

UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA  
IN STATISTICA E INFORMATICA

TESI DI LAUREA

**VEROSIMIGLIANZA A COPPIE IN  
MODELLI NORMALI MULTIVARIATI**

RELATORE: PROF. ALESSANDRA SALVAN

CORRELATORE: DOTT. NICOLA SARTORI

LAUREANDO: EULOGE CLOVIS KENNE PAGUI

ANNO ACCADEMICO 2008/2009



# Indice

<b>Introduzione</b>	<b>III</b>
<b>1 Verosimiglianza composta</b>	<b>1</b>
1.1 Introduzione . . . . .	1
1.2 Inferenza statistica . . . . .	2
1.3 Modelli statistici parametrici . . . . .	2
1.4 Funzione di verosimiglianza . . . . .	3
1.4.1 Informazione osservata . . . . .	4
1.4.2 Informazione attesa . . . . .	4
1.5 Verosimiglianza composta . . . . .	5
1.5.1 Introduzione . . . . .	5
1.5.2 Definizioni e proprietà . . . . .	5
1.5.3 Quantità associate alla verosimiglianza composta . . . . .	8
1.5.4 Verifica delle ipotesi . . . . .	9
<b>2 Verosimiglianza a coppie ed efficienza</b>	<b>11</b>
2.1 Introduzione . . . . .	11
2.2 Un esempio con parametro scalare: normale simmetrica . . . . .	12
2.2.1 Verosimiglianza completa . . . . .	13
2.2.2 Verosimiglianza a coppie . . . . .	18
2.2.3 Efficienza asintotica . . . . .	20
2.3 Un esempio con parametro multidimensionale . . . . .	20
2.3.1 Verosimiglianza completa . . . . .	22
2.3.2 Verosimiglianza a coppie . . . . .	27

---

2.3.3	Efficienza . . . . .	36
<b>3</b>	<b>Test basati sulla verosimiglianza a coppie</b>	<b>41</b>
3.1	Introduzione . . . . .	41
3.2	Le statistiche test . . . . .	42
3.3	Distribuzione normale simmetrica: studio di simulazione . .	44
3.3.1	Distribuzioni nulle . . . . .	45
3.3.2	Potenza dei test . . . . .	49
3.4	Modello ad effetti casuali: studio di simulazione . . . . .	50
3.4.1	Risultati analitici . . . . .	51
3.4.2	Distribuzioni nulle . . . . .	51
	<b>Conclusioni</b>	<b>55</b>
	<b>Appendice</b>	<b>57</b>
	<b>Riferimenti bibliografici</b>	<b>59</b>

# Introduzione

Il ruolo della funzione di verosimiglianza è centrale nell'inferenza statistica. Tuttavia, per molti problemi riguardanti l'inferenza in modelli parametrici e semiparametrici, la funzione di verosimiglianza risulta difficile da calcolare esplicitamente a causa di dipendenze complesse, di assunzioni fatte sul modello oppure nel caso di insiemi di dati di grandi dimensioni.

In queste situazioni, per risolvere il problema, vari autori hanno proposto diversi approcci basati su varianti della funzione di verosimiglianza. Tra gli esempi più conosciuti, si ricordano la pseudo-verosimiglianza di Besag (1974) per l'inferenza in modelli spaziali e la verosimiglianza parziale di Cox (1975), introdotta per l'inferenza in modelli con rischi proporzionali.

Un'idea alternativa alla verosimiglianza classica, è quella di costruire pseudo-verosimiglianze semplici (Cox and Reid, 2004). In questa classe di pseudo-verosimiglianze, troviamo la verosimiglianza composita marginale che appartiene alla classe più generale delle verosimiglianze composite (Lindsay, 1988). La verosimiglianza composita si ottiene combinando adeguatamente validi oggetti di verosimiglianza relativi a piccoli sottoinsiemi di dati.

I vantaggi sono la riduzione della complessità computazionale rispetto alla verosimiglianza completa e la possibilità di non precisare la densità congiunta in assenza di una sufficiente conoscenza del meccanismo che ha generato i dati.

L'impiego di questa metodologia è presente in una varietà di settori applicativi, per esempio nelle scienze ambientali, nel campo della genetica, nell'analisi delle serie storiche e nell'analisi dei dati di sopravvivenza. Si veda Varin (2008) per una rassegna recente.

Sotto adeguate condizioni di regolarità, le varie procedure inferenziali basate sulla verosimiglianza composita, hanno proprietà teoriche simili a quelle dei metodi basati sulla verosimiglianza completa, sebbene ci si aspetti una perdita di efficienza.

Questa tesi ha come obiettivo principale lo studio di un caso particolare di verosimiglianza composita marginale, detto verosimiglianza a coppie. Si focalizza l'attenzione sul comportamento della distribuzione dei test basati sulla verosimiglianza a coppie. In alcuni esempi in cui sia disponibile anche la verosimiglianza completa e sia possibile svolgere analiticamente gran parte dei calcoli necessari, si confrontano i risultati ottenuti utilizzando la verosimiglianza a coppie rispetto a quella completa e si valuta la perdita di efficienza asintotica quando si utilizza la verosimiglianza a coppie.

Più in dettaglio, si trattano due esempi di modelli normali multivariati con vincoli sulla matrice di covarianza. Alcuni aspetti del primo esempio, relativo ad un modello monoparametrico, sono già trattati in letteratura, in particolare in Cox e Reid (2004). Qui si analizza in dettaglio un aspetto non ancora esplorato relativo alle statistiche test, e alla corrispondente costruzione degli intervalli di confidenza, basati sulla verosimiglianza a coppie. Tramite un articolato studio di simulazione, si confrontano l'adeguatezza delle diverse approssimazioni delle distribuzioni. Il secondo esempio, relativo a un caso multiparametrico, è solo accennato in Cox e Reid (2004). Si mostra un risultato di efficienza della verosimiglianza a coppie e si analizza la struttura della verosimiglianza a coppie in relazione a quella completa. Come nell'esempio precedente, si valuta l'adeguatezza delle approssimazioni asintotiche disponibili per le statistiche di verosimiglianza a coppie. Si studia inoltre una semplice approssimazione per la

distribuzione della statistica log-rapporto di verosimiglianza a coppie.

Nel Capitolo 1 vengono richiamati il concetto di modello statistico parametrico e i principali risultati legati all'approccio di verosimiglianza. La teoria riassunta è utile per l'introduzione del concetto di verosimiglianza composita, che viene presentato assieme alle principali definizioni e proprietà del metodo.

Il Capitolo 2 è dedicato al caso particolare della verosimiglianza a coppie. Tramite due esempi viene studiata l'efficienza dello stimatore di massima verosimiglianza a coppie rispetto all'analogo stimatore ottenuto dalla verosimiglianza completa.

Nel Capitolo 3 si studia il comportamento delle statistiche test basate sulla verosimiglianza a coppie relativamente ai due esempi trattati nel Capitolo 2.





# Capitolo 1

## Verosimiglianza composita

### 1.1 Introduzione

In questo capitolo si concentra l'attenzione sul concetto e sulle proprietà della verosimiglianza composita. Viene inoltre introdotto un caso particolare della verosimiglianza composita chiamato **verosimiglianza composita marginale**. La verosimiglianza composita è una particolare pseudo-verosimiglianza (si veda ad esempio Pace e Salvan, 1997, capitolo 4; Molenberghs e Verbeke, 2005, capitolo 9). Si richiamano preliminarmente alcuni concetti fondamentali: inferenza statistica, modelli statistici parametrici, funzione di verosimiglianza, equazione di stima, informazione osservata e informazione di Fisher.

I primi quattro paragrafi richiamano alcuni risultati di base dell'inferenza basata sulla verosimiglianza, utili per comprendere meglio le proprietà della verosimiglianza composita.

Il quinto paragrafo è dedicato alla verosimiglianza composita ed alle sue proprietà. In seguito, viene introdotto un caso particolare della verosimiglianza composita, detta verosimiglianza composita marginale, perché costruita tramite adeguate densità marginali. Successivamente, vengono definite alcune quantità associate alla verosimiglianza composita come, ad esempio, l'equazione di stima, lo stimatore di massima verosimiglianza

composita. Infine vengono introdotti tre test basati sulla verosimiglianza composta.

## 1.2 Inferenza statistica

L'inferenza statistica è il procedimento attraverso il quale si ottengono informazioni sulle caratteristiche di una popolazione dall'osservazione di un campione, cioè un aggregato di unità selezionato solitamente mediante un esperimento casuale. L'obiettivo, è quello di fornire strumenti e metodi per ricavare dai dati campionari informazioni sulla popolazione e per quantificare la fiducia da accordare a tali informazioni.

## 1.3 Modelli statistici parametrici

L'idea di base, su cui poggia l'inferenza statistica, è che i dati osservati,  $y^{oss}$ , costituiscono una realizzazione di un vettore casuale  $Y$  con legge di probabilità  $P^0$  e associata funzione di probabilità o di densità  $f^0(y)$ . Si indica con  $\mathcal{Y}$  lo spazio campionario di  $y^{oss}$ . Ciò si esprime concisamente con la scrittura  $y^{oss}$  realizzazione di  $Y \sim P^0$  con spazio campionario  $\mathcal{Y}$ . La legge di probabilità  $P^0$  è almeno in parte ignota, e occorre utilizzare l'informazione portata dai dati stessi per ottenere una sua ricostruzione.

La famiglia

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

di funzioni di densità (rispetto ad una fissata misura dominante) indicizzate dal vettore di parametri  $\theta$  appartenente allo spazio parametrico  $\Theta \subseteq \mathbb{R}^p$  verrà chiamata **modello statistico parametrico**.

Se  $f^0 \in \mathcal{F}$ , si dice che il modello statistico è correttamente specificato. Se il modello è correttamente specificato, e se il parametro è identificabile, ossia la corrispondenza fra  $\Theta$  e  $\mathcal{F}$  è biunivoca, si ha  $f^0(y) = f(y; \theta^0)$  per uno e un solo valore  $\theta^0 \in \Theta$ , detto vero valore del parametro. Allora fare inferenza su  $P^0$ , o su  $f^0(\cdot)$ , è equivalente a farla su  $\theta^0$ .

## 1.4 Funzione di verosimiglianza

I metodi di verosimiglianza sono molto importanti per l'inferenza in modelli statistici parametrici. Considerando un modello statistico parametrico per i dati  $y = (y_1, \dots, y_n)$  riferiti ad un campione casuale da  $Y = (Y_1, \dots, Y_n)$ , con funzione di densità congiunta  $f_Y(y; \theta)$  e  $\theta \in \Theta \subseteq \mathbb{R}^p$  parametro non noto, la funzione  $L : \Theta \rightarrow \mathbb{R}^+$ , definita da

$$L(\theta) = L(\theta; y) = f_Y(y; \theta),$$

è detta **funzione di verosimiglianza** (*likelihood*) di  $\theta$  basata sui dati  $y$ .

Nel caso di osservazioni indipendenti, si ha

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

dove  $f_{Y_i}(y_i; \theta)$ ,  $i = 1, \dots, n$ , rappresenta la densità marginale della variabile casuale  $Y_i$ . Spesso viene utilizzato la trasformata logaritmica di  $L(\theta)$

$$l(\theta) = \log L(\theta; y) = \sum_{i=1}^n \log f_{Y_i}(y_i; \theta),$$

chiamata funzione di log-verosimiglianza.

Frequentemente, nei modelli statistici considerati nelle applicazioni,  $\Theta$  è un sottoinsieme aperto di  $\mathbb{R}^p$  e  $l(\theta)$  è una funzione differenziabile almeno tre volte, con derivate parziali continue in  $\Theta$ . In tali casi si parla di modelli parametrici con verosimiglianza regolare.

Il metodo della massima verosimiglianza ricerca il valore più verosimile di  $\theta$ , ossia ricerca, all'interno dello spazio  $\Theta$  di tutti i possibili valori di  $\theta$ , il valore, o i valori,  $\hat{\theta}$ , del parametro che massimizzano la probabilità, o densità di probabilità di riottenere un campione uguale al campione dato, in una ipotetica replicazione dell'esperimento. Questo valore, se unico, viene chiamato **stima di massima verosimiglianza** di  $\theta$  e, formalmente,

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} l(\theta; y).$$

Non è detto che tale massimo esista o che sia unico.

Il vettore di derivate parziali prime della funzione di log-verosimiglianza,

$$l_*(\theta) = \left[ \frac{\partial l(\theta; y)}{\partial \theta_r} \right] = [l_r(\theta; y)], r = 1, \dots, p,$$

è detto **funzione punteggio** o **funzione score** (*score function*).

Se il modello ha verosimiglianza regolare, spesso la stima di massima verosimiglianza si individua come unica soluzione dell'**equazione di verosimiglianza**

$$l_*(\theta) = 0,$$

dettata dalle condizioni del primo ordine per un massimo locale.

### 1.4.1 Informazione osservata

La matrice di dimensione  $p \times p$  delle derivate parziali seconde di  $l(\theta)$  cambiate di segno,

$$j(\theta) = -l_{**}(\theta) = \left[ -\frac{\partial^2 l(\theta; y)}{\partial \theta_r \partial \theta_s} \right] = [-l_{rs}(\theta; y)],$$

è detta **matrice di informazione osservata**, dove  $-\frac{\partial^2 l(\theta; y)}{\partial \theta_r \partial \theta_s}$  rappresenta l'elemento di posto  $(r, s)$  della matrice  $j$ . A livello interpretativo si può dire che l'informazione osservata calcolata nella stima di massima verosimiglianza,  $j(\hat{\theta})$ , che non dipende dal parametro ma dal campione, è una misura dell'informazione portata dal campione per il dato esperimento.

### 1.4.2 Informazione attesa

Si dice **informazione attesa** o **informazione di Fisher** la quantità

$$i(\theta) = E_\theta \{j(\theta)\} = \left[ -E_\theta \left( \frac{\partial^2 l(\theta; Y)}{\partial \theta_r \partial \theta_s} \right) \right],$$

ossia il valore atteso dell'informazione osservata. È una matrice di dimensione  $p \times p$ , dove  $-E_\theta \left( \frac{\partial^2 l(\theta; Y)}{\partial \theta_r \partial \theta_s} \right)$  rappresenta l'elemento di posto  $(r, s)$ .

Sotto ulteriori condizioni di regolarità, l'informazione di Fisher può essere calcolata tramite la funzione punteggio, ovvero

$$i(\theta) = \text{Var}_\theta \{l_*(\theta)\} = E_\theta \{l_*(\theta; Y)l_*(\theta; Y)^T\}. \quad (1.2)$$

Sotto campionamento casuale semplice di numerosità  $n$ , l'informazione attesa è proporzionale a  $n$ ,

$$i(\theta) = ni_1(\theta),$$

dove  $i_1(\theta)$  è l'informazione attesa per una singola osservazione.

Sotto certe condizioni di regolarità, lo stimatore di massima verosimiglianza è consistente e ha distribuzione asintotica  $\hat{\theta} \sim N(\theta, i(\theta)^{-1})$ .

## 1.5 Verosimiglianza composta

### 1.5.1 Introduzione

La verosimiglianza composta è costruita da una combinazione di validi oggetti di verosimiglianza, di solito legati a piccoli sottoinsiemi di dati. Il vantaggio della verosimiglianza composta è quello di ridurre la complessità computazionale in modo che sia possibile far fronte a grandi insiemi di dati e modelli complessi che presentano difficoltà a causa di interdipendenze di elevate dimensioni, che non rendono possibile l'utilizzo della verosimiglianza standard. Un altro aspetto vantaggioso della verosimiglianza composta è che si può fare inferenza sui parametri di interesse senza fare assunzioni sulla distribuzione congiunta dei dati.

### 1.5.2 Definizioni e proprietà

Sia  $Y = (Y_1, \dots, Y_n)$  un vettore di variabili casuali di dimensione  $n$  con densità congiunta  $f_Y(y; \theta)$  per un parametro incognito  $\theta \in \Theta \subseteq \mathbb{R}^p$ . Si suppone che  $f_Y(y; \theta)$  sia difficile da valutare, o da specificare, ma che la

definizione della verosimiglianza per alcuni sottoinsiemi di dati sia possibile. Può quindi essere conveniente considerare pseudo-verosimiglianze ottenute combinando, in modo opportuno, tali oggetti di verosimiglianza. Quest'idea risale probabilmente a Besag (1974) ed è stata definita **verosimiglianza composta** da Lindsay (1988).

Si considera un modello statistico parametrico (1.1). Date le osservazioni  $y = (y_1, \dots, y_n)$ , la verosimiglianza composta è definita tramite  $K$  eventi marginali o condizionati  $A_k(y)$  su  $\mathcal{Y}$ ,  $k = 1, \dots, K$ , con contributi di verosimiglianza dati da  $L_k(\theta; y) = L(\theta; A_k(y))$ . Allora, la verosimiglianza composta, ottenuta combinando queste singole verosimiglianze, è definita come

$$CL(\theta; y) = \prod_{k=1}^K L_k(\theta; y)^{w_k},$$

dove  $w_k, k = 1, \dots, K$ , sono pesi positivi. Si indica con

$$Cl(\theta; y) = \log CL(\theta; y),$$

la funzione di log-verosimiglianza composta.

Questa classe contiene e quindi generalizza, l'usuale verosimiglianza ordinaria, insieme a tante altre interessanti alternative. Una possibile classificazione delle verosimiglianze composte (Varin e Vidoni, 2005) prende in considerazione i seguenti due principali gruppi :

- **Metodo di omissione**

In questo gruppo, la pseudo-verosimiglianza si ottiene togliendo alcuni termini che complicano la valutazione della verosimiglianza completa. L'auspicio è che la parte omessa non sia molto informativa sul parametro di interesse e quindi la perdita di efficienza sia tollerabile. Alcuni esempi comprendono sia le pseudo-verosimiglianze 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. Anche la famosa verosimiglianza parziale (Cox, 1975) può essere inclusa in questa classe.

- **Verosimiglianza composita marginale**

In questo gruppo, la verosimiglianza composita viene costruita combinando verosimiglianze marginali (tramite distribuzioni congiunte di coppie, triplete,...). L'interesse è focalizzato sulle composizioni di distribuzioni marginali di piccole dimensioni, dato l'importante risparmio computazionale.

Un primo esempio riguarda le pseudo-verosimiglianze costruite utilizzando, le sole marginali univariate, come sotto l'ipotesi di indipendenza. Si ottiene quindi

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

che viene denominata verosimiglianza di indipendenza (*independence likelihood*); si veda ad esempio, Chandler e Bate (2007).

Un secondo esempio è quello in cui la verosimiglianza composita viene costruita utilizzando marginali bivariate. Si ottiene quindi

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

che viene denominata verosimiglianza a coppie (*pairwise likelihood*). Si veda ad esempio Le Cessie e Van Houwelingen (1994). Si indica con

$$Pl(\theta; y) = \log PL(\theta; y),$$

la funzione di log-verosimiglianza a coppie.

In alcune circostanze sarebbe utile considerare dei sottogruppi più grandi come triplete d'osservazioni (Varin e Vidoni, 2005) o combinare verosimiglianza di indipendenza e a coppie in un modo ottimale (Cox e Reid, 2004).

L'inferenza basata sulla verosimiglianza composita, può essere giustificata dal fatto che il valore atteso della log-verosimiglianza composita è massimo in  $\theta^0$ , ossia,

$$E_{\theta^0}[Cl(\theta; Y)] \leq E_{\theta^0}[Cl(\theta^0; Y)],$$

per ogni  $\theta \neq \theta^0$ . Infatti

$$\begin{aligned} E_{\theta^0} \left[ \sum_{k=1}^K w_k \log L_k(\theta; Y) \right] &= \sum_{k=1}^K w_k E_{\theta^0} [\log L_k(\theta; Y)] \\ &= \sum_{k=1}^K w_k E_{\theta^0} [\log L(\theta; A_k(Y))] \\ &< \sum_{k=1}^K w_k E_{\theta^0} [\log L(\theta^0; A_k(Y))], \end{aligned}$$

per ogni  $\theta \neq \theta^0$  in base alla disuguaglianza di Wald che vale per ogni singolo addendo della sommatoria (si veda ad esempio Pace e Salvan, 2001, §5.4.1).

### 1.5.3 Quantità associate alla verosimiglianza composta

La funzione punteggio basata sulla verosimiglianza composta, si ottiene calcolando la derivata prima della log-verosimiglianza composta

$$Cl_*(\theta; y) = \frac{\partial Cl(\theta; y)}{\partial \theta} = \sum_{k=1}^K w_k \frac{\partial \log L_k(\theta; y)}{\partial \theta}.$$

L'equazione di stima basata sulla verosimiglianza composta, detta equazione di verosimiglianza composta è data da,

$$Cl_*(\theta; y) = 0.$$

Lo stimatore di massima verosimiglianza composta è la soluzione, se esiste ed è unica, dell'equazione di verosimiglianza composta.

Essendo ogni singola componente della verosimiglianza composta una verosimiglianza propria, risulta che la funzione punteggio composta è una combinazione lineare di funzioni di punteggio associate ad ogni singolo termine della verosimiglianza composta. Quindi l'equazione di stima ad essa associata risulta non distorta e sotto le usuali condizioni di regolarità, lo stimatore  $\hat{\theta}_C$  che massimizza  $Cl(\theta; y)$  è consistente e asintoticamente normale

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



dove

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

con  $H(\theta) = E_{\theta} \left\{ -\frac{\partial Cl_*(\theta; Y)}{\partial \theta} \right\}$  e  $J(\theta) = \text{Var}_{\theta} \{Cl_*(\theta; Y)\}$  nota come **informazione di Godambe** (Godambe, 1960) o **informazione sandwich**.

La forma dell'informazione di Godambe è dovuta al fatto che la verosimiglianza composita non soddisfa un'identità analoga alla (1.2) valida per la verosimiglianza completa in modelli regolari. Il fatto che  $H(\theta)$  non coincida con  $J(\theta)$  può comportare in genere una perdita di efficienza asintotica di  $\hat{\theta}_C$  rispetto a  $\hat{\theta}$  del modello completo.

#### 1.5.4 Verifica delle ipotesi

Anche nel caso della verosimiglianza composita, per quanto riguarda la verifica delle ipotesi, esiste una opportuna procedura per verificare la conformità dei dati a un sottomodello  $\mathcal{F}_0$  di  $\mathcal{F}$ , dove  $\mathcal{F}$  è supposto correttamente specificato.

Sebbene i test di Wald e *score* basati sulla verosimiglianza composita siano semplici da ricavare, il loro uso può tuttavia essere discutibile a causa della ben nota mancanza di invarianza di riparametrizzazione (test di Wald) e possibile instabilità numerica (*score test*). Per queste ragioni, un analogo del test del rapporto di verosimiglianza basato sulla verosimiglianza composita può essere più interessante.

Si suppone di voler saggiare l'ipotesi nulla

$$H_0 : \theta = \theta_0,$$

dove  $\theta$  è un vettore  $p$ -dimensionale. Allora il test del **rapporto di verosimiglianza composita** risulta

$$W(\theta_0) = 2 \left\{ Cl(\hat{\theta}_C; y) - Cl(\theta_0; y) \right\},$$

In questo caso la distribuzione asintotica di  $W(\theta_0)$  non ha una forma standard. In particolare, la sua distribuzione asintotica nulla è una combi-

nazione lineare di variabili casuali  $\chi_1^2$  indipendenti, ossia,

$$W(\theta_0) \sim \sum_{i=1}^p \lambda_i(\theta_0) Z_i^2$$

dove  $Z_i^2$  sono variabili casuali  $\chi_1^2$  indipendenti e  $\lambda_i(\theta_0)$ ,  $i = 1, \dots, p$ , sono gli autovalori della matrice  $J(\theta_0)^{-1}H(\theta_0)$ ; si veda Foutz e Srivastava (1997) oppure Kent (1982).

Il test di Wald corrispondente è dato da

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

mentre il test score è

$$W_u(\theta_0) = Cl_*(\theta_0; y)^T J(\theta_0) Cl_*(\theta_0; y).$$

Sia il test di Wald che il test score hanno l'usale distribuzione asintotica  $\chi_p^2$ .

Il risultato si generalizza al test del rapporto di verosimiglianza composta su un sottoinsieme di componenti di  $\theta$  (si veda ad esempio, Varin, 2008, pagina 6).

# Capitolo 2

## Verosimiglianza a coppie ed efficienza

### 2.1 Introduzione

Sono presenti in letteratura vari esempi di applicazione della verosimiglianza a coppie, si veda ad esempio Lele and Taper (2002) o Varin (2008) per una rassegna recente.

In questo capitolo, saranno presentate due applicazioni della verosimiglianza a coppie. Particolare attenzione sarà posta sull'efficienza attraverso il confronto tra i risultati ottenuti con la verosimiglianza a coppie e quelli ottenuti con il modello completo.

In particolare, nel secondo paragrafo tratteremo il primo esempio considerato da Cox e Reid (2004), ripercorrendone in dettaglio tutti i passaggi. Si tratta di un'applicazione della verosimiglianza a coppie per un modello normale multivariato monoparametrico. L'obiettivo è valutare l'efficienza, confrontando i risultati con quelli ottenuti attraverso la verosimiglianza completa.

Nel terzo paragrafo viene trattato un esempio nuovo, in cui si applica la verosimiglianza a coppie ad un modello normale multivariato con tre parametri. Si sono ottenuti inoltre alcuni risultati analitici di rilevanza,

così sintetizzabili:

- l'informazione di Godambe e l'informazione di Fisher coincidono;
- vi è una relazione lineare tra la funzione *score* basata sulla verosimiglianza a coppie e la funzione *score* basata su quella completa, con matrice della trasformazione dipendente da  $\theta$ , ma non da  $y$ .

Nel capitolo seguente verranno ripresi gli stessi esempi, considerando le distribuzioni di vari test basati sulla verosimiglianza a coppie.

## 2.2 Un esempio con parametro scalare: normale simmetrica

Questo primo esempio è stato già trattato in Cox e Reid (2004). Vengono qui ricostruiti i vari passaggi, e aggiunte alcune spiegazioni. L'interesse è focalizzato sulla perdita di efficienza dovuta all'impiego della verosimiglianza a coppie per la stima di un parametro incognito rispetto alla verosimiglianza completa.

Sia  $y_i = (y_{i1}, \dots, y_{iq})$  realizzazione di una variabile casuale normale  $q$ -dimensionale  $Y_i = (Y_{i1}, \dots, Y_{iq})$  con vettore delle medie nullo, varianze unitarie e matrice di covarianza  $R$ , ovvero  $Y_i \sim N_q(0, R)$ , dove  $\text{Cor}_\rho(Y_{ir}, Y_{is}) = \rho$  per  $r \neq s$ . Inoltre, si ha che  $|R| > 0$  per  $-\frac{1}{q-1} < \rho < 1$ , condizione necessaria affinché  $R$  sia una matrice di covarianza. Dunque

$$R = (1 - \rho) \left( I_q + \frac{\rho}{1 - \rho} \mathbf{1}_q \mathbf{1}_q^T \right) = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix},$$

con  $\mathbf{1}_q = (1, \dots, 1)^T$ , vettore  $q$ -dimensionale e  $I_q$  che rappresenta la matrice identità di ordine  $q$ . Per facilitare il calcolo dell'inversa e del determinante della matrice  $R$ , si può riscrivere l'elemento di posto  $(r, s)$ ,  $r, s = 1, \dots, q$ ,

nel seguente modo

$$R_{r,s} = (1 - \rho) \left( \delta_{rs} + \frac{\rho}{1 - \rho} h_{rs} \right),$$

dove  $h_{rs} = 1$  per  $r, s = 1, \dots, q$  e  $\delta_{rs} = \begin{cases} 1 & \text{se } r = s \\ 0 & \text{se } r \neq s \end{cases}$ .

L'elemento  $(r, s)$  di  $R^{-1}$  risulta

$$R^{rs} = \frac{1}{1 - \rho} \left( \delta_{rs} - \frac{\frac{\rho}{1 - \rho}}{1 + \frac{\rho}{1 - \rho}} \right) = \frac{1}{1 - \rho} \left( \delta_{rs} - \frac{\rho}{1 + \rho(q - 1)} \right)$$

e quindi

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

Inoltre il determinante di  $R$  si può scrivere come

$$|R| = (1 - \rho)^q \left( 1 + \frac{\rho}{1 - \rho} \right) = (1 - \rho)^{q-1} \{1 + \rho(q - 1)\}.$$

### 2.2.1 Verosimiglianza completa

Considerando una singola osservazione, la funzione di densità della distribuzione normale  $q$ -dimensionale risulta

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

Sviluppando l'espressione  $y_i^T R^{-1} y_i$ , si ottiene

$$y_i^T R^{-1} y_i = \frac{1}{1 - \rho} \left( y_i^T y_i - \frac{\rho}{1 + \rho(q - 1)} y_i^T \mathbf{1}_q \mathbf{1}_q^T y_i \right).$$

Si nota che  $y_i^T y_i = \sum_{r=1}^q y_{ir}^2$  e  $y_i^T \mathbf{1}_q = \mathbf{1}_q^T y_i = q \bar{y}_i = \sum_{r=1}^q y_{ir}$ . Di conseguenza, la funzione di densità diventa

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

La verosimiglianza completa calcolata per  $n$  osservazioni indipendenti risulta

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

e la log-verosimiglianza completa è

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

L'espressione  $\sum_{i=1}^n \sum_{r=1}^q y_{ir}^2$  si può riscrivere nel modo seguente

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

dove  $SSW = \sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \bar{y}_i)^2$  e  $SSB = \sum_{i=1}^n y_i^2$  con  $y_i = \sum_{r=1}^q y_{ir}$ , sono componenti della statistica sufficiente. Sostituendo le varie quantità ricavate, si ottiene infine

$$l(\rho) = -\frac{n(q-1)}{2} \log(1-\rho) - \frac{n}{2} \log\{1+\rho(q-1)\} - \frac{1}{2(1-\rho)} SSW \\ - \frac{1}{2q\{1+\rho(q-1)\}} SSB.$$

La funzione di punteggio risulta

$$l_*(\rho) = \frac{dl(\rho)}{d\rho} \\ = \frac{n(q-1)}{2} \frac{1}{(1-\rho)} - \frac{n(q-1)}{2} \frac{1}{\{1+\rho(q-1)\}} \\ - \frac{1}{2} \frac{1}{(1-\rho)^2} SSW + \frac{(q-1)}{2q} \frac{1}{\{1+\rho(q-1)\}^2} SSB.$$

e l'informazione osservata è

$$j(\rho) = -\frac{d^2l(\rho)}{d\rho^2} \\ = -\frac{n(q-1)}{2} \frac{1}{(1-\rho)^2} - \frac{n(q-1)^2}{2} \frac{1}{\{1+\rho(q-1)\}^2} \\ + \frac{1}{(1-\rho)^3} SSW + \frac{(q-1)^2}{q} \frac{1}{\{1+\rho(q-1)\}^3} SSB.$$

Per calcolare l'informazione attesa  $i(\rho) = E_\rho\{j(\rho)\}$ , è necessario conoscere i valori attesi di SSW e SSB.

Si ha che,  $SSW = \sum_{i=1}^n \sum_{r=1}^q (Y_{ir} - \bar{Y}_i)^2 = \sum_{i=1}^n Q_i$ , dove

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

$Y_i \sim N_q(0, R)$ . Poichè la matrice  $R \left( I_q - \frac{1}{q} 1_q 1_q^T \right)$  non risulta idempotente, si opera una trasformazione su  $Y_i$ , ponendo  $Z_i = (1 - \rho)^{-\frac{1}{2}} Y_i$ .

Allora  $Z_i \sim N_q(0, \Sigma)$ , con  $\Sigma = I_q + \frac{\rho}{1-\rho} 1_q 1_q^T$ . Di conseguenza,

$$Y_i^T \left( I_q - \frac{1}{q} 1_q 1_q^T \right) Y_i = (1 - \rho) Z_i^T \left( I_q - \frac{1}{q} 1_q 1_q^T \right) Z_i.$$

Si nota ora che la matrice  $\Sigma \left( I_q - \frac{1}{q} 1_q 1_q^T \right)$  è idempotente, quindi secondo il teorema 8.6 di Severini (2005), si ha

$Z_i^T \left( I_q - \frac{1}{q} 1_q 1_q^T \right) Z_i \sim \chi_{tr\{\Sigma(I_q - \frac{1}{q} 1_q 1_q^T)\}}^2$ , con  $tr\left\{\Sigma \left( I_q - \frac{1}{q} 1_q 1_q^T \right)\right\} = q - 1$ . Si deduce che

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

Quindi la distribuzione di  $SSW$  risulta

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

Per quanto riguarda la distribuzione di  $SSB = \sum_{i=1}^n Y_i^2$ , si ha che

$$\begin{aligned} Y_i &= \sum_{r=1}^q Y_{ir} = 1_q^T Y_i \sim N(0, 1_q^T R 1_q) \\ &\sim N(0, q \{1 + \rho(q - 1)\}). \end{aligned}$$

Quindi  $SSB$  può essere riscritto in questo modo

$$\begin{aligned} SSB &= q \{1 + \rho(q - 1)\} \sum_{i=1}^n \left[ [q \{1 + \rho(q - 1)\}]^{-\frac{1}{2}} Y_i \right]^2 \\ &= q \{1 + \rho(q - 1)\} \sum_{i=1}^n Z_i^2, \end{aligned}$$

dove  $Z_i \sim N(0, 1)$ . Infine si ottiene che

$$SSB \sim q \{1 + \rho(q - 1)\} \chi_n^2.$$

Si può verificare che le variabili  $SSW$  e  $SSB$  sono indipendenti.

Infatti, visto che gli  $n$  campioni sono indipendenti e  $SSW$  è funzione solo delle quantità  $\sum_{r=1}^q (Y_{ir} - \bar{Y}_i)^2$  e  $SSB$  solo delle  $Y_i$ , dimostrare che  $SSW$  e  $SSB$  sono indipendenti, corrisponde a dimostrare che  $\sum_{r=1}^q (Y_{ir} - \bar{Y}_i)^2$



e  $Y_i$  sono indipendenti. Posto  $B = I_q - \frac{1}{q}1_q1_q^T$ , attraverso il risultato (vi-ii) a pagina 188 di Rao (1973), la condizione necessaria e sufficiente per l'indipendenza è che

$$RBR1_q = 0.$$

Ora

$$\begin{aligned} RBR1_q &= (1 - \rho) \left( I_q + \frac{\rho}{1 - \rho} 1_q 1_q^T \right) \left( I_q - \frac{1}{q} 1_q 1_q^T \right) \\ &\quad (1 - \rho) \left( I_q + \frac{\rho}{1 - \rho} 1_q 1_q^T \right) 1_q \\ &= (1 - \rho)^2 \left\{ I_q - \frac{1}{q} 1_q 1_q^T + \frac{\rho}{1 - \rho} \left( 1_q 1_q^T - \frac{1}{q} 1_q 1_q^T \right) \right\} \\ &\quad \left( 1_q + \frac{q\rho}{1 - \rho} 1_q \right) \\ &= (1 - \rho)^2 \left( I_q - \frac{1}{q} 1_q 1_q^T \right) \left( 1_q + \frac{q\rho}{1 - \rho} 1_q \right) \\ &= (1 - \rho)^2 \left\{ 1_q + \frac{q\rho}{1 - \rho} 1_q - \frac{1}{q} \left( q 1_q + \frac{q^2 \rho}{1 - \rho} 1_q \right) \right\} \\ &= (1 - \rho)^2 \left( 1_q + \frac{q\rho}{1 - \rho} 1_q - 1_q - \frac{q\rho}{1 - \rho} 1_q \right) \\ &= 0. \end{aligned}$$

Quindi, l'informazione di Fisher è

$$\begin{aligned} i(\rho) &= E_\rho \{ j(\rho) \} \\ &= -\frac{n(q-1)}{2} \frac{1}{(1-\rho)^2} - \frac{n(q-1)^2}{2} \frac{1}{\{1 + \rho(q-1)\}^2} \\ &\quad + \frac{1}{(1-\rho)^3} E_\rho(SSW) + \frac{(q-1)^2}{q} \frac{1}{\{1 + \rho(q-1)\}^3} E_\rho(SSB), \end{aligned}$$

dato che  $E_\rho(SSW) = n(1 - \rho)(q - 1)$  e  $E_\rho(SSB) = nq \{1 + \rho(q - 1)\}$ , si ottiene che

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

## 2.2.2 Verosimiglianza a coppie

Per quanto riguarda la verosimiglianza a coppie, questa si basa sulle distribuzioni marginali bivariate, ovvero

$$\begin{pmatrix} Y_{ir} \\ Y_{is} \end{pmatrix} \sim N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

Per una sola osservazione la verosimiglianza a coppie combina le

$$\binom{q}{2} = \frac{q!}{2!(q-2)!} = \frac{q(q-1)}{2}$$

coppie possibili e risulta pari a

$$PL(\rho) = \prod_{r=1}^{q-1} \prod_{s=r+1}^q \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} (y_{ir}^2 + y_{is}^2 - 2\rho y_{ir}y_{is}) \right\},$$

quindi, date  $n$  osservazioni, la log-verosimiglianza a coppie risulta

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

Si nota che

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

Sostituendo quest'ultima forma nella verosimiglianza e semplificando, la log-verosimiglianza a coppie risulta pari a

$$Pl(\rho) = -\frac{nq(q-1)}{4} \log(1-\rho^2) - \frac{q-1+\rho}{2(1-\rho^2)} SSW - \frac{(q-1)}{2q(1+\rho)} SSB,$$

e la funzione punteggio risulta

$$\begin{aligned} Pl_*(\rho) &= \frac{dPl(\rho)}{d\rho} \\ &= \frac{nq(q-1)\rho}{2(1-\rho^2)} - \frac{1+\rho^2+2(q-1)\rho}{2(1-\rho^2)^2}SSW + \frac{(q-1)}{2q(1+\rho)^2}SSB. \end{aligned}$$

Le quantità necessarie per il calcolo dell'informazione di Godambe sono:

$$\begin{aligned} \frac{d^2Pl(\rho)}{d\rho^2} &= \frac{1}{(1-\rho^2)^3} \left\{ \frac{nq(q-1)}{2}(1-\rho^2)(1+\rho^2) - [(q+\rho-1)(1-\rho^2) \right. \\ &\quad \left. + 2\rho(1+\rho^2+2(q-1)\rho)]SSW - \frac{(q-1)}{q}(1-\rho)^3SSB \right\}, \end{aligned}$$

$$\begin{aligned} H(\rho) &= E_\rho \left\{ -\frac{d^2Pl(\rho)}{d\rho^2} \right\} \\ &= -\frac{1}{(1-\rho^2)^3} \left\{ \frac{nq(q-1)}{2}(1-\rho^4) - [(q+\rho-1)(1-\rho^2) \right. \\ &\quad \left. + 2\rho(1+\rho^2+2(q-1)\rho)]E_\rho(SSW) - \frac{(q-1)}{q}(1-\rho)^3E_\rho(SSB) \right\}. \end{aligned}$$

Infine, dopo un po' di algebra, si ottiene

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

La quantità  $J(\rho)$  invece risulta pari a

$$\begin{aligned} J(\rho) &= \text{Var}_\rho \{Pl_*(\rho)\} \\ &= \left\{ \frac{(1+\rho^2+2(q-1)\rho)}{2(1-\rho^2)^2} \right\}^2 \text{Var}_\rho(SSW) \\ &\quad + \left( \frac{q-1}{2q(1+\rho)^2} \right)^2 \text{Var}_\rho(SSB). \end{aligned}$$

Dato che  $\text{Var}_\rho(SSW) = (1-\rho)^2 2n(q-1)$ ,  $\text{Var}_\rho(SSB) = [q\{1+\rho(q-1)\}]^2 2n$  e tenuto conto dell'indipendenza di  $SSW$  e  $SSB$ . Dopo vari passaggi si ottiene

$$J(\rho) = \frac{nq(q-1)}{2(1+\rho)^4(1-\rho)^2} c(q, \rho),$$

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

L'informazione di Godambe risulta quindi pari a

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

### 2.2.3 Efficienza asintotica

L'efficienza asintotica dello stimatore di massima verosimiglianza a coppie viene ottenuto dal rapporto tra le due varianze asintotiche basate, rispettivamente, sull'informazione di Fisher e Godambe,

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

Questo rapporto vale 1 per  $q = 2$ , in quanto la verosimiglianza a coppie coincide con quella standard. Risulta anche uguale a 1 anche se  $\rho = 0$  o  $\rho = 1$ . Il grafico in Figura 2.1 illustra la perdita di informazione all'aumentare di  $q$ . Gli stimatori  $\hat{\rho}$  e  $\hat{\rho}_P$  sono ottenuti rispettivamente con la verosimiglianza completa e la verosimiglianza a coppie.

## 2.3 Un esempio con parametro multidimensionale: modello ad effetti casuali

Il secondo esempio che si considera è un'estensione del precedente (si veda ad esempio, Pace e Salvan, 1996, pagina 485) e riguarda sempre osservazioni normali multivariate equicorrelate. Ora, tuttavia, si considerano ignote e uguali tra di loro anche le medie e le varianze delle distribuzioni marginali univariate. Il parametro è dunque tridimensionale, con  $\theta = (\mu, \sigma^2, \rho)$ .

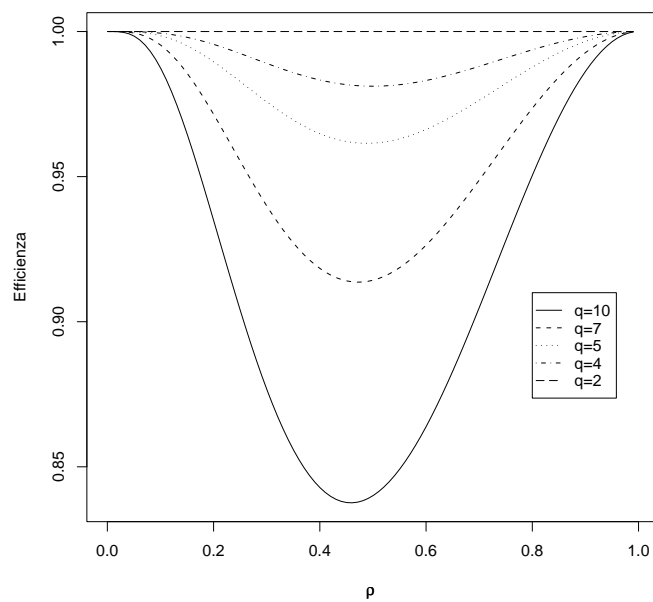


Figura 2.1: Rapporto asintotico delle varianze di  $\hat{\rho}$  e  $\hat{\rho}_P$  in funzione di  $\rho$  per  $q = 2, 4, 5, 7, 10$ .

Considerando una singola osservazione, si assume che  $Y_i \sim N_q(\mu 1_q, \Sigma)$ , dove  $\mu \in \mathbb{R}$ . Indicando con  $\Sigma_{r,s}$  l'elemento di posto  $(r, s)$  della matrice  $\Sigma$ , con

$$\Sigma_{r,s} = \sigma^2 \rho (1 - \rho) \left( \delta_{rs} + \frac{\rho}{1 - \rho} h_{rs} \right),$$

$$\text{dove } \delta_{rs} = \begin{cases} 1 & \text{se } r = s \\ 0 & \text{se } r \neq s \end{cases}$$

e  $h_{rs} = 1$ . Allora si ha che

$$\begin{aligned} \Sigma_{r,r} &= \sigma^2 (1 - \rho) \left( 1 + \frac{\rho}{1 - \rho} \right) \\ &= \sigma^2 (1 - \rho) \frac{1 - \rho + \rho}{(1 - \rho)} \\ &= \sigma^2 \\ \Sigma_{r,s} &= \sigma^2 (1 - \rho) \frac{\rho}{(1 - \rho)} \\ &= \rho \sigma^2. \end{aligned}$$

Quindi  $\text{Cor}_\rho(Y_{ir}, Y_{is}) = \rho$ , con  $-\frac{1}{q-1} < \rho < 1$  in quanto questa è condizione necessaria affinché  $|\Sigma| > 0$ , ossia che  $\Sigma$  sia una matrice di covarianza.

Un'applicazione importante si ha nell'ambito dell'analisi della varianza con effetti casuali, dove usualmente si assume  $Y_{ir} = \mu + a_i + e_{ir}$  ( $r = 1, \dots, q$ ,  $i = 1, \dots, n$ ), con l'effetto  $a_i$  realizzazione di una normale  $N(0, \sigma_a^2)$ , l'errore  $e_{ir}$  realizzazione di una normale  $N(0, \sigma_e^2)$  e indipendenza fra tali realizzazioni. Il modello statistico considerato è del tipo sopra considerato con  $\sigma^2 = \sigma_a^2 + \sigma_e^2$  e  $\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$ .

### 2.3.1 Verosimiglianza completa

Posto  $y_i = (y_{i1}, \dots, y_{iq})$ , la funzione di verosimiglianza per  $n$  osservazioni risulta

$$L(\mu, \sigma^2, \rho) = \prod_{i=1}^n \frac{1}{(2\pi\sigma^2)^{\frac{q}{2}} (1 - \rho)^{\frac{q-1}{2}} \{1 + \rho(q-1)\}^{\frac{1}{2}}} \exp\left(-\frac{1}{2\sigma^2(1-\rho)} \left[ \sum_{r=1}^q (y_{ir} - \mu)^2 - \frac{\rho q^2}{1 + \rho(q-1)} (\bar{y}_i - \mu)^2 \right] \right)$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{nq}{2}} (1-\rho)^{\frac{n(q-1)}{2}} \{1+\rho(q-1)\}^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2(1-\rho)} \sum_{i=1}^n \left[ \sum_{r=1}^q (y_{ir} - \mu)^2 - \frac{\rho q^2}{1+\rho(q-1)} (\bar{y}_i - \mu)^2 \right]\right),$$

dove, analogamente al caso precedente,  $\bar{y}_i = \sum_{r=1}^q y_{ir}/q$ .

La log-verosimiglianza risulta

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

Ponendo  $SSW = \sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \bar{y}_i)^2$  e  $SSB = \sum_{i=1}^n (\bar{y}_i - \bar{y})^2$ , si ottiene

$$\begin{aligned} l(\mu, \sigma^2, \rho) = & -\frac{1}{2\sigma^2(1-\rho)}SSW - \frac{q}{2\sigma^2\{1+\rho(q-1)\}}(SSB + n\bar{y}^2) \\ & + \frac{nq\mu}{\sigma^2\{1+\rho(q-1)\}}\bar{y} - \frac{nq\mu^2}{2\sigma^2\{1+\rho(q-1)\}} - \frac{nq}{2}\log\sigma^2 \\ & - \frac{n(q-1)}{2}\log(1-\rho) - \frac{n}{2}\log\{1+\rho(q-1)\}, \end{aligned}$$

dove  $\bar{y} = \frac{\sum_{i=1}^n \sum_{r=1}^q y_{ir}}{nq}$  e avendo sfruttato le identità

$$\begin{aligned} \sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \mu)^2 &= \sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \bar{y}_i)^2 + q \sum_{i=1}^n (\bar{y}_i - \mu)^2 \\ \sum_{i=1}^n (\bar{y}_i - \mu)^2 &= \sum_{i=1}^n (\bar{y}_i - \bar{y})^2 + n(\bar{y} - \mu)^2. \end{aligned}$$

La funzione punteggio ha componenti

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \mu} &= \frac{nq}{\sigma^2\{1+\rho(q-1)\}}\bar{y} - \frac{nq\mu}{\sigma^2\{1+\rho(q-1)\}} \\ \frac{\partial l(\theta)}{\partial \sigma^2} &= \frac{1}{2(\sigma^2)^2(1-\rho)}SSW + \frac{q}{2(\sigma^2)^2\{1+\rho(q-1)\}}(SSB + n\bar{y}^2) \\ &\quad - \frac{nq\mu}{(\sigma^2)^2\{1+\rho(q-1)\}}\bar{y} + \frac{nq\mu^2}{2(\sigma^2)^2\{1+\rho(q-1)\}} - \frac{nq}{2\sigma^2} \\ \frac{\partial l(\theta)}{\partial \rho} &= -\frac{1}{2\sigma^2(1-\rho)^2}SSW + \frac{q(q-1)}{2\sigma^2\{1+\rho(q-1)\}^2}(SSB + n\bar{y}^2) \\ &\quad - \frac{nq(q-1)\mu}{\sigma^2\{1+\rho(q-1)\}^2}\bar{y} + \frac{nq(q-1)\mu^2}{2\sigma^2\{1+\rho(q-1)\}^2} \\ &\quad + \frac{n(q-1)}{2(1-\rho)} - \frac{n(q-1)}{2\{1+\rho(q-1)\}}. \end{aligned}$$

La matrice d'informazione osservata

$$j(\mu, \sigma^2, \rho) = \begin{pmatrix} \dot{j}_{\mu\mu} & \dot{j}_{\mu\sigma^2} & \dot{j}_{\mu\rho} \\ \dot{j}_{\mu\sigma^2} & \dot{j}_{\sigma^2\sigma^2} & \dot{j}_{\sigma^2\rho} \\ \dot{j}_{\mu\rho} & \dot{j}_{\sigma^2\rho} & \dot{j}_{\rho\rho} \end{pmatrix}$$



è formata dai seguenti elementi

$$\begin{aligned}
j_{\mu\mu} &= -\frac{\partial^2 l(\theta)}{\partial \mu^2} = \frac{nq}{\sigma^2 \{1 + \rho(q-1)\}} \\
j_{\mu\sigma^2} &= -\frac{\partial^2 l(\theta)}{\partial \mu \partial \sigma^2} = \frac{nq}{(\sigma^2)^2 \{1 + \rho(q-1)\}} \bar{y} - \frac{nq\mu}{(\sigma^2)^2 \{1 + \rho(q-1)\}} \\
j_{\mu\rho} &= -\frac{\partial^2 l(\theta)}{\partial \mu \partial \rho} = \frac{nq(q-1)}{\sigma^2 \{1 + \rho(q-1)\}^2} \bar{y} - \frac{nq(q-1)\mu}{\sigma^2 \{1 + \rho(q-1)\}^2} \\
j_{\sigma^2\sigma^2} &= -\frac{\partial^2 l(\theta)}{\partial (\sigma^2)^2} = \frac{1}{(\sigma^2)^3 (1-\rho)} SSW + \frac{q}{(\sigma^2)^3 \{1 + \rho(q-1)\}} (SSB + n\bar{y}^2) \\
&\quad - \frac{2nq\mu}{(\sigma^2)^3 \{1 + \rho(q-1)\}} \bar{y} + \frac{nq\mu^2}{(\sigma^2)^3 \{1 + \rho(q-1)\}} - \frac{nq}{2(\sigma^2)^2} \\
j_{\sigma^2\rho} &= -\frac{\partial^2 l(\theta)}{\partial \sigma^2 \partial \rho} \\
&= -\frac{1}{2(\sigma^2)^2 (1-\rho)^2} SSW + \frac{q(q-1)}{2(\sigma^2)^2 \{1 + \rho(q-1)\}^2} (SSB + n\bar{y}^2) \\
&\quad - \frac{nq(q-1)\mu}{(\sigma^2)^2 \{1 + \rho(q-1)\}^2} \bar{y} + \frac{nq(q-1)\mu^2}{2(\sigma^2)^2 \{1 + \rho(q-1)\}^2} \\
j_{\rho\rho} &= -\frac{\partial^2 l(\theta)}{\partial \rho^2} = \frac{1}{\sigma^2 (1-\rho)^3} SSW + \frac{q(q-1)^2}{\sigma^2 \{1 + \rho(q-1)\}^3} (SSB + n\bar{y}^2) \\
&\quad - \frac{2nq(q-1)^2\mu}{\sigma^2 \{1 + \rho(q-1)\}^3} \bar{y} + \frac{nq(q-1)^2\mu^2}{\sigma^2 \{1 + \rho(q-1)\}^3} \\
&\quad - \frac{n(q-1)}{2(1-\rho)^2} - \frac{n(q-1)^2}{2 \{1 + \rho(q-1)\}^2}.
\end{aligned}$$

Prima di calcolare l'informazione attesa, è opportuno ricavare la distribuzione della statistica sufficiente  $(\bar{Y}, SSW, SSB)$

- Distribuzione marginale di  $\bar{Y}$

$$\bar{Y} = \frac{1}{nq} \sum_{i=1}^n \sum_{r=1}^q Y_{ir} \sim N\left(\mu, \frac{\sigma^2 \{1 + \rho(q-1)\}}{nq}\right), \text{ poiché } \sum_{r=1}^q Y_{ir} = \mathbf{1}_q^T Y_i \sim N(q\mu, q\sigma^2 [1 + \rho(q-1)]).$$

- Distribuzione marginale di  $SSW$

Analogamente all'esempio precedente risulta che  $SSW \sim \sigma^2 (1-\rho) \chi_{n(q-1)}^2$ .

- Distribuzione marginale di  $SSB$

Si ha che  $\bar{Y}_i = q^{-1} \sum_{r=1}^q Y_{ir} = q^{-1} 1_q^T Y_i \sim N(\mu, q^{-1} \sigma^2 [1 + \rho(q-1)])$ , pertanto  $\frac{SSB}{q^{-1} \sigma^2 \{1 + \rho(q-1)\}} \sim \chi_{n-1}^2$ , e quindi  $SSB \sim \frac{\sigma^2 \{1 + \rho(q-1)\}}{q} \chi_{n-1}^2$ .  
Tenendo conto che  $E_\theta(\bar{Y}) = \mu$ ,  $E_\theta(SSW) = \sigma^2(1-\rho)n(q-1)$  e  $E_\theta(SSB + n\bar{Y}^2) = \frac{n\sigma^2}{q} \{1 + \rho(q-1)\} + n\mu^2$  con

$$E_\theta(\bar{Y}^2) = \text{Var}_\theta(\bar{Y}) + \mu^2 = \frac{\sigma^2 \{1 + \rho(q-1)\}}{nq} + \mu^2,$$

la matrice di informazione di Fisher

$$i(\mu, \sigma^2, \rho) = E_\theta \{j(\mu, \sigma^2, \rho)\} = \begin{pmatrix} i_{\mu\mu} & i_{\mu\sigma^2} & i_{\mu\rho} \\ i_{\mu\sigma^2} & i_{\sigma^2\sigma^2} & i_{\sigma^2\rho} \\ i_{\mu\rho} & i_{\sigma^2\rho} & i_{\rho\rho} \end{pmatrix} \quad (2.1)$$

è formata dai seguenti elementi

$$\begin{aligned} i_{\mu\mu} &= \frac{nq}{\sigma^2 \{1 + \rho(q-1)\}} \\ i_{\mu\sigma^2} &= \frac{nq}{(\sigma^2)^2 \{1 + \rho(q-1)\}} E_\theta(\bar{Y}) - \frac{nq\mu}{(\sigma^2)^2 \{1 + \rho(q-1)\}} \\ &= \frac{nq}{(\sigma^2)^2 \{1 + \rho(q-1)\}} \mu - \frac{nq\mu}{(\sigma^2)^2 \{1 + \rho(q-1)\}} = 0 \\ i_{\mu\rho} &= \frac{nq(q-1)}{\sigma^2 \{1 + \rho(q-1)\}^2} E_\theta(\bar{Y}) - \frac{nq(q-1)\mu}{\sigma^2 \{1 + \rho(q-1)\}^2} \\ &= \frac{nq(q-1)}{\sigma^2 \{1 + \rho(q-1)\}^2} \mu - \frac{nq(q-1)\mu}{\sigma^2 \{1 + \rho(q-1)\}^2} = 0 \\ i_{\sigma^2\sigma^2} &= \frac{1}{(\sigma^2)^3(1-\rho)} E_\theta(SSW) + \frac{q}{(\sigma^2)^3 \{1 + \rho(q-1)\}} E_\theta[(SSB + n\bar{y}^2)] \\ &\quad - \frac{2nq\mu}{(\sigma^2)^3 \{1 + \rho(q-1)\}} E_\theta(\bar{Y}) + \frac{nq\mu^2}{(\sigma^2)^3 \{1 + \rho(q-1)\}} - \frac{nq}{2(\sigma^2)^2} \\ &= \frac{1}{(\sigma^2)^3(1-\rho)} \sigma^2(1-\rho)n(q-1) + \frac{q}{(\sigma^2)^3 \{1 + \rho(q-1)\}} \\ &\quad \left( \frac{n\sigma^2}{q} \{1 + \rho(q-1)\} + n\mu^2 \right) - \frac{2nq\mu}{(\sigma^2)^3 \{1 + \rho(q-1)\}} \mu \\ &\quad + \frac{nq\mu^2}{(\sigma^2)^3 \{1 + \rho(q-1)\}} - \frac{nq}{2(\sigma^2)^2} \\ &= \frac{nq}{2(\sigma^2)^2} \end{aligned}$$

$$\begin{aligned}
i_{\sigma^2\rho} &= -\frac{1}{2(\sigma^2)^2(1-\rho)^2}E_{\theta}(SSW) + \frac{q(q-1)}{2(\sigma^2)^2\{1+\rho(q-1)\}^2}E_{\theta}[(SSB+n\bar{y}^2)] \\
&\quad - \frac{nq(q-1)\mu}{(\sigma^2)^2\{1+\rho(q-1)\}^2}E_{\theta}(\bar{Y}) + \frac{nq(q-1)\mu^2}{2(\sigma^2)^2\{1+\rho(q-1)\}^2} \\
&= -\frac{1}{2(\sigma^2)^2(1-\rho)^2}\sigma^2(1-\rho)n(q-1) + \frac{q(q-1)}{2(\sigma^2)^2\{1+\rho(q-1)\}^2} \\
&\quad \left(\frac{n\sigma^2}{q}\{1+\rho(q-1)\} + n\mu^2\right) - \frac{nq(q-1)\mu}{(\sigma^2)^2\{1+\rho(q-1)\}^2}\mu \\
&\quad + \frac{nq(q-1)\mu^2}{2(\sigma^2)^2\{1+\rho(q-1)\}^2} \\
&= -\frac{nq(q-1)\rho}{2\sigma^2(1-\rho)\{1+\rho(q-1)\}} \\
i_{\rho\rho} &= \frac{1}{\sigma^2(1-\rho)^3}E_{\theta}(SSW) + \frac{q(q-1)^2}{\sigma^2\{1+\rho(q-1)\}^3}E_{\theta}[(SSB+n\bar{y}^2)] \\
&\quad - \frac{2nq(q-1)^2\mu}{\sigma^2\{1+\rho(q-1)\}^3}E_{\theta}(\bar{Y}) + \frac{nq(q-1)^2\mu^2}{\sigma^2\{1+\rho(q-1)\}^3} \\
&\quad - \frac{n(q-1)}{2(1-\rho)^2} - \frac{n(q-1)^2}{2\{1+\rho(q-1)\}^2} \\
&= \frac{1}{\sigma^2(1-\rho)^3}\sigma^2(1-\rho)n(q-1) + \frac{q(q-1)^2}{\sigma^2\{1+\rho(q-1)\}^3} \\
&\quad \left(\frac{n\sigma^2}{q}\{1+\rho(q-1)\} + n\mu^2\right) - \frac{2nq(q-1)^2\mu}{\sigma^2\{1+\rho(q-1)\}^3}\mu \\
&\quad + \frac{nq(q-1)^2\mu^2}{\sigma^2\{1+\rho(q-1)\}^3} - \frac{n(q-1)}{2(1-\rho)^2} \\
&\quad - \frac{n(q-1)^2}{2\{1+\rho(q-1)\}^2} \\
&= \frac{n(q-1)}{2} \left[ \frac{1}{(1-\rho)^2} + \frac{q-1}{\{1+\rho(q-1)\}^2} \right].
\end{aligned}$$

### 2.3.2 Verosimiglianza a coppie

La verosimiglianza a coppie, si basa sulle distribuzioni marginali bivari-  
ate, ovvero

$$\begin{pmatrix} Y_{ir} \\ Y_{is} \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \mu \\ \mu \end{pmatrix}, \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \quad r \neq s.$$

La densità bivariata è

$$p_{Y_{ir}, Y_{is}}(y_{ir}, y_{is}; \theta) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2\sigma^2(1-\rho^2)} \left[ (y_{ir} - \mu)^2 + (y_{is} - \mu)^2 - 2\rho(y_{ir} - \mu)(y_{is} - \mu) \right]\right).$$

La log-verosimiglianza per una singola coppia di osservazioni risulta

$$l(\theta; y_{ir}, y_{is}) = -\log \sigma^2 - \frac{1}{2} \log(1-\rho^2) - \frac{1}{2\sigma^2(1-\rho^2)} \left[ (y_{ir} - \mu)^2 + (y_{is} - \mu)^2 - 2\rho(y_{ir} - \mu)(y_{is} - \mu) \right],$$

e il contributo di  $y_i$  alla log-verosimiglianza a coppie è

$$\begin{aligned} Pl(\theta; y_{ir}, y_{is}) &= \sum_{r=1}^{q-1} \sum_{s=r+1}^q l(\theta; y_{ir}, y_{is}) \\ &= -\frac{q(q-1)}{2} \log \sigma^2 - \frac{q(q-1)}{4} \log(1-\rho^2) - \frac{1}{2\sigma^2(1-\rho^2)} \\ &\quad \sum_{r=1}^{q-1} \sum_{s=r+1}^q \left[ (y_{ir} - \mu)^2 + (y_{is} - \mu)^2 - 2\rho(y_{ir} - \mu)(y_{is} - \mu) \right]. \end{aligned}$$

Si nota che

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

e

$$\sum_{r=1}^{q-1} \sum_{s=r+1}^q (y_{is} - \mu)^2 = \sum_{r=1}^q (r-1)(y_{ir} - \mu)^2.$$

Dunque per  $n$  osservazioni, la log-verosimiglianza a coppie risulta

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

Inoltre, considerando che

$$\begin{aligned}
\sum_{r=1}^q (y_{ir} - \mu) &= \sum_{r=1}^q y_{ir} - q\mu = q\bar{y}_i - q\mu = q(\bar{y}_i - \mu), \\
\sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \mu)^2 &= \sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \bar{y} + \bar{y} - \mu)^2 \\
&= \sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \bar{y})^2 + nq(\bar{y} - \mu)^2 \\
&= \sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \bar{y}_i + \bar{y}_i - \bar{y})^2 + nq(\bar{y} - \mu)^2 \\
&= \sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \bar{y}_i)^2 + \sum_{i=1}^n \sum_{r=1}^q (\bar{y}_i - \bar{y})^2 + nq(\bar{y} - \mu)^2 \\
&= \sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \bar{y}_i)^2 + q \sum_{i=1}^n (\bar{y}_i - \bar{y})^2 + nq(\bar{y} - \mu)^2 \\
&= SSW + qSSB + nq(\bar{y} - \mu)^2, \\
\sum_{i=1}^n \left[ \sum_{r=1}^q (y_{ir} - \mu) \right]^2 &= \sum_{i=1}^n q^2 (\bar{y}_i - \mu)^2 = q^2 \sum_{i=1}^n (\bar{y}_i - \mu)^2 \\
&= q^2 \left[ \sum_{i=1}^n (\bar{y}_i - \bar{y})^2 + n(\bar{y} - \mu)^2 \right] \\
&= q^2 [SSB + n(\bar{y} - \mu)^2],
\end{aligned}$$

dove  $SSW = \sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \bar{y}_i)^2$ , e  $SSB = \sum_{i=1}^n (\bar{y}_i - \bar{y})^2$ , alla fine si può scrivere

$$\begin{aligned}
Pl(\theta; y) &= -\frac{q-1+\rho}{2\sigma^2(1-\rho^2)} SSW - \frac{q(q-1)}{2\sigma^2(1+\rho)} (SSB + n\bar{y}^2) + \frac{nq(q-1)\mu}{\sigma^2(1+\rho)} \bar{y} \\
&\quad - \frac{nq(q-1)\mu^2}{2\sigma^2(1+\rho)} - \frac{nq(q-1)}{2} \log \sigma^2 - \frac{nq(q-1)}{4} \log(1-\rho^2).
\end{aligned}$$

Gli elementi della funzione di punteggio risultano

$$\begin{aligned}
\frac{\partial Pl(\theta)}{\partial \mu} &= \frac{nq(q-1)}{\sigma^2(1-\rho)}(\bar{y} - \mu) \\
\frac{\partial Pl(\theta)}{\partial \sigma^2} &= \frac{q-1+\rho}{2(\sigma^2)^2(1-\rho^2)}SSW + \frac{q(q-1)}{2(\sigma^2)^2(1+\rho)}(SSB + n\bar{y}^2) - \frac{nq(q-1)\mu}{(\sigma^2)^2(1+\rho)}\bar{y} \\
&\quad + \frac{nq(q-1)\mu^2}{2(\sigma^2)^2(1+\rho)} - \frac{nq(q-1)}{2\sigma^2} \\
\frac{\partial Pl(\theta)}{\partial \rho} &= -\frac{\rho^2 + 2\rho(q-1) + 1}{2\sigma^2(1-\rho^2)^2}SSW + \frac{q(q-1)}{2\sigma^2(1+\rho)^2}(SSB + n\bar{y}^2) - \frac{nq(q-1)\mu}{\sigma^2(1+\rho)^2}\bar{y} \\
&\quad + \frac{nq(q-1)\mu^2}{2\sigma^2(1+\rho)^2} + \frac{nq(q-1)\rho}{2(1-\rho^2)}.
\end{aligned} \tag{2.2}$$

La matrice

$$H(\mu, \sigma^2, \rho) = \begin{pmatrix} H_{\mu\mu} & H_{\mu\sigma^2} & H_{\mu\rho} \\ H_{\mu\sigma^2} & H_{\sigma^2\sigma^2} & H_{\sigma^2\rho} \\ H_{\mu\rho} & H_{\sigma^2\rho} & H_{\rho\rho} \end{pmatrix}$$

è formata dai seguenti elementi

$$\begin{aligned}
H_{\mu\mu} &= E\left(-\frac{\partial^2 Pl}{\partial \mu^2}\right) = \frac{nq(q-1)}{\sigma^2(1+\rho)} \\
H_{\mu\sigma^2} &= E\left(-\frac{\partial^2 Pl}{\partial \mu \partial \sigma^2}\right) = \frac{nq(q-1)}{(\sigma^2)^2(1+\rho)}(E(y) - \mu) \\
&= \frac{nq(q-1)}{(\sigma^2)^2(1+\rho)}(\mu - \mu) = 0 \\
H_{\mu\rho} &= E\left(-\frac{\partial^2 Pl}{\partial \mu \partial \rho}\right) = \frac{nq(q-1)}{\sigma^2(1+\rho)^2}(E(y) - \mu) \\
&= \frac{nq(q-1)}{\sigma^2(1+\rho)^2}(\mu - \mu) = 0
\end{aligned}$$

$$\begin{aligned}
H_{\sigma^2\sigma^2} &= E\left(-\frac{\partial^2 Pl}{\partial(\sigma^2)^2}\right) \\
&= \frac{q-1+\rho}{(\sigma^2)^3(1-\rho^2)}E(SSW) + \frac{q(q-1)}{(\sigma^2)^3(1+\rho)}E(SSB+n\bar{y}^2) \\
&\quad - \frac{2nq(q-1)\mu}{(\sigma^2)^3(1+\rho)}E(\bar{y}) + \frac{nq(q-1)\mu^2}{(\sigma^2)^3(1+\rho)} - \frac{nq(q-1)}{2(\sigma^2)^2} \\
&= \frac{q-1+\rho}{(\sigma^2)^3(1-\rho^2)}\sigma^2(1-\rho)n(q-1) + \frac{q(q-1)}{(\sigma^2)^3(1+\rho)} \\
&\quad \left[\frac{n\sigma^2}{q}\{1+\rho(q-1)\} + n\mu^2\right] - \frac{2nq(q-1)\mu}{(\sigma^2)^3(1+\rho)}\mu + \frac{nq(q-1)\mu^2}{(\sigma^2)^3(1+\rho)} \\
&\quad - \frac{nq(q-1)}{2(\sigma^2)^2} \\
&= \frac{nq(q-1)}{2(\sigma^2)^2} \\
H_{\sigma^2\rho} &= E\left(-\frac{\partial^2 Pl}{\partial\sigma^2\partial\rho}\right) \\
&= -\frac{\rho^2+2\rho(q-1)+1}{2(\sigma^2)^2(1-\rho^2)^2}E(SSW) + \frac{q(q-1)}{2(\sigma^2)^2(1+\rho)^2}E(SSB+n\bar{y}^2) \\
&\quad - \frac{nq(q-1)\mu}{(\sigma^2)^2(1+\rho)^2}E(\bar{y}) + \frac{nq(q-1)\mu^2}{2(\sigma^2)^2(1+\rho)^2} \\
&= -\frac{\rho^2+2\rho(q-1)+1}{2(\sigma^2)^2(1-\rho^2)^2}\sigma^2(1-\rho)n(q-1) + \frac{q(q-1)}{2(\sigma^2)^2(1+\rho)^2} \\
&\quad \left[\frac{n\sigma^2}{q}\{1+\rho(q-1)\} + n\mu^2\right] - \frac{nq(q-1)\mu}{(\sigma^2)^2(1+\rho)^2}\mu \\
&\quad + \frac{nq(q-1)\mu^2}{2(\sigma^2)^2(1+\rho)^2} \\
&= -\frac{nq(q-1)\rho}{2\sigma^2(1-\rho^2)} \\
H_{\rho\rho} &= E\left(-\frac{\partial^2 Pl}{\partial\rho^2}\right) \\
&= \frac{2\rho^3-6\rho^2+6\rho^2q+6\rho+2q-2}{2\sigma^2(1-\rho^2)^3}E(SSW) + \frac{q(q-1)}{\sigma^2(1+\rho)^3} \\
&\quad E(SSB+n\bar{y}^2) - \frac{2nq(q-1)\mu}{\sigma^2(1+\rho)^3}E(\bar{y}) + \frac{nq(q-1)\mu^2}{\sigma^2(1+\rho)^3} \\
&\quad - \frac{nq(q-1)(1+\rho^2)}{2(1-\rho^2)^2}
\end{aligned}$$

$$\begin{aligned}
&= \frac{2\rho^3 - 6\rho^2 + 6\rho^2q + 6\rho + 2q - 2}{2\sigma^2(1 - \rho^2)^3} \sigma^2(1 - \rho)n(q - 1) + \frac{q(q - 1)}{\sigma^2(1 + \rho)^3} \\
&\quad \left[ \frac{n\sigma^2}{q} \{1 + \rho(q - 1)\} + n\mu^2 \right] - \frac{2nq(q - 1)\mu}{\sigma^2(1 + \rho)^3} \mu + \frac{nq(q - 1)\mu^2}{\sigma^2(1 + \rho)^3} \\
&\quad - \frac{nq(q - 1)(1 + \rho^2)}{2(1 - \rho^2)^2} \\
&= \frac{nq(q - 1)(1 + \rho^2)}{2(1 + \rho)^2}.
\end{aligned}$$

Per calcolare gli elementi della matrice  $J(\theta)$ , è necessario calcolare i valori attesi e varianze di alcune variabili. Le variabili  $SSW$ ,  $SSB$  e  $\bar{Y}$  sono indipendenti (si veda ad esempio, Davison, 2003, paragrafo 9.4).

- Valori attesi.

La variabile  $\bar{Y}$  si distribuisce come una normale con media  $\mu$  e varianza  $\frac{\sigma^2\{1+\rho(q-1)\}}{nq}$ . Quindi la sua funzione generatrice dei momenti risulta

$$M_{\bar{Y}}(t) = \exp\left(\frac{\sigma^2\{1 + \rho(q - 1)\}}{2nq}t^2 + \mu t\right).$$

Dunque

$$\begin{aligned}
E_{\theta}[\bar{Y}^2] &= M_{\bar{Y}}''(0) \\
&= \frac{\sigma^2\{1 + \rho(q - 1)\}}{nq} + \mu^2 \\
E_{\theta}[\bar{Y}^3] &= M_{\bar{Y}}'''(0) \\
&= \frac{3\mu\sigma^2\{1 + \rho(q - 1)\}}{nq} + \mu^3 \\
E_{\theta}[\bar{Y}^4] &= M_{\bar{Y}}''''(0) \\
&= \mu^4 + \frac{6\mu^2\sigma^2\{1 + \rho(q - 1)\}}{nq} + 3\left[\frac{\sigma^2\{1 + \rho(q - 1)\}}{nq}\right]^2
\end{aligned}$$

$$\begin{aligned}
E_{\theta}[SSW^2] &= \text{Var}_{\theta}(SSW) + [E_{\theta}(SSW)]^2 \\
&= 2n(q - 1)[\sigma^2(1 - \rho)]^2 + [n\sigma^2(q - 1)(1 - \rho)]^2 \\
&= 2n(q - 1)[\sigma^2(1 - \rho)]^2 \{2 + n(q - 1)\}
\end{aligned}$$



$$\begin{aligned}
E_\theta[SSB^2] &= \text{Var}_\theta(SSB) + [E_\theta(SSB)]^2 \\
&= 2(n-1) \left[ \frac{\sigma^2 \{1 + \rho(q-1)\}}{q} \right]^2 + \left[ \frac{(n-1)\sigma^2 \{1 + \rho(q-1)\}}{q} \right]^2 \\
&= (n^2 - 1) \left[ \frac{\sigma^2 \{1 + \rho(q-1)\}}{q} \right]^2 \\
E_\theta[(SSB + n\bar{Y}^2)^2] &= E_\theta[(SSB^2)^2 + 2nE_\theta[SSB]E_\theta[\bar{Y}^2] + n^2E_\theta[\bar{Y}^4] \\
&= (n^2 + 2n) \left[ \frac{\sigma^2 \{1 + \rho(q-1)\}}{q} \right]^2 + \frac{2n(n+2)\mu^2\sigma^2 \{1 + \rho(q-1)\}}{q} \\
&\quad + n^2\mu^4 \\
E_\theta[n\bar{Y}^2(\bar{y} - \mu)] &= \frac{2\mu\sigma^2 \{1 + \rho(q-1)\}}{q} \\
E_\theta[\bar{Y}(\bar{Y} - \mu)] &= \frac{\sigma^2 \{1 + \rho(q-1)\}}{nq} \\
E_\theta[\bar{Y}(SSB + n\bar{Y})] &= \frac{(n-1)\mu\sigma^2 \{1 + \rho(q-1)\}}{q} + \frac{3\mu\sigma^2 \{1 + \rho(q-1)\}}{q} + n\mu^3
\end{aligned}$$

- Varianze.

$$\begin{aligned}
\text{Var}_\theta(\bar{Y}^2) &= E[\bar{Y}^4] - (E[\bar{Y}^2])^2 \\
&= \frac{4\mu^2\sigma^2 \{1 + \rho(q-1)\}}{nq} + 2 \left[ \frac{\sigma^2 \{1 + \rho(q-1)\}}{nq} \right]^2 \\
\text{Var}_\theta(SSB + n\bar{Y}^2) &= \text{Var}_\theta(SSB) + n^2\text{Var}_\theta(\bar{Y}^2) \\
&= \frac{2n\sigma^2 \{1 + \rho(q-1)\}}{q} \left[ \frac{\sigma^2 \{1 + \rho(q-1)\}}{q} + 2\mu^2 \right] \\
\text{Cov}_\theta((SSB + n\bar{Y}^2), \bar{Y}) &= E_\theta(SSB\bar{Y} + n\bar{Y}^3) - E_\theta(SSB + n\bar{Y}^2)E_\theta(\bar{Y}) \\
&= \frac{2\mu\sigma^2 \{1 + \rho(q-1)\}}{q}.
\end{aligned}$$

La matrice

$$J(\mu, \sigma^2, \rho) = \begin{pmatrix} J_{\mu\mu} & J_{\mu\sigma^2} & J_{\mu\rho} \\ J_{\mu\sigma^2} & J_{\sigma^2\sigma^2} & J_{\sigma^2\rho} \\ J_{\mu\rho} & J_{\sigma^2\rho} & J_{\rho\rho} \end{pmatrix}$$

è formata dai seguenti elementi

$$\begin{aligned}
J_{\mu\mu} &= \text{Var}_\theta \left( \frac{\partial Pl}{\partial \mu} \right) \\
&= \left[ \frac{nq(q-1)}{\sigma^2(1+\rho)} \right]^2 \text{Var}_\theta(\bar{Y}) \\
&= \frac{nq(q-1)^2 \{1 + \rho(q-1)\}}{\sigma^2(1+\rho)^2} \\
J_{\sigma^2\sigma^2} &= \text{Var}_\theta \left( \frac{\partial Pl}{\partial \sigma^2} \right) \\
&= \left[ \frac{q-1+\rho}{2(\sigma^2)^2(1-\rho^2)} \right]^2 \text{Var}_\theta(SSW) + \left[ \frac{q(q-1)}{2(\sigma^2)^2(1+\rho)} \right]^2 \text{Var}_\theta(SSB + n\bar{Y}^2) \\
&\quad + \left[ \frac{nq(q-1)\mu}{(\sigma^2)^2(1+\rho)} \right]^2 \text{Var}_\theta(\bar{Y}) - 2 \frac{q(q-1)}{2(\sigma^2)^2(1+\rho)} \frac{nq(q-1)\mu}{(\sigma^2)^2(1+\rho)} \\
&\quad \text{Cov}_\theta((SSB + n\bar{Y}^2), \bar{Y}) \\
&= \left[ \frac{q-1+\rho}{2(\sigma^2)^2(1-\rho^2)} \right]^2 2n(q-1)[\sigma^2(1-\rho)]^2 + \left[ \frac{q(q-1)}{2(\sigma^2)^2(1+\rho)} \right]^2 \\
&\quad \frac{2n\sigma^2 \{1 + \rho(q-1)\}}{q} \left[ \frac{\sigma^2 \{1 + \rho(q-1)\}}{q} + 2\mu^2 \right] + \left[ \frac{nq(q-1)\mu}{(\sigma^2)^2(1+\rho)} \right]^2 \\
&\quad \frac{\sigma^2 \{1 + \rho(q-1)\}}{nq} - 2 \frac{q(q-1)}{2(\sigma^2)^2(1+\rho)} \frac{nq(q-1)\mu}{(\sigma^2)^2(1+\rho)} \\
&\quad \frac{2\mu\sigma^2 \{1 + \rho(q-1)\}}{q} \\
&= \frac{n(q-1)}{2(\sigma^2)^2(1+\rho)^2} \left\{ (q-1+\rho)^2 + (q-1) \{1 + \rho(q-1)\}^2 \right\} \\
J_{\rho\rho} &= \text{Var}_\theta \left( \frac{\partial Pl}{\partial \rho} \right) \\
&= \left[ \frac{\rho^2 + 2\rho(q-1) + 1}{2\sigma^2(1-\rho^2)^2} \right]^2 \text{Var}_\theta(SSW) + \left[ \frac{q(q-1)}{2\sigma^2(1+\rho)^2} \right]^2 \text{Var}_\theta(SSB + n\bar{Y}^2) \\
&\quad + \left[ \frac{nq(q-1)\mu}{\sigma^2(1+\rho)^2} \right]^2 \text{Var}_\theta(\bar{Y}) - 2 \frac{q(q-1)}{2\sigma^2(1+\rho)^2} \frac{nq(q-1)\mu}{\sigma^2(1+\rho)^2} \\
&\quad \text{Cov}_\theta((SSB + n\bar{Y}^2), \bar{Y})
\end{aligned}$$

$$\begin{aligned}
&= \left[ \frac{\rho^2 + 2\rho(q-1) + 1}{2\sigma^2(1-\rho^2)^2} \right]^2 2n(q-1)[\sigma^2(1-\rho)]^2 + \left[ \frac{q(q-1)}{2\sigma^2(1+\rho)^2} \right]^2 \\
&\quad \frac{2n\sigma^2 \{1 + \rho(q-1)\}}{q} \left[ \frac{\sigma^2 \{1 + \rho(q-1)\}}{q} + 2\mu^2 \right] + \left[ \frac{nq(q-1)\mu}{\sigma^2(1+\rho)^2} \right]^2 \\
&\quad \frac{\sigma^2 \{1 + \rho(q-1)\}}{nq} - 2 \frac{q(q-1)}{2\sigma^2(1+\rho)^2} \frac{nq(q-1)\mu}{\sigma^2(1+\rho)^2} \\
&\quad \frac{2\mu\sigma^2 \{1 + \rho(q-1)\}}{q} \\
&= \frac{n(q-1)}{2(1+\rho)^4} \left\{ \frac{(\rho^2 + 2\rho(q-1) + 1)^2}{(1-\rho)^2} + (q-1) \{1 + \rho(q-1)\}^2 \right\} \\
J_{\mu\sigma^2} &= E_{\theta} \left( \frac{\partial Pl}{\partial \mu} \frac{\partial Pl}{\partial \sigma^2} \right) \\
&= \frac{nq(q-1)}{\sigma^2(1+\rho)} \frac{q-1+\rho}{2(\sigma^2)^2(1-\rho^2)} E_{\theta}(\bar{Y} - \mu) E_{\theta}(SSW) + \frac{nq(q-1)}{\sigma^2(1+\rho)} \\
&\quad \frac{q(q-1)}{2(\sigma^2)^2(1+\rho)} E_{\theta}[(\bar{Y} - \mu)(SSB + n\bar{Y}^2)] - \frac{nq(q-1)}{\sigma^2(1+\rho)} \\
&\quad \frac{nq(q-1)\mu}{(\sigma^2)^2(1+\rho)} E_{\theta}[(\bar{Y} - \mu)\bar{Y}] \\
&\quad + \frac{nq(q-1)}{\sigma^2(1+\rho)} \frac{nq(q-1)\mu^2}{2(\sigma^2)^2(1+\rho)} E_{\theta}(\bar{Y} - \mu) - \frac{nq(q-1)}{\sigma^2(1+\rho)} \\
&\quad \frac{nq(q-1)}{2\sigma^2} E_{\theta}(\bar{Y} - \mu) \\
&= \frac{nq(q-1)}{\sigma^2(1+\rho)} \frac{q(q-1)}{2(\sigma^2)^2(1+\rho)} E_{\theta}[(\bar{Y} - \mu)(SSB + n\bar{Y}^2)] - \frac{nq(q-1)}{\sigma^2(1+\rho)} \\
&\quad \frac{nq(q-1)\mu}{(\sigma^2)^2(1+\rho)} E_{\theta}[(\bar{Y} - \mu)\bar{Y}] \\
&= \frac{nq(q-1)}{\sigma^2(1+\rho)} \frac{q(q-1)}{2(\sigma^2)^2(1+\rho)} E_{\theta}[(\bar{Y} - \mu)n\bar{Y}^2] - \frac{nq(q-1)}{\sigma^2(1+\rho)} \frac{nq(q-1)\mu}{(\sigma^2)^2(1+\rho)} \\
&\quad E_{\theta}[(\bar{Y} - \mu)\bar{Y}] \\
&= \frac{nq(q-1)}{\sigma^2(1+\rho)} \frac{q(q-1)}{2(\sigma^2)^2(1+\rho)} \frac{2\mu\sigma^2 \{1 + \rho(q-1)\}}{q} - \frac{nq(q-1)}{\sigma^2(1+\rho)} \\
&\quad \frac{nq(q-1)\mu}{(\sigma^2)^2(1+\rho)} \frac{\sigma^2 \{1 + \rho(q-1)\}}{nq} \\
&= 0
\end{aligned}$$

$$\begin{aligned}
J_{\mu\rho} &= E_{\theta} \left( \frac{\partial Pl}{\partial \mu} \frac{\partial Pl}{\partial \rho} \right) \\
&= -\frac{nq(q-1)}{\sigma^2(1+\rho)} \frac{\rho^2 + 2\rho(q-1) + 1}{2\sigma^2(1-\rho^2)^2} E_{\theta}(\bar{Y} - \mu) E_{\theta}(SSW) + \frac{nq(q-1)}{\sigma^2(1+\rho)} \\
&\quad \frac{q(q-1)}{2\sigma^2(1+\rho)^2} E_{\theta}[(\bar{Y} - \mu)(SSB + n\bar{Y}^2)] - \frac{nq(q-1)}{\sigma^2(1+\rho)} \frac{nq(q-1)\mu}{\sigma^2(1+\rho)^2} \\
&\quad E_{\theta}[(\bar{Y} - \mu)\bar{Y}] + \frac{nq(q-1)}{\sigma^2(1+\rho)} \frac{nq(q-1)\mu^2}{2\sigma^2(1+\rho)^2} E_{\theta}(\bar{Y} - \mu) \\
&\quad + \frac{nq(q-1)}{\sigma^2(1+\rho)} \frac{nq(q-1)\rho}{2(1-\rho^2)} E_{\theta}(\bar{Y} - \mu) \\
&= \frac{nq(q-1)}{\sigma^2(1+\rho)} \frac{q(q-1)}{2\sigma^2(1+\rho)^2} E_{\theta}[(\bar{Y} - \mu)(SSB + n\bar{Y}^2)] - \frac{nq(q-1)}{\sigma^2(1+\rho)} \\
&\quad \frac{nq(q-1)\mu}{\sigma^2(1+\rho)^2} E_{\theta}[(\bar{Y} - \mu)\bar{Y}] \\
&= \frac{nq(q-1)}{\sigma^2(1+\rho)} \frac{q(q-1)}{2\sigma^2(1+\rho)^2} \frac{2\mu\sigma^2 \{1 + \rho(q-1)\}}{q} - \frac{nq(q-1)}{\sigma^2(1+\rho)} \\
&\quad \frac{nq(q-1)\mu}{\sigma^2(1+\rho)^2} \frac{\sigma^2 \{1 + \rho(q-1)\}}{nq} \\
&= 0 \\
J_{\sigma^2\rho} &= E_{\theta} \left( \frac{\partial Pl}{\partial \sigma^2} \frac{\partial Pl}{\partial \rho} \right) \\
&= \frac{nq(q-1)\rho}{2(\rho-1)(\rho+1)^3\sigma^2} [q^2\rho^2 - 3q\rho^2 + 3\rho^2 - q^2\rho + 5q\rho - 4\rho + 1].
\end{aligned}$$

### 2.3.3 Efficienza

Mardia, Hughes e Charles (2007) mostrano che per questo esempio, le stime ottenute con la verosimiglianza a coppie sono uguali a quelle ottenute con la verosimiglianza completa, ossia,  $\hat{\theta}_P = \hat{\theta}$ .

Quest'esempio, è un caso particolare in cui non si ha perdita di efficienza utilizzando la verosimiglianza a coppie. In aggiunta, si può verificare analiticamente che l'informazione di Fisher e quella di Godambe coincidono, ossia,  $i(\theta) = G(\theta)$ .

**Dimostrazione.** Per il calcolo di  $G(\theta)$ , è necessario in primo luogo ottenere

l'espressione di  $J(\theta)^{-1}$ . La matrice

$$J(\theta)^{-1} = \begin{pmatrix} J^{\mu\mu} & J^{\mu\sigma^2} & J^{\mu\rho} \\ J^{\mu\sigma^2} & J^{\sigma^2\sigma^2} & J^{\sigma^2\rho} \\ J^{\mu\rho} & J^{\sigma^2\rho} & J^{\rho\rho} \end{pmatrix}$$

è formata dai seguenti elementi

$$J^{\mu\mu} = \frac{(\rho + 1)^2 \sigma^2}{n(q - 1)^2 q (q\rho - \rho + 1)}$$

$$J^{\mu\sigma^2} = 0$$

$$J^{\mu\rho} = 0$$

$$J^{\sigma^2\sigma^2} = \frac{2(q^2\rho^4 - 3q\rho^4 + 3\rho^4 - 2q^2\rho^3 + 8q\rho^3 - 6\rho^3 + q^2\rho^2 - 3q\rho^2 + 4\rho^2 + 2q\rho - 2\rho + 1)(\sigma^2)^2}{n(q - 1)^2 q (q\rho - \rho + 1)^2}$$

$$J^{\sigma^2\rho} = -\frac{2(\rho - 1)\rho(\rho + 1)(q^2\rho^2 - 3q\rho^2 + 3\rho^2 - q^2\rho + 5q\rho - 4\rho + 1)\sigma^2}{n(q - 1)^2 q (q\rho - \rho + 1)^2}$$

$$J^{\rho\rho} = \frac{2(\rho - 1)^2(\rho + 1)^2(q^2\rho^2 - 3q\rho^2 + 3\rho^2 + 2q\rho - 2\rho + q - 1)}{n(q - 1)^2 q (q\rho - \rho + 1)^2}.$$

Dunque, l'informazione di Godambe

$$G(\theta) = H(\theta)J(\theta)^{-1}H(\theta) = \begin{pmatrix} G_{\mu\mu} & G_{\mu\sigma^2} & G_{\mu\rho} \\ G_{\mu\sigma^2} & G_{\sigma^2\sigma^2} & G_{\sigma^2\rho} \\ G_{\mu\rho} & G_{\sigma^2\rho} & G_{\rho\rho} \end{pmatrix}$$

ha elementi

$$G_{\mu\mu} = \frac{nq}{\sigma^2 \{1 + \rho(q - 1)\}}$$

$$G_{\mu\sigma^2} = 0$$

$$G_{\mu\rho} = 0$$

$$G_{\sigma^2\sigma^2} = \frac{nq}{2(\sigma^2)^2}$$

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

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

Si può notare come gli elementi di  $G(\theta)$  coincidono con quelli di  $i(\theta)$  riportati dopo la (2.1). Quindi  $i(\theta) = G(\theta)$ .  $\square$

Le quantità  $J(\theta)^{-1}$  e  $G(\theta)$  sono state ottenute utilizzando il programma di calcolo simbolico Maxima; si veda l'Appendice.

Essendo le stime ottenute con la verosimiglianza a coppie e con quella completa esattamente uguali, si può congetturare che

$$Pl_*(\theta) = A(\theta)l_*(\theta), \quad (2.3)$$

con  $A(\theta)$  matrice definita positiva di dimensione  $3 \times 3$ .

La verifica di tale congettura è interessante perché suggerisce quale possa essere il collegamento tra la verosimiglianza a coppie e quella completa in condizioni di efficienza. Si desidera in primo luogo ottenere l'espressione di  $A(\theta)$  per poi verificare la (2.3).

Se vale l'equazione (2.3) allora

$$\begin{aligned} \text{Var}_\theta \{Pl_*(\theta)\} &= A(\theta)\text{Var}_\theta \{l_*(\theta)\} A(\theta)^T \\ &= A(\theta)i(\theta)A(\theta)^T = J(\theta). \end{aligned} \quad (2.4)$$

Avendo dimostrato che  $i(\theta) = G(\theta) = H(\theta)J(\theta)^{-1}H(\theta)$ , si ricava

$$J(\theta) = H(\theta)i(\theta)^{-1}H(\theta). \quad (2.5)$$

Mettendo insieme l'equazione (2.4) e (2.5), si ottiene

$$H(\theta)i(\theta)^{-1}H(\theta) = A(\theta)i(\theta)A(\theta)^T. \quad (2.6)$$

Da quest'ultima si ottiene

$$\begin{aligned} A(\theta) &= H(\theta)i(\theta)^{-1} \\ &= H(\theta)G(\theta)^{-1} \\ &= H(\theta)H(\theta)^{-1}J(\theta)H(\theta)^{-1} \\ &= J(\theta)H(\theta)^{-1}. \end{aligned}$$

Prima di ricavare gli elementi della matrice  $A(\theta)$ , è necessario calcolare  $i(\theta)^{-1}$  oppure  $H(\theta)^{-1}$ .

La matrice

$$i(\theta)^{-1} = \begin{pmatrix} i^{\mu\mu} & i^{\mu\sigma^2} & i^{\mu\rho} \\ i^{\mu\sigma^2} & i^{\sigma^2\sigma^2} & i^{\sigma^2\rho} \\ i^{\mu\rho} & i^{\sigma^2\rho} & i^{\rho\rho} \end{pmatrix}$$

è formata dai seguenti elementi

$$\begin{aligned} i^{\mu\mu} &= \frac{(q\rho - \rho + 1)\sigma^2}{nq} \\ i^{\mu\sigma^2} &= 0 \\ i^{\mu\rho} &= 0 \\ i^{\sigma^2\sigma^2} &= \frac{2(q\rho^2 - \rho^2 + 1)(\sigma^2)^2}{nq} \\ i^{\sigma^2\rho} &= -\frac{2(\rho - 1)\rho(q\rho - \rho + 1)\sigma^2}{nq} \\ i^{\rho\rho} &= \frac{2(\rho - 1)^2(q\rho - \rho + 1)^2}{nq(q - 1)}. \end{aligned}$$

Inoltre la matrice

$$H(\theta)^{-1} = \begin{pmatrix} H^{\mu\mu} & H^{\mu\sigma^2} & H^{\mu\rho} \\ H^{\mu\sigma^2} & H^{\sigma^2\sigma^2} & H^{\sigma^2\rho} \\ H^{\mu\rho} & H^{\sigma^2\rho} & H^{\rho\rho} \end{pmatrix}$$

ha elementi

$$\begin{aligned} H^{\mu\mu} &= \frac{(\rho + 1)\sigma^2}{n(q - 1)q} \\ H^{\mu\sigma^2} &= 0 \\ H^{\mu\rho} &= 0 \\ H^{\sigma^2\sigma^2} &= \frac{2(\rho^2 + 1)(\sigma^2)^2}{n(q - 1)q} \\ H^{\sigma^2\rho} &= -\frac{2(\rho - 1)\rho(\rho + 1)\sigma^2}{n(q - 1)q} \\ H^{\rho\rho} &= \frac{2(\rho - 1)^2(\rho + 1)^2}{n(q - 1)q}. \end{aligned}$$

Infine la matrice

$$A(\theta) = \begin{pmatrix} A_{\mu\mu} & A_{\mu\sigma^2} & A_{\mu\rho} \\ A_{\mu\sigma^2} & A_{\sigma^2\sigma^2} & A_{\sigma^2\rho} \\ A_{\mu\rho} & A_{\sigma^2\rho} & A_{\rho\rho} \end{pmatrix}$$

è formata dai seguenti elementi

$$A_{\mu\mu} = \frac{(q-1)(q\rho - \rho + 1)}{(\rho + 1)}$$

$$A_{\mu\sigma^2} = 0$$

$$A_{\mu\rho} = 0$$

$$A_{\sigma^2\sigma^2} = \frac{(q-1)(q\rho^2 - 2\rho^2 + \rho + 1)}{(\rho + 1)}$$

$$A_{\sigma^2\rho} = -\frac{(q-2)(\rho-1)\rho(q\rho - \rho + 1)}{(\rho + 1)\sigma^2}$$

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

Effettuando il prodotto  $A(\theta)l_*(\theta)$ , si ottiene un vettore di dimensione  $3 \times 1$  di cui gli elementi coincidono con quelli ottenuti in (2.2). Dunque  $Pl_*(\theta) = A(\theta)l_*(\theta)$ . Si è utilizzato il programma Maxima, per il calcolo delle inverse delle matrici ed alcuni prodotti matriciali.

Come conseguenza di quanto esposto sopra, il test di Wald completo e il test *score* coincidano con i corrispondenti test basati sulla verosimiglianza a coppie (cfr. paragrafo 3.4).



# Capitolo 3

## Test basati sulla verosimiglianza a coppie

### 3.1 Introduzione

Questo ultimo capitolo si pone l'obiettivo, di verificare, tramite simulazione, il comportamento distributivo delle statistiche test basate sulla verosimiglianza a coppie e i livelli di copertura effettivi delle corrispondenti regioni di confidenza. Si confronta il comportamento della statistica log-rapporto di verosimiglianza basata sulla verosimiglianza a coppie con quella basata sulla verosimiglianza completa. Viene inoltre prosposta una semplice approssimazione per la distribuzione della statistica log-rapporto di verosimiglianza a coppie alternativa a quella richiamata nel paragrafo 1.5.4. Si focalizzerà l'attenzione sugli stessi esempi trattati nel Capitolo 2. Per il primo esempio, viene analizzata anche la potenza dei rispettivi test, attraverso uno studio di simulazione.

Nel secondo paragrafo vengono costruiti test basati sulla verosimiglianza a coppie per il primo e secondo esempio trattato nel Capitolo 2. Questo aspetto non era stato esplorato da Cox e Reid (2004). Successivamente, si propone la semplice approssimazione per la distribuzione della statistica log-rapporto di verosimiglianza a coppie.

Nel terzo paragrafo, si presentano i risultati ottenuti tramite lo studio di simulazione per il caso della normale simmetrica. In particolare, si valuta il comportamento delle distribuzioni nulle e si stima la potenza dei diversi test.

Nel quarto paragrafo, attraverso uno studio di simulazione, si valuta il comportamento delle distribuzioni nulle dei diversi test relativi all'esempio multiparametrico. Infine, vengono presentati alcuni risultati analitici ottenuti. In particolare, si dimostra che il test di Wald e il test *score* basato sulla verosimiglianza a coppie coincidono con i rispettivi test basati sulla verosimiglianza completa.

## 3.2 Le statistiche test

La distribuzione asintotica nulla del test di rapporto di verosimiglianza a coppie, come richiamato nel paragrafo 1.5.4, non è quella usuale, ma è una combinazione lineare di variabili indipendenti chi-quadrato con un grado di libertà.

Si indicherà con

$$W(\theta) = 2 \left\{ l(\hat{\theta}) - l(\theta) \right\}, \quad (3.1)$$

la statistica test log-rapporto di verosimiglianza completa, con

$$W_e(\theta) = (\hat{\theta} - \theta)^T i(\theta) (\hat{\theta} - \theta), \quad (3.2)$$

il test di Wald basato sulla verosimiglianza completa e con

$$W_u(\theta) = l_*(\theta)^T i(\theta)^{-1} l_*(\theta), \quad (3.3)$$

il test *score* basato sulla verosimiglianza completa. Nel caso monoparametrico, le statistiche definite in (3.1), (3.2) e (3.3), sotto condizioni di regolarità, si approssimano ad un chi-quadrato con un grado di libertà, sotto  $\theta$ . Nel caso multiparametrico, con  $p$  parametri, si approssimano ad un chi-quadrato con  $p$  gradi di libertà, sotto  $\theta$ .

Considerando invece la verosimiglianza a coppie si indicherà con

$$W^P(\theta) = 2 \left\{ Pl(\hat{\theta}_P) - Pl(\theta) \right\},$$

la statistica test log-rapporto di verosimiglianza a coppie. Nel caso multi-parametrico, le sue distribuzioni si approssimano, sotto  $\theta$ , a  $\sum_{i=1}^p \lambda_i(\theta) Z_i^2$ ,  $Z_i \sim N(0, 1)$ , dove  $\lambda_1(\theta), \lambda_2(\theta), \dots, \lambda_p(\theta)$  sono gli autovalori della matrice  $H(\theta)^{-1} J(\theta)$ .

Una approssimazione più semplice può seguire la proposta di Satterthwaite (1946). L'idea è quella di determinare  $c = c(\theta)$  e  $\nu = \nu(\theta)$ , tali che

$$\sum_{i=1}^p \lambda_i(\theta) Z_i^2 \sim c \chi_{\nu}^2.$$

Si ricavano  $c$  e  $\nu$  in modo tale che le due distribuzioni abbiano la medesima media e varianza. Devono quindi essere soddisfatte le relazioni

$$\begin{aligned} E \left[ \sum_{i=1}^p \lambda_i(\theta) Z_i^2 \right] &= \sum_{i=1}^p \lambda_i(\theta) = c\nu \\ \text{Var} \left( \sum_{i=1}^p \lambda_i(\theta) Z_i^2 \right) &= \sum_{i=1}^p \lambda_i(\theta)^2 = 2c^2\nu \end{aligned}$$

Da queste due equazioni, si ricavano

$$\begin{aligned} c &= c(\theta) = \frac{\sum_{i=1}^p \lambda_i(\theta)^2}{\sum_{i=1}^p \lambda_i(\theta)} \\ \nu &= \nu(\theta) = \frac{(\sum_{i=1}^p \lambda_i(\theta))^2}{\sum_{i=1}^p \lambda_i(\theta)^2}. \end{aligned}$$

La distribuzione nulla della statistica riscalata

$$\frac{2 \left\{ Pl(\hat{\theta}_P) - Pl(\theta) \right\}}{c}, \quad (3.4)$$

si approssima con un chi-quadrato con  $\nu$  gradi di libertà, sotto  $\theta$ .

Siano  $\hat{c} = c(\hat{\theta}_P)$  e  $\hat{\nu} = \nu(\hat{\theta}_P)$ , rispettivamente, le stime delle costanti  $c$  e  $\nu$ , calcolate sostituendo a  $\theta$  la sua stima di massima verosimiglianza a coppie  $\hat{\theta}_P$ . Si indica con

$$\hat{W}^P(\theta) = \frac{2 \left\{ Pl(\hat{\theta}_P) - Pl(\theta) \right\}}{\hat{c}}, \quad (3.5)$$

la statistica log-rapporto di verosimiglianza a coppie riscalata con  $\hat{c}$ , con distribuzione approssimata  $\chi_{\hat{v}}^2$ .

Si indicano inoltre con

$$W_e^P(\theta) = (\hat{\theta}_P - \theta)^T G(\theta) (\hat{\theta}_P - \theta) \quad (3.6)$$

e

$$W_u^P(\theta) = Pl_*(\theta)^T J(\theta) Pl_*(\theta), \quad (3.7)$$

le statistiche di tipo Wald e *score* basate sulla verosimiglianza a coppie. Le statistiche definite in (3.6), (3.7) hanno distribuzione nulla approssimata chi-quadrato con  $p$  gradi di libertà.

Una approssimazione alternativa (si veda ad esempio, Varin, 2008, pagina 7) è quella che considera  $\bar{\lambda}(\theta) = \frac{\sum_{i=1}^p \lambda_i(\theta)}{p}$  e una distribuzione approssimata chi-quadrato con  $p$  gradi di libertà per la statistica riscalata

$$\bar{W}^P(\theta) = \frac{2 \left\{ Pl(\hat{\theta}_P) - Pl(\theta) \right\}}{\bar{\lambda}(\theta)}. \quad (3.8)$$

In particolare, nel caso di un solo parametro, le statistiche definite in (3.4), (3.5), (3.6) e (3.7), si approssimano ad un chi-quadrato con un grado di libertà, la costante  $c$ , essendo uguale a  $H(\theta)^{-1} J(\theta)$ .

### 3.3 Distribuzione normale simmetrica: studio di simulazione

Si considera l'esempio trattato nel paragrafo 2.2. Con l'obiettivo di valutare l'adeguatezza delle diverse approssimazioni per le distribuzioni dei sette test (3.1), (3.2), (3.3), (3.4), (3.4), (3.6), (3.7), si sono stimate le distribuzioni nulle tramite simulazione. I quantili dei diversi test sono stati ottenuti effettuando 10000 replicazioni. Su un unico grafico (quantile-quantile), vengono messi a confronto con i rispettivi quantili teorici di un chi-quadrato con un grado di libertà. Successivamente, lo stesso confronto

Tabella 3.1: Livelli di copertura empirici degli intervalli di confidenza di livello 0.90, 0.95, 0.99, per  $\rho = 0.5$ ,  $n = 10$ ,  $q = 3, 8, 15, 20$ , 10000 repliche.

	$q = 3$			$q = 8$			$q = 15$			$q = 20$		
	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$W(\rho)$	0.889	0.943	0.988	0.896	0.947	0.989	0.897	0.949	0.990	0.898	0.949	0.989
$W_e(\rho)$	0.836	0.878	0.932	0.879	0.927	0.973	0.893	0.939	0.981	0.895	0.943	0.984
$W_u(\rho)$	0.907	0.952	0.986	0.901	0.949	0.989	0.898	0.948	0.989	0.902	0.949	0.989
$W^P(\rho)$	0.892	0.945	0.989	0.902	0.950	0.990	0.900	0.950	0.990	0.900	0.951	0.990
$\hat{W}^P(\rho)$	0.882	0.933	0.979	0.892	0.941	0.983	0.892	0.944	0.986	0.896	0.945	0.985
$W_e^P(\rho)$	0.836	0.879	0.933	0.875	0.927	0.973	0.884	0.935	0.979	0.885	0.936	0.980
$W_u^P(\rho)$	0.906	0.950	0.986	0.901	0.948	0.987	0.897	0.948	0.987	0.902	0.951	0.986

avverrà tramite i livelli di copertura rispetto ad un quantile teorico fissato. Per esempio, considerando la statistica test  $W(\rho)$ , il suo livello di copertura con un livello nominale  $1 - \alpha$ , viene stimato da

$$\frac{\#(W(\rho) \leq \chi_{1,1-\alpha}^2)}{N_{\text{rep}}},$$

dove  $N_{\text{rep}}$  rappresenta il numero di repliche nella simulazione effettuata. Ci si aspetta che questo rapporto sia prossimo a  $1 - \alpha$ , se l'approssimazione è buona.

### 3.3.1 Distribuzioni nulle

Dai grafici in Figura 3.1 si può notare che, fissata la numerosità campionaria, le curve che descrivono la distribuzione empirica di tutti i test si avvicinano alla retta continua, all'aumentare di  $q$ . Tale retta evidenzia la situazione in cui i quantili campionari e i quantili teorici coincidono. L'approssimazione dei test alla distribuzione teorica sembra migliorare a partire dal grafico relativo a  $q = 8$ . Se l'obiettivo è la costruzione di regioni di confidenza, risulta essere di maggiore interesse la vicinanza delle curve alla retta in corrispondenza di valori dei quantili non maggiore di 7, valore relativo a più o meno il livello nominale di interesse 0.99.

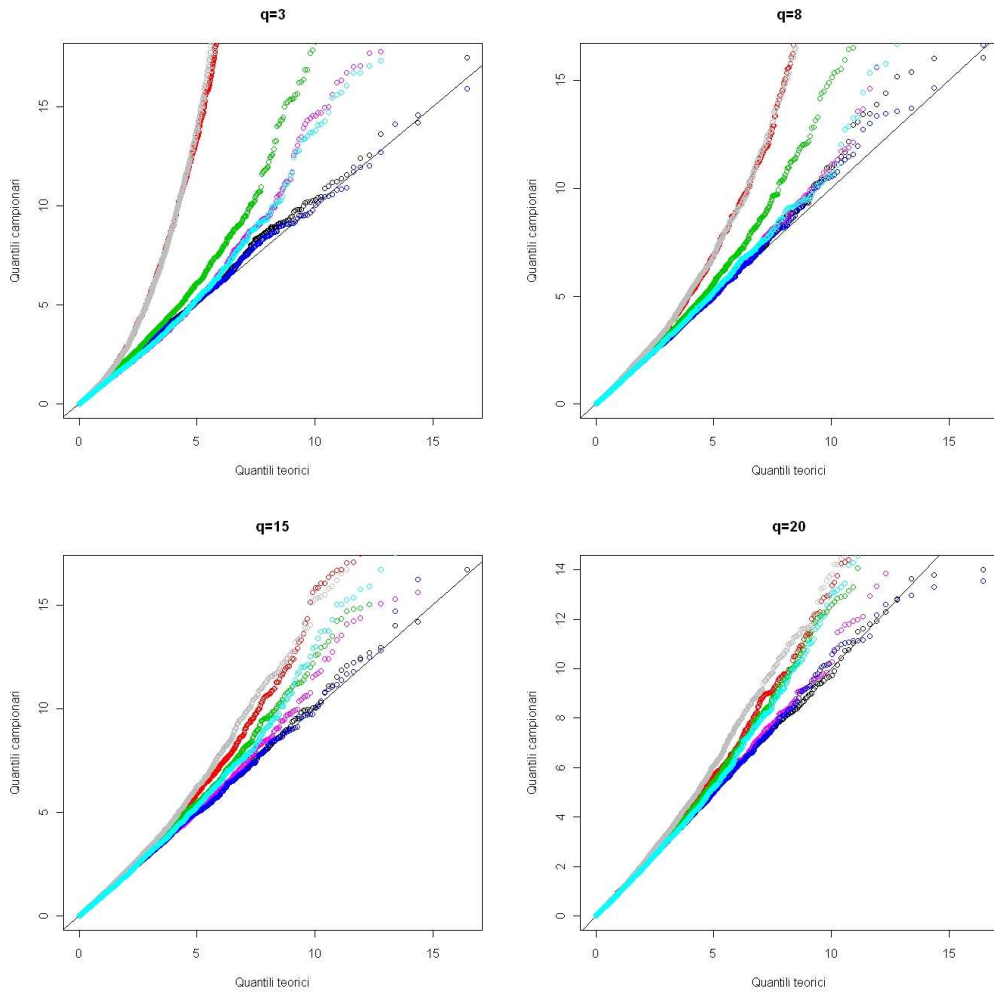


Figura 3.1: Confronto dei quantili empirici con quelli teorici di un chi-quadrato con 1 grado di libertà per  $\rho = 0.5$ ,  $n = 10$ ,  $q = 3, 8, 15, 20$ , 10000 repliche. In particolare  $W(\rho)$  nero,  $W_e(\rho)$  rosso,  $W_u(\rho)$  viola,  $W^P(\rho)$  blu,  $\hat{W}^P(\rho)$  verde,  $W_e^P(\rho)$  grigio,  $W_u^P(\rho)$  celeste.

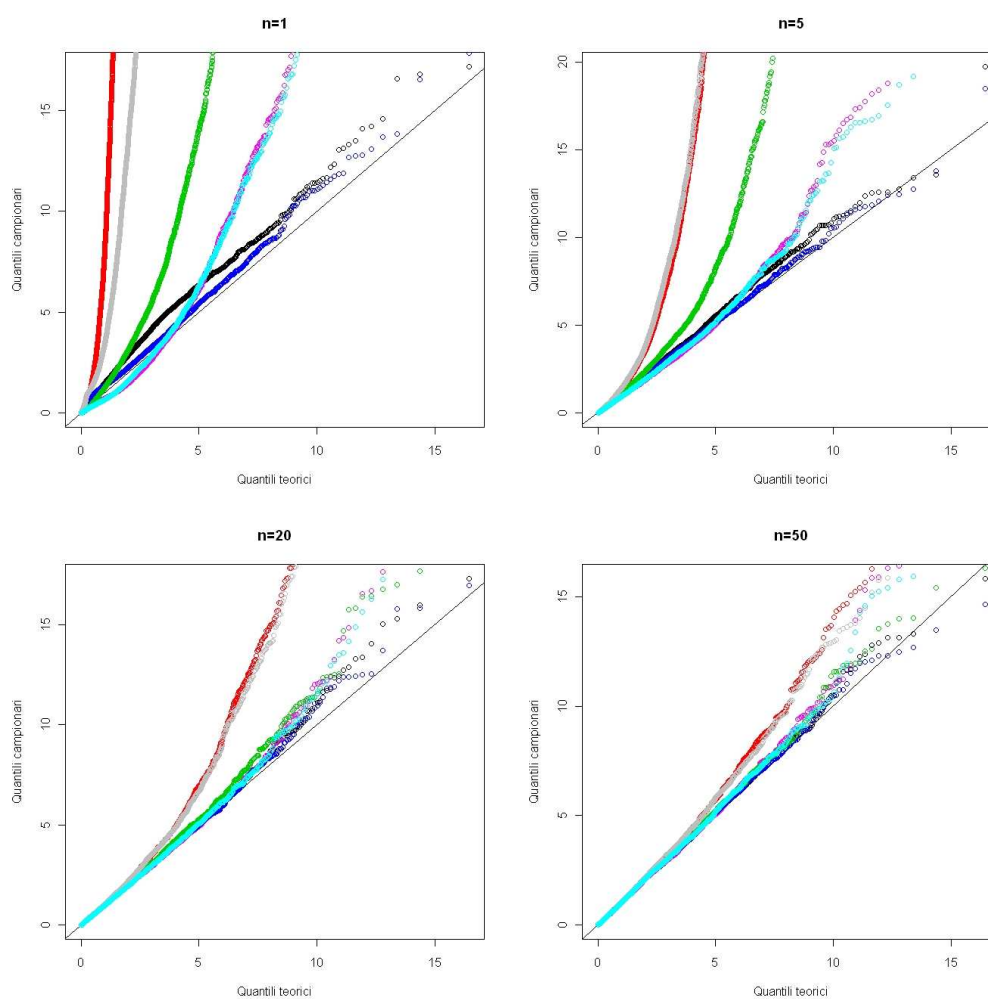


Figura 3.2: Confronto dei quantili empirici con quelli teorici di un chi-quadrato con 1 grado di libertà per  $\rho = 0.5$ ,  $q = 4$ ,  $n = 1, 5, 20, 50, 10000$  replicazioni. In particolare  $W(\rho)$  nero,  $W_e(\rho)$  rosso,  $W_u(\rho)$  viola,  $W^P(\rho)$  blu,  $\hat{W}^P(\rho)$  verde,  $W_e^P(\rho)$  grigio,  $W_u^P(\rho)$  celeste.

Tabella 3.2: Livelli di copertura empirici degli intervalli di confidenza di livello 0.90, 0.95, 0.99, per  $\rho = 0.5$ ,  $q = 4$ ,  $n = 1, 5, 20, 50, 10000$  repliche.

	$n = 1$			$n = 5$			$n = 20$			$n = 50$		
	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$W(\rho)$	0.830	0.902	0.979	0.883	0.940	0.986	0.900	0.948	0.988	0.898	0.949	0.989
$W_e(\rho)$	0.568	0.607	0.665	0.804	0.853	0.903	0.879	0.929	0.972	0.887	0.942	0.982
$W_u(\rho)$	0.918	0.948	0.978	0.906	0.953	0.986	0.904	0.951	0.988	0.896	0.949	0.988
$W^P(\rho)$	0.881	0.941	0.987	0.889	0.945	0.989	0.900	0.949	0.989	0.892	0.949	0.989
$\hat{W}^P(\rho)$	0.820	0.872	0.934	0.864	0.915	0.965	0.896	0.944	0.987	0.890	0.947	0.988
$W_e^P(\rho)$	0.646	0.699	0.768	0.798	0.845	0.896	0.878	0.929	0.974	0.884	0.939	0.983
$W_u^P(\rho)$	0.911	0.946	0.976	0.904	0.950	0.986	0.902	0.949	0.989	0.895	0.948	0.988

Un avvicinamento delle curve alla retta si può notare anche dai grafici in Figura 3.2, che descrivono la stessa situazione, per  $q$  fissato, all'aumentare della numerosità campionaria. In questo caso si può apprezzare la buona approssimazione per tutti i test con numerosità elevate.

In entrambe le figure si può notare un ordinamento nella bontà dell'approssimazione alla distribuzione teorica: le curve relative ai test  $W(\rho)$  e  $W^P(\rho)$  risultano quelle che più si avvicinano alla retta, quelle relative ai test  $W_u(\rho)$  e  $W_u^P(\rho)$  risultano essere leggermente più lontane,  $\hat{W}^P(\rho)$  ha uno scostamento ulteriore rispetto alla retta, mentre  $W_e(\rho)$  e  $W_e^P(\rho)$  riportano l'adattamento peggiore. Tali differenze risultano più marcate per valori bassi di  $q$  e  $n$ .

La bontà dell'approssimazione delle distribuzioni dei test può essere studiata anche tramite i livelli di copertura dei relativi intervalli di confidenza. Sono state stimate le probabilità di copertura per diversi valori nominali  $1 - \alpha$  con  $\alpha = 0.1, 0.05, 0.01$ . Il buon comportamento del test viene confermato se il valore stimato per la probabilità di copertura è prossimo ad  $1 - \alpha$ . I risultati fanno riferimento alla stessa simulazione con la quale sono stati prodotti i grafici delle Figure 3.1 e 3.2. Quindi dapprima si è fissato  $n = 10$  e si è fatto variare  $q = 3, 8, 15, 20$ , e successivamente si è considerato il caso con  $q = 4$  e variando  $n = 1, 5, 20, 50$ . In riferimento alla Tabella 3.1, per valori piccoli di  $q$ , si può notare come i livelli di copertura



Tabella 3.3: Potenza dei test per l'ipotesi nulla  $\rho_0 = 0.2$  contro, rispettivamente,  $\rho_1 = 0.3$  e  $\rho_1 = 0.7$ , per  $n = 10$ ,  $q = 3, 8, 15, 20$ ,  $\alpha = 0.05$ , 10000 repliche.

	$q = 3$		$q = 8$		$q = 15$		$q = 20$	
	$\rho_1 = 0.3$	$\rho_1 = 0.7$	$\rho_1 = 0.3$	$\rho_1 = 0.7$	$\rho_1 = 0.3$	$\rho_1 = 0.7$	$\rho_1 = 0.3$	$\rho_1 = 0.7$
$W(\rho_0)$	0.072	0.867	0.148	1.000	0.249	1.000	0.296	1.000
$W_e(\rho_0)$	0.060	0.832	0.069	1.000	0.133	1.000	0.185	1.000
$W_u(\rho_0)$	0.074	0.659	0.173	0.993	0.257	1.000	0.308	1.000
$W^P(\rho_0)$	0.089	0.886	0.199	1.000	0.274	1.000	0.304	1.000
$\hat{W}^P(\rho_0)$	0.040	0.691	0.041	0.989	0.073	1.000	0.092	1.000
$W_e^P(\rho_0)$	0.054	0.803	0.057	1.000	0.073	1.000	0.090	1.000
$W_u^P(\rho_0)$	0.075	0.647	0.179	0.964	0.248	0.992	0.287	0.997

stimati siano buoni per tutti i test considerati tranne  $W_e(\rho)$  e  $W_e^P(\rho)$ , casi in cui i valori ottenuti per le stime risultano inferiori ai livelli nominali. Lo stesso risultato può essere dedotto dalla Tabella 3.2, per valori piccoli di  $n$ ; in questo caso si nota anche, in aggiunta alla non buona copertura degli intervalli basati su  $W_e(\rho)$  e  $W_e^P(\rho)$ , un cattivo comportamento dell'intervallo di confidenza relativo a  $\hat{W}^P(\rho)$ . Tali risultati sono coerenti con le considerazioni fatte come commento ai grafici in Figure 3.1 e 3.2.

### 3.3.2 Potenza dei test

Con l'obiettivo di confrontare la potenza dei test, si è considerata l'ipotesi nulla  $\rho = 0.2$  contro le alternative  $\rho = 0.3$  e  $\rho = 0.7$ . Questa scelta è stata preferita a quella più usuale  $\rho = 0$  perché, per  $\rho = 0$ , l'informazione di Fisher e quella di Godambe sono uguali, pertanto, anche i corrispondenti test *score*.

Dai risultati prodotti nelle Tabelle 3.3 e 3.4, si può notare come le potenze di tutti i test aumentano all'aumentare sia di  $q$  che di  $n$ . In riferimento alla Tabella 3.3, per valori piccoli di  $q$ , le potenze dei test  $\hat{W}^P(\rho_0)$ ,  $W_u(\rho_0)$ ,  $W_u^P(\rho_0)$  risultano leggermente basse per l'ipotesi alternativa  $\rho_1 = 0.7$ . Dalla Tabella 3.4, per valori piccoli di  $n$ , si può evidenziare che, in aggiunta ai

Tabella 3.4: Potenza dei test per l'ipotesi nulla  $\rho_0 = 0.2$  contro, rispettivamente,  $\rho_1 = 0.3$  e  $\rho_1 = 0.7$ , per  $q = 4$ ,  $n = 1, 5, 20, 50$ ,  $\alpha = 0.05$ , 10000 replicazioni.

	$n = 1$		$n = 5$		$n = 20$		$n = 50$	
	$\rho_1 = 0.3$	$\rho_1 = 0.7$	$\rho_1 = 0.3$	$\rho_1 = 0.7$	$\rho_1 = 0.3$	$\rho_1 = 0.7$	$\rho_1 = 0.3$	$\rho_1 = 0.7$
$W(\rho_0)$	0.048	0.126	0.065	0.730	0.134	1.000	0.308	1.000
$W_e(\rho_0)$	0.041	0.019	0.043	0.488	0.103	1.000	0.285	1.000
$W_u(\rho_0)$	0.055	0.141	0.076	0.509	0.140	0.992	0.301	1.000
$W^P(\rho_0)$	0.063	0.231	0.098	0.799	0.161	1.000	0.330	1.000
$\hat{W}^P(\rho_0)$	0.036	0.041	0.031	0.314	0.084	0.998	0.259	1.000
$W_e^P(\rho_0)$	0.063	0.187	0.044	0.513	0.098	1.000	0.283	1.000
$W_u^P(\rho_0)$	0.058	0.149	0.079	0.495	0.142	0.986	0.302	1.000

test  $\hat{W}^P(\rho_0)$ ,  $W_u(\rho_0)$ ,  $W_u^P(\rho_0)$ , anche i test  $W_e(\rho_0)$  e  $W_e^P(\rho_0)$  hanno potenza leggermente bassa rispetto agli altri.

### 3.4 Modello ad effetti casuali: studio di simulazione

Si considera ora l'esempio del paragrafo 2.3. Diversamente dal caso mono-parametrico, le otto statistiche, (3.1), (3.2), (3.3), (3.4), (3.4), (3.6), (3.7), (3.8), non hanno la stessa distribuzione nulla. Quindi, per mettere in evidenza l'adeguatezza delle approssimazioni delle loro distribuzioni, vengono stimati, tramite 10000 replicazioni, i quantili empirici ad essi associati, e poi, su un unico grafico, vengono confrontati i valori della funzione di ripartizione teorica relativa a tali quantili empirici con i quantili di una distribuzione uniforme in (0,1). Successivamente, lo stesso confronto avviene tramite i livelli di copertura.

Tabella 3.5: Livelli di copertura per  $\theta = (4, 2, 0.7)$ ,  $n = 10$ ,  $q = 3, 8, 15, 20$ ,  $\alpha = 0.1, 0.05, 0.01$ , 10000 repliche.

	$q = 3$			$q = 8$			$q = 15$			$q = 20$		
	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$W(\theta)$	0.884	0.937	0.985	0.882	0.938	0.985	0.880	0.938	0.985	0.875	0.933	0.985
$W_e(\theta) = W_e^P(\theta)$	0.834	0.888	0.946	0.789	0.845	0.905	0.751	0.806	0.872	0.728	0.783	0.851
$W_u(\theta) = W_u^P(\theta)$	0.904	0.941	0.978	0.907	0.946	0.981	0.905	0.946	0.981	0.902	0.945	0.980
$W^P(\theta)$	0.858	0.924	0.984	0.906	0.955	0.992	0.920	0.963	0.993	0.920	0.961	0.993
$\hat{W}^P(\theta)$	0.868	0.929	0.979	0.857	0.913	0.970	0.858	0.912	0.968	0.845	0.902	0.964
$\bar{W}^P(\theta)$	0.879	0.937	0.986	0.876	0.930	0.981	0.874	0.926	0.980	0.867	0.922	0.976

### 3.4.1 Risultati analitici

In base ai risultati del paragrafo 2.3.3, le stime di massima verosimiglianza completa e a coppie coincidono. Inoltre, l'informazione di Fisher  $i(\theta)$  e di Godambe  $G(\theta)$  coincidono. Questi due risultati ci portano a concludere che il test di Wald basato sulla verosimiglianza a coppie è uguale al test di Wald basato sulla verosimiglianza completa. Ossia,

$$W_e(\theta) = W_e^P(\theta). \quad (3.9)$$

Inoltre si è dimostrato anche che  $Pl_*(\theta) = A(\theta)l_*(\theta)$ . Pertanto usando la (2.2) si ha

$$\begin{aligned} W_u^P(\theta) &= Pl_*(\theta)^T J^{-1}(\theta) Pl_*(\theta) \\ &= l_*(\theta)^T A(\theta)^T \{A(\theta)^T\}^{-1} i(\theta)^{-1} A(\theta)^{-1} A(\theta) l_*(\theta) \\ &= l_*(\theta)^T i(\theta)^{-1} l_*(\theta) = W_u(\theta). \end{aligned} \quad (3.10)$$

Si conclude quindi, che il test *score* basato sulla verosimiglianza a coppie e il test *score* basato sulla verosimiglianza completa coincidono.

### 3.4.2 Distribuzioni nulle

A differenza della precedente analisi, i grafici in Figura 3.3 evidenziano scostamenti delle curve dalla retta all'aumentare di  $q$  con numerosità campionaria fissata. In Figura 3.4, tenendo costante  $q$  e aumentando la numerosità campionaria, le curve tendono a coincidere con la retta. I test

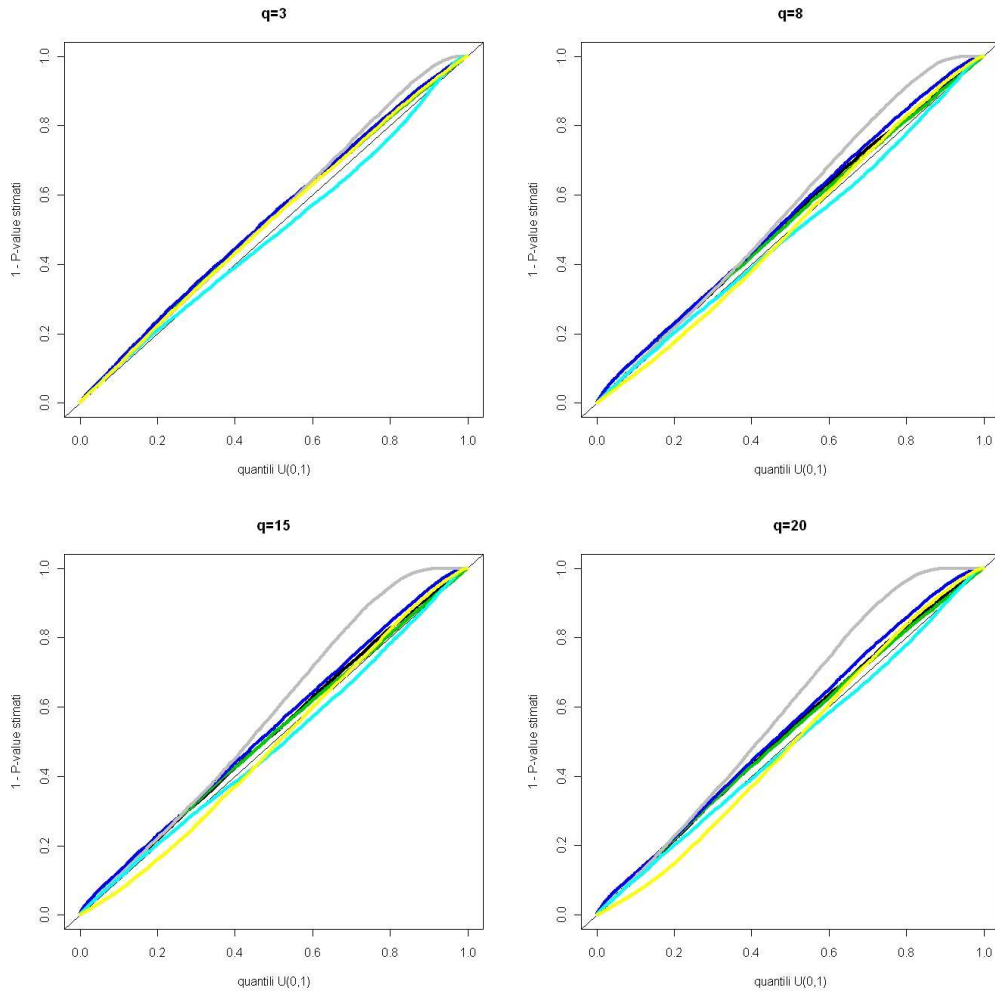


Figura 3.3: Confronto della trasformata tramite funzione di ripartizione teorica dei quantili empirici dei test con i quantili di una uniforme in  $(0,1)$  per  $\theta = (4, 2, 0.7)$ ,  $n = 10$ ,  $q = 3, 8, 15, 20$ , 10000 repliche. In particolare,  $W(\theta)$  nero,  $W_e(\theta)$  rosso,  $W_u(\theta)$  viola,  $W^P(\theta)$  blu,  $\hat{W}^P(\theta)$  verde,  $W_e^P(\theta)$  grigio,  $W_u^P(\theta)$  celeste,  $\bar{W}^P(\theta)$  giallo.

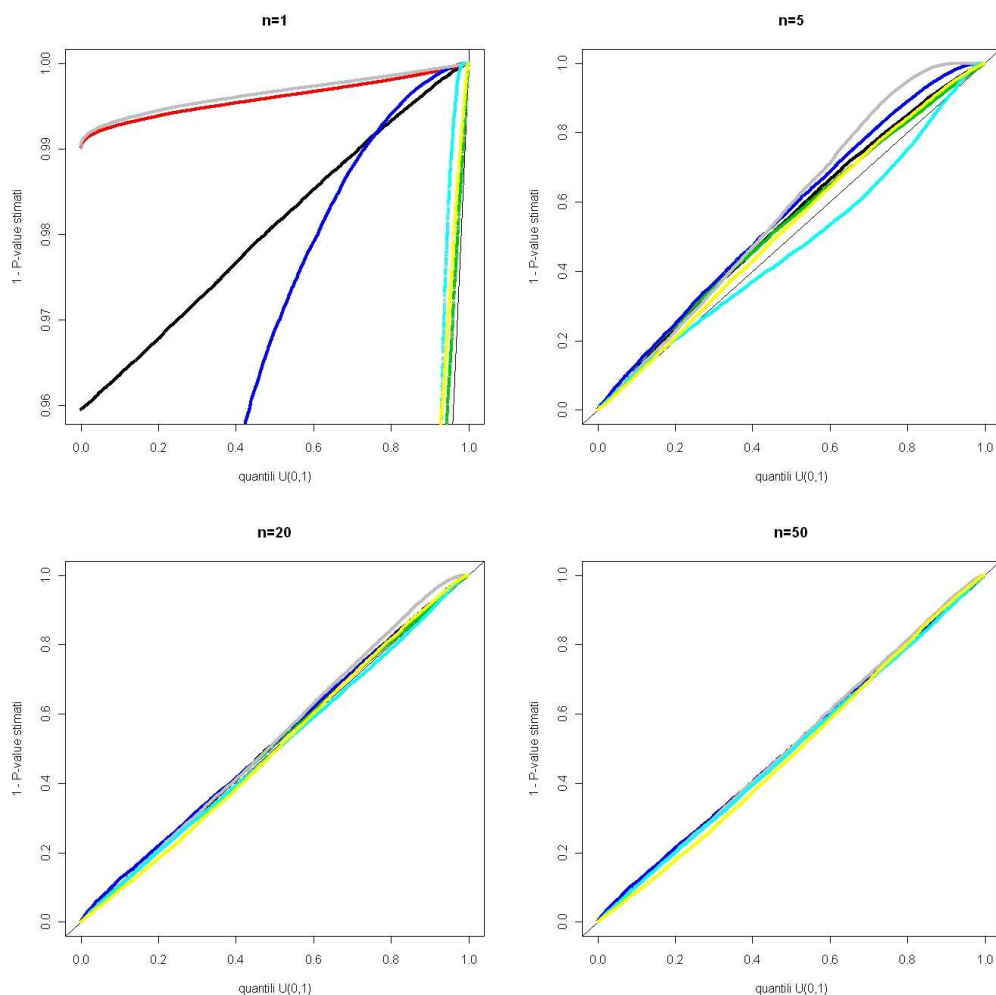


Figura 3.4: Confronto della trasformata tramite funzione di ripartizione teorica dei quantili empirici dei test con i quantili di una uniforme in  $(0,1)$  per  $\theta = (4, 2, 0.7)$ ,  $q = 5$ ,  $n = 1, 5, 20, 50, 10000$  replicazioni. In particolare,  $W(\theta)$  nero,  $W_e(\theta)$  rosso,  $W_u(\theta)$  viola,  $W^P(\theta)$  blu,  $\hat{W}^P(\theta)$  verde,  $W_e^P(\theta)$  grigio,  $W_u^P(\theta)$  celeste,  $\bar{W}^P(\theta)$  giallo.

Tabella 3.6: Livelli di copertura per  $\theta = (4, 2, 0.7)$ ,  $q = 5$ ,  $n = 1, 5, 20, 50$ ,  $\alpha = 0.1, 0.05, 0.01$ , 10000 repliche.

	$n = 1$			$n = 5$			$n = 20$			$n = 50$		
	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$W(\theta)$	0.000	0.000	0.717	0.855	0.921	0.980	0.890	0.941	0.987	0.899	0.950	0.989
$W_e(\theta) = W_e^P(\theta)$	0.000	0.000	0.000	0.749	0.804	0.872	0.853	0.903	0.959	0.884	0.933	0.978
$W_u(\theta) = W_u^P(\theta)$	0.900	0.925	0.957	0.904	0.941	0.974	0.905	0.946	0.983	0.904	0.950	0.987
$W^P(\theta)$	0.851	0.932	0.986	0.878	0.938	0.989	0.896	0.947	0.988	0.903	0.952	0.990
$\hat{W}^P(\theta)$	0.135	0.371	0.732	0.814	0.875	0.943	0.882	0.936	0.981	0.897	0.949	0.987
$\overline{W}^P(\theta)$	0.833	0.916	0.979	0.866	0.924	0.983	0.885	0.936	0.983	0.890	0.942	0.985

$W_e(\theta)$  e  $W_e^P(\theta)$  per alti valori di  $q$  e per basse numerosità campionarie sono i due che presentano scostamenti maggiori dalla retta di riferimento.

Le tabelle 3.5 e 3.6 evidenziano con maggiore precisione ciò che si è visto dai grafici nelle Figure 3.3 e 3.4.

Particolarmente, dai grafici nelle figure 3.3 e 3.4, dalle tabelle 3.5 e 3.6, si può notare come l'approssimazione basata sulla (3.4) per la distribuzione del test del rapporto di verosimiglianza a coppie  $W^P(\theta)$ , risulta tendenzialmente migliore di quella proposta in letteratura per  $\overline{W}^P(\theta)$ .

# Conclusioni

Il primo scopo di questa tesi è fare una breve rassegna dei concetti di base della verosimiglianza composita, costituita da una combinazione di validi oggetti di verosimiglianza, di solito legati a piccoli sottoinsiemi di dati. Essa può essere utilizzata al posto della verosimiglianza completa quando, ad esempio, si desidera ridurre la complessità computazionale in modo che sia possibile far fronte a grandi dimensioni di dati e modelli complessi che presentano difficoltà a causa di interdipendenze di elevate dimensioni, anche quando si desidera fare inferenza sui parametri di interesse senza fare assunzioni sulla distribuzione congiunta dei dati.

L'obiettivo principale è stato quello di studiare la verosimiglianza a coppie, un caso particolare della verosimiglianza composita marginale. In particolare, si è concentrata l'attenzione sulla eventuale perdita di efficienza nell'uso della verosimiglianza a coppie anziché quella completa.

L'inferenza basata sulla verosimiglianza a coppie può essere giustificata dal fatto che il valore atteso della log-verosimiglianza a coppie è massimo in  $\theta^0$ . Le procedure inferenziali basate sulla verosimiglianza a coppie hanno proprietà teoriche simili a quelle basate sulla verosimiglianza completa. L'uso dei metodi basati sulla verosimiglianza a coppie ci porta in genere ad una perdita di efficienza dovuta al fatto che non è più valida l'identità dell'informazione.

In questa tesi, applicando i metodi basati sulla verosimiglianza a coppie, particolarmente nel caso del modello ad effetti casuali, si ha che le stime dei parametri ottenute con la verosimiglianza a coppie sono uguali

a quelle ottenute con la verosimiglianza completa. Questo è stato ottenuto da Mardia, Hughes e Charles (2007). In aggiunta, si è mostrato che l'informazione di Fisher e l'informazione di Godambe coincidono, e che c'è una relazione lineare tra la funzione *score* basata sulla verosimiglianza a coppie e la funzione *score* basata sulla verosimiglianza completa, con matrice della trasformazione dipendente da  $\theta$ , ma non da  $y$ . Pertanto, si può dire per questo caso che le stime ottenute con la verosimiglianza a coppie sono efficienti tanto quanto quelle ottenute con la verosimiglianza composta. Da questi risultati, si è dimostrato che il test *score* e il test di Wald basati sulla verosimiglianza a coppie coincidono con i rispettivi test basati sulla verosimiglianza completa.

La distribuzione del test del rapporto di verosimiglianza a coppie non è più quella standard, quindi si è proposta una semplice approssimazione per la sua distribuzione. Tramite simulazione, si sono stimate le distribuzioni nulle e si è notato che l'approssimazione proposta risulta migliore di quelli disponibili in letteratura. In entrambi gli esempi considerati, sia per valori piccoli di  $q$  che per valori piccoli di  $n$ , il test di Wald basato sulla verosimiglianza a coppie e il test di Wald basato su quella completa, sono quelli per cui l'adeguatezza dell'approssimazione non è soddisfacente.



# Appendice

Per il controllo dei risultati analitici e il per calcolo di alcune quantità, si è fatto ricorso a un software di calcolo simbolico: Maxima versione 5.18.1 ([http:// maxima.sourceforge.net/](http://maxima.sourceforge.net/)). Maxima è un *computer algebra system*, ossia, un programma che permette di fare calcolo simbolico e numerico. Precisamente, Maxima è stato utile per la verifica di certi risultati ottenuti, e, soprattutto, per il calcolo di  $i(\theta)^{-1}$ ,  $J(\theta)^{-1}$ ,  $H(\theta)^{-1}$  e  $A(\theta)$ , nel Capitolo 2.

Nel seguito viene mostrato un esempio illustrativo di come funziona il programma. Innanzitutto, un valore viene assegnato ad una variabile nel seguente modo:

nomevariabile : valore ;

Si supponga di voler calcolare l'inversa della matrice  $i(\theta)$  (cfr. paragrafo 2.3.3). Si creano tutti i suoi elementi, ad esempio nel seguente modo

i11:(n \* qv)/(sig \* (1 + rho \* (qv - 1)));

i12:0;

i13:0;

i21:0;

i22:(n \* qv)/(2 \* sig<sup>2</sup>);

i23:-(rho \* n \* qv \* (qv - 1))/(2 \* sig \* (1 - rho) \* (1 + rho \* (qv - 1)));

i31:0;

i32:-(rho \* n \* qv \* (qv - 1))/(2 \* sig \* (1 - rho) \* (1 + rho \* (qv - 1)));

i33: (n \* (qv - 1)/2) \* (1/(1 - rho)<sup>2</sup> + (qv - 1)/(1 + rho \* (qv - 1))<sup>2</sup>);

dove  $sig = \sigma^2$ ,  $rho = \rho$  e  $qv = q$ .

```
I:matrix([i11,i12,i13],[i21,i22,i23],[i31,i32,i33]);
```

```
invI:invert(I);
```

```
invI;
```

$$\begin{bmatrix} \frac{(qv\rho-rho+1)*sig}{n*qv} & 0 & 0 \\ 0 & \frac{2*(qv\rho^2-rho^2+1)*sig^2}{n*qv} & \frac{-2*(rho-1)*rho*(qv\rho-rho+1)*sig}{n*qv} \\ 0 & \frac{-2*(rho-1)*rho*(qv\rho-rho+1)*sig}{n*qv} & \frac{2*(rho-1)^2*(qv\rho-rho+1)^2}{n*(qv-1)*qv} \end{bmatrix}.$$

Il programma è stato utilizzato per i calcoli nel paragrafo 2.3.3 e per il calcolo della quantità  $J_{\sigma^2\rho}$  del paragrafo 2.3.2.

# Bibliografia

- [1] Azzalini, A. (1983). Maximum likelihood of order  $m$  for stationary stochastic processes. *Biometrika*, **70**, 367-381.
- [2] Besag, J.E. (1974). Spatial interaction and the analysis of lattice systems. *Journal of the Royal Statistical Society*, **B, 34**, 192-236.
- [3] Besag, J.E. (1977). Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika*, **64**, 616-618.
- [4] Chandler, R.E., Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, **94**, 167-183.
- [5] Cox, D.R. (1975). Partial likelihood. *Biometrika*, **62**, 269-276.
- [6] Cox, D.R., Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91**, 729-737.
- [7] Davison A. C. (2003). *Statistical Models*, Cambridge University Press, Cambridge.
- [8] Foutz, R.V., Srivastava, R.C. (1977). The performance of the likelihood ratio test when the model is incorrect. *The Annals of Statistics*, **5**, 1183-1194.
- [9] Godambe, V.P. (1960). An optimum property of regular maximum likelihood equation. *Annals of Mathematical Statistics*, **31**, 1208-1211.

- [10] Le Cessie, S., Van Houwelingen, J.C. (1994). Logistic regression for correlated binary data. *Applied Statistics*, **43**, 95-108.
- [11] Lele, S., Taper, M.L. (2002). A composite likelihood approach to estimation of (co)variance components. *Journal of Statistical Planning and Inference*, **103**, 117-135.
- [12] Lindsay, B.G. (1988). Composite likelihood methods. *Contemporary Mathematics*, **80**, 221-240.
- [13] Mardia, K. V., Hugues, G. e Taylor, C. C. (2007). Efficiency of the pseudo-likelihood for multivariate normal and Von Mises distributions. *Technical Report 07-02 Department of Statistics, University of Leeds*.
- [14] Molenberghs, V., Verbeke, V. (2005). *Models for Discrete Longitudinal Data*, Springer-Verlag, Berlin.
- [15] Pace, L., Salvan, A. (1996). *Teoria della Statistica*, Cedam, Padova.
- [16] Pace, L., Salvan, A. (1997). *Principles of Statistical Inference from a Neo-Fisherian Perspective. Advanced Series on Statistical Science and Applied Probability*, World Scientific Publishing Co., Inc., River Edge, NJ.
- [17] Pace, L., Salvan, A. (2001). *Introduzione alla Statistica*, Cedam, Padova.
- [18] Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, Second edition. New York: Wiley.
- [19] Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, **2**, 110-114.
- [20] Severini, T. A. (2005). *Elements of Distribution Theory*, Cambridge University Press, Cambridge.
- [21] Stein, M.L. (2004). Approximating likelihoods for large spatial sets. *Journal of the Royal Statistical Society*, **B**, **66**, 275-296.

- 
- [22] Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis*, **92**, 1-28.
- [23] Varin, C., Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, **92**, 519-528.