intervalli di confidenza per ρ che non comprendano valori inammissibili per il parametro, ovvero intervalli basati sul test r .

Partendo dall'espressione della log-verosimiglianza profilo

$$\begin{aligned} l_P(\rho, \sigma^2) &= -\frac{qT}{2} \log \sigma^2 + \frac{q}{2} \log(1 - \rho^2) - \frac{1}{2\sigma^2} \sum_{i=1}^q \left\{ \sum_{t=1}^T (y_{it} - \hat{\mu}_{i\rho})^2 + \right. \\ &\quad \left. + \rho^2 \sum_{t=2}^{T-1} (y_{it} - \hat{\mu}_{i\rho})^2 - 2\rho \sum_{t=2}^T (y_{it} - \hat{\mu}_{i\rho})(y_{it-1} - \hat{\mu}_{i\rho}) \right\} \\ &= -\frac{qT}{2} \log \sigma^2 + \frac{q}{2} \log(1 - \rho^2) - \frac{1}{2\sigma^2} A(\rho, \hat{\mu}_{i\rho}), \end{aligned}$$

con

$$A(\rho, \hat{\mu}_{i\rho}) = \sum_{i=1}^q \left\{ \sum_{t=1}^T (y_{it} - \hat{\mu}_{i\rho})^2 + \rho^2 \sum_{t=2}^{T-1} (y_{it} - \hat{\mu}_{i\rho})^2 - 2\rho \sum_{t=2}^T (y_{it} - \hat{\mu}_{i\rho})(y_{it-1} - \hat{\mu}_{i\rho}) \right\},$$

e derivando rispetto a σ^2 , ora trattato come parametro di disturbo, si ottiene

$$l_{\sigma^2} = -\frac{qT}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} A(\rho, \hat{\mu}_{i\rho}).$$

Quindi, ponendo $l_{\sigma^2} = 0$ e risolvendo rispetto a σ^2 , si ottiene

$$\hat{\sigma}_\rho^2 = \frac{A(\rho, \hat{\mu}_{i\rho})}{qT}.$$

La log-verosimiglianza profilo per ρ , $l_P(\rho)$, è quindi pari alla log-verosimiglianza profilo per (ρ, σ^2) calcolata in $\hat{\sigma}_\rho^2$, $l_P(\rho, \hat{\sigma}_\rho^2)$.

Per il calcolo della log-verosimiglianza profilo della profilo modificata, i passaggi sono analoghi. Per cui, partendo dalla log-verosimiglianza profilo modificata

$$l_M(\rho, \sigma^2) = l_P(\rho, \sigma^2) + \frac{q}{2} \log \sigma^2 + \frac{q}{2} \log \{T - 2(T-1)\rho + (T-2)\rho^2\} - q \log K(\rho),$$

con

$$K(\rho) = A + (\hat{\rho} + \rho)[(\hat{\rho} + \rho)B - 2C - 2\hat{\rho}\rho] + 2\hat{\rho}\rho(B + 1 + \hat{\rho}^{T-1}) + \hat{\rho}^2 \rho^2 D,$$

e derivando rispetto a σ^2 , si ottiene

$$l_{M\sigma^2} = -\frac{qT}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} A(\rho, \hat{\mu}_{i\rho}) + \frac{q}{2\sigma^2}.$$

Quindi, ponendo $l_{M\sigma^2} = 0$ e risolvendo rispetto a σ^2 , si ottiene

$$\hat{\sigma}_{M\rho}^2 = \frac{A(\rho, \hat{\mu}_{i\rho})}{q(T-1)}.$$

La log-verosimiglianza profilo della profilo modificata è quindi pari a $l_{MP}(\rho) = l_M(\rho, \hat{\sigma}_{M\rho}^2)$, log-verosimiglianza profilo modificata calcolata in $\hat{\sigma}_{M\rho}^2$.

Riferimenti bibliografici

- Arellano, M. e Hahn, J. (2006). Understanding Bias in Nonlinear Panel Models.
- Azzalini, A. (2001). *Inferenza Statistica: una Presentazione Basata sul Concetto di Verosimiglianza*. Springer-Verlag Italia, Milano.
- Barndorff-Nielsen, O.E. (1980). Conditionality resolutions. *Biometrika*, **67**, 293-310.
- Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343-365.
- Barndorff-Nielsen, O.E. (1994). Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. *J. Roy. Statist. Soc., B*, **56**, 125-140.
- Barndorff-Nielsen, O.E. (1995). Stable and invariant adjusted profile likelihood for curved exponential models. *Biometrika*, **82**, 489-499.
- Cox, D.R. e Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Roy. Statist. Soc., B*, **49**, 1-39.
- Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc. A*, **144**, 285-307.
- Fisher, R.A. (1935). The logic of inductive inference. *J. Roy. Statist. Soc.*, **98**, 39-54.

- Fraser, D.A.S. e Reid N. (1988). On conditional inference for a real parameter: a differential approach on the sample space. *Biometrika*, **38**, 251-274.
- Fraser, D.A.S. e Reid N. (1989). Adjustments to profile likelihoods. *Biometrika*, **76**, 477-488.
- Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Journal of Econometrics*, **7**, 98-119.
- Honoré, Bo E. (2002). Nonlinear Models with Panel Data. Working Paper CWP13. Department of Economics. Princeton University.
- Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, **95**, 391-413.
- McCullagh, P. e Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *J. Roy. Statist. Soc., B*, **52**, 325-344.
- Neyman, J. e Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 1-32.
- Pace, L. e Salvan, A. (1996). *Teoria della Statistica*. Cedam, Padova.
- Pace, L. e Salvan, A. (2001). *Introduzione alla Statistica*. Cedam, Padova.
- Sartori, N. (2001). *Likelihood Treatment of Incidental Nuisance Parameters*. Tesi di Dottorato, Facoltà di Scienze Statistiche, Università di Padova.
- Sartori, N. (2003). Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika*, **90**, 533-549.
- Sartori, N. (2005). Modified maximum likelihood for stratified data. Department of Statistics. University of Venice.

Severini, T.A. (1998). An approximation to the modified profile likelihood function. *Biometrika*, **85**, 403-411.

Severini, T.A. (2000). *Likelihood Methods in Statistics*. Oxford University Press, Oxford.