

UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA
IN STATISTICA ECONOMIA E FINANZA

TESI DI LAUREA

**VEROSIMIGLIANZA COMPOSITA
MARGINALE**

RELATORE: PROF. ALESSANDRA SALVAN

LAUREANDA: ELISA GAIOFATTO

ANNO ACCADEMICO 2007/2008

Indice

Introduzione	III
1 Verosimiglianza composta	1
1.1 Introduzione	1
1.2 Specificazione del modello statistico	2
1.3 La funzione di verosimiglianza	2
1.4 Verosimiglianza composta	4
1.5 Verosimiglianza composta marginale	6
1.6 Equazioni di stima	7
1.7 Stimatore MCL	9
1.7.1 Verosimiglianza composta pesata	11
1.7.2 Errori standard	12
1.7.3 Verifica d'ipotesi	12
2 Verosimiglianza composta marginale	15
2.1 Introduzione	15
2.2 Verosimiglianza a coppie	16
2.2.1 Combinazione delle verosimiglianze	16
2.2.2 Verosimiglianza composta marginale combinata	16
2.3 Verosimiglianza d'indipendenza	19
2.3.1 Le equazioni di stima d'indipendenza	19
3 Applicazioni	23
3.1 Introduzione	23

3.2	Verosimiglianza completa e a coppie	23
3.3	Verosimiglianza completa e d'indipendenza	30
3.4	Un'applicazione all'analisi di dati di sopravvivenza	42
	Appendice: comandi R	59
	Conclusioni	65
	Riferimenti bibliografici	67

Introduzione

Nell'ambito della teoria dell'inferenza statistica, la funzione di verosimiglianza ha un ruolo di primaria importanza. Tuttavia, in numerose situazioni, fare inferenza in modelli parametrici complessi, in cui la funzione di verosimiglianza è difficile da calcolare o fortemente basata sulle assunzioni del modello, non risulta conveniente.

Nelle applicazioni statistiche recenti si hanno spesso molte informazioni, disponibili in insiemi di dati di dimensioni via via crescenti. Si tende inoltre sempre più ad utilizzare modelli realistici per descrivere fenomeni complessi che rendono impossibile il calcolo della funzione di verosimiglianza.

Esigenze computazionali hanno portato numerosi autori a studiare possibili ampliamenti alla definizione di funzione di verosimiglianza in modo tale da affrontare le difficoltà legate alla sua specificazione.

Un'alternativa ai metodi di verosimiglianza ordinaria consiste nell'adottare pseudo-verosimiglianze più semplici, come quelle appartenenti alla classe delle *verosimiglianze composite* (Lindsay, 1988). La verosimiglianza composta si ottiene combinando validi oggetti di verosimiglianza, relativi solitamente a piccoli sottoinsiemi di dati.

Alcuni esempi sono la *pseudo-verosimiglianza di Besag* (1974) per l'inferenza in modelli spaziali e la *verosimiglianza parziale di Cox* (1975) suggerita per l'inferenza in modelli con rischi proporzionali. L'utilizzo della verosimiglianza composta si riscontra anche nel campo della genetica, delle serie storiche e nell'analisi di dati di sopravvivenza, si veda Varin

(2007) per un resoconto recente.

Il merito della verosimiglianza composta risiede nella capacità di ridurre la complessità computazionale, cosicché sia possibile considerare grandi insiemi di dati e modelli statistici assai complessi. Inoltre, è utilizzabile anche quando non è possibile calcolare la verosimiglianza completa.

Un crescente interesse si sta formando attorno alle *verosimiglianze composite marginali*, ovvero a pseudo-verosimiglianze ottenute componendo adeguatamente densità marginali. Quando sono coinvolte distribuzioni marginali di piccole dimensioni, il calcolo degli oggetti di verosimiglianza risulta relativamente semplice. La verosimiglianza composta è utilizzata principalmente per la sua semplicità computazionale. Inoltre, sotto adeguate condizioni di regolarità, le varie procedure inferenziali hanno proprietà teoriche simili a quelle dei metodi basati sulla verosimiglianza completa, sebbene ci si aspetti una perdita di efficienza.

L'obiettivo principale della tesi consiste nello studio della verosimiglianza composta e nella sua applicazione ad alcuni semplici esempi. In questi si confrontano i risultati ottenuti utilizzando sia la verosimiglianza completa sia quella composta e si valuta la perdita asintotica di efficienza quando si considera la verosimiglianza composta marginale.

Il primo capitolo è un capitolo di rassegna che introduce gli strumenti e i concetti che verranno poi applicati nel seguito della tesi. Vengono richiamate le nozioni fondamentali dell'inferenza statistica e le equazioni di stima. Vengono trattati dettagliatamente i concetti di *verosimiglianza composta* e di *verosimiglianza composta marginale*. Infine, si considera lo stimatore di massima verosimiglianza composta e le sue principali proprietà statistiche.

Il capitolo 2 consiste in una trattazione di due specifici tipi di verosimiglianza composta marginale e relative applicazioni: *verosimiglianza a coppie* e *verosimiglianza d'indipendenza*. Viene introdotta una particolare forma di verosimiglianza composta marginale, detta *combinata*, la quale cerca di combinare in modo ottimale le informazioni provenienti dalle dis-

tribuzioni univariate e bivariate (Cox e Reid, 2004). Successivamente, si considera un'applicazione della verosimiglianza d'indipendenza a strutture di dati raggruppati.

Il terzo capitolo ha lo scopo di adattare la teoria esposta ad alcuni esempi, utilizzando sia la verosimiglianza completa sia quella composita marginale. In particolare vengono confrontati i risultati ottenuti con i due metodi e viene calcolata la perdita di efficienza che si ottiene quando si considera la verosimiglianza composita marginale.

Capitolo 1

Verosimiglianza composta

1.1 Introduzione

Il presente capitolo è essenzialmente di rassegna, avente l'obiettivo di introdurre gli strumenti e le metodologie utilizzati in seguito. Si richiamano le nozioni di base dell'inferenza statistica, il modello statistico, la funzione di verosimiglianza e le quantità di verosimiglianza ad essa associate, fissandone la notazione.

Il quarto paragrafo è dedicato alla verosimiglianza composta e alle sue caratteristiche. Successivamente, viene introdotto il concetto di verosimiglianza composta marginale, la quale è un caso particolare della verosimiglianza composta perché costruita a partire da particolari densità marginali.

Il sesto paragrafo verte sulle equazioni di stima, dove si evidenzia l'importanza della proprietà di non distorsione dell'equazione di stima per avere stimatori consistenti.

Infine, si considera lo stimatore di massima verosimiglianza composta e le sue principali proprietà asintotiche.

1.2 Specificazione del modello statistico

L'obiettivo dell'inferenza statistica è sintetizzare i dati osservati y attraverso la ricostruzione della distribuzione di probabilità $f^0(y)$ della variabile casuale Y che si assume li abbia generati. Si cerca di ricavare una descrizione sintetica del fenomeno oggetto di studio, che ne permetta l'interpretazione e successivamente la previsione, considerando le osservazioni y come una determinazione di un vettore casuale Y .

Inizialmente si delimitano le forme ritenute possibili per $f^0(y)$, ovvero si specifica una famiglia di distribuzioni di probabilità associata ai dati e ad essi qualitativamente compatibile. Questo insieme \mathcal{F} viene definito modello statistico.

Il modello statistico avrà una generalità più o meno ampia a seconda delle conoscenze preliminari e del volume di dati disponibili. Si possono determinare tre livelli di specificazione, a seconda del numero di assunzioni e di informazioni disponibili: parametrica, semiparametrica e non parametrica.

Nel seguito della tesi verranno considerati soltanto problemi inferenziali trattabili in ambito parametrico.

Un modello statistico è quindi una classe, i cui elementi sono delle funzioni di densità di probabilità indicizzati da un parametro θ , p -dimensionale

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

Assumendo y realizzazione di una variabile casuale $Y \sim f^0(y)$, se $f^0(y) \in \mathcal{F}$ il modello si dice **correttamente specificato**. In questo caso si ha $f^0(y) = f(y; \theta_0)$, dove con $\theta_0 \in \Theta$ si indica il vero valore del parametro.

1.3 La funzione di verosimiglianza

Assumendo valido il modello \mathcal{F} , la sintesi informativa dei dati osservati y su θ è la funzione di verosimiglianza

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

dove $c(y) > 0$ è una costante di proporzionalità arbitraria e per questo trascurabile.

Se i dati sono costituiti da n osservazioni, $y = (y_1, \dots, y_n)$, indipendenti ed identicamente distribuite (i.i.d.), la verosimiglianza assume la forma

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

dove $f(y_i; \theta)$, $i = 1, \dots, n$, indica la distribuzione marginale della variabile casuale Y_i , relativa all' i -esima osservazione. In questo caso la funzione di verosimiglianza complessiva è semplicemente il prodotto delle funzioni di verosimiglianza ottenute nei singoli esperimenti.

Spesso le procedure inferenziali basate su $L(\theta)$ vengono espresse tramite il logaritmo naturale della verosimiglianza, ottenendo così la funzione di log-verosimiglianza

$$l(\theta) = l(\theta; y) = \log L(\theta).$$

Convenzionalmente si definisce $l(\theta) = -\infty$ se $L(\theta) = 0$.

Sotto campionamento casuale semplice si ha

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

Assumendo che la log-verosimiglianza sia una funzione di θ regolare, ossia infinitamente derivabile, si richiamano le seguenti quantità di verosimiglianza con la notazione che verrà utilizzata in seguito.

- Dato il parametro θ con componenti $\theta_1, \dots, \theta_p$, la funzione di punteggio è il vettore delle derivate parziali prime di $l(\theta)$ rispetto alle componenti $\theta_1, \dots, \theta_p$ di θ e verrà indicato come

$$l_* = l_*(\theta; y) = (l_1, \dots, l_p)^\top,$$

dove $l_r = \partial l(\theta) / \partial \theta_r$, $r = 1, \dots, p$, indica la derivata parziale di $l(\theta)$ rispetto all' r -esima componente del vettore θ .

- La matrice di informazione osservata è la matrice di dimensione $(p \times p)$ delle derivate parziali seconde di $l(\theta)$ cambiate di segno. Viene così indicata

$$j = j(\theta) = \left[-\frac{\partial^2 l}{\partial \theta \partial \theta^\top} \right]$$

o, in notazione compatta, $j = [-l_{rs}]$, dove $l_{rs} = \partial^2 l(\theta) / (\partial \theta_r \partial \theta_s)$ rappresenta l'elemento di posto (r, s) nella matrice j .

- La matrice di informazione attesa di Fisher è il valore atteso, sotto θ , dell'informazione osservata, indicata con

$$i = i(\theta) = \mathbb{E}_\theta \{j(\theta)\},$$

Anche in questo caso il generico elemento di posto (r, s) viene indicato con i_{rs} .

Se si assumono soddisfatte le condizioni che permettono di scambiare le operazioni di derivazione e di integrazione, valgono alcuni risultati di particolare interesse inferenziale.

- La funzione di punteggio ha valore atteso nullo $\mathbb{E}_\theta \{l_*(\theta)\} = 0$, $\theta \in \Theta$. Questa proprietà stabilisce che l'equazione di verosimiglianza $l_*(\theta) = 0$ è un'equazione di stima non distorta.
- Vale l'identità dell'informazione, ossia

$$\mathbb{E}_\theta (l_*(\theta) l_*(\theta)^\top) = i(\theta), \quad \theta \in \Theta.$$

La matrice $i(\theta)$ risulta essere uguale alla matrice di varianze e covarianze della funzione di punteggio, e come tale è una matrice definita non negativa.

1.4 Verosimiglianza composta

Il presente paragrafo verte sul concetto di verosimiglianza composta, il cui termine indica una classe di pseudo-verosimiglianze basate su oggetti

di tipo verosimiglianza. Sia $Y = (Y_1, \dots, Y_q)^\top$ una variabile casuale di dimensione q con densità congiunta $f(y; \theta)$, dove $\theta \in \Theta \subseteq \mathbb{R}^p$. Si consideri un modello statistico parametrico $\{f(y; \theta), y \in Y \subseteq \mathbb{R}^q, \theta \in \Theta \subseteq \mathbb{R}^p\}$ e un insieme di eventi misurabili $\{A_i; i = 1, \dots, m\}$.

La verosimiglianza composita, CL , è il processo ponderato delle verosimiglianze corrispondenti a ogni singolo evento, ovvero è la combinazione di validi oggetti di verosimiglianza

$$CL(\theta; y) = \prod_{i=1}^m f(y \in A_i; \theta)^{w_i},$$

dove $f(y \in A_i; \theta)^{w_i}$ è la verosimiglianza relativa all'evento $y \in A_i$ e $\{w_i; i = 1, \dots, m\}$ sono pesi positivi. Ad essa è associata la log-verosimiglianza composita $cl(\theta; y) = \log CL(\theta; y)$.

Il suo massimo, se esiste, è detto **stimatore di massima verosimiglianza composita** (*MCLE*).

La classe delle verosimiglianze di tipo composito può essere suddivisa in 2 gruppi:

- **metodi di omissione**

La pseudo-verosimiglianza è ottenuta rimuovendo alcuni termini che rendono complicato il calcolo della verosimiglianza completa. Si desidera che tale riduzione di complessità avvenga con una perdita d'informazione su θ nulla o trascurabile, ovvero la parte omessa non dovrebbe essere molto informativa sul parametro d'interesse e così la perdita di efficienza di questo metodo risulta tollerabile.

Alcuni esempi riguardano la *pseudo-verosimiglianza di Besag* (Besag, 1974, 1977) per dati spaziali, la *verosimiglianza di ordine m* per processi stazionari (Azzalini, 1983) e la *verosimiglianza di Stein* (2004) per grandi insiemi di dati spaziali.

- verosimiglianze composite marginali

Si tratta di verosimiglianze composite costruite a partire dalla pseudo-verosimiglianza marginale. L'idea di base consiste nel costruire *modelli* a partire dalla specificazione diretta della distribuzione congiunta di coppie, triplette, ... ossia nel costruire pseudo verosimiglianze componendo distribuzioni marginali di piccole dimensioni. La verosimiglianza composta marginale differisce dalla pseudo verosimiglianza marginale utilizzata per fare inferenza in presenza di parametri di disturbo.

Questo metodo, consistente nella combinazione di densità marginali, risulta un'alternativa assai attraente soprattutto nel caso di applicazioni complesse, che vanno dai *cluster* e dai dati longitudinali o spaziali fino alle serie temporali e all'analisi di dati di sopravvivenza.

Nel seguito della tesi verrà approfondito il concetto di verosimiglianza composta marginale attraverso alcuni casi e relative applicazioni.

1.5 Verosimiglianza composta marginale

La verosimiglianza composta marginale comporta un notevole risparmio computazionale poiché non richiede la distribuzione congiunta ma solo una particolare combinazione di densità marginali fino a un certo ordine.

Utilizzando solo le marginali univariate si costruisce la verosimiglianza marginale

$$SL(\theta; y) = \prod_{i=1}^q f(y_i; \theta)^{w_i},$$

denominata verosimiglianza d'indipendenza, si veda ad esempio Chandler e Bate (2007). Utilizzando le marginali bivariate si ha

$$PL(\theta; y) = \prod_{i=1}^{q-1} \prod_{j=1+1}^q f(y_i, y_j; \theta)^{w_{i,j}},$$

detta verosimiglianza a coppie o *pairwise likelihood*, si veda ad esempio Le Cassie e Van Honwlingen (1994).

Talvolta può essere utile considerare sottoinsiemi composti da triplete di osservazioni (Varin e Vidoni, 2005; Engler et al., 2006).

Un problema interessante consiste nel combinare in modo ottimale le verosimiglianze basate sulle marginali univariate e bivariate (Cox e Reid, 2004). Tuttavia si deve prestare particolare attenzione alla fase di costruzione dell'adeguata combinazione delle verosimiglianze al fine di garantire la consistenza degli stimatori.

La semplicità computazionale non è l'unico motivo che porta a preferire l'utilizzo della verosimiglianza composita marginale in luogo della verosimiglianza completa. Si deve sempre tenere in considerazione anche la proprietà di robustezza, dato che per la verosimiglianza composita marginale è sufficiente specificare solo distribuzioni uni-, bi-, ... dimensionali e non è richiesta la distribuzione congiunta dei dati.

1.6 Equazioni di stima

Affinché l'inferenza basata sulla verosimiglianza composita sia sufficientemente accurata, l'equazione di stima ad essa associata deve essere almeno approssimativamente non distorta. In questo paragrafo si riassumono i principali concetti relativi alle equazioni di stima non distorte.

Un'equazione di stima è un'equazione del tipo

$$q(\theta; y) = 0$$

tale che, risolto rispetto a θ , fornisca uno stimatore $\hat{\theta}$. La funzione $q(\theta; y)$ è detta funzione di stima.

Dato un modello \mathcal{F} con parametro $\theta \in \Theta \subseteq \mathbb{R}^p$, l'equazione di stima $q(\theta; y) = 0$ è detta non distorta se vale

$$\mathbb{E}_\theta(q(\theta; Y)) = 0, \quad \text{per ogni } \theta \in \Theta.$$

Se il parametro θ è p -dimensionale allora $q(\theta; Y) = 0$ rappresenta un sistema di p equazioni in p incognite. Per un modello con verosimiglianza regolare, sotto opportune condizioni, di cui la principale è che il supporto di Y non dipenda da θ , tale proprietà è verificata per l'equazione di verosimiglianza $l_*(\theta) = 0$.

Nel caso si disponga di un campione di osservazioni $y = (y_1, \dots, y_n)$, indipendenti e con funzione di densità $p(y_i; \theta)$, le funzioni di stima assumono la forma

$$q(\theta; y) = \sum_{i=1}^n g(\theta; y_i),$$

dove si richiede che le funzioni di stima per una singola osservazione, $g(\theta; y_i)$, siano non distorte.

Se θ è scalare, la funzione di *punteggio* gode della proprietà di ottimalità tra tutte le funzioni di stima non distorte $q(\theta; y)$, poiché soddisfa il criterio di ottimalità di Godambe

$$\frac{V_\theta(l_*(\theta))}{\{\mathbb{E}_\theta(\partial l_*/\partial \theta)\}^2} \leq \frac{V_\theta(q(\theta; Y))}{\{\mathbb{E}_\theta(\partial q(\theta; Y)/\partial \theta)\}^2}, \quad \text{per ogni } \theta \in \Theta.$$

Il requisito di non distorsione della funzione di stima risulta fondamentale per dimostrare la consistenza dello stimatore corrispondente.

In genere, anche per equazioni di stima non distorte, non vale la seconda identità di Bartlett, ovvero

$$V_\theta(g(\theta; Y_i)) \neq -\mathbb{E}_\theta\left(\frac{\partial}{\partial \theta^\top} g(\theta; Y_i)\right).$$

Tuttavia è possibile definire, una nuova equazione di stima non distorta, basata sulla funzione modificata

$$\tilde{q}(\theta; y) = H_g(\theta)^\top V_g(\theta)^{-1} \sum_{i=1}^n g(\theta; y_i),$$

dove $H_g(\theta) = -\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta^\top}g(\theta; Y_i)\right)$ e $V_g(\theta) = \text{Var}_\theta(g(\theta; Y_i))$.

Questa equazione, sotto opportune condizioni, ha la stessa soluzione dell'equazione di stima basata su $q(\theta; y)$ e soddisfa l'identità $V_\theta(\tilde{q}(\theta; Y)) = -\mathbb{E}_\theta(\partial\tilde{q}(\theta; Y)/\partial\theta^\top)$, per ogni $\theta \in \Theta$.

1.7 Stimatore MCL

Data la log-verosimiglianza composta, si può costruire la funzione di punteggio calcolandone la derivata prima

$$U(\theta; y) = \nabla cl(\theta; y) = \sum_{i=1}^m w_i U_i(\theta; y),$$

dove $U_i(\theta; y) = \nabla \log f(y \in A_i; \theta)$, per $i = 1, \dots, m$.

Lo stimatore di massima verosimiglianza composta (MCLE) è tale da annullare $U(\theta; y)$. Pertanto lo stimatore è la soluzione, se esiste ed è unica, dell'equazione di stima basata sulla verosimiglianza composta.

Essendo ogni componente di $cl(\theta; y)$ un oggetto di tipo verosimiglianza, l'equazione di stima $U(\theta; y) = 0$ risulta non distorta, sotto le usuali condizioni di regolarità. Le proprietà asintotiche dello stimatore $\hat{\theta}$ sono ottenute dalla non distorsione dell'equazione di stima $U(\theta; y)$. Tuttavia, la non distorsione di $U(\theta; y)$ non è una condizione sufficiente per la consistenza dello stimatore di massima verosimiglianza composta. Sono necessarie appropriate assunzioni, come le condizioni di regolarità proprie della verosimiglianza completa, che devono essere soddisfatte per ogni termine che forma la pseudo-verosimiglianza.

Poiché la funzione di punteggio è una combinazione lineare di funzioni di punteggio associate a ogni singolo termine che forma la verosimiglianza composta, allora, sotto appropriate assunzioni, lo stimatore $\hat{\theta}_{MCLE}$ è consistente e asintoticamente normale

$$\hat{\theta}_{MCLE} \sim N(\theta, G(\theta)^{-1}),$$

dove $G(\theta; y)$ è la matrice di informazione di Godambe associata alla verosimiglianza composta. Questa matrice, detta anche *sandwich information matrix*, si può scrivere come

$$G(\theta) = H(\theta)J(\theta)^{-1}H(\theta),$$

con $H(\theta) = \mathbb{E}_\theta(-\nabla U(\theta; y))$ e $J(\theta) = \text{Var}_\theta(U(\theta; y))$.

La forma della matrice di Godambe è dovuta alla non validità dell'identità dell'informazione, ovvero $H(\theta) \neq J(\theta)$. Questo conduce generalmente ad una perdita di efficienza rispetto a quando si utilizza lo stimatore di massima verosimiglianza.

L'inefficienza delle verosimiglianze composte può essere ridotta attraverso un'adeguata modifica della funzione di punteggio composta $U(\theta; y)$. Nel caso di un parametro scalare, l'informazione di Godambe può essere riscritta come

$$G(\theta) = i(\theta)\text{Corr}_\theta(U(\theta; y), l_*(\theta; y))^2,$$

con $i(\theta)$ informazione di Fisher e $l_*(\theta; y)$ funzione di punteggio basata sulla verosimiglianza completa. L'efficienza è raggiunta in presenza di una relazione lineare tra la funzione di punteggio composta e quella ordinaria. Quindi la funzione punteggio composta modificata $\tilde{U}(\theta; y) = H(\theta)J(\theta)^{-1}l_*(\theta)$ risulta ottima in quanto soddisfa la seconda identità di Bartlett.

Al fine di migliorare l'efficienza di $\hat{\theta}_{MCLE}$, una possibile alternativa consiste nello scegliere un adeguato insieme di pesi w_i , imponendo un qualche criterio di ottimo.

Le proprietà asintotiche degli stimatori di massima verosimiglianza composta dipendono principalmente dalla velocità con cui l'informazione di Godambe diverge. Se diverge abbastanza velocemente, allora lo stimatore di massima verosimiglianza composta sarà consistente e con distribuzione asintotica normale. Ad esempio, quando si considerano

modelli per serie storiche con memoria particolarmente lunga, i metodi di verosimiglianza composita non conducono a stimatori consistenti del parametro θ .

Può anche succedere che lo stimatore di massima verosimiglianza composita marginale risulti consistente, mentre non lo sia lo stimatore ottenuto considerando la verosimiglianza completa. Questa situazione deriva dal fatto che le procedure che utilizzano la verosimiglianza composita sono solitamente più robuste di quelle basate sulla verosimiglianza completa (Liang, 1987; Lindsay, 1988).

Infine, si deve porre particolare attenzione al problema dell'identificabilità dei parametri perché la verosimiglianza composita marginale si basa unicamente su densità marginali di piccole dimensioni e non sull'intero insieme di dati.

1.7.1 Verosimiglianza composita pesata

Quando si considera la verosimiglianza composita si deve affrontare il problema riguardante la scelta ottima dei pesi, in modo tale da suggerire una strategia che migliori la funzione di stima composita. Sia $U(\theta; y) = (U_1(\theta; y), \dots, U_m(\theta; y))$ il vettore costituito dalle componenti della funzione punteggio e sia $w = (w_1, \dots, w_m)$ il vettore dei pesi associati, dove w_i è il peso attribuito al generico evento A_i . L'obiettivo è trovare una combinazione lineare $w(\theta)^\top U(\theta; y)$ che massimizzi la matrice di informazione di Godambe oppure che risolva un problema ai minimi quadrati. I pesi ottimi risultano

$$w_{ott}(\theta) = \text{Var}_\theta(U(\theta; Y))^{-1} \mathbb{E}_\theta(U^2(\theta; Y)).$$

Spesso la stima dei pesi risulta difficile in modo particolare quando il parametro θ non è unidimensionale (cfr. Lindsay, 1988). Inoltre i pesi non sono sempre scelti secondo criteri ottimi. Alcuni autori hanno proposto

dei pesi che possono essere considerati delle variabili indicatrici, le quali selezionano i termini da includere nella verosimiglianza composta.

1.7.2 Errori standard

Per calcolare gli errori standard è necessario disporre di stime consistenti delle matrici $H(\theta)$ e $J(\theta)$ che compongono la matrice d'informazione di Godambe. Uno stimatore empirico per $H(\theta)$ si ottiene sostituendo θ con la sua stima $\hat{\theta}_{MCLE}$ in $\mathbb{E}_\theta(-\nabla U(\theta; y))$ ottenendo

$$\widehat{H}(Y) = -\frac{1}{W_m} \sum_{i=1}^m w_i \nabla U_i(\hat{\theta}_{MCLE}; y),$$

dove $W_m = \sum_{i=1}^m w_i$, oppure, se l'Hessiano è difficile da trattare

$$\widehat{H}(Y) = -\frac{1}{W_m} \sum_{i=1}^m w_i U_i(\hat{\theta}_{MCLE}; y) U_i(\hat{\theta}_{MCLE}; y)^\top,$$

dato che la seconda identità di Bartlett rimane valida per ogni singolo termine di verosimiglianza.

La stima di $J(\theta)$ presenta maggiori difficoltà. Si può calcolare uno stimatore empirico solo se sono disponibili osservazioni indipendenti o pseudo-indipendenti.

1.7.3 Verifica d'ipotesi

Anche utilizzando verosimiglianze composte si possono effettuare dei test per la verifica d'ipotesi.

Se l'interesse è su un sottovettore γ del parametro θ , si possono considerare ipotesi del tipo $H_0 : \gamma = \gamma_0$ con γ sottovettore d -dimensionale di $\theta = (\gamma, \tau)$.

Il test di Wald risulta

$$W_a = (\hat{\gamma}_{MCLE} - \gamma_0)^\top I_{\gamma, \gamma}(\hat{\theta}_{MCLE})(\hat{\gamma}_{MCLE} - \gamma_0),$$

dove $I_{\gamma, \gamma}(\hat{\theta}_{MCLE})$ è la sottomatrice di informazione di Godambe relativa a γ , calcolata nella stima di massima verosimiglianza composta. Il test si

distribuisce asintoticamente sotto H_0 come un χ_d^2 a seguito della normalità asintotica dello stimatore di massima verosimiglianza composita.

E' necessario fare attenzione quando si utilizza questo test perché non gode della proprietà di equivarianza rispetto a riparametrazioni che non alterano l'interesse.

Talvolta è più ragionevole e opportuno utilizzare il test basato sul log-rapporto di verosimiglianza composita

$$W = 2[cl(\hat{\theta}_{MCLE}; y) - cl(\hat{\theta}_0; y)],$$

dove $\hat{\theta}_0 = (\gamma_0, \hat{\tau}_{MCLE}(\gamma_0))$ con $\hat{\tau}_{MCLE}(\gamma_0)$ stimatore di massima verosimiglianza composita di τ con γ noto e pari a γ_0 .

La distribuzione asintotica del test è una combinazione lineare di χ_1^2 indipendenti, ovvero

$$W \sim \sum_{i=1}^p \lambda_i Z_i^2,$$

con Z_i^2 variabili casuali indipendenti χ_1^2 e λ_i autovalori di $H_{\gamma,\gamma}(\theta_0)^{-1} J_{\gamma,\gamma}(\theta_0)$ dove $H_{\gamma,\gamma}$ e $J_{\gamma,\gamma}$ sono, rispettivamente, le sottomatrici di H e J relative a γ calcolate sotto l'ipotesi nulla.

Geys, Molenberghs e Ryan (1999) hanno proposto una versione del test aggiustata $\tilde{W} = W/\bar{\lambda}$, dove $\bar{\lambda}$ è la media degli autovalori $\lambda_1, \dots, \lambda_d$. Così facendo, si può ripristinare in modo approssimato l'usuale distribuzione asintotica χ_p^2 ; tuttavia questa semplice correzione si rivela spesso poco accurata.

Capitolo 2

Verosimiglianza composta marginale

2.1 Introduzione

In questo capitolo vengono spiegati dettagliatamente i due tipi di verosimiglianza composta marginale considerati finora: verosimiglianza a coppie e verosimiglianza d'indipendenza.

Inizialmente viene trattata la verosimiglianza composta marginale combinata, considerata da Cox e Reid (2004). Questa particolare verosimiglianza consiste nel combinare adaguatamente le distribuzioni univariate e bivariate in un'unica funzione.

Il terzo paragrafo è dedicato alla verosimiglianza d'indipendenza per fare inferenza su dati raggruppati, secondo l'approccio di Chandler e Bate (2007).

2.2 Verosimiglianza a coppie

2.2.1 Combinazione delle verosimiglianze

Per risolvere problemi di inferenza che riguardano distribuzioni in cui sono presenti dipendenze complesse tra le osservazioni, è spesso utile ricorrere a verosimiglianze composite marginali. Si costruiscono particolari verosimiglianze basate su distribuzioni marginali univariate e bivariate per trattare le interdipendenze coinvolte nella verosimiglianza completa.

Alcuni autori hanno proposto delle strategie in due passi: la stima del vettore delle medie si ottiene dalla verosimiglianza d'indipendenza, mentre la stima dei parametri di associazione viene calcolata utilizzando la verosimiglianza a coppie (Anderson e Pemberton, 1985; Anderson, 2004).

Studi di simulazione hanno evidenziato che la stima simultanea di tutti i parametri attraverso la verosimiglianza a coppie risulta più efficiente (Zhao e Joe, 2005).

Tuttavia, se si scelgono appropriati pesi per la funzione punteggio della verosimiglianza d'indipendenza, è possibile migliorare in modo sostanziale l'approccio a due livelli.

Un'interessante possibilità consiste nel combinare la verosimiglianza d'indipendenza e quella a coppie in una nuova funzione di verosimiglianza (Cox e Reid, 2004). Nel prossimo paragrafo viene illustrato quanto discusso dai due autori.

2.2.2 Verosimiglianza composita marginale combinata

Sia Y una variabile casuale di dimensione q con densità $f(y; \theta)$ e parametro θ inizialmente supposto unidimensionale. Date n osservazioni $Y^{(1)}, \dots, Y^{(n)}$ indipendenti della variabile casuale Y , si può calcolare lo stimatore di massima verosimiglianza $\hat{\theta}$ il quale, sotto le usuali condizioni di

regolarità, si distribuisce asintoticamente come una normale di media θ e varianza pari al reciproco dell'informazione di Fisher.

Spesso risulta difficile specificare la distribuzione congiunta q -dimensionale, ma è possibile specificare tutte le distribuzioni univariate e bivariate specificando le densità univariate $f(y_s; \theta)$ e bivariate $f(y_s, y_t; \theta)$ con $s \neq t$ per tutti gli $s, t = 1, \dots, q$.

Considerando il vettore Y si possono determinare i contributi del primo e secondo ordine della log-verosimiglianza

$$pl_1(\theta; Y) = \sum_{s=1}^q \log f(Y_s; \theta),$$

$$pl_2(\theta; Y) = \sum_{s>t}^q \log f(Y_s, Y_t; \theta) - aq \sum_{s=1}^q \log f(Y_s; \theta),$$

dove a è una costante che viene scelta come soluzione di un problema di ottimalità.

Ponendo $a = 0$, ovvero considerando unicamente tutte le distribuzioni bivariate, si ottiene la verosimiglianza a coppie. Porre $a = 0$ risulta adeguato se $pl_1(\theta; Y)$ è indipendente da θ , cioè se le distribuzioni marginali univariate non contengono informazioni sul parametro θ .

Con $a = 1/2$ si considerano tutte le distribuzioni condizionate di una componente data un'altra. La costante a assume un valore negativo quando, raramente, la maggior parte dell'informazione è in pl_1 e niente o poca in pl_2 .

La scelta di a permette di combinare la verosimiglianza a coppie e quella basata sulle marginali univariate.

Al fine di garantire la consistenza degli stimatori si deve prestare particolare attenzione alla costruzione dell'opportuna combinazione delle verosimiglianze composite marginali.

Dati n vettori $Y^{(i)}$ indipendenti e identicamente distribuiti, si possono

costruire le log-verosimiglianze

$$pl_1(\theta; Y^{(1)}, \dots, Y^{(n)}) = \sum_{i=1}^n pl_1(\theta; Y^{(i)}),$$

$$pl_2(\theta; Y^{(1)}, \dots, Y^{(n)}) = \sum_{i=1}^n pl_2(\theta; Y^{(i)}),$$

e le relative funzioni di punteggio

$$U_1(\theta; Y^{(1)}, \dots, Y^{(n)}) = \sum_{i=1}^n U_1(\theta; Y^{(i)})$$

essendo $U_1(\theta; Y) = \sum_{s=1}^q U(\theta; Y_s)$,

$$U_2(\theta; Y^{(1)}, \dots, Y^{(n)}) = \sum_{i=1}^n U_2(\theta; Y^{(i)})$$

con $U_2(\theta; Y) = \sum_{s>t}^q U(\theta; Y_s, Y_t) - aq \sum_{s=1}^q U(\theta; Y_s)$.

Le equazioni di stima $U_v(\hat{\theta}_{MCLE}; Y^{(1)}, \dots, Y^{(n)}) = 0$ per $v = 1, 2$ risultano non distorte sotto le usuali condizioni di regolarità. Lo stimatore così ottenuto si distribuisce approssimativamente come una normale con media θ e varianza

$$[\mathbb{E}_\theta(-U'_v)]^{-2} \mathbb{E}_\theta(U_v^2),$$

dove $\mathbb{E}_\theta(-U'_v)$ può essere stimato da $\frac{-l''_v(\theta)}{n}$

e $\mathbb{E}_\theta(U_v^2)$ può essere stimato da $\frac{\sum_{i=1}^n U_v^2(\hat{\theta}_{MCLE}; Y^{(i)})}{n}$.

Finora, si è considerata la situazione in cui il vettore Y ha dimensione q fissata e la numerosità n aumenta.

Un caso particolare di interesse si presenta se la numerosità n del campione resta fissata ma cresce la dimensione q del vettore Y . Questo problema sorge quando si devono analizzare situazioni in cui si dispone di un campione di piccola dimensione. Ad esempio, quando si può osservare soltanto una replica singola di un processo o fenomeno dove sono presenti dipendenze complesse tra le varie componenti del vettore Y , come a volte si riscontra nelle serie temporali o nei dati spaziali.

2.3 Verosimiglianza d'indipendenza

In questo paragrafo viene trattato quanto studiato da Chandler e Bate (2007) relativamente all'utilizzo della verosimiglianza composta d'indipendenza per fare inferenza su dati raggruppati.

Come si è detto, vi sono numerose applicazioni in cui si riscontrano strutture di dati dipendenti per le quali è utile considerare la verosimiglianza composta marginale.

Un esempio può essere quello di dati longitudinali in cui la dipendenza deriva da misurazioni ripetute fatte sulle stesse unità statistiche nel tempo. Ogni individuo costituisce un raggruppamento di misure che sono reciprocamente dipendenti, ma indipendenti da quelle dei restanti gruppi. Si parla infatti di dipendenza entro i gruppi.

Quando si analizzano dati raggruppati, si può modellare esplicitamente la dipendenza, utilizzando ad esempio i minimi quadrati generalizzati e le equazioni di stima generalizzate, in modo tale da sostituire nelle equazioni di stima la matrice di covarianza diagonale con una matrice di covarianza che riflette le dipendenze entro i gruppi. In alcune situazioni le equazioni di stima generalizzate producono stimatori distorti che possono mancare di robustezza e le funzioni di stima possono soffrire di non identificabilità dei parametri.

Un altro approccio consiste nell'assumere come ipotesi di lavoro l'indipendenza. Viene poi calcolata la matrice di informazione di Godambe per tenere conto della dipendenza. Lo svantaggio di questo metodo risiede nella perdita di efficienza, compensata però dall'aumento della robustezza e da notevoli vantaggi computazionali riscontrabili in grandi e complessi insiemi di dati.

2.3.1 Le equazioni di stima d'indipendenza

Si considerino n gruppi indipendenti di osservazioni che possono avere lunghezze diverse $y_s^{(i)}, i = 1, \dots, n$ e $s = 1, \dots, q_i$ dove q_i sono le osser-

vazioni presenti nell' i -esimo gruppo. Si assume che le osservazioni siano tratte da una famiglia parametrica di distribuzioni congiunte, indicizzate dal parametro θ e con densità congiunta

$$\prod_{i=1}^n f(Y^{(i)}; \theta),$$

con $f(Y^{(i)}; \theta)$ densità congiunta dell' i -esimo gruppo. Conoscendo la distribuzione congiunta, l'inferenza si può basare sulla funzione di verosimiglianza completa $l(\theta; y) = \sum_{i=1}^n \log f(y^{(i)}; \theta)$. La funzione di punteggio $l_*(\theta; y)$ si può scrivere come somma dei contributi provenienti da ogni gruppo

$$l_*(\theta; y) = \sum_{i=1}^n l_i(\theta; y).$$

Sotto condizioni generali si può mostrare che $\mathbb{E}_\theta(l_i) = 0$ per $i = 1, \dots, n$ e $Cov_\theta(l_i(\theta; Y), l_k(\theta; Y)) = 0$ per $i \neq k$.

Poiché in genere la formulazione della distribuzione congiunta è difficile, si può utilizzare una strategia basata sulla verosimiglianza composita marginale ottenuta pesando opportunamente le verosimiglianze a coppie di ogni gruppo

$$pl(\theta; y) = \sum_{i=1}^n w_i pl_i(\theta; y) = \sum_{i=1}^n w_i \sum_{s>t}^{q_i} \log f(y_s^{(i)}, y_t^{(i)}; \theta),$$

dove w_i sono i pesi specifici per ogni singolo raggruppamento.

Chandler e Bate (2007) propongono l'utilizzo della verosimiglianza d'indipendenza quando la formulazione della distribuzione congiunta risulta complessa. Si procede come se le osservazioni dentro i vari gruppi fossero indipendenti. Il parametro θ viene stimato massimizzando la *funzione di log-verosimiglianza d'indipendenza*

$$l_I(\theta) = \sum_{i=1}^n \sum_{s=1}^{q_i} \log f(y_s^{(i)}; \theta).$$

Si può focalizzare l'attenzione sullo stimatore risultante $\hat{\theta}_I$, assumendo per il momento che sia l'unica radice dell'equazione di stima d'indipendenza

$$U(\theta) = \frac{\partial l_I}{\partial \theta} = \sum_{i=1}^n U_i(\theta; Y^{(i)}),$$

con $U_i(\theta; Y^{(i)}) = \sum_{s=1}^{q_i} U(\theta; Y_s^{(i)})$. Le proprietà asintotiche dello stimatore d'indipendenza $\hat{\theta}_I$ si basano sullo sviluppo al primo ordine di $U(\theta)$ o sull'approssimazione quadratica di l_I . Quando $n \rightarrow \infty$, lo stimatore $\hat{\theta}_I$ è asintoticamente normale

$$\hat{\theta}_I \sim N(\theta, G(\theta)^{-1}),$$

dove $G(\theta) = H(\theta)J(\theta)^{-1}H(\theta)$ con $H(\theta)$ Hessiano di l_I e $J(\theta)$ matrice di covarianza di $U(\theta)$. Poiché i contributi alla funzione di punteggio provenienti dai vari gruppi sono incorrelati si ha

$$J(\theta) = \sum_{i=1}^n \text{Var}[U_i(\theta)] = \sum_{i=1}^n \mathbb{E}_\theta[U_i(\theta)U_i(\theta)^\top],$$

quindi la stima di J risulta $\hat{J} = \sum_{i=1}^n U_i(\hat{\theta}_I)U_i(\hat{\theta}_I)^\top$.

Sotto le usuali condizioni di regolarità e sotto l'assunzione che $\lim_{n \rightarrow +\infty} \frac{J}{g}$ sia finito e diverso da zero, la matrice $H^{-1}JH^{-1}$ può essere stimata in modo consistente dallo stimatore robusto

$$R = \widehat{H}^{-1} \widehat{J} \widehat{H}^{-1}.$$

Al contrario, lo stimatore della matrice di covarianza $-\widehat{H}^{-1}$ risulta non adeguato perché non considera la dipendenza entro i gruppi.

Capitolo 3

Applicazioni

3.1 Introduzione

Il presente capitolo ha l'obiettivo di sostenere le giustificazioni teoriche che portano all'utilizzo della verosimiglianza composita marginale, discusse nei capitoli precedenti. Si cercherà di adattare la teoria esposta ad alcuni esempi, dimostrando che sotto adeguate condizioni di regolarità le procedure inferenziali associate alla verosimiglianza composita marginale hanno proprietà simili a quelle dei metodi della verosimiglianza completa, sebbene ci si aspetti una perdita di efficienza.

L'obiettivo principale è confrontare per ogni esempio i risultati ottenuti considerando la verosimiglianza completa e quelli ottenuti utilizzando la verosimiglianza composita.

3.2 Verosimiglianza completa e a coppie

Il seguente esempio, già trattato da Cox e Reid (2004), riguarda il confronto tra la varianza dello stimatore di massima verosimiglianza completa e quella dello stimatore di massima verosimiglianza a coppie. Nell'esempio

la distribuzione congiunta delle n osservazioni risulta $Y_i \sim N_q(0, R)$, con

$$R = (1 - \rho) \left(I_q + \frac{\rho}{1 - \rho} 1_q 1_q^\top \right).$$

Ciò equivale ad assumere che l' i -esima osservazione sia realizzazione di una variabile casuale Y_i distribuita come una normale standard con $\text{Corr}_\rho(Y_{ir}, Y_{is}) = \rho$.

Nel caso di una singola osservazione la funzione di densità della distribuzione normale risulta

$$f_{Y_i}(y_i; \rho) = \frac{1}{(2\pi)^{q/2} |R|^{1/2}} \exp \left\{ -\frac{1}{2} y_i^\top R^{-1} y_i \right\},$$

dove $y_i = (y_{i1}, \dots, y_{iq})^\top$, $1_q = (1, \dots, 1)^\top$ e I_q matrice identità di ordine q . Per la matrice R valgono le seguenti relazioni

$$\begin{aligned} R^{-1} &= \frac{1}{1 - \rho} \left(I_q - \frac{\rho}{1 + \rho(q - 1)} 1_q 1_q^\top \right), \\ |R| &= (1 - \rho)^{q-1} (1 + \rho(q - 1)). \end{aligned}$$

Sviluppando l'espressione $y_i^\top R^{-1} y_i$ nella funzione di densità, si ottiene

$$\begin{aligned} y_i^\top R^{-1} y_i &= \frac{1}{1 - \rho} y_i^\top \left(I_q - \frac{\rho}{1 + \rho(q - 1)} 1_q 1_q^\top \right) y_i \\ &= \frac{1}{1 - \rho} \left\{ y_i^\top y_i - \frac{\rho}{1 + \rho(q - 1)} y_i^\top 1_q 1_q^\top y_i \right\}. \end{aligned}$$

Poiché $y_i^\top y_i = \sum_{r=1}^q y_{ir}^2$ e $1_q^\top y_i = \sum_{r=1}^q y_{ir} = q\bar{y}_i$ si può riscrivere la funzione di densità in questo modo

$$\begin{aligned} f_{Y_i}(y_i; \rho) &= \frac{1}{(2\pi)^{\frac{q}{2}} (1 - \rho)^{\frac{q-1}{2}} (1 + \rho(q - 1))^{\frac{1}{2}}} \\ &\quad \exp \left\{ -\frac{1}{2(1 - \rho)} \left[\sum_{r=1}^q y_{ir}^2 - \frac{\rho q^2}{1 + \rho(q - 1)} \bar{y}_i^2 \right] \right\}. \end{aligned}$$

La verosimiglianza completa calcolata per n osservazioni risulta dunque

$$\begin{aligned} L(\rho) &= (2\pi)^{-\frac{nq}{2}} (1 - \rho)^{-\frac{n(q-1)}{2}} (1 + \rho(q - 1))^{-\frac{n}{2}} \\ &\quad \exp \left\{ -\frac{1}{2(1 - \rho)} \sum_{i=1}^n \left[\sum_{r=1}^q y_{ir}^2 - \frac{\rho q^2}{1 + \rho(q - 1)} \bar{y}_i^2 \right] \right\}. \end{aligned}$$

La log-verosimiglianza completa è

$$l(\rho) = -\frac{n(q-1)}{2} \log(1-\rho) - \frac{n}{2} \log(1+\rho(q-1)) - \frac{1}{2(1-\rho)} \left\{ \sum_{i=1}^n \sum_{r=1}^q y_{ir}^2 - \frac{\rho q^2}{1+\rho(q-1)} \sum_{i=1}^n \bar{y}_i^2 \right\}.$$

L'espressione $\sum_{i=1}^n \sum_{r=1}^q y_{ir}^2$ può essere così sviluppata

$$\begin{aligned} \sum_{i=1}^n \sum_{r=1}^q y_{ir}^2 &= \sum_{i=1}^n \left(\sum_{r=1}^q y_{ir}^2 \pm q \bar{y}_i^2 \right) \\ &= \sum_{i=1}^n \left(\sum_{r=1}^q (y_{ir}^2 - \bar{y}_i^2) + q \bar{y}_i^2 \right) \\ &= ss_W + \sum_{i=1}^n q \left(\frac{1}{q} \sum_{r=1}^q y_{ir}^2 \right)^2 \\ &= ss_W + \frac{1}{q} ss_B, \end{aligned}$$

con $ss_B = \sum_{i=1}^n y_i^2$ e $y_i = \sum_{r=1}^q y_{ir}$. Pertanto

$$l(\rho) = -\frac{n(q-1)}{2} \log(1-\rho) - \frac{n}{2} \log(1+\rho(q-1)) - \frac{1}{2(1-\rho)} \left\{ ss_W + \frac{1}{q} - \frac{\rho}{1+\rho(q-1)} ss_B \right\}.$$

Sviluppando i calcoli, si ottiene la seguente espressione per la log-verosimiglianza completa

$$l(\rho) = -\frac{n(q-1)}{2} \log(1-\rho) - \frac{n}{2} \log(1+\rho(q-1)) - \frac{1}{2(1-\rho)} ss_W - \frac{1}{2q[1+\rho(q-1)]} ss_B.$$

La funzione di punteggio risulta

$$l_* = \frac{\partial l}{\partial \rho} = \frac{n(q-1)}{2} \frac{1}{(1-\rho)} - \frac{n(q-1)}{2} \frac{1}{(1+\rho(q-1))} - \frac{1}{2} \frac{1}{(1-\rho)^2} ss_W + \frac{q-1}{2q} \frac{1}{[1+\rho(q-1)]^2} ss_B.$$

L'informazione osservata vale

$$j(\rho) = -\frac{\partial^2 l}{\partial \rho^2} = -\frac{n(q-1)}{2} \frac{1}{(1-\rho)^2} - \frac{n(q-1)^2}{(1+\rho(q-1))^2} + \frac{1}{(1-\rho)^3} ss_W + \frac{(q-1)^2}{q} \frac{1}{[1+\rho(q-1)]^3} ss_B.$$

Per il calcolo dell'informazione attesa è necessario conoscere la distribuzione esatta di ss_W e ss_B . L'espressione ss_W si può riscrivere

$$\begin{aligned} ss_W &= \sum_{i=1}^n \sum_{r=1}^q (Y_{ir} - \bar{Y}_i)^2 \\ &= \sum_{i=1}^n Q_i, \end{aligned}$$

con

$$Q_i = \sum_{r=1}^q (Y_{ir} - \bar{Y}_i)^2 = Y_i^\top \left(I_q - \frac{1}{q} \mathbf{1}_q \mathbf{1}_q^\top \right) Y_i.$$

Poiché la variabile casuale Y_i si distribuisce come $N_q(0, R)$, allora

$$Z_i = (1-\rho)^{-1/2} Y_i \sim N_q(0, \Sigma),$$

con $\Sigma = I_q - \frac{\rho}{1-\rho} \mathbf{1}_q \mathbf{1}_q^\top$. Si ricava $Q_i = (1-\rho) Z_i^\top (I_q - q^{-1} \mathbf{1}_q \mathbf{1}_q^\top) Z_i$ e quindi

$$\frac{Q_i}{1-\rho} \sim \chi_m^2$$

se $\Sigma(I_q - q^{-1} \mathbf{1}_q \mathbf{1}_q^\top)$ è idempotente e $m = \text{tr}(\Sigma(I_q - q^{-1} \mathbf{1}_q \mathbf{1}_q^\top))$. Poiché $m = \text{tr}(\Sigma(I_q - q^{-1} \mathbf{1}_q \mathbf{1}_q^\top)) = \text{tr}((I_q - q^{-1} \mathbf{1}_q \mathbf{1}_q^\top)) = q-1$, si deduce che

$$\frac{Q_i}{1-\rho} \sim \chi_{q-1}^2,$$

quindi la distribuzione di ss_W risulta

$$ss_W = \sum_{i=1}^n Q_i \sim (1-\rho) \chi_{n(q-1)}^2.$$

Si vuole determinare anche la distribuzione di $ss_B = \sum_{i=1}^n Y_i^2$, dove

$$Y_i = \sum_{r=1}^q Y_{ir} = \mathbf{1}_q^\top Y_i \sim N(0, \mathbf{1}_q^\top R \mathbf{1}_q).$$

Poiché $1_q^\top R 1_q = q(1 - \rho + q\rho)$, si ricava

$$ss_B = \sum_{i=1}^n q(1 - \rho + q\rho) Z_i^2 \sim q(1 - \rho + q\rho) \chi_n^2.$$

L'informazione di Fisher risulta

$$\begin{aligned} i(\rho) = \mathbb{E}_\rho \left(- \frac{\partial^2 l}{\partial \rho^2} \right) &= - \frac{n(q-1)}{2} \frac{1}{(1-\rho)^2} - \frac{n(q-1)}{2} \frac{1}{(1+\rho(q-1))} \\ &+ \frac{1}{(1-\rho)^3} (1-\rho)n(q-1) + \frac{(q-1)^2}{q} \frac{nq}{(1+\rho(q-1))^2}, \end{aligned}$$

dato che $ss_W \sim (1-\rho)\chi_{n(q-1)}^2$ e $ss_B \sim q(1+\rho(q-1))\chi_n^2$.

Semplificando si ottiene

$$i(\rho) = \frac{nq(q-1)}{2} \frac{(1+\rho^2(q-1))}{(1-\rho)^2(1+\rho(q-1))^2}.$$

Date n osservazioni, la funzione di log-verosimiglianza a coppie ottenuta considerando la distribuzione normale bivariata risulta

$$\begin{aligned} pl_2(\rho) &= - \frac{nq(q-1)}{4} \log(1-\rho^2) - \frac{1}{2(1-\rho^2)} \\ &\sum_{i=1}^n \sum_{r=1}^q \sum_{s=r+1}^q (y_{ir}^2 + y_{is}^2 - 2\rho y_{ir} y_{is}). \end{aligned}$$

Si può riscrivere

$$\begin{aligned} \sum_{r=1}^q \sum_{s=r+1}^q (y_{ir}^2 + y_{is}^2 - 2\rho y_{ir} y_{is}) &= \sum_{r=1}^q (q-r) y_{ir}^2 + \sum_{r=1}^q (r-1) y_{ir}^2 \\ &- \rho \left(\sum_{r=1}^q y_{ir} \right)^2 + \rho \sum_{r=1}^q y_{ir}^2, \end{aligned}$$

e, sostituendo nella funzione di log-verosimiglianza, si ha

$$\begin{aligned} pl_2(\rho) &= - \frac{nq(q-1)}{4} \log(1-\rho^2) - \frac{1}{2(1-\rho^2)} \\ &\left[(q-1+\rho) \sum_{i=1}^n \sum_{r=1}^q y_{ir}^2 - \rho \sum_{i=1}^n (y_{ir})^2 \right]. \end{aligned}$$

La log-verosimiglianza a coppie risulta

$$\begin{aligned}
 pl_2(\rho) &= -\frac{nq(q-1)}{4} \log(1-\rho^2) - \frac{1}{2(1-\rho^2)} \left[(q-1+\rho) \right. \\
 &\quad \left. \left(ss_W + \frac{1}{q} ss_B \right) - \rho ss_B \right] \\
 &= -\frac{nq(q-1)}{4} \log(1-\rho^2) - \frac{1}{2(1-\rho^2)} \left[(q-1+\rho) ss_W \right. \\
 &\quad \left. + \frac{(q-1)(1-\rho)}{q} ss_B \right] \\
 &= -\frac{nq(q-1)}{4} \log(1-\rho^2) - \frac{q-1+\rho}{2(1-\rho^2)} ss_W - \frac{(q-1)(1-\rho)}{2q(1-\rho^2)} ss_B.
 \end{aligned}$$

La funzione di punteggio associata è

$$U_2(\rho; Y^{(1)}, \dots, Y^{(n)}) = -\frac{nq(q-1)\rho}{2(1-\rho^2)} - \frac{1+\rho^2+2(q-1)\rho}{2(1-\rho^2)^2} ss_W + \frac{(q-1)}{2q(1+\rho)^2} ss_B,$$

e la sua derivata rispetto al parametro ρ risulta

$$\begin{aligned}
 \nabla U_2(\rho) &= -\frac{nq(q-1)}{2} \frac{1+\rho^2}{(1-\rho^2)^2} - \frac{1}{(1-\rho^2)^3} \left[(\rho+q-1)(1-\rho^2) \right. \\
 &\quad \left. + 2\rho(1+\rho^2+2(q-1)\rho) \right] ss_W + \frac{(q-1)}{q} \frac{-1}{(1+\rho)^3} ss_B \\
 &= \frac{1}{(1-\rho^2)^3} \left[\frac{nq(q-1)}{2} (1+\rho^2)(1-\rho^2) \right. \\
 &\quad \left. - \left[(q+\rho-1)(1-\rho^2) + 2\rho(1+\rho^2+2(q-1)\rho) \right] ss_W \right. \\
 &\quad \left. - \frac{(q-1)}{q} (1-\rho)^3 ss_B \right].
 \end{aligned}$$

Per calcolare la varianza dello stimatore di massima verosimiglianza composita, occorre conoscere le seguenti quantità che permettono di ottenere l'informazione di Godambe

$$\begin{aligned}
 H(\rho) &= \mathbb{E}_\rho(-\nabla U_2) = \frac{1}{(1-\rho^2)^3} \left[-\frac{nq(q-1)}{2} (1-\rho^4) + \left[(q+\rho-1)(1-\rho^2) \right. \right. \\
 &\quad \left. \left. + 2\rho(1+\rho^2+2(q-1)\rho) \right] (1-\rho)n(q-1) \right. \\
 &\quad \left. + \frac{(q-1)}{q} (1-\rho)^3 q(1+\rho(q-1))n \right]
 \end{aligned}$$

$$\begin{aligned}
&= \frac{nq(q-1)}{2q} \frac{1-\rho}{(1-\rho^2)^3} \left[-q(1+\rho^2)(1+\rho) + 2 \left[(q+\rho-1)(1-\rho^2) \right. \right. \\
&\quad \left. \left. + 2\rho(1+\rho^2+2(q-1)\rho) \right] + 2(1-\rho^2)(1+\rho(q-1)) \right] \\
&= \frac{nq(q-1)}{2} \frac{1+\rho^2}{(1-\rho^2)^2}, \\
J(\rho) &= \text{Var}_\rho(U_2) = \left(\frac{1+\rho^2+2(q-1)\rho}{2(1-\rho^2)^2} \right)^2 (1-\rho)^2 2n(q-1) \\
&\quad + \left(\frac{q-1}{2q(1+\rho)^2} \right)^2 (q(1+q\rho-\rho))^2 2n \\
&= \frac{n(q-1)}{2(1+\rho)^4(1-\rho)^2} \left[(1+\rho^2+2(q-1)\rho)^2 \right. \\
&\quad \left. + (q-1)(1+\rho(q-1))^2 \right] \\
&= \frac{nq(q-1)}{2(1+\rho)^4(1-\rho)^2} \left[\rho^4(q^2-3q+3) + \rho^3(-2q^2+8q-6) \right. \\
&\quad \left. + \rho^2(q^2-3q+4) + \rho(2q-2) + 1 \right].
\end{aligned}$$

L'informazione di Godambe risulta

$$G(\rho) = \frac{H(\rho)^2}{J(\rho)} = \frac{nq(q-1)}{2} \frac{(1-\rho)^2 c(q, \rho)}{(1+\rho^2)^2},$$

dove $c(q, \rho) = \rho^4(q^2-3q+3) + \rho^3(-2q^2+8q-6) + \rho^2(q^2-3q+4) + \rho(2q-2) + 1$.

La varianza asintotica dello stimatore di massima verosimiglianza composita è data dal reciproco dell'informazione di Godambe, ovvero

$$G(\rho)^{-1} = \frac{2}{nq(q-1)} \frac{(1+\rho^2)^2}{(1-\rho)^2 c(q, \rho)},$$

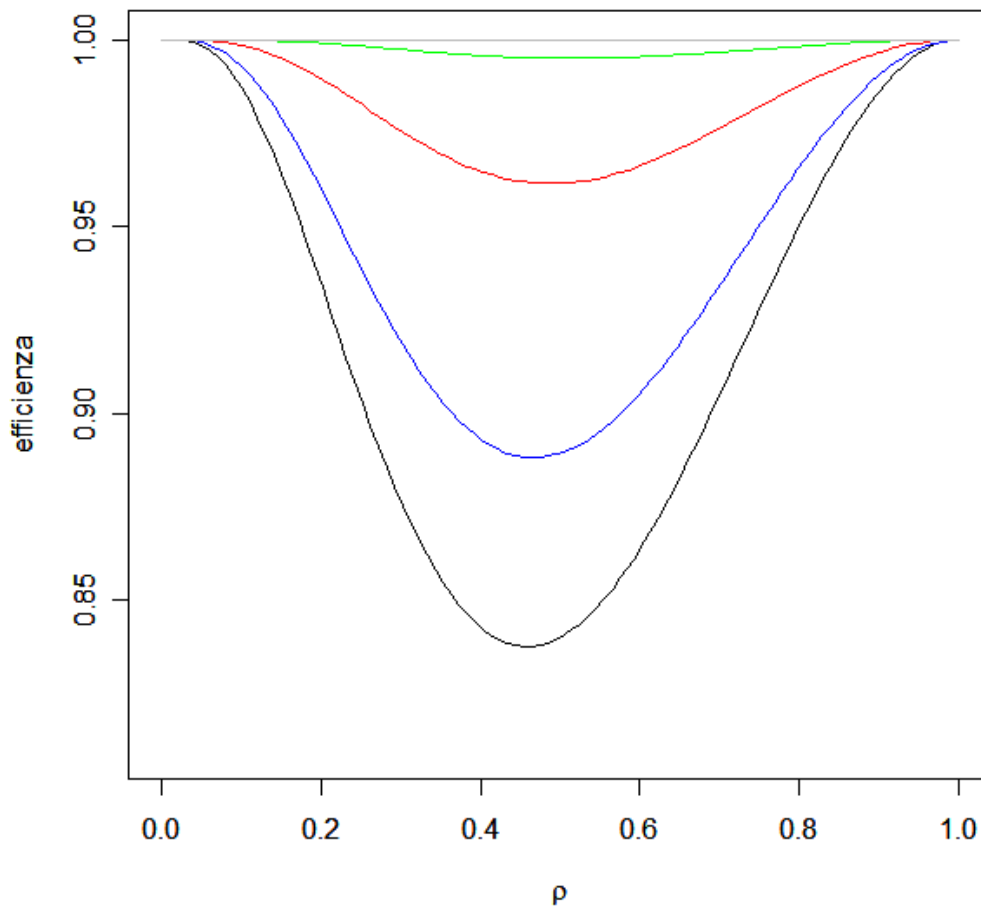
mentre la varianza asintotica dello stimatore ottenuto considerando la verosimiglianza completa è

$$i(\rho)^{-1} = \frac{2}{nq(q-1)} \frac{(1-\rho)^2 (1+\rho(q-1))^2}{(1+\rho^2(q-1))}.$$

Le due varianze possono essere confrontate per ottenere l'efficienza dello stimatore di massima verosimiglianza composita

$$\frac{i(\rho)^{-1}}{G(\rho)^{-1}} = \frac{(1+\rho^2)^2 (1+\rho(q-1))^2}{c(q, \rho) (1+\rho^2(q-1))}.$$

Quando $q = 2$ il rapporto è ovviamente pari a 1, altrettanto quando la correlazione ρ è uguale a 0 oppure a 1, per ogni valore di q . All'aumentare di q si ha una perdita di efficienza.



3.3 Verosimiglianza completa e d'indipendenza

Esempio 3.1

Si consideri un campione casuale semplice di numerosità n proveniente da una distribuzione normale bivariata. Precisa-

mente, sia (y_{i1}, y_{i2}) realizzazione di una variabile casuale

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix} \right),$$

con parametro d'interesse μ e con parametro $\rho_0 = Cov_\mu(Y_{i1}, Y_{i2})$ assunto noto.

La funzione di densità di una normale bivariata è

$$f_{Y_1, Y_2}(y_1, y_2; \mu) = \frac{1}{2\pi\sqrt{1-\rho_0^2}} \exp \left\{ -\frac{1}{2(1-\rho_0^2)} [(y_1 - \mu)^2 + (y_2 - \mu)^2 - 2\rho_0(y_1 - \mu)(y_2 - \mu)] \right\}.$$

Per una singola osservazione, la funzione di log-verosimiglianza è

$$\begin{aligned} l(\mu) &= \log f_{Y_1, Y_2}(y_1, y_2; \mu) \\ &= -\frac{1}{2(1-\rho_0^2)} [y_1^2 + y_2^2 - 2\mu y_1 - 2\mu y_2 + 2\mu^2 \\ &\quad - 2\rho_0 y_1 y_2 + 2\rho_0 \mu y_1 + 2\rho_0 \mu y_2 - 2\rho_0 \mu^2]. \end{aligned}$$

Per n osservazioni, la log-verosimiglianza completa risulta

$$\begin{aligned} l(\mu; Y^{(1)}, \dots, Y^{(n)}) &= -\frac{1}{2(1-\rho_0^2)} \sum_{i=1}^n [(y_{i1}^2 + y_{i2}^2 - 2\mu y_{i1} \\ &\quad - 2\mu y_{i2} - 2\rho_0 y_{i1} y_{i2} + 2\mu \rho_0 y_{i1} + 2\mu \rho_0 y_{i2}) \\ &\quad + 2\mu^2 - 2\rho_0 \mu^2], \end{aligned}$$

la cui funzione di punteggio è

$$\begin{aligned} l_*(\mu; Y^{(1)}, \dots, Y^{(n)}) &= -\frac{1}{(1-\rho_0^2)} \left[2n\mu - 2n\mu\rho_0 - \sum_{i=1}^n y_{i1} \right. \\ &\quad \left. - \sum_{i=1}^n y_{i2} + \rho_0 \sum_{i=1}^n y_{i1} + \rho_0 \sum_{i=1}^n y_{i2} \right]. \end{aligned}$$

L'informazione osservata $j(\mu)$ è ottenuta come derivata parziale seconda di $l(\mu)$ cambiata di segno, ovvero

$$j(\mu) = \frac{1}{(1-\rho_0^2)} (2n - 2n\rho_0) = \frac{2n}{(1+\rho_0)}.$$

L'informazione di Fisher è il valore atteso, sotto μ , dell'opposto della derivata della funzione di punteggio. In questo caso, si ha

$$i(\mu) = \mathbb{E}_\mu(-\nabla l_*) = \frac{2n}{(1 + \rho_0)}.$$

Si può verificare che l'informazione di Fisher coincide con l'informazione di Godambe $G(\mu) = \frac{H(\mu)^2}{J(\mu)}$, perché si è sotto il vero modello. Infatti

$$\begin{aligned} J(\mu) &= \text{Var}_\mu(l_*) = \frac{1}{(1 - \rho_0^2)^2} \left[(1 - \rho_0^2)^2 \text{Var}_\mu(\sum (y_{i1} + y_{i2})) \right] \\ &= \frac{1}{(1 + \rho_0)^2} \text{Var}_\mu(\sum y_{i1} + \sum y_{i2}) \\ &= \frac{1}{(1 + \rho_0)^2} \left[\text{Var}_\mu(\sum y_{i1}) + \text{Var}_\mu(\sum y_{i2}) + 2\text{Cov}_\mu(\sum y_{i1}, \sum y_{i2}) \right] \\ &= \frac{1}{(1 + \rho_0)^2} (n + n + 2n\rho_0) \\ &= \frac{2n}{(1 + \rho_0)}, \\ G(\mu) &= \frac{H(\mu)^2}{J(\mu)} = \frac{4n^2}{(1 + \rho_0)^2} \frac{1 + \rho_0}{2n} \\ &= \frac{2n}{1 + \rho_0}. \end{aligned}$$

La stima di massima verosimiglianza si ottiene uguagliando a 0 la funzione di punteggio e risolvendo rispetto al parametro μ , ricavando

$$\begin{aligned} \hat{\mu} &= \frac{\sum y_{i1} + \sum y_{i2} - \rho_0(\sum y_{i1} + \sum y_{i2})}{2n(1 - \rho_0)} \\ &= \frac{\sum y_{i1} + \sum y_{i2}}{2n}. \end{aligned}$$

Si può osservare come la stima di massima verosimiglianza $\hat{\mu}$ non dipenda da ρ_0 . Assumendo come ipotesi di lavoro l'indipendenza, si può calcolare la verosimiglianza d'indipendenza per la normale bivariata. Le distribuzioni marginali

sono

$$\begin{aligned} f_{Y_1}(y_1; \mu) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_1 - \mu)^2\right), \\ f_{Y_2}(y_2; \mu) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_2 - \mu)^2\right). \end{aligned}$$

La log-verosimiglianza d'indipendenza per una singola coppia di osservazioni risulta

$$pl_1 = -\frac{1}{2}(y_1 - \mu)^2 - \frac{1}{2}(y_2 - \mu)^2 = -\frac{1}{2}(y_1^2 + y_2^2) + \mu(y_1 + y_2) - \mu^2.$$

Per n osservazioni le quantità di verosimiglianza sono

$$\begin{aligned} pl_1(\mu; Y^{(1)}, \dots, Y^{(n)}) &= \sum_{i=1}^n \left[-\frac{1}{2}(y_{i1}^2 + y_{i2}^2) + \mu(y_{i1} + y_{i2}) - \mu^2 \right] \\ &= -\frac{1}{2} \sum y_{i1}^2 - \frac{1}{2} \sum y_{i2}^2 \\ &\quad + \mu \sum y_{i1} - n\mu^2, \\ U_1 &= \frac{\partial l_1}{\partial \mu} = \sum y_{i1} + \sum y_{i2} - 2n\mu, \\ \nabla U_1 &= \frac{\partial^2 l_1}{\partial \mu^2} = -2n. \end{aligned}$$

L'informazione di Godambe si ottiene calcolando

$$\begin{aligned} H(\mu) &= \mathbb{E}_\mu(-\nabla U_1) = 2n, \\ J(\mu) &= \text{Var}_\mu\left(\frac{\partial pl_1}{\partial \mu}\right) = \text{Var}_\mu\left(\sum y_{i1} + \sum y_{i2}\right) \\ &= \text{Var}_\mu\left(\sum y_{i1}\right) + \text{Var}_\mu\left(\sum y_{i2}\right) + 2\text{Cov}_\mu\left(\sum y_{i1}, \sum y_{i2}\right) \\ &= 2n + 2n\rho_0, \\ G(\mu) &= \frac{H(\mu)^2}{J(\mu)} = \frac{4n^2}{2n(1 + \rho_0)} = \frac{2n}{1 + \rho_0}, \end{aligned}$$

che coincide con $i(\mu)$ sotto il vero modello.

La stima di massima verosimiglianza composita marginale si ottiene risolvendo rispetto a μ la funzione U_1 uguagliata a 0

$$\tilde{\mu} = \frac{\sum y_{i1} + \sum y_{i2}}{2n}.$$

Si può notare come la stima di massima verosimiglianza ottenuta sotto l'assunzione di indipendenza risulti uguale a quella ottenuta sotto il vero modello. In questo caso lo stimatore di massima verosimiglianza composta risulta dunque perfettamente efficiente.

Esempio 3.2 Si consideri il vettore (y_{i1}, y_{i2}) realizzazione di una variabile casuale

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix},$$

avente distribuzione normale bivariata con vettore delle medie $\begin{pmatrix} \mu_1 \\ \mu_1 + \varphi \end{pmatrix}$ e matrice di covarianza $\begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix}$,

dove ρ_0 e μ_1 vengono assunti noti. L'obiettivo dell'analisi è fare inferenza sul parametro φ che rappresenta la differenza delle medie delle variabili Y_1 e Y_2 . La funzione di densità congiunta per Y_1 e Y_2 risulta

$$f_{Y_1, Y_2}(y_1, y_2; \varphi) = \frac{1}{2\pi\sqrt{1-\rho_0^2}} \exp \left\{ -\frac{1}{2(1-\rho_0^2)} \left[(y_1 - \mu_1)^2 + (y_2 - \mu_1 - \varphi)^2 - 2\rho_0(y_1 - \mu_1)(y_2 - \mu_1 - \varphi) \right] \right\}.$$

Data una singola osservazione, la log-verosimiglianza è

$$l(\varphi) = -\frac{1}{2(1-\rho_0^2)} \left[y_1^2 - 2\mu_1 y_1 + \mu_1^2 + y_2^2 + \mu_1^2 + \varphi^2 - 2\mu_1 y_2 - 2\varphi y_2 + 2\mu_1 \varphi - 2\rho_0(y_1 y_2 - \mu_1 y_1 - \varphi y_1 - \mu_1 y_2 + \mu_1^2 + \varphi \mu_1) \right].$$

Considerate n osservazioni, la log-verosimiglianza completa risulta

$$\begin{aligned} l(\varphi; Y^{(1)}, \dots, Y^{(n)}) &= -\frac{1}{2(1-\rho_0^2)} \left[2n\mu_1^2 + n\varphi^2 + 2n\mu_1\varphi - 2n\rho_0\mu_1^2 - 2n\rho_0\varphi\mu_1 + \sum y_{i1}^2 \right. \\ &\quad - 2\mu_1 \sum y_{i1} + \sum y_{i2}^2 - 2\mu_1 \sum y_{i2} - 2\varphi \sum y_{i2} - 2\rho_0 \sum y_{i1}y_{i2} \\ &\quad \left. + 2\rho_0\mu_1 \sum y_{i1} + 2\rho_0\varphi \sum y_{i1} + 2\rho_0\varphi \sum y_{i2} \right]. \end{aligned}$$

La funzione di punteggio è

$$l_*(\varphi; Y^{(1)}, \dots, Y^{(n)}) = -\frac{1}{(1-\rho_0^2)} \left(n\varphi + n\mu_1 - n\mu_1\rho_0 - \sum y_{i2} + \rho_0 \sum y_{i1} + \rho_0 \sum y_{i2} \right),$$

e l'informazione osservata risulta

$$j(\varphi) = -\nabla l_* = \frac{n}{(1-\rho_0^2)}.$$

L'informazione attesa vale

$$i(\varphi) = \mathbb{E}_\varphi(-\nabla l_*) = \frac{n}{(1-\rho_0^2)}.$$

Si può verificare che l'informazione di Fisher e quella di Godambe, calcolata sulla verosimiglianza completa, coincidono perché sotto il vero modello

$$\begin{aligned} J(\varphi) &= \text{Var}_\varphi(l_*) \\ &= \frac{1}{(1-\rho_0^2)^2} \text{Var}_\varphi\left(-\sum y_{i2} + \rho_0 \sum y_{i1}\right) \\ &= \frac{1}{(1-\rho_0^2)^2} \left[\text{Var}_\varphi\left(\sum y_{i2}\right) + \rho_0^2 \text{Var}_\varphi\left(\sum y_{i1}\right) - 2\rho_0 \text{Cov}_\varphi\left(\sum y_{i1}, \sum y_{i2}\right) \right] \\ &= \frac{1}{(1-\rho_0^2)^2} (n + n\rho_0^2 - 2n\rho_0^2) \\ &= \frac{n}{1-\rho_0^2}, \\ G(\varphi) &= \frac{H(\varphi)^2}{J(\varphi)} = \frac{n^2}{(1-\rho_0^2)^2} \frac{1-\rho_0^2}{n} = \frac{n}{1-\rho_0^2} = i(\varphi). \end{aligned}$$

La stima di massima verosimiglianza ottenuta considerando la verosimiglianza completa è

$$\hat{\varphi} = \frac{n\mu_1(\rho_0 - 1) + \sum y_{i2} - \rho_0(\sum y_{i1} + \sum y_{i2})}{n}.$$

Per calcolare la verosimiglianza d'indipendenza si devono considerare le distribuzioni marginali delle variabili casuali Y_1 e Y_2 , ovvero

$$\begin{aligned} f_{Y_1}(y_1) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_1 - \mu_1)^2\right), \\ f_{Y_2}(y_2) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_2 - \mu_1 - \varphi)^2\right), \end{aligned}$$

aventi le seguenti funzioni di log-verosimiglianza

$$\begin{aligned} \log f_{Y_1} &= -\frac{1}{2}(y_1 - \mu_1)^2, \\ \log f_{Y_2} &= -\frac{1}{2}(y_2 - \mu_1 - \varphi)^2. \end{aligned}$$

La log-verosimiglianza d'indipendenza per una coppia di osservazioni è

$$\begin{aligned} pl_1 &= -\frac{1}{2}(y_1 - \mu_1)^2 - \frac{1}{2}(y_2 - \mu_1 - \varphi)^2 \\ &= -\frac{1}{2}(y_1^2 + y_2^2) + \mu_1(y_1 + y_2) + \varphi y_2 - \mu_1\varphi - \mu_1^2 - \frac{1}{2}\varphi^2. \end{aligned}$$

Date n osservazioni, si ha

$$\begin{aligned} pl_1(\varphi; Y^{(1)}, \dots, Y^{(n)}) &= \sum_{i=1}^n \left[-\frac{1}{2}(y_{i1}^2 + y_{i2}^2) + \mu_1(y_{i1} + y_{i2}) + \varphi y_{i2} - \mu_1\varphi - \mu_1^2 - \frac{1}{2}\varphi^2 \right] \\ &= -\frac{1}{2} \sum y_{i1}^2 - \frac{1}{2} \sum y_{i2}^2 + \mu_1 \sum y_{i1} + (\mu_1 + \varphi) \sum y_{i2} \\ &\quad - n\mu_1\varphi - n\mu_1^2 - \frac{1}{2}n\varphi^2. \end{aligned}$$

Le quantità di verosimiglianza che si devono considerare sono

$$\begin{aligned} U_1 &= \frac{\partial pl_1}{\partial \varphi} \\ &= \sum_{i=1}^n y_{i2} - n\mu_1 - n\varphi, \\ \nabla U_1 &= \frac{\partial^2 pl_1}{\partial \varphi^2} \\ &= -n. \end{aligned}$$

Per ottenere l'informazione di Godambe si devono calcolare le seguenti quantità

$$\begin{aligned} H(\varphi) &= \mathbb{E}_\varphi(-\nabla U_1) = n, \\ J(\varphi) &= \text{Var}_\varphi\left(\frac{\partial l_1}{\partial \varphi}\right) = \text{Var}_\varphi\left(\sum_{i=1}^n y_{i2}\right) = n. \end{aligned}$$

Essendo il parametro φ unidimensionale, l'informazione di Godambe risulta

$$G(\varphi) = \frac{H(\varphi)^2}{J(\varphi)} = \frac{n^2}{n} = n.$$

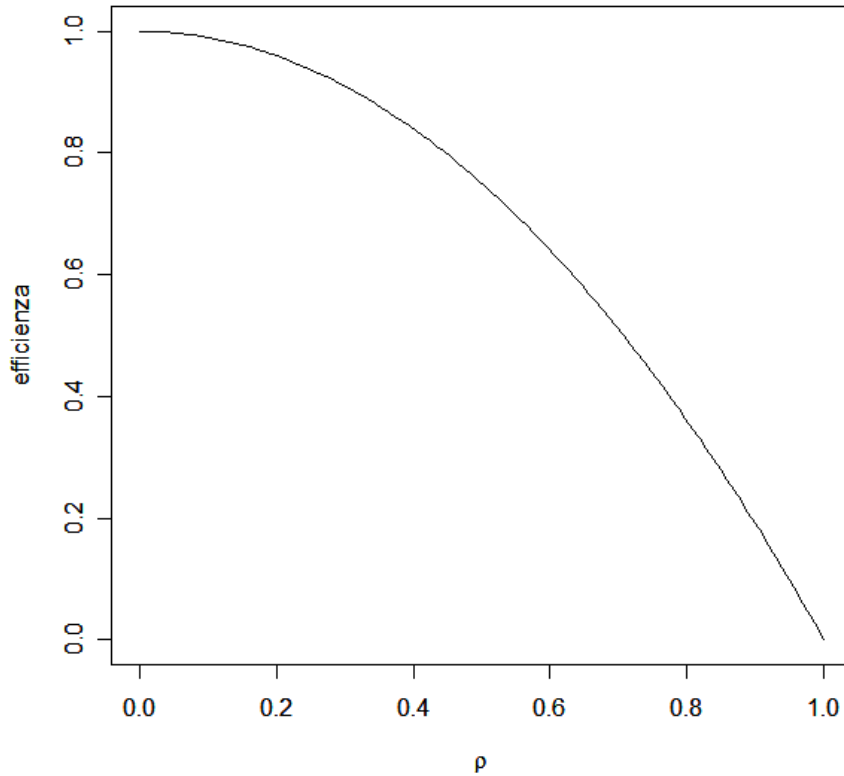
La stima di massima verosimiglianza composta marginale risulta

$$\tilde{\varphi} = \frac{\sum y_{i2} - n\mu_1}{n}.$$

La perdita di efficienza risulta

$$\frac{i(\mu)^{-1}}{G(\mu)^{-1}} = \frac{1 - \rho_0^2}{n} n = 1 - \rho_0^2.$$

Il valore ottenuto è uguale a 1 quando $\rho_0 = 0$. Questo significa che la perdita di efficienza è nulla quando le variabili Y_1 e Y_2 risultano incorrelate.



Esempio 3.3 Sia (y_{i1}, y_{i2}) realizzazione di una variabile casuale

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix},$$

avente distribuzione normale bivariata con vettore delle medie $\begin{pmatrix} \mu \\ \mu \end{pmatrix}$ e matrice di covarianza $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$,

dove il parametro $\theta = (\mu, \rho)$ è bidimensionale e μ è il parametro d'interesse. L'obiettivo consiste nel fare inferenza sul parametro μ attraverso la funzione di verosimiglianza a coppie.

Data una singola osservazione, la funzione di

log-verosimiglianza completa risulta

$$l(\mu, \rho) = -\frac{1}{2} \log(1 - \rho^2) - \frac{1}{2(1 - \rho^2)} [y_1^2 + y_2^2 - 2\mu y_1 - 2\mu y_2 + 2\mu^2 - 2\rho y_1 y_2 + 2\rho \mu y_1 + 2\rho \mu y_2 - 2\rho \mu^2].$$

Per n osservazioni, si ha

$$l(\mu, \rho; Y^{(1)}, \dots, Y^{(n)}) = -\frac{1}{2} \log(1 - \rho^2) - \frac{1}{2(1 - \rho^2)} \sum_{i=1}^n [y_{i1}^2 + y_{i2}^2 - 2\mu y_{i1} - 2\mu y_{i2} + 2\mu^2 - 2\rho y_{i1} y_{i2} + 2\rho \mu y_{i1} + 2\rho \mu y_{i2} - 2\rho \mu^2].$$

La funzione di punteggio risulta

$$\begin{aligned} \frac{\partial l}{\partial \mu} &= -\frac{1}{1 - \rho^2} [2n\mu - 2n\mu\rho - \sum y_{i1} - \sum y_{i2} + \rho \sum y_{i1} + \rho \sum y_{i2}] \\ \frac{\partial l}{\partial \rho} &= \frac{\rho}{1 - \rho^2} - \frac{1}{(1 - \rho^2)^2} [-\sum y_{i1} \sum y_{i2} + \mu \sum y_{i1} + \mu \sum y_{i2} - \rho^2 \sum y_{i1} \sum y_{i2} + \mu\rho^2 \sum y_{i1} + \mu\rho^2 \sum y_{i2} + \rho \sum y_{i1}^2 + \rho \sum y_{i2}^2 - 2\mu\rho \sum y_{i1} - 2\mu\rho \sum y_{i2} + 2n\mu^2\rho - n\mu^2 - n\mu^2\rho^2]. \end{aligned}$$

La matrice d'informazione osservata $j(\mu, \rho)$ è una matrice (2×2) con elementi

$$\begin{aligned} -\frac{\partial^2 l}{\partial \mu^2} &= \frac{4n - 4n\rho}{2(1 - \rho^2)} = \frac{2n}{1 + \rho}, \\ -\frac{\partial^2 l}{\partial \rho^2} &= -\frac{1 + \rho^2}{(1 - \rho^2)^2} + \frac{1}{(1 - \rho^2)^3} [-2\rho \sum y_{i1} \sum y_{i2} + \sum y_{i1}^2 + \sum y_{i2}^2 - 2\mu \sum y_{i1} - 2\mu \sum y_{i2} + 2\rho^3 \sum y_{i1} \sum y_{i2} + 2n\mu^2 - 2\mu\rho^3 \sum y_{i1} - 2\mu\rho^3 \sum y_{i2} - \rho^2 \sum y_{i1}^2 - \rho^2 \sum y_{i2}^2 + 2n\mu^2\rho^3 - 8\rho^2 \sum y_{i1} \sum y_{i2} + 4\rho \sum y_{i1}^2 + 4\rho \sum y_{i2}^2 - 6\mu\rho \sum y_{i1} - 6\mu\rho \sum y_{i2} + 6n\mu^2\rho - 10n\mu^2\rho^2 + 10\mu\rho^2 \sum y_{i1} + 10\mu\rho^2 \sum y_{i2}], \\ -\frac{\partial^2 l}{\partial \mu \partial \rho} &= \frac{1}{(1 - \rho^2)^2} [\sum y_{i1} + \sum y_{i2} - 2n\mu + \rho^2 \sum y_{i1} + \rho^2 \sum y_{i2} \end{aligned}$$

$$\begin{aligned}
& +4n\mu\rho - 2n\mu\rho^2 - 2\rho \sum y_{i1} - 2\rho \sum y_{i2} \\
& = -\frac{\partial^2 l}{\partial \rho \partial \mu}.
\end{aligned}$$

La matrice d'informazione attesa $i(\theta) = \mathbb{E}_\theta(-\nabla l_*)$ risulta

$$i(\theta) = \begin{pmatrix} i_{\mu\mu} & i_{\mu\rho} \\ i_{\rho\mu} & i_{\rho\rho} \end{pmatrix},$$

dove

$$\begin{aligned}
i_{\mu\mu} &= \mathbb{E}_\mu \left(-\frac{2n}{1+\rho} \right) = \frac{2n}{1+\rho}, \\
i_{\mu\rho} &= \frac{1}{(1-\rho^2)^2} [2n\mu - 2n\mu + 2n\mu\rho^2 + 4n\mu\rho - 2n\mu\rho^2 - 4n\mu\rho] \\
&= 0 = i_{\rho\mu}, \\
i_{\rho\rho} &= -\frac{1+\rho^2}{(1-\rho^2)^2} + \frac{2n}{(1-\rho^2)^3} [\rho^4 + \rho^3(n\mu^2 - \mu^2 - 4) \\
&\quad + \rho^2(4\mu^2 - 4n\mu^2 - 2) + \rho(\mu^2 - n\mu^2 + 4) + 1],
\end{aligned}$$

dato che

$$\begin{aligned}
\mathbb{E}_\theta(\sum Y_{i1}) &= n\mu, \\
\mathbb{E}_\theta(\sum Y_{i1}^2) &= \mathbb{E}_\theta(\sum Y_{i2}^2) = n(1 + \mu^2), \\
\mathbb{E}_\theta(\sum Y_{i1} \sum Y_{i2}) &= n^2\mu^2 + n\rho.
\end{aligned}$$

La stima di massima verosimiglianza per μ si ottiene ponendo

$\frac{\partial l}{\partial \mu} = 0$ e risolvendo rispetto a μ

$$\begin{aligned}
\hat{\mu} &= \frac{(\sum y_{i1} + \sum y_{i2})(1-\rho)}{2n(1-\rho)} \\
&= \frac{\sum y_{i1} + \sum y_{i2}}{2n}.
\end{aligned}$$

Si può notare come la stima ottenuta non dipenda dal parametro ρ .

La varianza asintotica dello stimatore risulta

$$i^{\mu\mu} = i(\theta)_{11}^{-1} = \frac{1}{i(\theta)_{11}} = \frac{1+\rho}{2n},$$

essendo la matrice dell'informazione di Fisher una matrice diagonale.

Assumendo come ipotesi di lavoro l'indipendenza, ci si attende che non vi sia in questo caso perdita di efficienza per l'inferenza su μ .

Le distribuzioni marginali sono

$$\begin{aligned} f_{Y_1}(y_1; \mu) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_1 - \mu)^2\right), \\ f_{Y_2}(y_2; \mu) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_2 - \mu)^2\right). \end{aligned}$$

La log-verosimiglianza d'indipendenza per una singola coppia di osservazioni risulta

$$pl_1 = -\frac{1}{2}(y_1 - \mu)^2 - \frac{1}{2}(y_2 - \mu)^2 = -\frac{1}{2}(y_1^2 + y_2^2) + \mu(y_1 + y_2) - \mu^2.$$

Per n osservazioni, le quantità di verosimiglianza sono

$$\begin{aligned} pl_1(\mu; Y^{(1)}, \dots, Y^{(n)}) &= \sum_{i=1}^n \left[-\frac{1}{2}(y_{i1}^2 + y_{i2}^2) + \mu(y_{i1} + y_{i2}) - \mu^2 \right] \\ &= -\frac{1}{2} \sum y_{i1}^2 - \frac{1}{2} \sum y_{i2}^2 + \mu \sum y_{i1} - n\mu^2, \\ U_1 &= \frac{\partial l_1}{\partial \mu} = \sum y_{i1} + \sum y_{i2} - 2n\mu, \\ \nabla U_1 &= \frac{\partial^2 l_1}{\partial \mu^2} = -2n. \end{aligned}$$

L'informazione di Godambe si ottiene calcolando

$$\begin{aligned} H(\mu) &= \mathbb{E}_\mu(-\nabla U_1) = 2n, \\ J(\mu) &= \text{Var}_\mu\left(\frac{\partial pl_1}{\partial \mu}\right) = \text{Var}_\mu\left(\sum y_{i1} + \sum y_{i2}\right) \\ &= \text{Var}_\mu\left(\sum y_{i1}\right) + \text{Var}_\mu\left(\sum y_{i2}\right) \\ &\quad + 2\text{Cov}_\mu\left(\sum y_{i1}, \sum y_{i2}\right) = 2n + 2n\rho, \\ G(\mu) &= \frac{H(\mu)^2}{J(\mu)} = \frac{4n^2}{2n(1 + \rho)} = \frac{2n}{1 + \rho}. \end{aligned}$$

La stima di massima verosimiglianza composita marginale si ottiene risolvendo rispetto a μ la funzione U_1 uguagliata a 0

$$\tilde{\mu} = \frac{\sum y_{i1} + \sum y_{i2}}{2n}.$$

Si può notare che questo valore coincide con quello ottenuto per la stima di massima verosimiglianza sotto il vero modello. Lo stimatore di massima verosimiglianza composita risulta perfettamente efficiente, essendo uguali le due stime.

3.4 Un'applicazione all'analisi di dati di sopravvivenza

Il seguente esempio, trattato da Bjarnason e Hougaard (2000), riguarda i modelli di fragilità, *frailty models*, per dati di sopravvivenza multivariati utilizzati per rappresentare il tempo trascorso per ciascuna unità dall'inizio dell'esperimento o dell'osservazione, fino al verificarsi di più eventi d'interesse.

Si considera un modello di fragilità, ovvero un modello a effetti casuali per dati di sopravvivenza, applicato a n coppie indipendenti e identicamente distribuite, per le quali si assume un modello di fragilità bivariato.

L'assunzione fondamentale del modello di fragilità è la presenza di dipendenza causata dalla fragilità, Z , che rappresenta fattori di rischio non osservabili. Condizionatamente a Z , le osservazioni di una coppia hanno tempi di vita T_1, T_2 indipendenti con tasso di guasto $Z\mu_j(t)$, dove $\mu_j(t)$ tiene conto della dipendenza da t della distribuzione di T_j , per $j = 1, 2$.

In questo esempio si assume $Z \sim \text{Gamma}(\delta, \delta)$ con funzione di densità

$$f(z; \delta) = \frac{\delta^\delta z^{\delta-1} e^{-\delta z}}{\Gamma(\delta)},$$

con $z > 0$ e $\delta > 0$. Condizionatamente a Z , la distribuzione del tempo di

vita T_j è una *Weibull* $(Z\lambda_j, \gamma)$, $j = 1, 2$, con funzione di densità

$$f(t; Z\lambda_j, \gamma) = Z\lambda_j\gamma t^{\gamma-1}e^{-z\lambda_j t^\gamma},$$

e tasso di guasto $Z\mu_j(t) = Z\lambda_j\gamma t^{\gamma-1}$.

La funzione di sopravvivenza bivariata condizionata risulta

$$P(T_1 > t_1, T_2 > t_2 \mid Z = z) = e^{-z(\lambda_1 t_1^\gamma + \lambda_2 t_2^\gamma)},$$

e quella non condizionata è

$$P(T_1 > t_1, T_2 > t_2) = \left(1 + (\lambda_1 t_1^\gamma + \lambda_2 t_2^\gamma)\right)^{-\delta}.$$

La funzione di densità congiunta per una singola coppia (T_1, T_2) è

$$f(t_1, t_2) = \eta_1\eta_2\gamma^2\delta(\delta + 1)\frac{t_1^{\gamma-1}t_2^{\gamma-1}}{(1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma)^{\delta+2}},$$

e la corrispondente funzione di log-verosimiglianza completa risulta

$$\begin{aligned} l(\delta, \eta_1, \eta_2, \gamma) &= 2\log\gamma + \log\delta + \log(\delta + 1) + \log\eta_1 + \log\eta_2 \\ &\quad + (\gamma - 1)(\log t_1 + \log t_2) - (\delta + 2)\log(1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma), \end{aligned}$$

dove si è posto $\eta_j = \lambda_j/\delta$ per $j = 1, 2$.

Le componenti della funzione di punteggio sono

$$\begin{aligned} \frac{\partial l}{\partial \delta} &= \frac{1}{\delta} + \frac{1}{\delta + 1} - \log(1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma), \\ \frac{\partial l}{\partial \eta_1} &= \frac{1}{\eta_1} - (\delta + 2)\frac{t_1^\gamma}{1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma}, \\ \frac{\partial l}{\partial \eta_2} &= \frac{1}{\eta_2} - (\delta + 2)\frac{t_2^\gamma}{1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma}, \\ \frac{\partial l}{\partial \gamma} &= \frac{2}{\gamma} + \log t_1 + \log t_2 - (\delta + 2)\frac{\eta_1 t_1^\gamma \log t_1 + \eta_2 t_2^\gamma \log t_2}{1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma}. \end{aligned}$$

Si svolgono nel seguito in dettaglio i calcoli per ottenere la matrice di informazione attesa così come riportata da Bjarnason e Hougaard (2000).

La matrice d'informazione osservata $j(\delta, \eta_1, \eta_2, \gamma)$ è formata dai seguenti elementi

$$\begin{aligned} -\frac{\partial^2 l}{\partial \delta^2} &= \frac{1}{\delta^2} + \frac{1}{(\delta + 1)^2}, \\ -\frac{\partial^2 l}{\partial \eta_1 \partial \delta} &= \frac{t_1^\gamma}{1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma}, \\ -\frac{\partial^2 l}{\partial \eta_1^2} &= \frac{1}{\eta_1^2} - (\delta + 2) \frac{t_1^{2\gamma}}{(1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma)^2}, \\ -\frac{\partial^2 l}{\partial \eta_1 \partial \eta_2} &= -(\delta + 2) \frac{t_1^\gamma t_2^\gamma}{(1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma)^2}, \\ -\frac{\partial^2 l}{\partial \gamma \partial \delta} &= \frac{\eta_1 t_1^\gamma \log t_1 + \eta_2 t_2^\gamma \log t_2}{1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma}, \\ -\frac{\partial^2 l}{\partial \eta_1 \partial \gamma} &= (\delta + 2) \left[\frac{t_1^\gamma \log t_1}{1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma} - \frac{(\eta_1 t_1^{2\gamma} \log t_1 + \eta_2 t_2^{2\gamma} \log t_2)^2}{(1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma)^2} \right], \\ -\frac{\partial l}{\partial \gamma^2} &= \frac{2}{\gamma^2} + (\delta + 2) \left[\frac{\eta_1 t_1^\gamma \log^2 t_1 + \eta_2 t_2^\gamma \log^2 t_2}{1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma} - \frac{(\eta_1 t_1^\gamma \log t_1 + \eta_2 t_2^\gamma \log t_2)^2}{(1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma)^2} \right]. \end{aligned}$$

Risulta conveniente, per il calcolo dei valori attesi, considerare la trasformazione $T_j^\gamma = U_j$ con $j = 1, 2$, la quale, condizionatamente a Z , si distribuisce come un'esponenziale di media $(Z\lambda_j)^{-1}$. Si può riscrivere la funzione di densità congiunta per (U_1, U_2)

$$f_U(u_1, u_2) = f_T(t_1(u_1, u_2), t_2(u_1, u_2)) |J|,$$

dove $|J|$ è il determinante della matrice jacobiana

$$J = \begin{pmatrix} \frac{1}{\gamma} u_1^{\frac{1}{\gamma}-1} & 0 \\ 0 & \frac{1}{\gamma} u_2^{\frac{1}{\gamma}-1} \end{pmatrix}.$$

Si ricava

$$\begin{aligned} f(u_1, u_2) &= \frac{\eta_1 \eta_2 \delta (\delta + 1) \gamma^2}{(1 + \eta_1 t_1^\gamma + \eta_2 t_2^\gamma)^{\delta+2}} \left(u_1^{1-\frac{1}{\gamma}} u_2^{1-\frac{1}{\gamma}} \right) \frac{1}{\gamma^2} u_1^{\frac{1}{\gamma}-1} u_2^{\frac{1}{\gamma}-1} \\ &= \frac{\eta_1 \eta_2 \delta (\delta + 1)}{(1 + \eta_1 u_1 + \eta_2 u_2)^{\delta+2}}. \end{aligned}$$

Per il calcolo del valore atteso è utile considerare il seguente lemma (Bjarnason e Hougaard, 2000, Lemma 1).

Se, condizionatamente a Z , U_1 e U_2 sono indipendenti e seguono la distribuzione $\text{Gamma}(1, Z\lambda_1)$ e $\text{Gamma}(1, Z\lambda_2)$, rispettivamente, allora per $q > -\delta$ e per una funzione h con $\mathbb{E}_{\delta+q, \theta} |h(U_1, U_2)| < \infty$, si ha

$$\mathbb{E}_{\delta, \theta} \left[\frac{h(U_1, U_2)}{(1 + \eta_1 U_1 + \eta_2 U_2)^q} \right] = \frac{\delta(\delta + 1)}{(\delta + q)(\delta + q + 1)} \mathbb{E}_{\delta+q, \theta} h(U_1, U_2)$$

dove $\eta_j = \lambda_j / \delta$ e $\mathbb{E}_{\delta, \theta}[\cdot]$ indica che Z è $\text{Gamma}(\delta, \theta)$.

La matrice d'informazione di Fisher è il valore atteso dell'opposto delle derivate seconde della funzione di log-verosimiglianza.

Nel seguito vengono riportati alcuni risultati utili ai fini del calcolo del valore atteso.

$$\mathbb{E}_{\theta}(U \mid Z = z) = \frac{1}{z\lambda}, \quad (3.1)$$

$$\mathbb{E}_{\theta}(U^2 \mid Z = z) = \text{Var}_{\theta}(U_1 \mid Z = z) + \mathbb{E}_{\theta}(U_1 \mid Z = z)^2 = \frac{2}{z^2\lambda^2}, \quad (3.2)$$

$$\mathbb{E}_{\theta}(U \log U \mid Z = z) = \int_0^{+\infty} u \log u z\lambda e^{-z\lambda u} du,$$

si ponga $x = z\lambda u$ e $du = \frac{dx}{z\lambda}$, ottenendo

$$\begin{aligned} \mathbb{E}_{\theta}(U \log U \mid Z = z) &= \frac{1}{z\lambda} \left(\int_0^{+\infty} x \log x e^{-x} dx \right. \\ &\quad \left. - \log(z\lambda) \int_0^{+\infty} x e^{-x} dx \right) \\ &= \frac{\psi(2) - \log \lambda}{z\lambda} - \frac{\log z}{z\lambda}, \end{aligned} \quad (3.3)$$

$$\begin{aligned} \mathbb{E}_{\theta}(U^2 \log U \mid Z = z) &= \int_0^{+\infty} u^2 \log u z\lambda e^{-z\lambda u} du \\ &= \frac{1}{z^2\lambda^2} \left(\int_0^{+\infty} x^2 \log x e^{-x} dx - \log z \int_0^{+\infty} x^2 e^{-x} dx \right. \\ &\quad \left. - \log \lambda \int_0^{+\infty} x^2 e^{-x} dx \right) \\ &= \frac{2\psi(2) + 1 - 2 \log \lambda}{z^2\lambda^2} - \frac{2 \log z}{z^2\lambda^2}, \end{aligned} \quad (3.4)$$

$$\begin{aligned}
\mathbb{E}_\theta(U \log^2 U \mid Z = z) &= \int_0^{+\infty} u \log^2 u z \lambda e^{-z\lambda u} du \\
&= \frac{1}{z\lambda} \int_0^{+\infty} x (\log x - \log z - \log \lambda)^2 e^{-x} dx \\
&= \frac{1}{z\lambda} \left(\int_0^{+\infty} x \log^2 x e^{-x} dx + \log^2 z \int_0^{+\infty} x e^{-x} dx \right. \\
&\quad \left. + \log^2 \lambda \int_0^{+\infty} x e^{-x} dx - 2 \log z \int_0^{+\infty} x \log x e^{-x} dx \right. \\
&\quad \left. - 2 \log \lambda \int_0^{+\infty} x \log x e^{-x} dx + 2 \log z \log \lambda \int_0^{+\infty} x e^{-x} dx \right) \\
&= \frac{\log^2 z}{z\lambda} + \frac{\log z}{z\lambda} (-2\psi(2) + 2 \log \lambda) \\
&\quad + \frac{1}{z\lambda} (\psi'(2) + \psi(2)^2 + \log^2 \lambda - 2\psi(2) \log \lambda), \tag{3.5} \\
\mathbb{E}_\theta(U^2 \log^2 U \mid Z = z) &= \int_0^{+\infty} u^2 \log^2 u z \lambda e^{-z\lambda u} du \\
&= \frac{1}{z^2 \lambda^2} \int_0^{+\infty} x^2 (\log x - \log z - \log \lambda)^2 e^{-x} dx \\
&= \frac{1}{z^2 \lambda^2} \left(\int_0^{+\infty} x^2 \log^2 x e^{-x} dx + \log^2 z \int_0^{+\infty} x^2 e^{-x} dx \right. \\
&\quad \left. + \log^2 \lambda \int_0^{+\infty} x^2 e^{-x} dx - 2 \log z \int_0^{+\infty} x^2 \log x e^{-x} dx \right. \\
&\quad \left. - 2 \log \lambda \int_0^{+\infty} x^2 \log x e^{-x} dx + 2 \log z \log \lambda \int_0^{+\infty} x^2 e^{-x} dx \right) \\
&= \frac{2 \log^2 z}{z^2 \lambda^2} + \frac{\log z}{z^2 \lambda^2} (-4\psi(2) - 2 + 4 \log \lambda) \\
&\quad + \frac{1}{z^2 \lambda^2} (2\psi'(2) + 2\psi(2)^2 + 2\psi(2) + 2 \log^2 \lambda \\
&\quad - 4\psi(2) \log \lambda - 2 \log \lambda). \tag{3.6}
\end{aligned}$$

Altri risultati utili per i successivi calcoli dell'informazione di Fisher sono

$$\mathbb{E}_{\delta+1,\delta}(Z^{-1}) = \delta \frac{\Gamma(\delta+1-1)}{\Gamma(\delta+1)} = 1, \tag{3.7}$$

$$\mathbb{E}_{\delta+1,\delta}(Z^{-2}) = \delta \frac{\Gamma(\delta+1-2)}{\Gamma(\delta+1)} = \frac{\delta}{\delta-1}, \tag{3.8}$$

$$\begin{aligned}
\mathbb{E}_{\delta+1,\delta}\left(\frac{\log Z}{Z}\right) &= \int_0^{+\infty} \frac{\log z}{z} \frac{\delta e^{-\delta z} (\delta z)^\delta}{\Gamma(\delta+1)} dz \\
&= \frac{\delta^{\delta+1}}{\Gamma(\delta+1)} \int_0^{+\infty} z^{\delta-1} \log z e^{-\delta z} dz,
\end{aligned}$$

si ponga $y = \delta z$ e $dz = \frac{dy}{\delta}$, ottenendo

$$\begin{aligned}\mathbb{E}_{\delta+1,\delta}\left(\frac{\log Z}{Z}\right) &= \frac{\delta}{\Gamma(\delta+1)}\left(\int_0^{+\infty} y^{\delta-1} \log y e^{-y} dy\right. \\ &\quad \left.- \log \delta \int_0^{+\infty} y^{\delta-1} e^{-y} dy\right) \\ &= \psi(\delta) - \log \delta,\end{aligned}\tag{3.9}$$

$$\begin{aligned}\mathbb{E}_{\delta+1,\delta}\left(\frac{\log^2 Z}{Z}\right) &= \int_0^{+\infty} \frac{\log^2 z}{z} \frac{\delta e^{-\delta z} (\delta z)^\delta}{\Gamma(\delta+1)} dz \\ &= \frac{\delta^{\delta+1}}{\Gamma(\delta+1)} \int_0^{+\infty} z^{\delta-1} \log^2 z e^{-\delta z} dz \\ &= \frac{\delta}{\Gamma(\delta+1)}\left(\int_0^{+\infty} y^{\delta-1} \log^2 y e^{-y} dy\right. \\ &\quad \left.- 2 \log \delta \int_0^{+\infty} y^{\delta-1} \log y e^{-y} dy + \log^2 \delta \int_0^{+\infty} y^{\delta-1} e^{-y} dy\right) \\ &= (\psi'(\delta) + \psi(\delta)^2 - 2\psi(\delta) \log \delta + \log^2 \delta),\end{aligned}\tag{3.10}$$

$$\mathbb{E}_{\delta+2,\delta}(Z^{-1}) = \delta \frac{\Gamma(\delta+2-1)}{\Gamma(\delta+2)} = \frac{\delta}{\delta+1},\tag{3.11}$$

$$\mathbb{E}_{\delta+2,\delta}(Z^{-2}) = \delta^2 \frac{\Gamma(\delta+2-2)}{\Gamma(\delta+2)} = \frac{\delta}{\delta+1},\tag{3.12}$$

$$\begin{aligned}\mathbb{E}_{\delta+2,\delta}\left(\frac{\log Z}{Z}\right) &= \int_0^{+\infty} \frac{\log z}{z} \frac{\delta e^{-\delta z} (\delta z)^{\delta+1}}{\Gamma(\delta+2)} dz \\ &= \frac{\delta^{\delta+2}}{\Gamma(\delta+2)} \int_0^{+\infty} z^\delta \log z e^{-\delta z} dz \\ &= \frac{\delta}{\Gamma(\delta+2)}\left(\int_0^{+\infty} y^\delta \log y e^{-y} dy\right. \\ &\quad \left.- \log \delta \int_0^{+\infty} y^\delta e^{-y} dy\right) \\ &= \frac{\delta}{\delta+1}(\psi(\delta+1) - \log \delta),\end{aligned}\tag{3.13}$$

$$\mathbb{E}_{\delta+2,\delta}\left(\frac{\log Z}{Z^2}\right) = \frac{\delta}{\delta+1}(\psi(\delta) - \log \delta),\tag{3.14}$$

$$\begin{aligned}\mathbb{E}_{\delta+2,\delta}\left(\frac{\log^2 Z}{Z}\right) &= \int_0^{+\infty} \frac{\log^2 z}{z} \frac{\delta e^{-\delta z} (\delta z)^{\delta+1}}{\Gamma(\delta+2)} dz \\ &= \frac{\delta^{\delta+2}}{\Gamma(\delta+2)} \int_0^{+\infty} z^\delta \log^2 z e^{-\delta z} dz\end{aligned}$$

$$\begin{aligned}
&= \frac{\delta}{\Gamma(\delta+2)} \left(\int_0^{+\infty} y^\delta \log^2 y e^{-y} dy \right. \\
&\quad \left. - 2 \log \delta \int_0^{+\infty} y^\delta \log y e^{-y} dy + \log^2 \delta \int_0^{+\infty} y^\delta e^{-y} dy \right) \\
&= \frac{\delta}{\delta+1} \left(\psi'(\delta+1) + \psi(\delta+1)^2 + \log^2 \delta \right), \quad (3.15) \\
\mathbb{E}_{\delta+2,\delta} \left(\frac{\log^2 Z}{Z^2} \right) &= \int_0^{+\infty} \frac{\log^2 z}{z^2} \frac{\delta e^{-\delta z} (\delta z)^{\delta+1}}{\Gamma(\delta+2)} dz \\
&= \frac{\delta^{\delta+2}}{\Gamma(\delta+2)} \int_0^{+\infty} z^{\delta-1} \log^2 z e^{-\delta z} dz \\
&= \frac{\delta}{\Gamma(\delta+2)} \left(\int_0^{+\infty} y^{\delta-1} \log^2 y e^{-y} dy \right. \\
&\quad \left. - 2 \log \delta \int_0^{+\infty} y^{\delta-1} \log y e^{-y} dy + \log^2 \delta \int_0^{+\infty} y^{\delta-1} e^{-y} dy \right) \\
&= \frac{\delta}{\delta+1} \left(\psi'(\delta) + \psi(\delta)^2 - 2\psi(\delta) \log \delta + \log^2 \delta \right). \quad (3.16)
\end{aligned}$$

Applicando i risultati precedenti e il lemma, è possibile calcolare gli elementi della matrice dell'informazione di Fisher

$$\begin{aligned}
\mathbb{E}_\theta \left(- \frac{\partial^2 l}{\partial \delta^2} \right) &= \frac{1}{\delta^2} + \frac{1}{(\delta+1)^2}, \\
\mathbb{E}_\theta \left(- \frac{\partial^2 l}{\partial \eta_1 \partial \delta} \right) &= \frac{\delta(\delta+1)}{(\delta+1)(\delta+2)} \mathbb{E}_{\delta+1,\delta} \left(\mathbb{E}_\theta(U_1 | Z) \right) = \frac{\delta}{\delta+2} \mathbb{E}_{\delta+1,\delta} \left(\frac{1}{Z \lambda_1} \right) \\
&= \frac{1}{\eta_1(\delta+2)}, \\
\mathbb{E}_\theta \left(- \frac{\partial^2 l}{\partial \eta_1 \partial \eta_2} \right) &= -(\delta+2) \frac{\delta(\delta+1)}{(\delta+2)(\delta+3)} \mathbb{E}_{\delta+2,\delta} \left(\mathbb{E}_\theta(U_1 | Z) \mathbb{E}_\theta(U_2 | Z) \right) \\
&= -\frac{\delta(\delta+1)}{\delta+3} \mathbb{E}_{\delta+2,\delta} \left(\frac{1}{Z^2 \lambda_1 \lambda_2} \right) = -\frac{\delta(\delta+1)}{(\delta+3) \lambda_1 \lambda_2 (\delta+1)} \\
&= -\frac{1}{\eta_1 \eta_2 (\delta+3)}, \\
\mathbb{E}_\theta \left(- \frac{\partial^2 l}{\partial \eta_1^2} \right) &= \frac{1}{\eta_1^2} - (\delta+2) \frac{\delta(\delta+1)}{(\delta+2)(\delta+3)} \mathbb{E}_{\delta+2,\delta} \left(\mathbb{E}_\theta(U_1^2 | Z) \right) \\
&= \frac{1}{\eta_1^2} - (\delta+2) \frac{\delta(\delta+1)}{(\delta+2)(\delta+3)} \frac{2}{\lambda_1^2} \mathbb{E}_{\delta+2,\delta} \left(\frac{1}{Z^2} \right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\eta_1^2} - \frac{\delta(\delta+1)}{(\delta+3)} \frac{2}{\lambda_1^2} \frac{\delta}{(\delta+1)} \\
&= \frac{\delta+1}{\eta_1^2(\delta+3)}, \\
\mathbb{E}_\theta \left(-\frac{\partial^2 l}{\partial \gamma \partial \delta} \right) &= \frac{\delta(\delta+1)}{\gamma(\delta+2)(\delta+3)} \mathbb{E}_{\delta+1,\delta} \left(\mathbb{E}_\theta(\eta_1 U_1 \log U_1 + \eta_2 U_2 \log U_2 \mid Z) \right) \\
&= \frac{\delta}{\gamma(\delta+2)} \mathbb{E}_{\delta+1,\delta} \left[\frac{\eta_1}{\lambda_1} \left(\frac{\psi(2) - \log \lambda_1 - \log Z}{Z} \right) \right. \\
&\quad \left. + \frac{\eta_2}{\lambda_2} \left(\frac{\psi(2) - \log \lambda_2 - \log Z}{Z} \right) \right] \\
&= \frac{\delta}{\gamma(\delta+2)} \left[\frac{\eta_1}{\lambda_1} (\psi(2) - \log \lambda_1 - \psi(\delta) + \log \delta) \right. \\
&\quad \left. + \frac{\eta_2}{\lambda_2} (\psi(2) - \log \lambda_2 - \psi(\delta) + \log \delta) \right] \\
&= \frac{2h_1(\delta) - \log \eta_1 - \log \eta_2}{\gamma(\delta+2)}, \\
\text{dove } h_1(\delta) &= \psi(2) - \psi(\delta), \\
\mathbb{E}_\theta \left(-\frac{\partial^2 l}{\partial \eta_1 \partial \gamma} \right) &= \frac{(\delta+2)}{\gamma} \frac{\delta(\delta+1)}{(\delta+1)(\delta+2)} \mathbb{E}_{\delta+1,\delta} \mathbb{E}_\theta(U_1 \log U_1 \mid Z) \\
&\quad - \frac{(\delta+2)}{\gamma} \frac{\delta(\delta+1)}{(\delta+2)(\delta+3)} \mathbb{E}_{\delta+2,\delta} \mathbb{E}_\theta(\eta_1 U_1^2 \log U_1 \\
&\quad + \eta_2 U_1 U_2 \log U_2 \mid Z) \\
&= \frac{\delta}{\gamma} \mathbb{E}_{\delta+1,\delta} \left[\frac{1}{Z \lambda_1} (\psi(2) - \log \lambda_1 - \log Z) \right] - \frac{\delta(\delta+1)}{\gamma(\delta+3)} \\
&\quad \left[\eta_1 \mathbb{E}_{\delta+2,\delta} \left(\frac{1}{Z^2 \lambda_1^2} (2\psi(2) + 1 - 2 \log \lambda_1 - 2 \log Z) \right) \right. \\
&\quad \left. + \eta_2 \mathbb{E}_{\delta+2,\delta} \left(\frac{1}{Z^2 \lambda_1 \lambda_2} (\psi(2) - \log \lambda_2 - \log Z) \right) \right] \\
&= \frac{\delta}{\gamma} \left[\frac{1}{\eta_1} (\psi(2) - \log \lambda_1 - \psi(\delta) + \log \delta) \right] - \frac{1}{\gamma(\delta+3)} \\
&\quad \left[\frac{\delta}{\eta_1(\delta+1)} (2\psi(2) + 1 - 2 \log \lambda_1 - 2\psi(\delta) + 2 \log \delta) \right. \\
&\quad \left. + \frac{\delta}{\eta_1(\delta+1)} (\psi(2) - \log \lambda_2) - \frac{\delta}{\eta_1(\delta+1)} (\psi(\delta) - \log \delta) \right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{h_1(\delta) - \log \eta_1}{\gamma \eta_1} - \frac{1}{\gamma \eta_1 (\delta + 3)} \left(2\psi(2) + 1 - 2 \log \lambda_1 - 2\psi(\delta) \right. \\
&\quad \left. + 2 \log \delta + \psi(2) - \log \lambda_2 - \psi(\delta) + \log \delta \right) \\
&= \frac{\delta h_1(\delta + 1) - (\delta + 1) \log \eta_1 + \log \eta_2}{\gamma \eta_1 (\delta + 3)}, \\
\mathbb{E}_\theta \left(- \frac{\partial l}{\partial \gamma^2} \right) &= \frac{2}{\gamma^2} + \frac{(\delta + 2)}{\gamma^2} \frac{\delta(\delta + 1)}{(\delta + 1)(\delta + 2)} \left[\mathbb{E}_{\delta+1, \delta} \left(\eta_1 \mathbb{E}_\theta (U_1 \log^2 U_1 \mid Z) \right. \right. \\
&\quad \left. \left. + \eta_2 \mathbb{E}_\theta (U_2 \log^2 U_2 \mid Z) \right) \right] - \frac{(\delta + 2)}{\gamma^2} \frac{\delta(\delta + 1)}{(\delta + 2)(\delta + 3)} \\
&\quad \left[\mathbb{E}_{\delta+2, \delta} \left(\mathbb{E}_\theta (\eta_1^2 U_1^2 \log^2 U_1 + 2\eta_1 \eta_2 U_1 \log U_1 U_2 \log U_2 \right. \right. \\
&\quad \left. \left. + \eta_2^2 U_2^2 \log^2 U_2 \mid Z) \right) \right] \\
&= \frac{2}{\gamma^2} + \frac{\delta}{\gamma^2} \mathbb{E}_{\delta+1, \delta} \left[\eta_1 \left(\frac{\log^2 Z}{Z \lambda_1} + \frac{\log Z}{Z \lambda_1} (-2\psi(2) + 2 \log \lambda_1) \right. \right. \\
&\quad \left. \left. + \frac{1}{Z \lambda_1} (\psi'(2) + \psi(2)^2 + \log^2 \lambda_1 - 2\psi(2) \log \lambda_1) \right) \right. \\
&\quad \left. + \eta_2 \left(\frac{\log^2 Z}{Z \lambda_2} + \frac{\log Z}{Z \lambda_2} (-2\psi(2) + 2 \log \lambda_2) \right. \right. \\
&\quad \left. \left. + \frac{1}{Z \lambda_2} (\psi'(2) + \psi(2)^2 + \log^2 \lambda_2 - 2\psi(2) \log \lambda_2) \right) \right] \\
&\quad - \frac{\delta(\delta + 1)}{\gamma^2 (\delta + 3)} \mathbb{E}_{\delta+2, \delta} \left[\eta_1^2 \left(\frac{2 \log^2 Z}{Z^2 \lambda_1^2} + \frac{\log Z}{Z^2 \lambda_1^2} (-4\psi(2) - 2 \right. \right. \\
&\quad \left. \left. + 4 \log \lambda_1) + \frac{1}{Z^2 \lambda_1^2} (2\psi'(2) + 2\psi(2)^2 + 2\phi(2) + 2 \log^2 \lambda_1 \right. \right. \\
&\quad \left. \left. - 4\psi(2) \log \lambda_1 - 2 \log \lambda_1) \right) + 2\eta_1 \eta_2 \left(\frac{\psi(2) - \log \lambda_1 \log Z}{Z \lambda_1} \right. \right. \\
&\quad \left. \left. \frac{\psi(2) - \log \lambda_2 \log Z}{Z \lambda_2} \right) + \eta_2^2 \left(\frac{2 \log^2 Z}{Z^2 \lambda_2^2} + \frac{\log Z}{Z^2 \lambda_2^2} (-4\psi(2) - 2 \right. \right. \\
&\quad \left. \left. + 4 \log \lambda_2) + \frac{1}{Z^2 \lambda_2^2} (2\psi'(2) + 2\psi(2)^2 + 2\phi(2) + 2 \log^2 \lambda_2 \right. \right. \\
&\quad \left. \left. - 4\psi(2) \log \lambda_2 - 2 \log \lambda_2) \right) \right] \\
&= \frac{2}{\gamma^2} + \frac{1}{\gamma^2} \left[(\psi'(2) + \psi(2)^2 + \log^2 \lambda_1 - 2\psi(2) \log \lambda_1 + \psi'(\delta) \right.
\end{aligned}$$

$$\begin{aligned}
& +\psi(\delta)^2 - 2\psi(\delta) \log \delta + \log^2 \delta + 2\psi(\delta) \log \delta - 2 \log \lambda_1 \log \delta \\
& - 2\psi(2)\psi(\delta) + 2\psi(2) \log \delta) + (\psi'(2) + \psi(2)^2 + \log^2 \lambda_2 \\
& - 2\psi(2) \log \lambda_2 + \psi'(\delta) + \psi(\delta)^2 - 2\psi(\delta) \log \delta + \log^2 \delta \\
& + 2\psi(\delta) \log \delta - 2 \log \lambda_2 \log \delta - 2\psi(2)\psi(\delta) + 2\psi(2) \log \delta) \Big] \\
& - \frac{1}{\gamma^2(\delta + 3)} \Big[(2\psi'(2) + 2\psi(2)^2 + 2\psi(2) + 2 \log^2 \lambda_1 \\
& - 4\psi(2) \log \lambda_1 - 2 \log \lambda_1 + 2\psi'(\delta) + 2\psi(\delta)^2 - 4\psi(\delta) \log(\delta) \\
& + 2 \log^2 \delta + 4\psi(\delta) \log \lambda_1 - 4\psi(2)\psi(\delta) - 2\psi(\delta) - 4 \log \lambda_1 \log \delta \\
& + 4\psi(2) \log \delta + 2 \log \delta) + (\psi(2)^2 - \psi(2) \log \lambda_1 - \psi(2) \log \lambda_2 \\
& + \log \lambda_1 \log \lambda_2 + \psi(\delta) \log \lambda_2 - 2\psi(2)\psi(\delta) + \psi(\delta) \log \lambda_1 \\
& - \log \lambda_2 \log \delta + 2\psi(2) \log \delta - \log \lambda_1 \log \delta + \psi'(\delta) + \psi(\delta)^2 \\
& - 2\psi(\delta) \log \delta + \log^2 \delta) + (2\psi'(2) + 2\psi(2)^2 + 2\psi(2) + 2 \log^2 \lambda_2 \\
& - 4\psi(2) \log \lambda_2 - 2 \log \lambda_2 + 2\psi'(\delta) + 2\psi(\delta)^2 - 4\psi(\delta) \log(\delta) \\
& + 2 \log^2 \delta + 4\psi(\delta) \log \lambda_2 - 4\psi(2)\psi(\delta) - 2\psi(\delta) - 4 \log \lambda_2 \log \delta \\
& + 4\psi(2) \log \delta + 2 \log \delta) \Big] \\
& = \frac{1}{\gamma^2(\delta + 3)} \Big[(\delta + 1)(\log^2 \eta_1 + \log^2 \eta_2) - 2 \log \eta_1 \log \eta_2 \\
& - 2\delta h_1(\delta + 1)(\log \eta_1 + \log \eta_2) + 2h_2(\delta) \Big],
\end{aligned}$$

$$\text{dove } h_2(\delta) = \delta + 3 + \delta h_1^2(\delta) - 2h_1(\delta) + (\delta + 1)\psi'(2) + \delta\psi'(\delta).$$

La matrice calcolata coincide con quella riportata in Bjarnason e Hougaard (2000).

Si considera ora il caso particolare nel quale si pone $\gamma = 1$ e $\lambda_1 = \lambda_2 = \lambda$, ovvero $\eta_1 = \eta_2 = \eta$. Il parametro $\theta = (\delta, \eta)$ è dunque bidimensionale.

La funzione di densità congiunta per una singola coppia (T_1, T_2) , sotto queste nuove ipotesi, diventa

$$f(t_1, t_2) = \frac{\eta^2 \delta (\delta + 1)}{(1 + \eta t_1 + \eta t_2)^{\delta+2}},$$

e la funzione di log-verosimiglianza completa risulta

$$l(\delta, \eta) = 2 \log \eta + \log \delta + \log(\delta + 1) - (\delta + 2) \log(1 + \eta t_1 + \eta t_2).$$

Le componenti della funzione di punteggio sono

$$\begin{aligned} \frac{\partial l}{\partial \delta} &= \frac{1}{\delta} - \log(1 + \eta t_1 + \eta t_2), \\ \frac{\partial l}{\partial \eta} &= \frac{2}{\eta} - (\delta + 2) \frac{t_1 + t_2}{1 + \eta t_1 + \eta t_2}. \end{aligned}$$

La matrice d'informazione osservata $j(\delta, \eta)$ risulta

$$\begin{aligned} -\frac{\partial^2 l}{\partial \delta^2} &= \frac{1}{\delta^2} + \frac{1}{(\delta + 1)^2}, \\ -\frac{\partial^2 l}{\partial \eta \partial \delta} &= \frac{t_1 + t_2}{1 + \eta t_1 + \eta t_2}, \\ -\frac{\partial^2 l}{\partial \eta^2} &= \frac{2}{\eta^2} - (\delta + 2) \frac{t_1^2 + 2t_1 t_2 + t_2^2}{(1 + \eta t_1 + \eta t_2)^2}. \end{aligned}$$

Gli elementi della matrice di informazione di Fisher $i(\theta)$ sono

$$\begin{aligned} \mathbb{E}_\theta \left(-\frac{\partial^2 l}{\partial \delta^2} \right) &= \frac{1}{\delta^2} + \frac{1}{(\delta + 1)^2}, \\ \mathbb{E}_\theta \left(-\frac{\partial^2 l}{\partial \eta \partial \delta} \right) &= \frac{\delta(\delta + 1)}{(\delta + 1)(\delta + 2)} \mathbb{E}_{\delta+1, \delta} \left(\mathbb{E}_\theta(T_1 | Z) + \mathbb{E}_\theta(T_2 | Z) \right) \\ &= \frac{\delta}{\delta + 2} \mathbb{E}_{\delta+1, \delta} \left(\frac{1}{Z\lambda} + \frac{1}{Z\lambda} \right) \\ &= \frac{2}{\eta(\delta + 2)}, \\ \mathbb{E}_\theta \left(-\frac{\partial^2 l}{\partial \eta^2} \right) &= \frac{2}{\eta^2} - (\delta + 2) \frac{\delta(\delta + 1)}{(\delta + 2)(\delta + 3)} \mathbb{E}_{\delta+2, \delta} \left(\mathbb{E}_\theta(T_1^2 + 2T_1 T_2 + T_2^2 | Z) \right) \\ &= \frac{2}{\eta^2} - \frac{\delta(\delta + 1)}{(\delta + 3)} \left[\frac{2}{\lambda^2} \mathbb{E}_{\delta+2, \delta} \left(\frac{1}{Z^2} \right) \right. \\ &\quad \left. + \frac{2}{\lambda^2} \mathbb{E}_{\delta+2, \delta} \left(\frac{1}{Z^2} \right) + \frac{2}{\lambda^2} \mathbb{E}_{\delta+2, \delta} \left(\frac{1}{Z^2} \right) \right] \\ &= \frac{2}{\eta^2} - \frac{\delta(\delta + 1)}{(\delta + 3)} \frac{6}{\lambda^2} \frac{\delta}{(\delta + 1)} \\ &= \frac{2\delta}{\eta^2(\delta + 3)}. \end{aligned}$$

Essendo δ il parametro d'interesse, la varianza asintotica dello stimatore risulta

$$\begin{aligned} i^{\delta\delta} &= (i(\theta)^{-1})_{11} = \frac{1}{\det i(\theta)} i(\theta)_{22} \\ &= \left[\left(\frac{1}{\delta^2} + \frac{1}{(\delta+1)^2} \right) \frac{2\delta}{\eta^2(\delta+3)} - \frac{4}{\eta^2(\delta+2)^2} \right]^{-1} \frac{2\delta}{\eta^2(\delta+3)}. \end{aligned}$$

Si può semplificare η^2 , ottenendo una varianza asintotica che non dipende da η .

Si desidera ora valutare la perdita di efficienza asintotica nella stima del parametro δ , quando si assume come ipotesi di lavoro l'indipendenza all'interno delle coppie e si pone $\gamma = 1$ e $\lambda_1 = \lambda_2 = \lambda$. La distribuzione marginale per una coppia di osservazioni risulta

$$f(t_1, t_2; \delta, \eta) = \delta^2 \eta^2 \frac{1}{(1 + \eta t_1)^{\delta+1}} \frac{1}{(1 + \eta t_2)^{\delta+1}},$$

e la log-verosimiglianza d'indipendenza è

$$pl_1(\delta, \eta) = 2 \log \delta + 2 \log \eta - (\delta + 1) \log(1 + \eta t_1) - (\delta + 1) \log(1 + \eta t_2).$$

Le quantità di verosimiglianza sono

$$\begin{aligned} \frac{\partial pl_1}{\partial \delta} &= \frac{2}{\delta} - \log(1 + \eta t_1) - \log(1 + \eta t_2), \\ \frac{\partial pl_1}{\partial \eta} &= \frac{2}{\eta} - (\delta + 1) \frac{t_1}{1 + \eta t_1} - (\delta + 1) \frac{t_2}{1 + \eta t_2}. \end{aligned}$$

La matrice d'informazione osservata $j(\delta, \eta)$ è formata dai seguenti elementi

$$\begin{aligned} -\frac{\partial^2 pl_1}{\partial \delta^2} &= \frac{2}{\delta^2}, \\ -\frac{\partial^2 pl_1}{\partial \eta \partial \delta} &= \frac{t_1}{1 + \eta t_1} + \frac{t_2}{1 + \eta t_2}, \\ -\frac{\partial^2 pl_1}{\partial \eta^2} &= \frac{2}{\eta^2} - (\delta + 1) \frac{t_1^2}{(1 + \eta t_1)^2} - (\delta + 1) \frac{t_2^2}{(1 + \eta t_2)^2}. \end{aligned}$$

L'informazione di Godambe si ottiene calcolando

$$H(\theta) = \mathbb{E}_\theta \left(-\frac{\partial^2 pl_1}{\partial \theta \partial \theta^\top} \right) \quad \text{e} \quad J(\theta) = \text{Var}_\theta \left(\frac{\partial pl_1}{\partial \theta} \right).$$

Sono qui riportati gli elementi della matrice $H(\theta)$

$$\begin{aligned}\mathbb{E}_\theta\left(-\frac{\partial^2 p l_1}{\partial \delta^2}\right) &= \frac{2}{\delta^2}, \\ \mathbb{E}_\theta\left(-\frac{\partial^2 p l_1}{\partial \eta \partial \delta}\right) &= \frac{\delta}{\delta+1} \left[\mathbb{E}_{\delta+1, \delta} \left(\mathbb{E}_\theta(T_1 | Z) + \mathbb{E}_\theta(T_2 | Z) \right) \right] \\ &= \frac{\delta}{\delta+1} \mathbb{E}_{\delta+1, \delta} \left(\frac{1}{Z\lambda} + \frac{1}{Z\lambda} \right) \\ &= \frac{\delta}{(\delta+1)\lambda} \frac{2}{\lambda} \\ &= \frac{2}{\eta(\delta+1)}, \\ \mathbb{E}_\theta\left(-\frac{\partial^2 p l_1}{\partial \eta^2}\right) &= \frac{2}{\eta^2} - (\delta+1) \frac{\delta}{\delta+2} \left[\mathbb{E}_{\delta+2, \delta} \left(\mathbb{E}_\theta(T_1^2 | Z) + \mathbb{E}_\theta(T_2^2 | Z) \right) \right] \\ &= \frac{2}{\eta^2} - (\delta+1) \frac{\delta}{\delta+2} \mathbb{E}_{\delta+2, \delta} \left(\frac{2}{Z^2 \lambda^2} + \frac{2}{Z^2 \lambda^2} \right) \\ &= \frac{2}{\eta^2} - (\delta+1) \frac{\delta}{(\delta+2)} \frac{4\delta}{(\delta+1)\lambda^2} \\ &= \frac{2}{\eta^2} - \frac{4}{\eta^2(\delta+2)} \\ &= \frac{2\delta}{\eta^2(\delta+2)}.\end{aligned}$$

Per il calcolo della matrice $J(\theta)$ è conveniente considerare l'osservazione seguente. La distribuzione marginale di T_j risulta

$$f(t_j; \delta, \eta) = \delta \eta \frac{1}{(1 + \eta t_j)^{\delta+1}}$$

e una conveniente trasformazione è $Y_j = \log(1 + \eta T_j)$. Poiché

$$T_j = \frac{1}{\eta} (e^{Y_j} - 1) \quad \text{e} \quad \frac{\partial t_j}{\partial y_j} = \frac{1}{\eta} e^{y_j},$$

la distribuzione marginale per il tempo di vita T_j diventa

$$\begin{aligned}f(y_j; \delta, \eta) &= \delta \eta e^{-(\delta+1)y_j} \frac{1}{\eta} e^{y_j} \\ &= \delta e^{-\delta y_j}.\end{aligned}$$

Si deduce che la variabile Y_j si distribuisce come un esponenziale di media δ^{-1} .

La distribuzione congiunta di (Y_1, Y_2) risulta

$$f_Y(y_1, y_2) = f_T(t_1(y_1, y_2), t_2(y_1, y_2)) | J |,$$

dove $| J |$ è il determinante della matrice jacobiana

$$J = \begin{pmatrix} \frac{e^{y_1}}{\eta} & 0 \\ 0 & \frac{e^{y_2}}{\eta} \end{pmatrix}.$$

Si ricava

$$f(y_1, y_2) = \delta(\delta + 1) \frac{e^{y_1 + y_2}}{(e^{y_1} + e^{y_2} - 1)^{\delta + 2}},$$

e dunque tale distribuzione dipende solo da δ .

Gli elementi della matrice $J(\theta)$, calcolati sotto il vero modello di dipendenza all'interno delle coppie, sono

$$\begin{aligned} \text{Var}_\theta \left(\frac{\partial p_{l_1}}{\partial \delta} \right) &= \text{Var}_\theta \left(\log(1 + \eta T_1) \right) + \text{Var}_\theta \left(\log(1 + \eta T_2) \right) \\ &\quad + 2 \text{Cov}_\theta \left(\log(1 + \eta T_1), \log(1 + \eta T_2) \right), \\ \text{Var}_\theta \left(\frac{\partial p_{l_1}}{\partial \eta} \right) &= (\delta + 1)^2 \left[\text{Var}_\theta \left(\frac{T_1}{1 + \eta T_1} \right) + \text{Var}_\theta \left(\frac{T_2}{1 + \eta T_2} \right) \right. \\ &\quad \left. + 2 \text{Cov}_\theta \left(\frac{T_1}{1 + \eta T_1}, \frac{T_2}{1 + \eta T_2} \right) \right], \\ \text{Cov}_\theta \left(\frac{\partial p_{l_1}}{\partial \delta}, \frac{\partial p_{l_1}}{\partial \eta} \right) &= \mathbb{E}_\theta \left(\frac{\partial p_{l_1}}{\partial \delta} \frac{\partial p_{l_1}}{\partial \eta} \right) - \mathbb{E}_\theta \left(\frac{\partial p_{l_1}}{\partial \delta} \right) \mathbb{E}_\theta \left(\frac{\partial p_{l_1}}{\partial \eta} \right). \end{aligned}$$

Poiché

$$\mathbb{E}_\theta \left(\frac{\partial p_{l_1}}{\partial \delta} \right) \mathbb{E}_\theta \left(\frac{\partial p_{l_1}}{\partial \eta} \right) = 0,$$

si ha

$$\begin{aligned} \text{Cov}_\theta \left(\frac{\partial p_{l_1}}{\partial \delta}, \frac{\partial p_{l_1}}{\partial \eta} \right) &= \mathbb{E}_\theta \left(\frac{\partial p_{l_1}}{\partial \delta} \frac{\partial p_{l_1}}{\partial \eta} \right) \\ &= \mathbb{E}_\theta \left[\frac{4}{\delta \eta} - \frac{2(\delta + 1)}{\delta} \left(\frac{T_1}{1 + \eta T_1} + \frac{T_2}{1 + \eta T_2} \right) \right] \end{aligned}$$

$$\begin{aligned}
& +(\delta + 1) \left(\log(1 + \eta T_1) \frac{T_1}{1 + \eta T_1} + \log(1 + \eta T_2) \frac{T_2}{1 + \eta T_2} \right) \\
& +(\delta + 1) \left(\log(1 + \eta T_1) \frac{T_2}{1 + \eta T_2} + \log(1 + \eta T_2) \frac{T_1}{1 + \eta T_1} \right) \\
& - \frac{2}{\eta} \left(\log(1 + \eta T_1) + \log(1 + \eta T_2) \right) \Big].
\end{aligned}$$

Alcuni degli integrali richiesti per valutare tali valori attesi non hanno un'espressione semplice.

Un esempio è

$$\begin{aligned}
\mathbb{E}_{\delta, \delta} \mathbb{E}_{\theta} \left(\log(1 + \eta T_1) \log(1 + \eta T_2) \mid Z \right) &= \mathbb{E}_{\delta, \delta} \left(\mathbb{E}_{\theta}(\log(1 + \eta T_1) \mid Z) \right. \\
&\quad \left. \mathbb{E}_{\theta}(\log(1 + \eta T_2) \mid Z) \right),
\end{aligned}$$

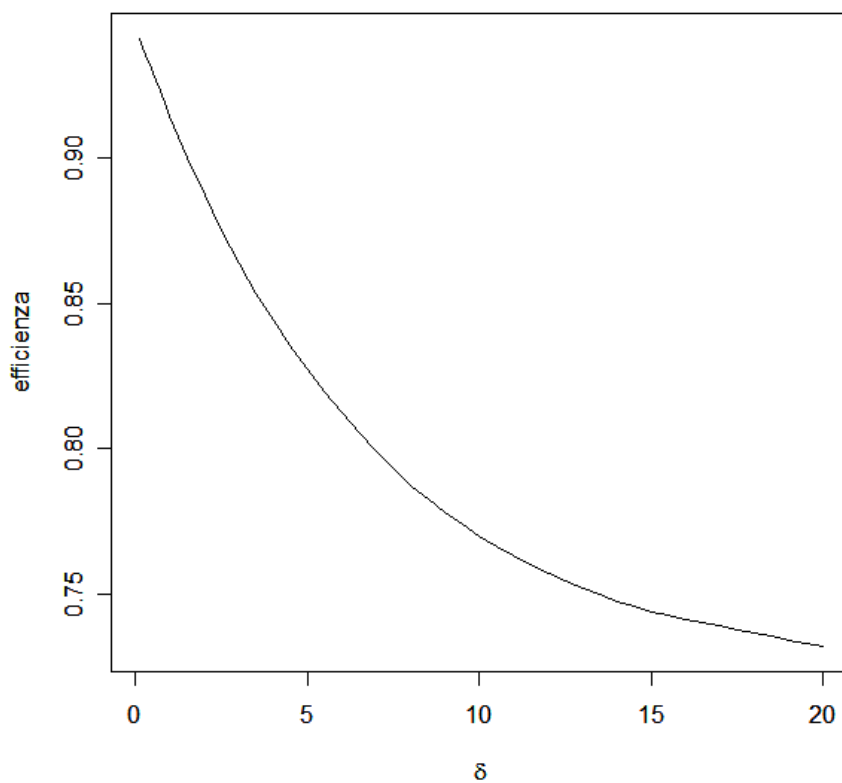
il cui calcolo richiede la risoluzione di integrali del tipo

$$\int_0^{+\infty} \log(1 + \eta t_j) z \lambda e^{-z \lambda t_j} dt_j, \quad j = 1, 2.$$

A sua volta, il prodotto dei risultati va integrato rispetto alla densità di Z .

Si è dunque deciso di valutare $J(\theta)$ tramite simulazione utilizzando l'ambiente di calcolo statistico R. Pertanto, ricorrendo alla simulazione Monte Carlo è possibile generare realizzazioni pseudo-casuali dalla distribuzione $Gamma(\delta, \delta)$ e dall' $Exp(z\lambda)$ e valutare gli elementi della matrice $J(\theta)$.

L'obiettivo finale è di confrontare la varianza asintotica dello stimatore di massima verosimiglianza con quella dello stimatore ottenuto in ipotesi di indipendenza. Viene riportato il grafico che illustra la perdita di efficienza per vari valori del parametro δ , fissato $\eta = 1$.



Come si può notare, la perdita di efficienza aumenta all'aumentare del valore di δ . Tuttavia, per valori piccoli di δ , la perdita di efficienza che si ottiene lavorando sotto l'ipotesi di indipendenza sembra essere trascurabile. Alcuni valori della perdita di efficienza sono riportati nella tabella sottostante.

δ	0.1	1	2	4	6	8
efficienza	0.9820	0.9024	0.8561	0.8429	0.8138	0.7869

δ	10	12	14	16	18	20
efficienza	0.7721	0.7679	0.7438	0.7385	0.7395	0.7321

Appendice: comandi R

Vengono riportati alcuni comandi di R utilizzati nel paragrafo 3.4.

```
# numero di simulazioni
nr=10000
eta=1
lambda=eta*delta
delta=1

#varianza asintotica dello stimatore di
#massima verosimiglianza
var0=2*delta*((1/delta^2+1/(delta+1)^2)*(2*delta/(delta+3))
-4/(delta+2)^2)^{-1}/(delta+3)

# elementi della matrice H(theta)
h11=2/delta^2
h12=2/(eta*(delta+1))
h22=2*delta/(eta^2*(delta+2))
h=matrix(c(h11,h12,h12,h22), byrow=TRUE, ncol=2)

mu1=0
mu2=0
mu1q=0
mu2q=0
```

```
mulu2=0
for(i in 1:nr){z=rgamma(1,delta,1)/delta
  par=lambda*z
  t1=rexp(1,par)
  t2=rexp(1,par)
  u1=2/delta-log(1+eta*t1)-log(1+eta*t2)
  u2=2/eta-(delta+1)*t1/(1+eta*t1)-(delta +1)*t2/(1+eta*t2)
  mu1=mu1+u1
  mu2=mu2+u2
  mu1q=mu1q+u1^2
  mu2q=mu2q+u2^2
  mulu2=mulu2+u1*u2
}

# valori attesi
mu1=mu1/nr
mu2=mu2/nr
mu1q=mu1q/nr
mu2q=mu2q/nr
mulu2=mulu2/nr

#elementi della matrice J(theta)
j11=mu1q-mu1^2
j22=mu2q-mu2^2
j12=mulu2-mu1*mu2
j=matrix(c(j11,j12,j12,j22), byrow=TRUE, ncol=2)

#matrice di informazione di Godambe
g=h%*%solve(j)%*%h

#inversa della matrice di Godambe
```

```
ginv=solve(g)

# varianza asintotica dello stimatore sotto
#l'ipotesi di indipendenza
var1=ginv[1,1]

#perdita di efficienza
eff=var0/var1
eff
```

Sono state eseguite 10000 simulazioni per diversi valori del parametro δ . I risultati ottenuti sono

```
#valori assegnati al parametro delta
delta<-scan()
1: 0.1
2: 0.2
3: 0.3
4: 0.4
5: 0.5
6: 0.7
7: 0.9
8: 1
9: 1.5
10: 2
11: 2.5
12: 3
13: 3.5
14: 4
15: 4.5
16: 5
17: 5.5
```

```
18: 6
19: 6.5
20: 7
21: 7.5
22: 8
23: 8.5
24: 9
25: 9.5
26: 10
27: 10.5
28: 11
29: 11.5
30: 12
31: 12.5
32: 13
33: 13.5
34: 14
35: 14.5
36: 15
37: 15.5
38: 16
39: 16.5
40: 17
41: 17.5
42: 18
43: 18.5
44: 19
45: 19.5
46: 20
```

```
#valori di efficienza ottenuti
```

```
eff<-scan()  
1: 0.982  
2: 0.9677  
3: 0.9539  
4: 0.923  
5: 0.9207  
6: 0.9172  
7: 0.9053  
8: 0.9024  
9: 0.8936  
10: 0.8561  
11: 0.8576  
12: 0.8523  
13: 0.8517  
14: 0.8429  
15: 0.846  
16: 0.843  
17: 0.8372  
18: 0.8138  
19: 0.8008  
20: 0.7888  
21: 0.7823  
22: 0.7869  
23: 0.7867  
24: 0.7778  
25: 0.7755  
26: 0.7721  
27: 0.7681  
28: 0.7616  
29: 0.7575  
30: 0.7679
```

```
31: 0.7515
32: 0.7539
33: 0.75
34: 0.7438
35: 0.7471
36: 0.7366
37: 0.7401
38: 0.7385
39: 0.7354
40: 0.7482
41: 0.7473
42: 0.7395
43: 0.7287
44: 0.7366
45: 0.7305
46: 0.7321
```

```
# grafico
plot(delta, eff, xlab=expression(delta), ylab="efficienza",
type="l")
```

Per ottenere un andamento più liscio della curva è stato utilizzato il comando *smooth.spline*. Si è stimata sui dati una spline di lisciamiento cubica, dove *spar* è il parametro di lisciamiento.

```
plot(smooth.spline(delta, eff, spar=0.9), xlab=expression
(delta), ylab="efficienza", type="l")
```


Conclusioni

In questa tesi è stata considerata la verosimiglianza composta ottenuta combinando validi oggetti di verosimiglianza, utilizzata per fare inferenza al posto della verosimiglianza completa.

L'obiettivo principale era quello di analizzare le proprietà della verosimiglianza composta, mettendo in evidenza gli aspetti che la caratterizzano e la distinguono dall'usuale verosimiglianza. In particolare, si è fatto riferimento alla verosimiglianza composta marginale, particolare tipo di verosimiglianza composta costruita componendo distribuzioni marginali di piccole dimensioni. Le procedure inferenziali di questo nuovo approccio hanno proprietà teoriche simili a quelle dei metodi basati sulla verosimiglianza completa, ma per lo stimatore di massima verosimiglianza composta non vale l'identità dell'informazione, portando a una perdita di efficienza.

Nel Capitolo 2 sono stati trattati due specifici tipi di verosimiglianza composta marginale: la verosimiglianza a coppie, ottenuta considerando la distribuzione bivariata, e la verosimiglianza d'indipendenza, che utilizza solo le marginali univariate.

Nel Capitolo 3 è stata adattata la teoria esposta nei precedenti capitoli ad alcuni esempi in cui si sono confrontati i risultati ottenuti con la verosimiglianza completa e con quella composta marginale, in termini di perdita asintotica di efficienza. I risultati ottenuti nelle varie applicazioni evidenziano che l'utilizzo della verosimiglianza composta marginale consente di ottenere calcoli più semplici nelle procedure inferenziali, ma con-

duce a un aumento della varianza asintotica dello stimatore di massima verosimiglianza composita.

Riferimenti bibliografici

- Andersen, E.W. (2004). Composite likelihood and two-stage estimation in family studies. *Biostatistics*, **5**, 15-30.
- Anderson, J.A. e Pemberton, J.D. (1985). The grouped continuous model for multivariate ordered categorical variables and covariate adjustment. *Biometrics* **41**, 875-885.
- Azzalini, A. (1983). Maximum likelihood of order m for stationary stochastic processes. *Biometrika*, **70**, 381-367.
- Barndorff-Nielsen, O.E. e Cox, D.R. (1994). *Inference and Asymptotics*. Chapman and Hall, Londra.
- Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*, **B 34**, 192-236.
- Besag, J.E. (1977). Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika*, **64**, 616-618.
- Bjarnason, H. e Hougaard, P. (2000). Fisher information for two Gamma frailty bivariate Weibull models. *Lifetime Data Analysis*, **6**, 59-71.
- Chandler, R.E. e Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, **94**, 167-183.
- Cox, D.R. (1975). Partial likelihood. *Biometrika*, **62**, 269-276.

- Cox, D.R. e Reid, N. (2004). A note for clustered data using the independence loglikelihood. *Biometrika*, **91**, 729-737.
- Engler, D.A., Mohapatra, M., Louis, D.N. e Betensky, R.A. (2006). A pseudolikelihood approach for simultaneous analysis of array comparative genomic hybridization. *Biostatistics*, **7**, 399-421.
- Geys, H., Molenberghs, G. e Ryan, L.M. (1999). Pseudolikelihood modelling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association*, **94**, 734-745.
- Le Cessie, S. e Van Houwelingen, J.C. (1994). Logistic regression for correlated binary data. *Applied Statistics*, **43**, 95-108.
- Liang, K.-Y. (1987). Extended Mantel-Haenszel estimating procedure for multivariate logistic regression models. *Biometrics*, **43**, 289-299.
- Lindsay, B.G. (1988). Composite likelihood methods. *Contemporary Mathematics*, **80**, 221-240.
- Pace, L. e Salvan, A. (1996). *Teoria della Statistica*. Cedam, Padova.
- Stein, M.L. (2004). Approximating likelihoods for large spazial sets. *Journal of the Royal Statistical Society*, **B 66**, 275-296.
- Varin, C. (2007). On composite marginal likelihoods. *Asta Advances in Statistical Analysis*, in corso di stampa.
- Varin, C. e Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, **92**, 519-528.
- Varin, C. e Vidoni, P. (2007). Pairwise likelihood inference for general state space models. *Econometric Reviews*, in corso di stampa.
- Zhao, Y. e Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics*, **33**, 335-356.