

Università degli studi di Padova

Facoltà di Scienze Statistiche

Tesi di Laurea in Statistica e Tecnologie Informatiche

**MIGLIORAMENTI ASINTOTICI DEL
LOG-RAPPORTO DI VEROSIMIGLIANZA CON
PARAMETRO DI INTERESSE
MULTIDIMENSIONALE**

A.A. 2004/2005

Relatore: Prof.ssa VENTURA Laura

Soriani Nicola - 471044

23 settembre 2004

Indice

1	Teoria asintotica del primo ordine e verosimiglianza profilo	3
1.1	Richiami	3
1.2	Verosimiglianza e quantità collegate	4
1.3	Verosimiglianza Profilo	7
2	Teoria asintotica di ordine superiore	11
2.1	Basi Teoriche	11
2.2	La formula p^*	13
2.3	Inferenza su un parametro di interesse scalare	14
2.3.1	La statistica radice con segno modificata r^*	14
2.3.2	Presenza di parametri di disturbo	15
2.4	Inferenza su un parametro di interesse multiparametrico	16
3	Un'applicazione nelle famiglie esponenziali	19
3.1	Famiglie esponenziali	19
3.1.1	La Verosimiglianza nelle famiglie esponenziali e Quantità Collegate	21
3.2	Applicazioni	25
3.2.1	Esempio 1: Parametri incidentali	25
3.2.2	Esempio 2: Regressione Logistica	26
3.3	Conclusioni	31
A	Chiarimenti	35
A.1	Giustificazione per la sufficienza di $(\hat{\theta}, a)$	35

A.2	Giustificazione per la normalità asintototica della r^*	36
B	Codici Esempi	37
B.1	Esempio 1	37
B.2	Esempio 2	45

Introduzione

In tutti gli approcci all'inferenza di ispirazione frequentista, tra i problemi centrali si colloca il calcolo di distribuzioni campionarie ed il trattamento di parametri di disturbo. I due problemi, anche se diversi, sono collegati, dal momento che l'eliminazione dei parametri di disturbo idealmente è ottenuta tramite quantità la cui distribuzione campionaria non dipende da essi, mentre è massimamente informativa sul parametro di interesse.

Le soluzioni cosiddette esatte costituiscono una parte rilevante della teoria classica, ma lasciano parzialmente insoddisfatti, in quanto valide sotto condizioni molto specifiche, anche se estremamente rilevanti ai fini pratici, come per i modelli lineari in ipotesi di normalità. Il termine soluzioni esatte si riferisce sia alla distribuzione campionaria sia alla eliminazione dei parametri di disturbo tramite il passaggio ad un modello '*ridotto*', usualmente ottenuto per marginalizzazione o per condizionamento.

Al di fuori di tali situazioni, l'inferenza basata sulla verosimiglianza propone un insieme di procedure dotate di ampia applicabilità, grazie al riferimento a risultati di approssimazione asintotica (*del primo ordine*), quali la normalità della distribuzione campionaria dello stimatore di massima verosimiglianza e la distribuzione chi-quadrato del log-rapporto di verosimiglianza. Inoltre, l'eliminazione dei parametri di disturbo avviene con il passaggio alla verosimiglianza profilo, facendo affidamento sulla similarità asintotica del log-rapporto di verosimiglianza profilo.

Da una ventina di anni, una parte rilevante della ricerca nell'ambito dell'inferenza basata sulla verosimiglianza si è concentrata sullo sviluppo di

metodologie che vanno sotto il nome di **metodi asintotici di ordine superiore** (*higher-order asymptotics*). I raffinamenti hanno riguardato in particolare la precisione delle approssimazioni distributive delle quantità di verosimiglianza e modificazioni della verosimiglianza profilo per tener conto in modo più accurato della presenza di parametri di disturbo.

Obiettivo di questa tesi è presentare e discutere la teoria che sta alla base di questi metodi di ordine superiore, con particolare riferimento alle famiglie esponenziali nel caso multiparametrico, dove la letteratura al riguardo è abbastanza limitata.

La tesi si articola in tre capitoli principali. Nel primo vengono presentate la teoria della verosimiglianza e le tecniche di primo ordine. Nel secondo viene introdotta la teoria e i risultati generali alla base dei metodi di ordine superiore. Infine nel terzo e ultimo capitolo verranno ripresentate le tecniche introdotte nel capitolo due, nell'ambito delle famiglie esponenziali, applicandole e riportandone i risultati ad un esempio su un dataset reale e a un esempio più teorico, così da testarne l'efficacia nella pratica.

Capitolo 1

Teoria asintotica del primo ordine e verosimiglianza profilo

1.1 Richiami

Lo scopo dell'analisi statistica è la ricostruzione del modello probabilistico $p_0(\cdot)$ attraverso la sintesi dei dati osservati y^{oss} . Spesso alcune delle assunzioni che agevolano la ricostruzione del modello probabilistico sono sintetizzate tramite una delimitazione delle forme possibili per $p_0(\cdot)$, ossia attraverso la specificazione di una famiglia \mathcal{F} di distribuzioni di probabilità almeno qualitativamente compatibili con y^{oss} . La famiglia \mathcal{F} è detta **modello statistico**. Se la funzione di probabilità che ha generato i dati è un elemento di \mathcal{F} , ossia se $p_0(\cdot) \in \mathcal{F}$, si dice allora che il modello statistico è **correttamente specificato**.

Il modello statistico \mathcal{F} è pensato come una collezione di funzioni di densità. In generale, il modello \mathcal{F} è specificato una volta che sia assegnata la terna

$$\{\mathcal{Y}, P_\theta, \Theta\},$$

dove \mathcal{Y} rappresenta lo spazio campionario, P_θ una misura di probabilità su una σ -algebra \mathcal{B} di sottoinsiemi di \mathcal{Y} , e Θ è lo spazio parametrico in cui può avere valori il parametro θ che indicizza la distribuzione di P_θ .

Ci troviamo in presenza di un modello statistico parametrico quando le informazioni a disposizione conducono ad una classe \mathcal{F} piuttosto ristretta, in cui gli elementi che la compongono sono indicizzabili con un numero finito p di parametri reali:

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

ove $p_Y(y; \theta)$ è una densità rispetto ad una comune misura dominante, $\theta = (\theta^1, \dots, \theta^p)$, e $\Theta \subseteq \mathbb{R}^p$ è lo spazio parametrico. Se il modello è correttamente specificato si ha $p_0(y) = p(y; \theta_0)$ per un valore $\theta_0 \in \Theta$, detto vero valore del parametro.

1.2 Verosimiglianza e quantità collegate

Il concetto di verosimiglianza è centrale per l'inferenza statistica nella visione fisheriana. Nell'ambito di un modello statistico parametrico \mathcal{F} specificato nella forma (1.1), la funzione di verosimiglianza di un campione casuale semplice (*c.c.s*) $y = (y_1, \dots, y_n)$ è definita come:

$$L(\theta) = L(\theta; y) = c(y) \cdot p_Y(y; \theta), \quad (1.2)$$

ove $c(y) > 0$ è una costante di proporzionalità. Per maggiori chiarimenti si rimanda al Pace e Salvan [12] o all'Azzalini [1]

Viene definita, e spesso molto più conveniente considerare, funzione di log-verosimiglianza, il logaritmo naturale della funzione di verosimiglianza

$$l(\theta) = l(\theta; y) = \ln L(\theta) \quad (1.3)$$

determinata stavolta a meno di una costante additiva dipendente solo da y .

Nel seguito si assumerà, che la log-verosimiglianza sia una funzione di θ sufficientemente regolare, cioè che ammetta derivate parziali fino agli ordini richiesti.

Si indicheranno con

$$l_r = l_r(\theta; y) = \frac{\partial}{\partial \theta^r} l(\theta),$$

$$l_{rs} = l_{rs}(\theta; y) = \frac{\partial}{\partial \theta^r \partial \theta^s} l(\theta), \quad (1.4)$$

$$l_{rst} = l_{rst}(\theta; y) = \frac{\partial}{\partial \theta^r \partial \theta^s \partial \theta^t} l(\theta),$$

eccetera, le derivate parziali della log-verosimiglianza, con $r, s, t, \dots = 1, \dots, p$. Tali derivate, e più in generale le quantità definite a partire dalla funzione di verosimiglianza, sono dette **quantità di verosimiglianza**. Le derivate di $l(\theta)$ fino al secondo ordine rivestono un ruolo centrale per l'inferenza, mentre le derivate successive saranno importanti per i raffinamenti della teoria asintotica dell'inferenza, che saranno trattati nei prossimi capitoli.

Il vettore delle derivate parziali prime di $l(\theta)$, che indicheremo con $l_\theta = l_\theta(\theta; y) = (l_1, \dots, l_p)$, è detto **funzione score** o **punteggio**. Vale

$$\mathbb{E}_\theta[l_\theta(\theta)] = 0,$$

ossia il momento primo nullo della score è zero.

Il valore $\hat{\theta}$ di θ che massimizza $L(\theta; y)$, o equivalentemente $l(\theta; y)$, su Θ , ossia $L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta)$, è detto **stima di massima verosimiglianza (s.m.v)** di θ , e indica il valore a cui i dati offrono massimo sostegno empirico. Nei casi regolari, la s.m.v va cercata tra le soluzioni della funzione score

$$l_\theta(\theta) = 0.$$

La matrice contenente le derivate seconde della log-verosimiglianza cambiate di segno è detta matrice di **informazione osservata** di Fisher, ed è definita come

$$j = j(\theta) = [-l_{rs}] = \begin{pmatrix} -l_{11} & \dots & -l_{1p} \\ \vdots & \ddots & \vdots \\ -l_{p1} & \dots & -l_{pp} \end{pmatrix}. \quad (1.5)$$

La quantità $i = i(\theta) = \mathbb{E}_\theta[-l_{rs}(\theta)] = 0$, è detta matrice di **informazione attesa** di Fisher.

La funzione di verosimiglianza, così come la log-verosimiglianza, non risentono della parametrizzazione scelta per \mathcal{F} . Sia $\psi = \psi(\theta)$, con $\psi(\cdot)$ funzione biunivoca e regolare da $\Theta \subseteq \mathbb{R}^p$ in $\Psi \subseteq \mathbb{R}^p$. La scrittura $L^\Psi(y)$ equivale a

$L^\Theta(\theta(\psi))$, in quanto θ e $\psi(\theta)$ individuano lo stesso elemento di \mathcal{F} . Ciò implica naturalmente che se $\hat{\theta}$ è la s.m.v per gli elementi di Θ , allora $\hat{\psi} = \psi(\hat{\theta})$ è la s.m.v per gli elementi di Ψ .

Uno sviluppo di Taylor nella s.m.v $\hat{\theta}$, ponendo θ come il vero valore del parametro, fornisce

$$l(\theta) = l(\hat{\theta}) + l_\theta(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{2}l_{\theta\theta}(\hat{\theta})(\theta - \hat{\theta})^2 + \dots$$

In altre parole abbiamo

$$l(\theta) = l(\hat{\theta}) - \frac{1}{2}j(\hat{\theta})(\theta - \hat{\theta})^2 + \dots$$

da questa formula otteniamo per la log-verosimiglianza normalizzata:

$$l(\theta) - l(\hat{\theta}) = -\frac{1}{2}j(\hat{\theta})(\theta - \hat{\theta})^2 + \dots$$

Tale formula esprime l'approssimazione parabolica per la log-verosimiglianza normalizzata.

La distribuzione approssimata per lo stimatore di massima verosimiglianza $\hat{\theta}$, risulta essere:

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

Da cui consegue che l'errore di stima risulta essere distribuito come

$$\hat{\theta} - \theta \sim \mathcal{N}_p(0, j(\hat{\theta})^{-1}).$$

Indifferentemente, si potrebbe usare l'informazione attesa $i(\hat{\theta})$. A partire da questo risultato possiamo costruire:

- Intervalli di confidenza per θ , nella forma:

$$\hat{\theta} \pm Z_{1-\alpha/2} \cdot \sqrt{j(\hat{\theta})}, \tag{1.6}$$

dove $Z_{1-\alpha/2}$ rappresenta il quantile $1-\alpha/2$ di una $\mathcal{N}(0, 1)$.

- Test di verifica per l'ipotesi nulla $H_0 : \theta = \theta_0$

$$T = (\hat{\theta} - \theta_0) \cdot \sqrt{j(\hat{\theta})}, \tag{1.7}$$

con $T \sim \mathcal{N}(0, 1)$, sotto l'ipotesi H_0 .

Gli intervalli basati sulla distribuzione approssimata possono risultare poco precisi, in quanto non sempre l'approssimazione parabolica è perfettamente simmetrica. A questo scopo risulta molto più utile e preciso, utilizzare per costruire intervalli di confidenza ed effettuare test di verifica d'ipotesi, la statistica log-rapporto di verosimiglianza, definita come

$$W(\theta) = 2(l(\hat{\theta}) - l(\theta)). \quad (1.8)$$

Si può mostrare che $W(\theta_0) \simeq j(\hat{\theta}) \cdot (\theta_0 - \hat{\theta})(\theta_0 - \hat{\theta})^T \sim \chi_p^2$ con $\dim(\theta) = p$ sotto H_0 . A partire da $W(\theta)$ si può costruire un intervallo di confidenza della forma

$$\Theta(y) = \{\theta \in \Theta : W(\theta) \leq \chi_{p,1-\alpha}^2\}, \quad (1.9)$$

dove $\chi_{p,1-\alpha}^2$ è il percentile $1 - \alpha$ di una distribuzione Chi-quadrato con p gradi di libertà.

Nei casi in cui θ è scalare, è possibile utilizzare una statistica, detta **Radice con segno** di $W(\theta)$; definita come

$$r(\theta) = \text{sgn}(\hat{\theta} - \theta) \cdot \sqrt{W(\theta)}. \quad (1.10)$$

Si ha $r(\theta_0) \sim \mathcal{N}(0, 1)$ sotto H_0 .

1.3 Verosimiglianza Profilo

Come abbiamo detto $\theta = (\theta_1, \dots, \theta_p)$ è formato da p parametri. Nella pratica però, accade che si sia interessati a fare inferenza solamente su un numero limitato dei componenti di θ . In questo caso, il parametro θ è partizionato come $\theta = (\tau, \zeta)$, con il parametro τ , di dimensione k detto **parametro d'interesse**, e il parametro ζ , di dimensione $p - k$, detto **parametro di disturbo**.

Per l'inferenza sul parametro di interesse τ l'idea è sostituire nella verosimiglianza originaria il parametro di disturbo ζ con una sua stima consistente. In un modello statistico parametrico con funzione di verosimiglianza $L(\theta) = L(\tau, \zeta)$, viene detta verosimiglianza profilo per il parametro di

interesse τ la funzione

$$L_p(\theta) = L(\theta, \hat{\zeta}_\tau), \quad (1.11)$$

ove $\hat{\zeta}_\tau$ è la s.m.v di ζ per τ fissato.

Di conseguenza la funzione score si distinguerà

$$l_\theta = (l_\tau, l_\zeta),$$

con $l_\tau = \frac{\partial}{\partial \tau} l(\tau, \zeta)$ vettore k dimensionale, e $l_\zeta = \frac{\partial}{\partial \zeta} l(\tau, \zeta)$ vettore di dimensione $p - k$. La verosimiglianza profilo pur non essendo una verosimiglianza in senso proprio, in quanto non dedotta da una funzione di densità, gode tuttavia di alcune proprietà che l'assimilano ad una verosimiglianza propria. In particolare:

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

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

- La statistica log-rapporto di verosimiglianza profilo coincide con la statistica basata su $L(\tau, \zeta)$ usualmente considerata per la verifica di ipotesi su τ , con ζ ignoto. Infatti

$$W_p = W_p(\tau) = 2 \{l_p(\hat{\tau}) - l_p(\tau)\} = 2 \{l(\hat{\tau}, \hat{\zeta}) - l(\tau, \hat{\zeta}_\tau)\}. \quad (1.12)$$

Sotto le usuali assunzioni sotto H_0 , W_p ha distribuzione asintotica nulla χ_k^2 . Se τ è scalare, la versione con segno di W_p è

$$r_p = \text{sgn}(\hat{\tau} - \tau) \cdot \sqrt{W_p(\tau)} = \text{sgn}(\hat{\tau} - \tau) \cdot \sqrt{2 \{l_p(\hat{\tau}) - l_p(\tau)\}} \quad (1.13)$$

- Poichè

$$P_{\tau, \zeta} \{2(l_p(\hat{\tau}) - l(\tau)) \leq \chi_{k; 1-\alpha}^2\} \approx 1 - \alpha,$$

le regioni di confidenza basate sul test rapporto di verosimiglianza profilo hanno livello nominale asintotico $1 - \alpha$.

- L'informazione osservata profilo è

$$j_p(\tau) = - \left(\frac{\partial^2}{\partial \tau \partial \tau^T} \right) l_p(\tau) = - \left(\frac{\partial^2}{\partial \tau \partial \tau^T} \right) l(\tau, \hat{\zeta}_\tau) \quad (1.14)$$

Va notato che

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

Ma il secondo membro è uguale a zero poichè $l_\zeta(\tau, \hat{\zeta}_\tau) = 0$, per cui

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

La matrice delle derivate seconde è uguale a

$$\left(\frac{\partial^2}{\partial \tau \partial \tau^T} \right) l_p(\tau) = l_{\tau\tau}(\tau, \hat{\zeta}_\tau) + l_{\zeta\tau}(\tau, \hat{\zeta}_\tau) \cdot \frac{\partial}{\partial \tau} \hat{\zeta}_\tau.$$

La quantità $\left(\frac{\partial}{\partial \tau} \hat{\zeta}_\tau \right)$ è ottenuta considerando la derivata rispetto a τ dell'equazione di verosimiglianza $l_\zeta(\tau, \hat{\zeta}_\tau) = 0$, ottenendo l'uguaglianza

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

da cui

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

Da ciò ne consegue che

$$j_p(\tau) = - (l_{\tau\tau} - l_{\tau\zeta} \cdot l_{\zeta\tau} \cdot l_{\zeta\zeta}^{-1})$$

dove tutte le derivate sono calcolate in $(\tau, \hat{\zeta}_\tau)$, pertanto

$$[j_p(\tau)]^{-1} = j_{\tau\tau}(\tau, \hat{\zeta}_\tau),$$

ove $j_{\tau\tau}(\tau, \hat{\zeta}_\tau)$ rappresenta il blocco (τ, τ) dell'inversa della matrice di informazione osservata complessiva.

$$j(\theta) = \begin{pmatrix} j_{\tau\tau} & j_{\tau\zeta} \\ j_{\zeta\tau} & j_{\zeta\zeta} \end{pmatrix}$$

ossia

$$j(\theta)^{-1} = \begin{pmatrix} j^{\tau\tau} & j^{\tau\zeta} \\ j^{\zeta\tau} & j^{\zeta\zeta} \end{pmatrix}$$

Queste proprietà rendono la verosimiglianza profilo uno strumento indubbiamente interessante, ma tuttavia come abbiamo detto essa non costituisce una verosimiglianza in senso proprio. Ad esempio, la funzione score profilo non ha valore atteso nullo pari a zero, ma di ordine $O(n)$.

Capitolo 2

Teoria asintotica di ordine superiore

Come il teorema del limite centrale permette di ottenere le usuali approssimazioni asintotiche per quantità di verosimiglianza, i risultati sulle approssimazioni di ordine elevato presentate in questo capitolo consentono di sviluppare metodi asintotici di ordine superiore per l'inferenza basata sulla verosimiglianza.

In questo capitolo saranno presentati alcuni dei risultati principali, partendo dalla formula p^* , presentata nel §2.2, per arrivare ad ottenere dei raffinamenti per le distribuzioni dei combinanti di tipo r , nel §2.3, anche in presenza di parametri di disturbo, e W , nel §2.4. Per maggiori chiarimenti si rimanda al Pace e Salvan [12], o Barndorff-Nielsen & Cox [3] oppure Severini [18].

2.1 Basi Teoriche

Nel capitolo precedente, si è evidenziato che la score calcolata sulla verosimiglianza profilo non ha valore atteso nullo pari a zero. Esiste un risultato (cfr. Barndorff-Nielsen & Cox [3], formula 6.28) che afferma che

$$\mathbb{E}_{\tau, \zeta} \left(l_{\tau}(\tau, \hat{\zeta}_{\tau}) \right) = -\frac{1}{2} \text{tr} \left\{ i_{\zeta \zeta}^{-1} (\gamma_{\zeta \zeta, \tau} + \gamma_{\zeta, \zeta, \tau}) \right\} + O(n^{-1}), \quad (2.1)$$

dove

$$\gamma_{\zeta\zeta,\tau} = \nu_{\zeta\zeta,\tau} - \nu_{\zeta\zeta,\zeta} i_{\zeta\zeta}^{-1} i_{\zeta\tau}, \quad \text{e} \quad \gamma_{\zeta,\zeta,\tau} = \nu_{\zeta,\zeta,\tau} - \nu_{\zeta,\zeta,\zeta} i_{\zeta\zeta}^{-1} i_{\zeta\tau} \quad .$$

il simbolo tr indica la funzione traccia. I valori attesi sono indicati con $\nu_\theta = \mathbb{E}_\theta(\frac{\partial}{\partial\theta}l(\theta)) = \mathbb{E}_\theta(l_\theta)$, $\nu_{\theta\theta} = \mathbb{E}_\theta(\frac{\partial^2}{\partial\theta^2}l(\theta)) = \mathbb{E}_\theta(l_{\theta\theta})$; mentre, i simboli $\nu_{\theta,\theta}$ e $\nu_{\theta,\theta\theta}$ indicano le collezioni ordinate dei valori attesi rispetto a $p_Y(y; \theta)$ rispettivamente dei prodotti di due elementi di l_θ , e dei prodotti di un elemento di l_θ con uno di $l_{\theta\theta}$. Il simbolo ν con tre o più indici a deponente, sta ad indicare una matrice a tre o più dimensioni.

Il termine dominante dello sviluppo 2.1 è di ordine $O(1)$, purchè il parametro di disturbo ζ abbia dimensione non dipendente da n . Se la dimensione di ζ dipende da n , ed in tal caso si parla di parametri incidentali, tale termine dà luogo alla non consistenza dello stimatore di τ basato sulla verosimiglianza profilo. Qualora $p - k$ sia elevato rispetto a n , la distorsione di l_τ potrebbe non essere trascurabile, tanto da rendere meno accurate le argomentazioni di linearizzazione locale di l_τ in un intorno dello zero che sta alla base delle giustificazioni teoriche per l'inferenza con la verosimiglianza profilo.

Un noto esempio estremo è quello del problema di Neyman e Scott, in cui si considerano n coppie indipendenti tratte da una distribuzione $\mathcal{N}(\mu_i, \sigma^2)$, $i = 1, \dots, n$. Se $\tau = \sigma^2$ è il parametro di interesse, allora l'inferenza basata su $L_p(\tau)$ conduce ad una stima non consistente del parametro di interesse.

Queste argomentazioni ci spingono a ricercare termini di correzione per le usuali statistiche di verosimiglianza con lo scopo di:

- Rendere più accurate le usuali approssimazioni normale o chi-quadrato per le statistiche di tipo r e W , nel caso di numerosità campionaria piccola o moderata.
- Tener conto della presenza di parametri di disturbo.

2.2 La formula p^*

Il risultato principale sul quale si appoggiano tutte le teorie asintotiche di ordine superiore è la formula p^* di Barndorff-Nielsen [7]. Questo risultato consiste in una approssimazione per la densità dello stimatore di massima verosimiglianza condizionatamente ad una statistica ancillare. Sia $s = s(y)$ una statistica sufficiente minimale per θ e sia $a = a(y)$ una statistica ancillare, ossia una statistica con distribuzione non dipendente da θ e tale che $(\hat{\theta}, a)$ sia in corrispondenza biunivoca con s . Esprimendo s in funzione di $(\hat{\theta}, a)$, si può scrivere $l(\theta; s) = l(\theta; \hat{\theta}, a)$. La formula p^* è data da

$$p_{\hat{\theta}|a}^*(\hat{\theta}; \theta, a) = c(\theta, a) \left| j(\hat{\theta}; \hat{\theta}, a) \right|^{1/2} e^{l(\theta; \hat{\theta}, a) - l(\hat{\theta}; \hat{\theta}, a)}, \quad (2.2)$$

dove $c(\theta, a)$ è una costante di normalizzazione che rende pari a 1 l'integrale della (2.2) rispetto a $\hat{\theta}$, e mantenendo costanti a e θ .

La (2.2) costituisce un'approssimazione con errore di ordine $O(n^{-1})$ per la densità condizionata di $\hat{\theta}$ dato a

$$p_{\hat{\theta}|a}(\hat{\theta}; \theta) = p_{\hat{\theta}|a}^*(\hat{\theta}; \theta, a) \{1 + O(n^{-1} \|\hat{\theta} - \theta\|)\}, \quad (2.3)$$

Per maggiori dettagli si rimanda a §A.1 .

Generalmente la bontà dell'approssimazione della (2.2) è analoga a quella delle approssimazioni secondo il metodo del punto di sella: l'errore relativo di ordine $O(n^{-1})$ è uniformemente maggiorabile in una regione di grande scarto, di forma $\hat{\theta} - \theta_0 = O(1)$, mentre in una regione di scarto normale, di forma $\hat{\theta} - \theta_0 = O(n^{-1/2})$, l'errore è di ordine $O(n^{-3/2})$.

La statistica a nella (2.2) può essere anche ancillare in senso approssimato, ad esempio con densità $p_A(a; \theta) = p_0(a) \{1 + O(n^{-1/2})\}$. In tal caso la (2.3) vale con errore relativo di ordine $O(n^{-1})$, per a in una regione di scarto normale, $(a - \mathbb{E}(A))/\sqrt{\mathbb{V}(A)} = O(1)$.

Per la formula p^* valgono le seguenti proprietà:

- è invariante rispetto a trasformazioni biunivoche dei dati y ;

- si trasforma in modo regolare per effetto di riparametrizzazione; cioè se $\psi = \psi(\theta)$ è una riparametrizzazione allora:

$$p_{\psi|a}^*(\hat{\psi}; \psi, a) = p_{\hat{\theta}|a}^*(\theta(\hat{\psi}); \theta(\psi), a) \left| \partial\theta(\hat{\psi})/\partial\hat{\psi} \right|;$$

- la costante di normalizzazione $c(\theta, a)$ è invariante rispetto alla parametrizzazione.

Per dettagli e relative dimostrazioni si rimanda al Barndorff-Nielsen & Cox [4], §6.2 .

È da evidenziare il fatto che il calcolo della costante di normalizzazione $c(\theta, a)$ richieda di disporre esplicitamente della rappresentazione di s come $(\hat{\theta}, a)$. Tuttavia vale il risultato di approssimazione $c(\theta, a)(2\pi)^{p/2} = 1 + O(n^{-1})$ che da luogo ad un'approssimazione con errore di ordine $O(n^{-1})$

$$p_{\hat{\theta}|a}^\dagger(\hat{\theta}; \theta, a) = (2\pi)^{-p/2} \left| j(\hat{\theta}; \hat{\theta}, a) \right|^{1/2} e^{l(\theta; \hat{\theta}, a) - l(\hat{\theta}; \hat{\theta}, a)}, \quad (2.4)$$

L'approssimazione fornita dalla (2.4) è detta *stabile* in quanto fornisce una approssimazione della densità di $\hat{\theta}$ condizionatamente a qualunque ragionevole statistica ancillare, senza la necessità che quest'ultima venga espressa.

2.3 Inferenza su un parametro di interesse scalare

2.3.1 La statistica radice con segno modificata r^*

La formula p^* permette di ottenere una versione modificata per la statistica *radice con segno*, (si rimanda al Pace e Salvan [12]§11.5) con distribuzione nulla $\mathcal{N}(0, 1)$ con errore di ordine $O(n^{-1})$ in una regione di grande scarto per $\hat{\theta}$, e di ordine $O(n^{-3/2})$ in una regione di scarto normale. La statistica in questione, introdotta da Barndorff-Nielsen [8], è chiamata **radice con segno modificata** (*modified directed likelihood*) ed è definita come

$$r^*(\theta) = r(\theta_0) - \frac{1}{r(\theta)} \log \frac{r(\theta)}{u(\theta)}, \quad (2.5)$$

dove

$$u(\theta) = j(\hat{\theta})^{-1/2}(l_{;\hat{\theta}}(\hat{\theta}) - l_{;\hat{\theta}}(\theta)),$$

con

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

dove quest'ultima quantità è detta *derivata nello spazio campionario*.

2.3.2 Presenza di parametri di disturbo

Se $\theta = (\tau, \zeta)$, con τ parametro di interesse *scalare*, si utilizzerà come radice con segno modificata per τ la statistica

$$r_p^*(\tau) = r_p(\tau) - \frac{1}{r_p(\tau)} \log \frac{r_p(\tau)}{u_p(\tau)}, \quad (2.6)$$

dove, posto $\tilde{\theta} = (\tau, \hat{\zeta}_\tau)$,

$$u_p(\tau) = \left| l_{;\hat{\theta}}(\hat{\theta}) - l_{;\hat{\theta}}(\tilde{\theta}) l_{\zeta;\hat{\theta}}(\tilde{\theta}) \right| / \left\{ \left| j_{\zeta\zeta}(\tilde{\theta}) \right| \left| j(\hat{\theta}) \right| \right\}^{1/2}. \quad (2.7)$$

Una forma equivalente è

$$r_p^*(\tau) = r_p(\tau) + NP + INF,$$

dove NP è la **correzione per i parametri di disturbo**, data da

$$NP = -\frac{1}{r_p(\tau)} \log C(\tau),$$

con

$$C(\tau) = \frac{\left\{ \left| j_{\zeta\zeta}(\hat{\theta}) \right| \left| j_{\zeta\zeta}(\tilde{\theta}) \right| \right\}^{1/2}}{\left| l_{\zeta;\hat{\theta}}(\tilde{\theta}) \right|},$$

mentre INF è la **correzione per la precisione**, data da

$$INF = \frac{1}{r_p(\tau)} \log \frac{\tilde{u}_p(\tau)}{r_p(\tau)},$$

dove

$$\tilde{u}_p(\tau) = j_p(\tau)^{-1/2} \frac{\partial}{\partial \hat{\tau}} \{ l_p(\hat{\tau}) - l_p(\tau) \}.$$

Osservazioni:

- Nell'espressione (2.7) il numeratore è il determinante della matrice ($p \times p$) che ha come prima colonna $l_{;\hat{\theta}}(\hat{\theta}) - l_{;\hat{\theta}}(\tilde{\theta})$ e le cui $p - 1$ colonne si ottengono derivando rispetto a $\hat{\theta}$ il vettore riga $l_{\zeta}(\theta; \hat{\theta}, a)$ e valutando le derivate in $\theta = \tilde{\theta}$.
- Nell'espressione di $\tilde{u}_p(\tau)$, le derivate rispetto a $\hat{\tau}$ sono calcolate considerando $l_p(\hat{\tau}) - l_p(\tau)$ come funzione di τ , $\hat{\tau}$, $\hat{\zeta}_{\tau}$ e a .
- Il miglioramento nell'approssimazione normale che si ha passando da r a r^* è sorprendente. I risultati empirici in Pierce e Peters [14] indicano che la correzione NP è spesso cospicua e preponderante rispetto la correzione INF , in particolare quando la dimensione del parametro di disturbo ζ è elevata.

2.4 Inferenza su un parametro di interesse multiparametrico

La statistica r^* in (2.5) è destinata a test di ipotesi unidimensionali, anche in presenza di parametri di disturbo. Accade spesso però che si vogliono verificare ipotesi su più parametri di interesse, per esempio testando gli effetti di un trattamento in un modello misto lineare o non lineare.

Una generalizzazione per la (2.5) nel caso multiparametrico, con proprietà simili, è la seguente (si veda Skovgaard [19])

$$W^* = W \left(1 - \frac{1}{W} \log \gamma \right), \quad (2.8)$$

oppure la versione asintoticamente equivalente

$$W^{**} = W - 2 \log \gamma, \quad (2.9)$$

dove $W = r^2$ è l'usuale log-rapporto di verosimiglianza, e

$$\gamma = \left| \hat{j} \right|^{1/2} \left| \tilde{D}'_1 \right|^{-1} \left| \tilde{j}_{\zeta\zeta} \right|^{1/2} \left| \tilde{j}_{\zeta\zeta} \right|^{-1/2} \left| \tilde{j} \right|^{1/2} \frac{\left\{ \tilde{D}'_1 \tilde{j}^{-1} \tilde{D}_1 \right\}^{q/2}}{W^{q/2-1} \left(\hat{v} - \tilde{v} \right)^T \left(\tilde{D}'_1 \right)^{-1} \tilde{D}_1}.$$

Le quantità di verosimiglianza presenti sono così definite

$$\hat{j} = j(\hat{\theta}; \hat{\theta}, a) = j(\hat{\tau}, \hat{\zeta}; \hat{\tau}, \hat{\zeta}, a),$$

$$\tilde{j} = j(\theta; \tilde{\theta}, a) = j(\tau, \zeta; \tau, \hat{\zeta}_\tau, a),$$

$$\tilde{\tilde{j}} = j(\tilde{\theta}; \tilde{\theta}, a) = j(\tau, \hat{\zeta}_\tau; \tau, \hat{\zeta}_\tau, a),$$

dove $\tilde{j}_{\zeta\zeta}$ e $\tilde{\tilde{j}}_{\zeta\zeta}$ indicano i blocchi di coordinate (ζ, ζ) delle rispettive matrici, e

$$\hat{l}' = l_{;\hat{\theta}}(\hat{\theta}) \quad \text{e} \quad \tilde{l}' = l_{;\tilde{\theta}}(\tilde{\theta})$$

sono le derivate della log-verosimiglianza nello spazio campionario calcolate rispettivamente in $\hat{\theta}$ e $\tilde{\theta}$. Infine

$$\tilde{D}_1 = \frac{\partial}{\partial \theta} l(\theta; \hat{\theta}, a) \quad \text{e} \quad \tilde{D}'_1 = \frac{\partial}{\partial \hat{\theta}} \left(l_\theta(\theta; \hat{\theta}, a) \right) = \frac{\partial}{\partial \hat{\theta}} \left(\frac{\partial}{\partial \theta} l(\theta; \hat{\theta}, a) \right)$$

sono rispettivamente la funzione *score* e la derivata della *score* nello spazio campionario, calcolate entrambe in $\tilde{\theta} = (\tau, \hat{\zeta}_\tau)$.

Le statistiche test (2.9) e (2.8) hanno distribuzione approssimata \mathcal{X}_k^2 , per $k = \dim(\tau)$, con elevato grado di accuratezza. La versione W^{**} è la prima che si ricava naturalmente attraverso lo sviluppo teorico. La versione W^* ha invece il vantaggio di essere non negativa e di coincidere con $(r^*)^2$, cioè con il quadrato della radice con segno modificata, nel caso $k = 1$.

Capitolo 3

Un'applicazione nelle famiglie esponenziali

Tra i modelli statistici parametrici possiamo individuare una classe nettamente caratterizzata per struttura formale e peculiarità inferenziali: la classe della **famiglia esponenziale**. Essa costituisce un serbatoio di modelli statistici utili per le applicazioni, direttamente o come elementi di un modello più complesso. Per di più, se il modello generatore dei dati appartiene alla famiglia esponenziale è possibile definire procedure inferenziali generali, fondamentalmente semplici, e accurate.

In questo capitolo della tesi, si procederà ad applicare le tecniche asintotiche di ordine superiore W^* e W^{**} , viste nel capitolo precedente, e a verificarne l'efficacia nel caso multiparametrico, in alcuni modelli facenti parte della famiglia esponenziale.

3.1 Famiglie esponenziali

Si consideri un'osservazione y da una variabile casuale Y . La funzione di densità o probabilità (f.d.p) $p(y; \phi)$ associata a Y appartiene ad una famiglia

esponenziale di ordine p , se può essere scritta nella forma

$$p(y; \phi) = \exp \left\{ \sum_{j=1}^p a_j(\phi) b_j(y) - c(\phi) \right\} h(y) = \exp \left\{ a(\phi) b(y) - c(\phi) \right\} h(y); \quad (3.1)$$

in cui ϕ è un parametro (vettoriale o scalare) che assume valori in $\Phi \subseteq \mathbb{R}^p$, e $a(\phi) = (a_1(\phi), \dots, a_p(\phi))$ e $c(\phi)$ sono funzioni in ϕ che non dipendono da y . Invece $b(y) = (b_1(y), \dots, b_p(y))$ e $h(y) > 0$ sono funzioni in y che non dipendono da ϕ . Considerando la trasformazione parametrica $\theta = a(\phi)$, la (3.1) può essere scritta come

$$p(y; \theta) = \exp \left\{ \theta b(y) - K(\theta) \right\} h(y),$$

in cui θ viene chiamato parametro **canonico** o **naturale**, il quale assume valori nello spazio parametrico $\tilde{\Theta} \subseteq \mathbb{R}^p$. Mentre $K(\theta) = K(a(\phi))$ è un'opportuna funzione nel parametro naturale.

In corrispondenza ad un campione i.i.d. $y = (y_1, \dots, y_n)$ la f.d.p congiunta è data da

$$p(y; \theta) = \exp \left\{ \theta \sum_{i=1}^n b(y_i) - n K(\theta) \right\} \left(\prod_{i=1}^n h(y_i) \right) = \exp \left\{ \theta \cdot s - n K(\theta) \right\} H(y),$$

dove $s = s(y) = \sum_{i=1}^n b(y_i)$ definisce la statistica sufficiente minimale per θ .

Pensiamo ora di poter scomporre il parametro naturale $\theta = (\tau, \zeta)$, nelle componenti τ di dimensione k , e ζ di dimensione $r = p - k$. La f.d.p di una singola osservazione diviene

$$p(y; \tau, \zeta) = \exp \left\{ \sum_{i=1}^k \tau_i f_i(y) + \sum_{j=1}^r \zeta_j g_j(y) - K(\tau, \zeta) \right\} h(y).$$

Allora valgono i seguenti risultati:

1. la distribuzione marginale di $g = (g_1, \dots, g_r)$ è ancora di tipo esponenziale, con f.d.p.

$$p(g; \tau, \zeta) = \exp \left\{ \sum_{j=1}^r \zeta_j g_j(y) - K^*(\tau, \zeta) \right\} h^*(g; \tau)$$

2. la distribuzione condizionata di $f = (f_1, \dots, f_k)$ dato g è ancora di tipo esponenziale, con f.d.p.

$$p(f|g; \tau) = \exp \left\{ \sum_{i=1}^k \tau_i f_i(y) - K^{**}(\tau; g) \right\} h^{**}(f|g)$$

Tra i principali modelli che fanno parte della famiglia esponenziale troviamo la Normale, la Binomiale, la Poisson e la Gamma (si veda Tabella 3.1).

In una famiglia esponenziale regolare la (2.2) coincide con l'approssimazione secondo il metodo del punto di sella normalizzata, con errore di ordine $O(n^{-3/2})$. Inoltre in questo caso il condizionamento ad una statistica ancillare a non è operante, in quanto la stima di massima verosimiglianza è statistica sufficiente.

3.1.1 La Verosimiglianza nelle famiglie esponenziali e Quantità Collegate

La funzione di verosimiglianza è

$$L(\theta; s) = L(\tau, \zeta; t, u) = \exp \left\{ \theta \cdot s - n K(\theta) \right\} = \exp \left\{ \tau \cdot t + \zeta \cdot u + n K(\tau, \zeta) \right\}, \quad (3.2)$$

dove s è statistica sufficiente minimale per θ , $\theta = (\tau, \zeta)$, $s = (t, u)$, con t e u statistiche sufficienti minimali rispettivamente per τ e ζ . La log-verosimiglianza è

$$l(\theta; s) = \theta \cdot s - n K(\theta) = \tau \cdot t + \zeta \cdot u + n K(\tau, \zeta). \quad (3.3)$$

Le quantità di verisimiglianza di primo ordine

La funzione score è data da

$$l_\theta = s - n K_\theta(\theta) = (l_\tau, l_\zeta) \implies \begin{cases} l_\tau = t - n K_\tau(\tau, \zeta) \\ l_\zeta = u - n K_\zeta(\tau, \theta) \end{cases}. \quad (3.4)$$

Ricordando che s è in corrispondenza biunivoca con $\hat{\theta}$, quindi t e u saranno in corrispondenza biunivoca con $\hat{\tau}$ e $\hat{\zeta}$. Questo ci porta a dire che la s.m.v

sarà data da

$$\hat{\theta} : s = nK_{\theta}(\hat{\theta}) \implies \begin{cases} \hat{\tau} : t = nK_{\tau}(\hat{\tau}, \hat{\zeta}) \\ \hat{\zeta} : u = nK_{\zeta}(\hat{\tau}, \hat{\zeta}) \end{cases}.$$

Così otteniamo, rispettivamente, che

$$l_{\theta} = nK'(\hat{\theta}) - nK'(\theta) \implies \begin{cases} l_{\tau} = nK_{\tau}(\hat{\tau}, \hat{\zeta}) - nK_{\tau}(\tau, \zeta) \\ l_{\zeta} = nK_{\zeta}(\hat{\tau}, \hat{\zeta}) - nK_{\zeta}(\tau, \zeta) \end{cases}.$$

L'informazione osservata

$$j(\theta; s) = -l_{\theta\theta} = nK_{\theta\theta}(\theta) = \begin{bmatrix} nK_{\tau\tau} & nK_{\tau\zeta} \\ nK_{\zeta\tau} & nK_{\zeta\zeta} \end{bmatrix} \quad (3.5)$$

ma poichè $j(\theta; s)$ non dipende da s , essa coincide con l'informazione attesa $i(\theta)$.

La funzione di verosimiglianza, o log-verosimiglianza, calcolata nel punto $\tilde{\theta} = (\tau, \hat{\zeta}_{\tau})$ corrisponde alla rispettiva funzione profilo per τ ed è definita da

$$l_p(\tau) = l_p(\tau; \hat{\tau}, \hat{\zeta}) = l(\tilde{\theta}; \hat{\theta}) = l(\tau, \hat{\zeta}_{\tau}; \hat{\tau}, \hat{\zeta}) = \tau nK_{\tau}(\hat{\tau}, \hat{\zeta}) + \hat{\zeta}_{\tau} nK_{\zeta}(\hat{\tau}, \hat{\zeta}) - K(\tau, \hat{\zeta}_{\tau}).$$

Le quantità di verosimiglianza per i raffinamenti di ordine superiore

La derivata nello spazio campionario della log-verosimiglianza è facilmente calcolabile ed è data da:

$$l_{;\hat{\theta}} = \theta nK_{\theta;\hat{\theta}}(\hat{\theta}; \hat{\theta}) = \tau nK_{\tau;\hat{\theta}}(\hat{\tau}, \hat{\zeta}) + \zeta nK_{\zeta;\hat{\theta}}(\hat{\tau}, \hat{\zeta}).$$

La derivata nello spazio campionario della funzione score è

$$D'_1 = nK_{\theta;\hat{\theta}}(\hat{\theta}) \implies \begin{cases} l_{\tau;\hat{\theta}} = nK_{\tau;\hat{\theta}}(\hat{\tau}, \hat{\zeta}) \\ l_{\zeta;\hat{\theta}} = nK_{\zeta;\hat{\theta}}(\hat{\tau}, \hat{\zeta}) \end{cases}$$

La quantità u e u_p diventano rispettivamente

$$u = \frac{\hat{\theta} nK_{\theta;\hat{\theta}}(\hat{\theta}) - \theta nK_{\theta;\hat{\theta}}(\hat{\theta})}{\sqrt{nK_{\theta\theta}(\hat{\theta})}}$$

e

$$u_p(\tau) = \frac{\left| \hat{\theta} n K_{\theta; \hat{\theta}}(\hat{\theta}) - n K_{\zeta; \hat{\theta}}(\hat{\theta}) \tilde{\theta} n K_{\theta; \hat{\theta}}(\hat{\theta}) \right|}{\sqrt{|n K_{\zeta\zeta}(\tilde{\theta})| |n K_{\theta\theta}(\hat{\theta})|}}$$

Mentre

$$\gamma = \frac{\left\{ \tilde{D}_1^T \tilde{J}^{-1} \tilde{D}_1 \right\}^{\frac{k}{2}}}{W^{\frac{k}{2}-1} \cdot (\hat{\theta} - \tilde{\theta})^T \tilde{D}_1} \cdot \frac{|\tilde{J}|^{1/2}}{|\hat{J}|^{1/2}},$$

questo perchè nel caso esponenziale si osservano le seguenti relazioni; $\tilde{\tilde{J}} = \tilde{J}$, $\hat{l}' - \tilde{l}' = \hat{J} \cdot (\hat{\theta} - \tilde{\theta})$ e $\tilde{D}'_1 = \hat{J}$. Questo significa che il valore di γ sarà pari a

$$\gamma = \sqrt{\frac{|n K_{\theta\theta}(\tilde{\theta})|}{|n K_{\theta\theta}(\hat{\theta})|}} \cdot \frac{\{[n K_{\theta}(\hat{\theta}) - n K_{\theta}(\tilde{\theta})]^T [n K_{\theta\theta}(\tilde{\theta})]^{-1} [n K_{\theta}(\hat{\theta}) - n K_{\theta}(\tilde{\theta})]\}^{\frac{k}{2}}}{W^{\frac{g}{2}-1} \cdot (\hat{\theta} - \tilde{\theta})^T \cdot [n K_{\theta}(\hat{\theta}) - n K_{\theta}(\tilde{\theta})]}.$$

È interessante osservare che tutti gli ingredienti necessari per calcolare le quantità W^* e W^{**} nell'ambito delle famiglie esponenziali sono ottenibili semplicemente a partire dalla 3.1. In particolare, per il calcolo di W^* e W^{**} è sufficiente calcolare la s.m.v. $\hat{\theta}$, la s.m.v. vincolata $\tilde{\theta}$, e le derivate prime e seconde della funzione $K(\theta)$.

Un'altra famiglia di funzioni, in cui il calcolo delle statistiche 2.8 e 2.9 è relativamente semplice, è la classe dei modelli di gruppo (si veda Skovgaard [19]).

Tabella 3.1: Le principali famiglie esponenziali

\mathcal{F}_e^p	$\mathcal{N}(\mu, \sigma^2)$	$\text{Bi}(n, \pi)$	$\text{P}(\lambda)$	$\text{Ga}(\omega, \lambda)$	$\text{GI}(\omega, \lambda)$
$\mathbf{p}(\mathbf{y})$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}$	$\binom{n}{y} \pi^y (1-\pi)^{n-y}$	$\frac{e^{-\lambda} \lambda^y}{y!}$	$\frac{\lambda^\omega y^{\omega-1} e^{-\lambda y}}{\Gamma(\omega)}$	$\frac{\sqrt{\lambda e^{\lambda \omega}}}{\sqrt{2\pi y^3}} e^{-\left(\frac{\lambda}{y} + \omega y\right)/2}$
θ	$\theta_1 : \frac{\mu}{\sigma^2} \quad \theta_2 : -\frac{1}{\sigma^2}$	$\log \frac{\pi}{1-\pi}$	$\log \lambda$	$\theta_1 : \omega - 1 \quad \theta_2 : -\lambda$	$\theta_1 : -\frac{\lambda}{2} \quad \theta_2 : -\frac{\omega}{2}$
$\mathbf{K}(\theta)$	$-\frac{\theta^2}{2\theta_2} - 2 \log(-2\theta_2)$	$\log(1 - e^\theta)$	e^θ	$\log \Gamma(\theta_1 + 1) - (\theta_1 + 1) \log(-\theta_2)$	$-\log \sqrt{-2\theta_1} - \sqrt{4\theta_1 \theta_2}$
$\mathbf{b}(\mathbf{y})$	$b_1 : y \quad b_2 : y^2$	y	y	$b_1 : \log y \quad b_2 : y$	$b_1 : \frac{1}{y} \quad b_2 : y$
$\mathbf{h}(\mathbf{y})$	$(2\pi)^{-\frac{1}{2}}$	$\binom{n}{y}$	$(y!)^{-1}$		$(2\pi y^3)^{-1/2}$

3.2 Applicazioni

Obiettivo di questo paragrafo, è applicare le tecniche finora viste a due particolari esempi. Il primo più teorico considera la situazione di parametri di disturbo incidentali, mentre il secondo è più pratico, una regressione logistica, anche perchè verrà applicato ad un dataset reale.

3.2.1 Esempio 1: Parametri incidentali

L'esempio che consideriamo è simile al proplema di Neyman e Scott, con distribuzione esponenziale (cfr. al Pace e Salvan [12], esempio 4.12). Siano y_{ij} , $i = 1, \dots, n$, $j = 1, 2$, determinazioni indipendenti di variabili esponenziali con media $(\tau_j + \zeta_i)^{-1}$. La f.d.p congiunta risulta essere uguale ad

$$p(y_{ij}, \tau, \zeta) = \exp \left\{ \sum_{i=1}^n \sum_{j=1}^2 \log(\tau_j + \zeta_i) - \sum_j \tau_j \sum_i y_{ij} - \sum_{ij} \zeta_i y_{ij} \right\}$$

che porta alla seguente funzione di log-verosimiglianza

$$l(\tau, \zeta) = \tau \cdot t + \zeta \cdot u + \sum_i \log(\tau_1 + \zeta_i) + \sum_i \log(\tau_2 + \zeta_i),$$

dove $\tau = (\tau_1, \tau_2)$, $\zeta = (\zeta_1, \dots, \zeta_n)$, $t = (t_1, t_2)$, con $t_j = -\sum_i y_{ij}$ e $u = (u_1, \dots, u_n)$, con $u_i = -\sum_j y_{ij}$. Sia τ il nostro parametro di interesse, e ζ quello di disturbo, e si desidera verificare l'ipotesi $H_0 : \tau_1 = \tau_2 = 0$. La tabella 3.2, riporta i risultati ottenuti da una simulazione basata su 10.000 replicazioni Monte Carlo. Obiettivo è confrontare il livello di significatività effettivo con quello empirico della statistica test W_p e W^* , i risultati di W^{**} sono stati omessi in quanto in gran parte negativi. Entrambi i test accettano l'ipotesi nulla se il valore osservato è inferiore al percentile di livello $1 - \alpha$ di un Chi-quadrato con 2 gradi di libertà, con $\alpha = 0.1, 0.05, 0.01$.

I risultati preliminari di simulazione forniscono le seguenti indicazioni: per un numero elevato di parametri di disturbo ($n = 50$), il comportamento dell'usuale statistica W_p degenera. Anche la statistica W^* non fornisce risultati del tutto soddisfacenti. Una ragione per giustificare questo comportamento inatteso è dovuto essenzialmente a problemi di instabilità numerica

Tabella 3.2: Risultati preliminari di simulazione

n	Stat.	$1 - \alpha$		
		0.90	0.95	0.99
50	W_p	0.1451	0.2314	0.5320
	W^*	0.3859	0.4613	0.5943

legati alla s.m.v. e alle derivate seconde della verosimiglianza (cioè all'inversa dell'informazione osservata), forse a causa dell'elevato numero dei parametri di disturbo e alla complessità delle formule in oggetto. Ovviamente ulteriori studi di simulazione per valutare il comportamento di W_p e W^* devono essere effettuati.

3.2.2 Esempio 2: Regressione Logistica

In questo esempio, applicheremo le statistiche test di interesse ad un dataset di dati reali. Esattamente si considerano i dati **birthwt** contenuti nella libreria MASS di R, che però abbiamo ridefinito nel dataset *peso*, in cui alcune variabili sono state opportunamente ridefinite (cfr. Bortot, Ventura, Salvan [2], Es 5.3.2). I dati si riferiscono al peso di neonati di 189 nascite in un ospedale americano .

Il dataset è così formato:

```

      low age lwt smoke.fat ptl.fat ht.fat ui.fat ftv.fat race.fat
1      0  19 182          0      0      0      1      0      2
2      0  33 155          0      0      0      0      2      3
3      0  20 105          1      0      0      0      1      1
4      0  21 108          1      0      0      1      2      1
5      0  18 107          1      0      0      1      0      1
...
187    1  23  94          1      0      0      0      0      3
188    1  17 142          0      0      1      0      0      2

```

189 1 21 130 1 0 1 0 2 1

dove le variabili sono:

low	peso minore di 2,5 kg: 1 sì, 0 no;
age	età della madre in anni;
lwt	peso della madre al tempo dell'ultimo periodo metruale (libre);
smoke.fat	madre fumatrice: 1 sì, 0 no;
ptl.fat	precedenti parti prematuri: 1 presenti, 0 assenti;
ht.fat	iperstensione: 1 sì, 0 no;
ui.fat	irritabilità uterina: 1 sì, 0 no;
ftv.fat	livelli per il numero di visite mediche nel primo trimestre: 0 nessuna, 1 una, 2 due o più;
race.fat	razza: 1 bianca, 2 nera, 3 altro.

La variabile risposta è la variabile dicotomica **low**. Per modellare tale variabile in funzione dei regressori si può utilizzare una famiglia esponenziale di tipo binomiale con funzione legame canonico (*logit*), la cui funzione di verosimiglianza è

$$L(\phi) = \exp\left\{ \sum_{i=1}^n y_i \log \phi_i + \sum_{i=1}^n (1 - y_i) \log(1 - \phi_i) \right\},$$

con $\phi_i = \frac{e^{\beta^T x_i}}{1 + e^{\beta^T x_i}}$, $i = 1, \dots, n$. Operate le opportune sostituzioni si ottiene che la log-verosimiglianza per β è

$$l(\beta) = \sum_{j=1}^p \beta_j \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n \log(1 + e^{\beta^T x_i}) = \beta \cdot t - \sum_{i=1}^n \log(1 + e^{\beta^T x_i}),$$

dove $\beta = (\beta_1, \dots, \beta_p)$ e $t = (t_1, \dots, t_p)$ con $t_j = \sum_{i=1}^n y_i x_{ij}$.

Come prima cosa allora utilizziamo un GLM logistico, per modellare i dati, con formula del modello

CAPITOLO 3. UN'APPLICAZIONE NELLE FAMIGLIE
ESPONENZIALI

```
low ~ age + lwt + smoke.fat + ptl.fat + ht.fat + ui.fat +  
      + ftv.fat + race.fat
```

Il summary di tale modello fornisce:

Call:

```
glm(formula = low ~ age + lwt + smoke.fat + ptl.fat + ht.fat +  
      ui.fat + ftv.fat + race.fat, family = binomial, data = peso)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7038	-0.8068	-0.5008	0.8836	2.2152

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.82302	1.24471	0.661	0.50848
age	-0.03723	0.03870	-0.962	0.33602
lwt	-0.01565	0.00708	-2.211	0.02705 *
smoke.fat1	0.75553	0.42502	1.778	0.07546 .
ptl.fat1	1.34376	0.48062	2.796	0.00518 **
ht.fat1	1.91317	0.72074	2.654	0.00794 **
ui.fat1	0.68019	0.46434	1.465	0.14296
ftv.fat1	-0.43638	0.47939	-0.910	0.36268
ftv.fat2	0.17901	0.45638	0.392	0.69488
race.fat2	1.19241	0.53597	2.225	0.02609 *
race.fat3	0.74069	0.46174	1.604	0.10869

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 234.67 on 188 degrees of freedom
Residual deviance: 195.48 on 178 degrees of freedom

CAPITOLO 3. UN'APPLICAZIONE NELLE FAMIGLIE ESPONENZIALI

AIC: 217.48

Number of Fisher Scoring iterations: 4

Potrebbe essere di interesse testare la nullità dei fattori della variabile **race**. A tale scopo si adatta il modello ridotto e il summary fornisce:

Call:

```
glm(formula = low ~ age + lwt + smoke.fat + ptl.fat + ht.fat +  
      ui.fat + ftv.fat, family = binomial, data = peso)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.8464	-0.7782	-0.5696	0.9196	2.2400

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.596633	1.123174	1.422	0.15516
age	-0.047471	0.037303	-1.273	0.20317
lwt	-0.014603	0.006776	-2.155	0.03114 *
smoke.fat1	0.390995	0.358584	1.090	0.27554
ptl.fat1	1.449682	0.473270	3.063	0.00219 **
ht.fat1	1.925831	0.712348	2.703	0.00686 **
ui.fat1	0.640657	0.463715	1.382	0.16710
ftv.fat1	-0.589018	0.466832	-1.262	0.20704
ftv.fat2	0.079322	0.445762	0.178	0.85876

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 234.67 on 188 degrees of freedom
Residual deviance: 201.23 on 180 degrees of freedom
AIC: 219.23

Number of Fisher Scoring iterations: 4

Per testare l'ipotesi $H_0 : \beta_{rece.fat2} = \beta_{race.fat3} = 0$, il valore osservato di W_p è pari a 5.751, come si può constatare dalla seguente analisi della devianza

Analysis of Deviance Table

Model 1: low ~ age + lwt + smoke.fat + ptl.fat + ht.fat
+ ui.fat + ftv.fat

Model 2: low ~ age + lwt + smoke.fat + ptl.fat + ht.fat + ui.fat
+ ftv.fat + race.fat

	Resid. Df	Resid. Dev	Df	Deviance
1	180	201.227		
2	178	195.476	2	5.751

Il valore $W_p^{oss} = 5.751$ conduce all'accettazione dell'ipotesi nulla per livello di significatività $1 - \alpha = 0.95$. Invece, applicando le correzioni di ordine superiore (2.8) e (2.9) si ottengono i seguenti valori osservati delle statistiche test:

W*	W**
7.133743	7.059364

che conducono invece al rifiuto dell'ipotesi nulla. Si può osservare che il test alla Wald, sulla significatività dei parametri dei fattori di race, non portava all'accettazione dell'ipotesi $H_0 : \beta_{race.fat2} = 0$.

La Tabella 3.3, riassume i risultati di un preliminare studio di simulazione basato su 1000 repliche Monte Carlo. I dati sono generati sotto il modello ridotto.

Tabella 3.3: Risultati delle replicazioni

Stat.	$1 - \alpha$		
	0.90	0.95	0.99
W_p	0.0331	0.0927	0.3532
W^*	0.2759	0.4105	0.6357
W^{**}	0.2803	0.3973	0.6048

Anche in questo caso i risultati della simulazione non sono soddisfacenti. Ciò può essere dovuto a problemi di instabilità numeriche incontrate nel calcolo di γ , che definisce le statistiche 2.8 e 2.9. Questi problemi sono riscontrabili anche nei grafici riportati in Figura 3.1, in cui l'andamento di W^* e W^{**} è più irregolare di quello di W_p .

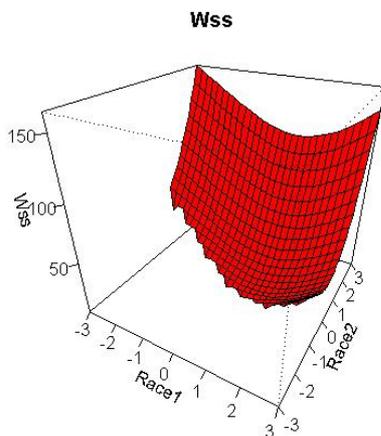
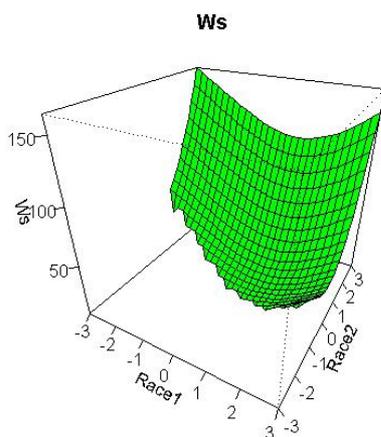
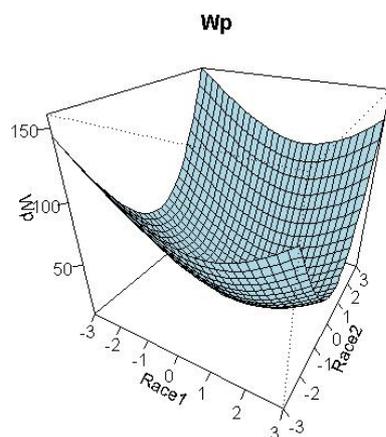
3.3 Conclusioni

L'obiettivo di questa tesi era di implementare computazionalmente e di studiare il comportamento di tecniche basate su metodi asintotici di ordine superiore nel ambito delle famiglie esponenziali con parametro di interesse multidimensionale. I risultati raggiunti nei due esempi presentati evidenziano delle difficoltà nel calcolo del termine di correzione γ , essenziale per la stima delle statistiche W^* (2.8) e W^{**} (2.9). Le difficoltà sono legate soprattutto:

- Instabilità numerica, delle componenti di γ , che nel ordine del 50% dei casi considerati rende impossibile la stima di (2.8) e (2.9).
- Valori assunti da γ portano a valori negativi delle statistiche (2.8) e (2.9), ma mentre per la statistica (2.8) questo caso era stato contemplato (si veda Skovgaard [19]), per la (2.9) questa cosa è del tutto inattesa.

Alla luce di tutto ciò si ritiene che non sia così immediato e computazionalmente conveniente applicare le statistiche W^* e W^{**} .

Figura 3.1: Grafici della W_p , della W^* e della W^{**}



Appendice A

Chiarimenti

A.1 Giustificazione per la sufficienza di $(\hat{\theta}, a)$

Per $\theta, \theta_0 \in \Theta$,

$$p_{\hat{\theta}, A}(\hat{\theta}, a; \theta_0) = \exp\{l(\theta_0; \hat{\theta}, a) - l(\theta; \hat{\theta}, a)\} p_{\hat{\theta}, A}(\hat{\theta}, a; \theta).$$

Dividendo ambo i membri per $p_A(a)$, otteniamo

$$\frac{p_{\hat{\theta}, A}(\hat{\theta}, a; \theta_0)}{p_A(a)} = \exp\{l(\theta_0; \hat{\theta}, a) - l(\theta; \hat{\theta}, a)\} \frac{p_{\hat{\theta}, A}(\hat{\theta}, a; \theta)}{p_A(a)}.$$

e pertanto

$$p_{\hat{\theta}|A=a}(\hat{\theta}; \theta_0) = \exp\{l(\theta_0; \hat{\theta}, a) - l(\theta; \hat{\theta}, a)\} p_{\hat{\theta}, A=a}(\hat{\theta}; \theta).$$

Si può ora approssimare $p_{\hat{\theta}|A=a}(\hat{\theta}; \theta)$ con la densità normale con media θ e matrice di covarianza $\hat{\mathbf{j}}^{-1}$. E dato che θ è arbitrario, si può porre $\theta = \hat{\theta}$, ottenendo

$$p_{\hat{\theta}|A=a}(\hat{\theta}; \theta_0) = e^{\{l(\theta_0; \hat{\theta}, a) - l(\hat{\theta}; \hat{\theta}, a)\}} (2\pi)^{-p/2} |\hat{\mathbf{j}}|^{1/2} \{1 + O(n^{-1})\}.$$

A.2 Giustificazione per la normalità asintotica della r^*

Se la distribuzione di $\hat{\theta}|A = a$ è continua, integrando la formula (2.2) rispetto a $\hat{\theta}$ si ottiene, con $c = c(\theta_0, a)$ e $\hat{j} = j(\hat{\theta}; \hat{\theta}, a)$,

$$Pr_{\theta_0}(\hat{\theta} \leq \hat{\theta}_0 | A = a) \doteq c \int_{-\infty}^{\hat{\theta}_0} \exp \left\{ -\frac{1}{2} \left[r - \frac{1}{r} \log(r/u) \right]^2 \right\} dr.$$

Con l'ulteriore cambiamento di variabile $r^* = r - \frac{1}{r} \log \frac{r}{u}$, il cui jacobiano risulta $1 + O(n^{-1})$ in una regione i scarto normale, si ha

$$Pr_{\theta_0}(r \leq r_0 | A = a)_{\theta_0}(r^* \leq r_0^* | A = a) \doteq c \int_{-\infty}^{r_0^*} \exp \left\{ -\frac{1}{2} (r^*)^2 \right\} dr^*.$$

Pertanto r^* ha distribuzione nulla approssimata $\mathcal{N}(0, 1)$, sia condizionata-mente ad a sia marginalmente.

Appendice B

Codici Esempi

B.1 Esempio 1

```
## Funzione di verosimiglianza
lv.exp<-function(par,s1,s2){
  tau<-par[1:length(s1)]
  zeta<-par[-(1:length(s1))]
  c<-0
  for (i in 1:length(zeta)){
    for(j in 1:length(tau)) c<-c+log(tau[j] + zeta[i])
  }
  sum(tau*s1)+sum(zeta*s2)+c
}

lvn.exp<-function(par,s1,s2){
  -lv.exp(par,s1,s2)
}

##### funzioni per la stima vincolata di zeta dato tao
```

```
lv2.exp<-function(zeta,tau,s1,s2){
c<-0
for (i in 1:length(zeta)){
for(j in 1:length(tau)) c<-c+log(tau[j] + zeta[i])
}
sum(tau*s1)+sum(zeta*s2)+c
}
```

```
nlv2.exp<-function(zeta,tau,s1,s2){
-lv2.exp(zeta,tau,s1,s2)
}
```

```
max.lv.tau<-function(tau.0,zeta,s1,s2){
stime<-optim(zeta,fn=nlv2.exp,tau=tau.0,s1=s1,s2=s2)
stime$par
}
```

funzioni per la verosimiglianza profilo di tau

```
lvptau.exp<-function(tau,zeta,s1,s2){
c<-0
for (i in 1:length(zeta)){
for(j in 1:length(tau)) c<-c+log(tau[j] + zeta[i])
}
sum(tau*s1)+sum(zeta*s2)+c
}
```

```
nlvptau.exp<-function(tau,zeta,s1,s2){
-lvptau.exp(tau,zeta,s1,s2)
}
```

funzioni che calcolano le matrici delle derivate prime e seconde della veros.

```
der1a<-function(t,z,st,sz){
p<-length(t)
n<-length(z)
x<-rep(0,n+p)
for (j in 1:p){
for(i in 1:n){
x[j]<-x[j]+(1/(t[j]+z[i]))-sum(1/(st[j]+sz[i])))
}
}
```

```
for(i in 1:n){
for (j in 1:p){
x[p+i]<-x[p+i]+1/(t[j]+z[i])-1/(st[j]+sz[i])
}
}
x
}
```

```
der2a<-function(t,z){
p<-length(t)
n<-length(z)
x<-array(0,dim=c(n+p,n+p))
for (j in 1:p){
for(i in 1:n){
x[j,j]<-x[j,j]-1/(t[j]+z[i])^2
x[p+i,j]<-x[j,(p+i)]<--1/(t[j]+z[i])^2
}}
}
```

```
for (i in (p+1):(p+n)) {
for(j in 1:p){
x[i,i]<-x[i,i]-1/(t[j]+z[(i-p)])^2
}
}
```

```

x
}

jmat <- function(t,z){-der2a(t,z)}

jmatinv <- function(t,z){chol2inv(jmat(t,z))}

der1a.camp<-function(st,sz,t,z){
p<-length(st)
n<-length(sz)
x<-rep(0,n+p)
for (j in 1:p){
for(i in 1:n)
x[j]<-x[j]+st[j]*(1/(t[j]+z[i])^2)+sz[i]*1/(t[j]+z[i])^2
}

for(i in 1:n){
for(j in 1:p){
x[p+i]<-x[p+i]+st[j]*1/(t[j]+z[i])^2+sz[i]*1/(t[j]+z[i])^2
}
}
x
}

##### Funzione Principale #####

Wexp<-function(dati,interesse=c(0,1),tau0=NULL,zeta0=NULL){

n<-dim(dati)[1]
p<-dim(dati)[2]
t<-rep(0,p)
u<-rep(0,n)
t<--apply(dati,2,sum)

```

APPENDICE B. CODICI ESEMPI

```
u<--apply(dati,1,sum)
tau.inz<--n/t
zeta.inz<--1/u
theta<-c(tau.inz,zeta.inz)
smv.theta<-optim(theta,fn=lvn.exp,s1=t,s2=u)$par
smv.tau<-smv.theta[1:p]
smv.zeta<-smv.theta[(p+1):(n+p)]
if(interesse==0){
W<-2*(lvptau.exp(smv.tau,smv.zeta,t,u)-lvptau.exp(tau0,zeta0,t,u))
W
}
if(interesse==1){
zeta.tilde<-optim(zeta.inz,fn=nlv2.exp,tau=tau0,s1=t,s2=u)$par
theta.tilde<-c(tau0,zeta.tilde)
Wp<-2*(lv.exp(smv.theta,t,u)-lvptau.exp(tau0,zeta.tilde,t,u))

gamma<-sqrt(1/det(jmat(smv.tau,smv.zeta)) * det(jmat(tau0,zeta.tilde))) *
  (der1a(tau0,zeta.tilde,smv.tau,smv.zeta)%%jmatinv(tau0,zeta.tilde)%%
  der1a(tau0,zeta.tilde,smv.tau,smv.zeta))^(p/2)/(Wp^(p/2-1) *
  (smv.theta-theta.tilde)%%der1a(tau0,zeta.tilde,smv.tau,smv.zeta))

Ws<-Wp*(1-log(gamma)/Wp)^2
Wss<-Wp-2*log(gamma)
c(Wp,Ws,Wss)
}
}

creadati<-function(n,p,tau,zeta){
data<-matrix(0,n,p)
for(j in 1:p){
for(i in 1:n){
data[i,j]<-rexp(1,(tau[j]+zeta[i]))
}
}
}
```

```

}
}
data
}

#####
##### LA SIMULAZIONE #####
#####

nr<-10000  ### numerosità della simulazione
n<-10      ### numerosità del campione
p<-2      ### n. delle popolazioni a confronto

Wp10<-Wp30<-Wp50<-rep(0,nr)
Ws10<-Ws30<-Ws50<-rep(0,nr)
Wss10<-Wss30<-Wss50<-rep(0,nr)

tau<-rep(1,p)
zeta<-1:n

for(r in 1:nr){
print(r)
dati<-creadati(n,p,tau,zeta)
Wp10[r]<-Wexp(dati,int=1,tau,zeta)[1]
Ws10[r]<-Wexp(dati,int=1,tau,zeta)[2]
Wss10[r]<-Wexp(dati,int=1,tau,zeta)[3]
}

n<-30
for(r in 1:nr){
print(r)
dati<-creadati(n,p,tau,zeta)
Wp30[r]<-Wexp(dati,int=1,tau,zeta)[1]

```

APPENDICE B. CODICI ESEMPI

```
Ws30[r]<-Wexp(dati,int=1,tau,zeta)[2]
Wss30[r]<-Wexp(dati,int=1,tau,zeta)[3]
}

n<-50
for(r in 1:nr){
print(r)
dati<-creadati(n,p,tau,zeta)
Wp50[r]<-Wexp(dati,int=1,tau,zeta)[1]
Ws50[r]<-Wexp(dati,int=1,tau,zeta)[2]
Wss50[r]<-Wexp(dati,int=1,tau,zeta)[3]
}

perc90<-qchisq(.9,p)
perc95<-qchisq(.95,p)
perc99<-qchisq(.99,p)

### Risultati con n=10

cond<- !is.na(Ws10)&Ws10>0 ### individua quali scartare

l<-length(Ws10[cond])
a<-length(Wp10[cond & Wp10>0 & Wp10<perc90])/l
b<-length(Ws10[cond & Ws10>0 & Ws10<perc90])/l
c<-length(Wss10[cond & Wss10>0 & Wss10<perc90])/l
d<-length(Wp10[cond & Wp10>0 & Wp10<perc95])/l
e<-length(Ws10[cond & Ws10>0 & Ws10<perc95])/l
f<-length(Wss10[cond & Wss10>0 & Wss10<perc95])/l
g<-length(Wp10[cond & Wp10>0 & Wp10<perc99])/l
h<-length(Ws10[cond & Ws10>0 & Ws10<perc99])/l
i<-length(Wss10[cond & Wss10>0 & Wss10<perc99])/l

tabella10<-data.frame("stat."=c("Wp","Ws","Wss"),
```

```

"0.90"=c(a,b,c),"0.95"=c(d,e,f),"0.99"=c(g,h,i))

### Risultati con n=30

cond<- !is.na(Ws30)&Ws30>0  ### individua quali scartare

l<-length(Ws30[cond])
a<-length(Wp30[cond & Wp30>0 & Wp30<perc90])/l
b<-length(Ws30[cond & Ws30>0 & Ws30<perc90])/l
c<-length(Wss30[cond & Wss30>0 & Wss30<perc90])/l
d<-length(Wp30[cond & Wp30>0 & Wp30<perc95])/l
e<-length(Ws30[cond & Ws30>0 & Ws30<perc95])/l
f<-length(Wss30[cond & Wss30>0 & Wss30<perc95])/l
g<-length(Wp30[cond & Wp30>0 & Wp30<perc99])/l
h<-length(Ws30[cond & Ws30>0 & Ws30<perc99])/l
i<-length(Wss30[cond & Wss30>0 & Wss30<perc99])/l

tabella30<-data.frame("stat."=c("Wp","Ws","Wss"),
  "0.90"=c(a,b,c),"0.95"=c(d,e,f),"0.99"=c(g,h,i))

### Risultati con n=50

cond<- !is.na(Ws50)&Ws50>0

l<-length(Ws50[cond])
a<-length(Wp50[cond & Wp50>0 & Wp50<perc90])/l
b<-length(Ws50[cond & Ws50>0 & Ws50<perc90])/l
c<-length(Wss50[cond & Wss50>0 & Wss50<perc90])/l
d<-length(Wp50[cond & Wp50>0 & Wp50<perc95])/l
e<-length(Ws50[cond & Ws50>0 & Ws50<perc95])/l
f<-length(Wss50[cond & Wss50>0 & Wss50<perc95])/l

```

```
g<-length(Wp50[cond & Wp50>0 & Wp50<perc99])/1
h<-length(Ws50[cond & Ws50>0 & Ws50<perc99])/1
i<-length(Wss50[cond & Wss50>0 & Wss50<perc99])/1

tabella50<-data.frame("stat."=c("Wp", "Ws", "Wss"),
  "0.90"=c(a,b,c), "0.95"=c(d,e,f), "0.99"=c(g,h,i))
```

B.2 Esempio 2

```
## Esempio Binomiale con regr. Logistica
```

```
lv.bin<-function(par,ydat,xdat){
n<-length(ydat)
t<-t(xdat)%*%ydat
c<-0
for(i in 1:n){
c<-c+log(1+exp(par%*%x[i,]))
}
par%*%t-c
}
```

```
nlv.bin<-function(par,ydat,xdat){
-lv.bin(par,ydat,xdat)
}
```

```
der1a<-function(x,stima,par){ ###stima=smv , par=valore parametro
n<-dim(x)[1]
p<-dim(x)[2]
a<-rep(0,p)
for(j in 1:p){
for(i in 1:n){
a[j]<-a[j]+(exp(stima%*%t(x[i,]))/(1+exp(stima%*%t(x[i,])))-
  exp(par%*%t(x[i,]))/(1+exp(par%*%t(x[i,]))))*x[i,j]
```

```

}
}
a
}

der2a<-function(par,x){
n<-dim(x)[1]
p<-dim(x)[2]
a<-matrix(0,p,p)
for(j in 1:p){
for(r in 1:p){
for(i in 1:n){
a[j,r]<-a[j,r]+(exp(t(par)%*%x[i,])/(1+exp(t(par)%*%x[i,]))-
exp(2*t(par)%*%x[i,])/(1+exp(t(par)%*%x[i,]))^2
)*x[i,j]*x[i,r]
}
}
}
-a
}

jmat <- function(stima,x){-der2a(stima,x)}

jmatinv<-function(stima,x){solve(jmat(stima,x))}

der1a.camp<-function(stima,par,x){ ###stima=smv , par=valore parametro
a<-par%*%jmat(stima,x)
a
}

#####

library(MASS)

```

APPENDICE B. CODICI ESEMPI

```
attach(birthwt)
smoke.fat<-as.factor(smoke)
race.fat<-as.factor(race)
ptl.fat<-as.factor((ptl>0)*1)

inter<-ftv
inter[inter>=2]<-2
ftv.fat<-as.factor(inter)

ht.fat<-as.factor(ht)
ui.fat<-as.factor(ui)
peso<-data.frame(low,age,lwt,smoke.fat,ptl.fat,ht.fat,ui.fat,ftv.fat,race.fat)
detach("birthwt")

y<-peso[,1]

stima<-lm(low~age+lwt+smoke.fat+ptl.fat+ht.fat+ui.fat+ftv.fat+race.fat,
  data=peso,x=T)

x<-stima$x

theta<-stima$coeff

stima1<-glm(low~age+lwt+smoke.fat+ptl.fat+ht.fat+ui.fat+ftv.fat+race.fat,
  family=binomial,data=peso)

summary(stima1)

smv.theta<-stima1$coeff

##### Fattore race nullo!!!
k<-2
```

```

stima2<-glm(low~age+lwt+smoke.fat+ptl.fat+ht.fat+ui.fat+ftv.fat,
  family=binomial,data=peso)

summary(stima2)

anova(stima2,stima1)

tilde<-stima2$coeff

theta.tilde<-c(tilde,race.fat2=0,race.fat3=0)

Wp<-2*(lv.bin(smv.theta,y,x)-lv.bin(theta.tilde,y,x))

gamma<-sqrt(1/det(jmat(smv.theta,x)) * det(jmat(theta.tilde,x)))*
  (der1a(x,smv.theta,theta.tilde)%*%jmatinv(theta.tilde,x)%*%
  der1a(x,smv.theta,theta.tilde))^(k/2)/(Wp^(k/2-1) *
  (smv.theta-theta.tilde)%*%der1a(x,smv.theta,theta.tilde))

c(gamma,gamma2)
Ws<-Wp*(1-log(gamma2)/Wp)^2
Wss<-Wp-2*log(gamma2)

c(Wp,Ws,Wss)

#####
##### Graffici #####
#####

n<-50 ### numerosità del campione

race1<-seq(-3,3,length=n)
race2<-race1

```

APPENDICE B. CODICI ESEMPI

```
Wliv<-Wsliv<-Wssliv<-matrix(0,n,n)

for(i in 1:n){
  print(i)
  for(j in 1:n){
    theta.tilde<-c(tilde,race1[i],race2[j])
    Wliv[i,j]<-2*(lv.bin(smv.theta,y,x)-lv.bin(theta.tilde,y,x))
    gamma<-sqrt(1/det(jmat(smv.theta,x)) * det(jmat(theta.tilde,x)))*
      (der1a(x,smv.theta,theta.tilde)%*%jmatinv2(theta.tilde,x)%*%
      der1a(x,smv.theta,theta.tilde))^(k/2)/(Wliv[i,j]^(k/2-1) *
      (smv.theta-theta.tilde)%*%der1a(x,smv.theta,theta.tilde))
    Wsliv[i,j]<-Wliv[i,j]*(1-log(gamma)/Wliv[i,j])^2
    Wssliv[i,j]<-Wliv[i,j]-2*log(gamma)

  }
}

jpeg("Wp.jpg")
persp(race1,race2,Wliv,theta=30,phi=30,col="lightblue",xlab="Race1",
      ylab="Race2",zlab="Wp",main="Wp",ticktype="detailed")
dev.off()

jpeg("Wpliv.jpg")
contour(race1,race2,Wliv,xlab="Race1",ylab="Race2",main="Wp")
dev.off()

jpeg("Ws.jpg")
persp(race1,race2,Wsliv,col="green",theta=30,phi=30,xlab="Race1",
      ylab="Race2",zlab="Ws",main="Ws",ticktype="detailed")
dev.off()

jpeg("Wsliv.jpg")
```

```

contour(race1,race2,Wsliv,xlab="Race1",ylab="Race2",main="Ws")
dev.off()

jpeg("Wss.jpg")
persp(race1,race2,Wssliv,theta=30,phi=30,col="red",xlab="Race1",
      ylab="Race2",zlab="Wss",main="Wss",ticktype="detailed")
dev.off()

jpeg("Wssliv.jpg")
contour(race1,race2,Wssliv,xlab="Race1",ylab="Race2",main="Wss")
dev.off()

#####
#### Simulazione #####
#####

n<-length(y)
nr<-1000
k<-2
m<-exp(theta.tilde*%t(x))/(1+exp(theta.tilde*%t(x)))

Wp<-Ws<-Wss<-rep(0,nr)

for(r in 1:nr){
print(r)
y.dat<-NULL
for(i in 1:n) y.dat[i]<-rbinom(1,1,m[i])

stimal<-glm(y.dat~age+lwt+smoke.fat+ptl.fat+ht.fat+ui.fat+ftv.fat+
+ race.fat,family=binomial,data=peso)

```

APPENDICE B. CODICI ESEMPI

```
s.theta<-stima1$coeff

stima2<-glm(low~age+lwt+smoke.fat+ptl.fat+ht.fat+ui.fat+ftv.fat,
  family=binomial,data=peso)

tilde<-stima2$coeff

theta.t<-c(tilde,race.fat2=0,race.fat3=0)

Wp[r]<-2*(lv.bin(s.theta,y.dat,x)-lv.bin(theta.t,y.dat,x))

gamma<-sqrt(1/det(jmat(s.theta,x)) * det(jmat(theta.t,x)))*
  (der1a(x,s.theta,theta.t)%*%jmatinv(theta.t,x)%*%
    der1a(x,s.theta,theta.t))^(k/2)/
  (Wp[r]^(k/2-1) * (s.theta-theta.t)%*%der1a(x,s.theta,theta.t))

Ws[r]<-Wp[r]*(1-log(gamma)/Wp[r])^2
Wss[r]<-Wp[r]-2*log(gamma)
}

perc90<-qchisq(.9,2)
perc95<-qchisq(.95,2)
perc99<-qchisq(.99,2)

cond<- !is.na(Ws)&Ws>0 ### individua quali scartare

l<-length(Ws[cond])
a<-length(Wp[cond & Wp>0 & Wp<perc90])/l
b<-length(Ws[cond & Ws>0 & Ws<perc90])/l
c<-length(Wss[cond & Wss10>0 & Wss<perc90])/l
d<-length(Wp[cond & Wp>0 & Wp<perc95])/l
e<-length(Ws[cond & Ws>0 & Ws<perc95])/l
```

```
f<-length(Wss[cond & Wss>0 & Wss<perc95])/1
g<-length(Wp[cond & Wp>0 & Wp<perc99])/1
h<-length(Ws[cond & Ws>0 & Ws<perc99])/1
i<-length(Wss[cond & Wss>0 & Wss<perc99])/1

tabella<-data.frame("stat."=c("Wp", "Ws", "Wss"),
  "0.90"=c(a,b,c), "0.95"=c(d,e,f), "0.99"=c(g,h,i))
```

Bibliografia

- [1] Azzalini A. (2000). Inferenza statistica: Una presentazione basata sul concetto di verosimiglianza. Springer
- [2] Bortot, P. Ventura, L. Salvan, A (2000). Inferenza statistica: Applicazioni con S-PLUS e R. CEDAM.
- [3] Barndorff-Nielsen, O.E. & D.R. Cox (1994). Inference and Asymptotics. Chapman and Hall. London.
- [4] Barndorff-Nielsen, O.E. & D.R. Cox (1995). Prediction and Asymptotics. Chapman and Hall. London.
- [5] Barndorff-Nielsen, O.E.(1978). Information and Exponential Families in Statistical Theory. Wiley Sons.
- [6] Barndorff-Nielsen, O.E. (1991). Modified signed log ratio. *Biometrika*, 78, 557-563.
- [7] Barndorff-Nielsen, O.E.(1983). On formula for the distribution of the maximum likelihood estimator. *Biometrika*, 70, 343-365.
- [8] Barndorff-Nielsen, O. E. (1986). Inference on Full or Partial Parameters Based on the Standardized Signed Log Likelihood Ratio, *Biometrika*, 73, 307-322.
- [9] Barndorff-Nielsen, O. E. (1994). Adjusted Versions of Profile Likelihood and Directed Likelihood, and Extended Likelihood. *Journal of the Royal Statistical Society. B*, 56, 125-140.

- [10] Goossens, M. Mittelbach, F. Saramin, A. (1994). The Latex Companion. Addison-Wesley Publishing Company.
- [11] Iacus, S.M. Masarotto, G. (2003). Laboratorio di Statistica con R. McGraw-Hill.
- [12] Pace, L. & Salvan A. (1996). Teoria della Statistica: metodi, modelli, approssimazioni asintotiche. Cedam.
- [13] Pace, L. & Salvan A. (1999). Point estimation based on confidence intervals: Exponential Families. *J.Statist. Comput. Simul.* 64: 1-21.
- [14] Pierce, D.A. & Peters, D.(1992). Practical use of higher order asymptotics for multiparameter exponential families. *J.Roy.Statist.Soc.Ser. B*,54, 701-737.
- [15] Pierce, D.A. & Peters, D. (1994) Higher order asymptotics and the likelihood principle: one parameter models. *Biometrika*, 81, 1-10.
- [16] Sartori, N. & Bellio, R.(2003). Extending conditional likelihood in models for stratified binary data. *Statistical Methods & Applications.* 12: 121-132.
- [17] Sartori, N.(2003). Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika*, 90,533-549.
- [18] Severini, T.A. (2000).Likelihood Methods in Statistics. University Press. Oxford.
- [19] Skovgaard, I.M. (2001). Likelihood asymptotics. *Scand. J. Statist.* 28, 3-32.