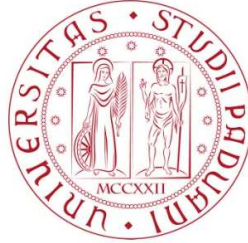


UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE
CORSO DI LAUREA TRIENNALE IN
STATISTICA PER L'ECONOMIA E L'IMPRESA



RELAZIONE FINALE

Verosimiglianza robusta in modelli lineari generalizzati

Relatrice Prof.ssa Alessandra Salvan
Dipartimento di Scienze Statistiche

Laureanda Marianna Viti
Matricola 1220249

Anno Accademico 2021/2022

Indice

Introduzione	1
1 Inferenza di verosimiglianza	3
1.1 Introduzione	3
1.2 Modello statistico	3
1.3 Funzione di verosimiglianza	4
1.4 Stima di massima verosimiglianza e quantità di verosimiglianza	4
1.5 Rapporto di verosimiglianza	6
1.6 Verosimiglianza profilo	6
2 Funzione di verosimiglianza robusta	9
2.1 Introduzione	9
2.2 Proprietà della funzione di verosimiglianza	9
2.3 Condizioni di robustezza della funzione di verosimiglianza	10
2.4 Trasformazione robusta della funzione di verosimiglianza	12
2.5 Trasformazione robusta della funzione di verosimiglianza profilo	13
2.6 Stima sandwich della varianza asintotica	14
2.7 Estensioni della trasformazione robusta della log-verosimiglianza	15
3 Applicazione ad alcuni modelli di campionamento casuale semplice	17
3.1 Introduzione	17
3.2 Applicazione a dati di conteggio	17
3.2.1 Dati di conteggio indipendenti ed identicamente distribuiti	18
3.3 Applicazione a dati binari	19
3.3.1 Dati binari raggruppati indipendenti e identicamente distribuiti	19
3.4 Studi di simulazione	21
3.4.1 Assunzione di distribuzione Poisson per dati generati da una bi- nomiale negativa	21
3.4.2 Assunzione di distribuzione binomiale per dati generati da una beta-binomiale	28
4 Applicazione ad alcuni modelli lineari generalizzati	33
4.1 Introduzione	33
4.2 Modelli lineari generalizzati	33

4.2.1	Famiglie di dispersione esponenziale	34
4.2.2	Specificazione dei GLM	35
4.3	Funzioni di verosimiglianza robusta nei GLM	35
4.4	Regressione semplice di Poisson con legame logaritmico	36
4.5	Studio di simulazione per la regressione semplice di Poisson con legame logaritmico	38
4.6	Regressione logistica semplice	41
Conclusioni		43
Appendice		47
Bibliografia		67

Introduzione

L'analisi statistica ricopre un importante ruolo all'interno della scienza, in termini di interpretazione dei dati osservati come evidenza. In modelli parametrici, la funzione di verosimiglianza, e in particolare il rapporto di verosimiglianza, forniscono una misura della forza dell'evidenza statistica a supporto di una distribuzione di probabilità rispetto ad un'altra. Le osservazioni, però, anche quando propriamente interpretate, possono indurre in errore, supportando una distribuzione quando l'alternativa è corretta.

Dopo aver proposto dei criteri per determinare la robustezza della funzione di verosimiglianza associata ad un modello statistico parametrico, Royall & Tsou (2003) e, precedentemente, Stafford (1996), hanno introdotto una tecnica per correggere le funzioni di verosimiglianza e di verosimiglianza profilo per renderle robuste, dunque adeguate anche in caso di errata specificazione del modello.

Obiettivo di questo studio è analizzare la trasformazione robusta proposta dagli studiosi sopra citati e ripresa da altri, approfondirne l'utilizzo per alcuni modelli di campionamento casuale semplice e nell'ambito di alcuni modelli lineari generalizzati e verificarne l'efficacia tramite degli studi di simulazione. Si confrontano inoltre le proprietà campionarie della statistica log-rapporto di verosimiglianza robusta con quelle della statistica di Wald che utilizza la stima sandwich della varianza.

Nel primo capitolo vengono richiamati i principali argomenti riguardanti l'inferenza basata sulla funzione di verosimiglianza. Nel secondo viene analizzato approfonditamente quanto introdotto da Royall & Tsou (2003) circa la robustezza e la trasformazione delle funzioni di verosimiglianza. Nel terzo capitolo la correzione proposta viene applicata al caso di campioni di osservazioni indipendenti ed identicamente distribuite attraverso due esempi: il primo si basa sull'assunzione distributiva di tipo Poisson per dei dati di conteggio, nel secondo si analizza il caso di dati binari raggruppati. Nel quarto capitolo la correzione viene calata nell'ambito di alcuni modelli lineari generalizzati, quali il modello di regressione di Poisson con legame logaritmico e il modello di

regressione logistica. Per ogni applicazione, si ricavano le quantità necessarie per calcolare il fattore di correzione e si conducono degli studi di simulazione, con il supporto del software **R**, con l'obiettivo di verificare le proprietà illustrate. Nell'ultimo capitolo si riportano le principali conclusioni dello studio.

Capitolo 1

Inferenza di verosimiglianza

1.1 Introduzione

In questo capitolo viene fornita una breve introduzione all'inferenza statistica basata sulla funzione di verosimiglianza, con particolare attenzione alle quantità di verosimiglianza, alcune loro proprietà e alla funzione di verosimiglianza profilo. Per maggiori approfondimenti si rimanda a Pace & Salvan (2001).

1.2 Modello statistico

Sia $y = (y_1, \dots, y_n)$ un vettore di osservazioni, realizzazione del vettore casuale $Y = (Y_1, \dots, Y_n)$, il cui spazio campionario è un sottoinsieme di uno spazio euclideo, $\mathcal{Y} \subseteq \mathbb{R}^n$. La distribuzione di probabilità di Y , F^0 , costituisce il modello probabilistico generatore dei dati. Al fine di ricostruire tale distribuzione sulla base dei dati e delle conoscenze preliminari, nella fase iniziale dell'inferenza statistica, si specifica un modello statistico \mathcal{F} , cioè una famiglia di distribuzioni di probabilità compatibile con i dati osservati.

Quando gli elementi di \mathcal{F} sono individuati da un parametro θ di dimensione finita, che rappresenta una o più caratteristiche della popolazione da cui provengono i dati, il modello si dice parametrico e si può indicare come

$$\mathcal{F} = \{f_Y(y; \theta), y \in \mathcal{Y}, \theta \in \Theta\}$$

se $\Theta \subseteq \mathbb{R}^p$, con $p \in \mathbb{N}^+$. La funzione $f_Y(\cdot; \cdot)$ da (\mathcal{Y}, Θ) in \mathbb{R}^+ è detta funzione del modello e, per θ fissato, rappresenta una funzione di probabilità nel caso discreto e una funzione di densità nel caso continuo.

Se la densità f_0 di F^0 appartiene a \mathcal{F} , si dice che il modello statistico è correttamente specificato e, se θ è identificabile, il valore θ_0 tale che $f_Y(y; \theta_0) = f^0$ è detto vero valore del parametro.

1.3 Funzione di verosimiglianza

Dato \mathcal{F} , modello statistico parametrico per i dati y con funzione del modello $f_Y(y; \theta)$, dove θ è un parametro p -dimensionale con valori in $\Theta \subseteq \mathbb{R}^p$, fissato il valore osservato y , $f_Y(y; \theta)$ è una funzione di θ . La funzione $L : \Theta \rightarrow \mathbb{R}^+$ definita come

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

con $c(y) > 0$ costante moltiplicativa non dipendente da θ , è detta funzione di verosimiglianza di θ basata sui dati y . Nel caso in cui si assume che le variabili di cui i dati sono realizzazioni siano indipendenti, la funzione di verosimiglianza si ottiene come prodotto delle n distribuzioni marginali $f_{Y_i}(y_i, \theta)$, cioè $L(\theta) = \prod_{i=1}^n f_{Y_i}(y_i, \theta)$.

La trasformazione logaritmica di $L(\theta; y)$ è detta funzione di log-verosimiglianza

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

Per convenzione, $l(\theta) = -\infty$ se $L(\theta) = 0$. Valori del parametro con elevata verosimiglianza hanno anche elevata log-verosimiglianza. Nel caso di indipendenza, $l(\theta) = \sum_{i=1}^n \log f_{Y_i}(y_i, \theta)$.

1.4 Stima di massima verosimiglianza e quantità di verosimiglianza

Si definisce stima di massima verosimiglianza un valore $\hat{\theta} \in \Theta$ tale che $L(\hat{\theta}) \geq L(\theta)$ per ogni $\theta \in \Theta$. Se $\hat{\theta}(y)$ esiste unico con probabilità uno, la variabile casuale $\hat{\theta}(Y)$ è detta stimatore di massima verosimiglianza. Sotto tenui condizioni di regolarità, esso è consistente. Un modello statistico si dice con verosimiglianza regolare se il supporto di Y non dipende dal parametro θ , lo spazio parametrico Θ è un sottoinsieme aperto di \mathbb{R}^p e la funzione di log-verosimiglianza $l(\theta)$ è differenziabile almeno tre volte, con derivate parziali continue in Θ . In questi casi, la stima di massima verosimiglianza si individua

come soluzione dell'equazione di verosimiglianza,

$$l_\theta = 0,$$

dove $l_\theta = (\partial l(\theta)/\partial \theta_1, \dots, \partial l(\theta)/\partial \theta_p)^T$ è il vettore delle derivate parziali prime della funzione di log-verosimiglianza, detto funzione score o di punteggio.

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

$$j(\theta) = j(\theta; y) = -l_{\theta\theta} = \left(-\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \right),$$

è detta matrice dell'informazione osservata. Il suo valore atteso,

$$i(\theta) = E_\theta[j(\theta; Y)],$$

è una matrice $p \times p$, detta informazione attesa o informazione di Fisher. Tale quantità è rilevante poiché ricorre nelle approssimazioni asintotiche delle distribuzioni di diverse quantità. In particolare, per n sufficientemente grande e sotto θ , vero valore del parametro, valgono le distribuzioni asintotiche

$$l_\theta \sim N_p(0, i(\theta))$$

e

$$\hat{\theta} - \theta \sim N_p(0, i(\theta)^{-1}).$$

Nei modelli con verosimiglianza regolare, valgono inoltre le seguenti identità, note come prima e seconda identità di Bartlett:

$$E_\theta[l_\theta(\theta; Y)] = 0 \quad \forall \theta \in \Theta, \quad (1.1)$$

$$E_\theta[l_\theta(\theta; Y)l_\theta(\theta; Y)^T] = i(\theta) \quad \forall \theta \in \Theta. \quad (1.2)$$

Si tratta di risultati esatti: il primo indica che l'equazione di verosimiglianza $l_\theta(\theta, Y) = 0$ è un'equazione di stima non distorta; il secondo, noto anche come identità dell'informazione, indica che la matrice di covarianza della funzione di punteggio coincide con l'informazione attesa.

1.5 Rapporto di verosimiglianza

La funzione di verosimiglianza sintetizza l'informazione sul parametro θ alla luce dei dati osservati. In particolare, non fornisce una valutazione assoluta a sostegno o contro un singolo valore dello spazio parametrico Θ , ma permette di confrontare coppie di valori dello stesso.

Lo strumento che fornisce tale misura è il rapporto di verosimiglianza: dati i valori θ_0 e θ_1 appartenenti a Θ e un'ipotesi H_0 che definisce un modello \mathcal{M}_0 , identificato da θ_0 , e la cui alternativa è H_1 , a cui corrisponde \mathcal{M}_1 , identificato da θ_1 , si afferma che i dati y supportano H_0 contro H_1 se $L(\theta_0) > L(\theta_1)$ e il rapporto di verosimiglianza $L(\theta_0)/L(\theta_1)$ è una statistica test che fornisce una misura di tale evidenza. Nella pratica, si fa spesso riferimento alla trasformata logaritmica della suddetta statistica test, $l(\theta_0) - l(\theta_1)$.

In particolare, data l'ipotesi nulla $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$, si fa riferimento alla statistica test, detta statistica log-rapporto di verosimiglianza,

$$W(\theta_0) = 2\{l(\hat{\theta}) - l(\theta_0)\} \quad (1.3)$$

la cui distribuzione asintotica nulla è, sotto condizioni di regolarità, χ_p^2 . Si può inoltre vedere $W(\theta)$ come una quantità asintoticamente pivotale che permette di costruire regioni di confidenza con livello nominale approssimato pari a $1 - \alpha$ definite dall'insieme

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

dove $\chi_{p;1-\alpha}^2$ è il quantile di ordine $1 - \alpha$ della distribuzione χ_p^2 .

Alcune varianti della statistica test del rapporto di verosimiglianza sono date dalla statistica test di Wald, definita come $W_e(\theta) = (\hat{\theta} - \theta)^T i(\theta) (\hat{\theta} - \theta) \sim \chi_p^2$, e dalla statistica test punteggio o score, $W_u(\theta) = l_\theta^T i(\theta)^{-1} l_\theta \sim \chi_p^2$. Le quantità $W(\theta)$, $W_e(\theta)$ e $W_u(\theta)$ sono asintoticamente equivalenti e, nelle ultime due, la stima di $i(\theta)$ tramite $i(\hat{\theta})$ o $j(\hat{\theta})$ non altera l'approssimazione asintotica.

1.6 Verosimiglianza profilo

Sia \mathcal{F} un modello statistico parametrico per i dati y con funzione del modello $f_Y(y; \theta)$, con $\theta \in \Theta \subseteq \mathbb{R}^p$. Nell'ambito di problemi inferenziali con $p > 1$, si può essere interessati all'inferenza su una parte delle componenti di θ . Consideriamo allora il parametro complessivo come costituito da due componenti: $\theta = (\psi, \lambda)$, dove ψ è un parametro q -dimensionale, $1 \leq q < p$, su cui si desidera fare inferenza, dunque di interesse, e λ

è $(p - q)$ -dimensionale e rappresenta il parametro di disturbo. Si definisce funzione di verosimiglianza profilo per ψ la funzione

$$L_P(\psi) = L_P(\psi; y) = L(\psi, \hat{\lambda}_\psi),$$

dove $\hat{\lambda}_\psi$ è la stima di massima verosimiglianza di λ nel modello con ψ fissato. La funzione $l_P(\psi) = \log L_P(\psi)$ è detta funzione di log-verosimiglianza profilo.

La stima di massima verosimiglianza profilo coincide con la stima di massima verosimiglianza basata sulla funzione $L(\psi, \lambda)$, cioè $L_P(\hat{\psi}) \geq L_P(\psi)$ per ogni ψ . Inoltre, la verosimiglianza profilo è assimilabile, sotto alcuni aspetti, a una verosimiglianza propria.

Capitolo 2

Funzione di verosimiglianza robusta

2.1 Introduzione

Royall & Tsou (2003) esaminano come si comporta la funzione di verosimiglianza quando il modello statistico parametrico non è correttamente specificato. Vengono analizzati criteri per determinare la robustezza della funzione di verosimiglianza e viene quindi introdotta una tecnica per correggere la funzione di verosimiglianza e la funzione di verosimiglianza profilo affinché siano robuste.

In questo capitolo vengono analizzate alcune delle osservazioni mosse dai due studiosi; vengono riportate poi le quantità necessarie al fine di definire i fattori di correzione per le funzioni di verosimiglianza da essi proposti.

2.2 Proprietà della funzione di verosimiglianza

Siano y_1, \dots, y_n realizzazioni di variabili casuali indipendenti Y_1, \dots, Y_n e sia

$$\mathcal{F} = \{f(y; \theta), y \in \mathcal{Y}, \theta \in \Theta\}, \quad (2.1)$$

il modello statistico parametrico per la singola osservazione, dove θ è un parametro p -dimensionale. La funzione di verosimiglianza $L(\theta) \propto \prod_{i=1}^n f(y_i; \theta)$ è la misura dell'evidenza presente nei dati a supporto di θ . Se \mathcal{F} è correttamente specificato e il parametro è identificabile, indichiamo con θ_0 il vero valore del parametro.

In tale contesto, la funzione di verosimiglianza soddisfa due importanti proprietà:

1. per ogni valore $\theta \neq \theta_0$, $P_{\theta_0}\{L(\theta_0)/L(\theta) \rightarrow \infty \text{ per } n \rightarrow \infty\} = 1$;

2. in grandi campioni, se la dimensione del parametro è $p = 1$, la probabilità funzione di θ , $P_{\theta_0}\{L(\theta)/L(\theta_0) \geq k\}$, è approssimata dalla *bump function* (funzione a gobbe), $\Phi\{-c/2 - \log k/c\}$, con $k \geq 1$ costante, Φ funzione di ripartizione della normale standard e $c > 0$ proporzionale alla distanza $\Delta = |\theta_0 - \theta|$ tra θ_0 e θ , secondo la relazione $\Delta = c(n i(\theta_0))^{-1/2}$, dove $i(\theta_0)$ è l'informazione di Fisher calcolata in corrispondenza del vero valore del parametro θ_0 . Se $p > 1$, sia $\theta = \theta_0 + (n i(\theta_0))^{-1/2} c$, dove c è ora un vettore p -dimensionale e $i(\theta_0)$ è la matrice dell'informazione attesa calcolata in θ_0 , la probabilità $P_{\theta_0}\{L(\theta)/L(\theta_0) \geq k\}$ è approssimata da $\Phi\{-\|c\|/2 - \log k/\|c\|\}$ (Royall, 2000).

La prima proprietà si applica a tutti i modelli statistici ed implica che, per ogni valore alternativo di θ fissato e per ogni $k > 1$, la probabilità di forte evidenza in favore del vero valore, $P_{\theta_0}\{L(\theta_0)/L(\theta) \geq k\}$, tende a 1. Quindi, le probabilità di errata evidenza statistica, cioè $P_{\theta_0}\{L(\theta)/L(\theta_0) \geq k\}$, e di debole evidenza in favore di θ_0 , $P_{\theta_0}\{1/k \leq L(\theta_0)/L(\theta) \leq k\}$, tendono entrambe a 0 al divergere di n .

La seconda proprietà si applica ai modelli statistici la cui funzione di verosimiglianza è regolare ed implica che, quando la numerosità campionaria è sufficientemente elevata, la probabilità di errata evidenza statistica con forza almeno pari a k è massimizzata sull'intero spazio parametrico da un valore costante pari a $\Phi[-\{2 \log k\}^{1/2}]$, valore massimo della *bump function*.

2.3 Condizioni di robustezza della funzione di verosimiglianza

Sia Y_i la generica componente di $Y = (Y_1, \dots, Y_n)$, con Y_1, \dots, Y_n indipendenti e identicamente distribuite. Consideriamo il caso in cui la vera distribuzione di Y_i non sia contenuta nel modello (2.1), cioè il caso in cui non vi sia un valore θ_0 tale che $f(\cdot; \theta_0)$ sia la vera distribuzione di Y_i . Supponiamo che la densità di probabilità di Y_i sia realmente $g(\cdot)$ e che θ_g sia il valore di θ che massimizza $E_g[\log\{f(Y_i; \theta)\}]$. Allora $f(\cdot; \theta_g)$ è la densità del modello che è più vicina alla vera densità $g(\cdot)$, nel senso che minimizza la distanza di Kullback-Leibler tra $f(\cdot; \theta)$ e $g(\cdot)$, e θ_g è l'oggetto di inferenza quando la vera densità è $g(\cdot)$. Se, invece, il modello è correttamente specificato e la vera densità di Y_i , $f(\cdot; \theta_0)$, è contenuta nel modello, allora $E_\theta[\log\{f(Y_i; \theta)\}]$ è massimizzato in θ_0 , dunque l'oggetto di inferenza è il vero valore del parametro.

Affinché una funzione di verosimiglianza rimanga valida per l'interpretazione dell'evidenza presente nei dati, anche se \mathcal{F} non è correttamente specificato, è necessario che

l'oggetto dell'inferenza rimanga l'oggetto di interesse anche quando le osservazioni derivano da una distribuzione che non è inclusa nel modello. Questa rappresenta la prima delle due condizioni per definire una verosimiglianza robusta.

Se, ad esempio, l'oggetto di interesse è il valore atteso di Y_i , $E(Y_i)$, e il modello è correttamente specificato, allora esso contiene un valore $\theta_0 = E(Y_i)$, che rappresenta l'oggetto dell'inferenza, e si interpreta la funzione di verosimiglianza associata al modello, $L(\theta_0)$, come misura dell'evidenza presente nei dati a supporto del valore θ_0 . Se, invece, la vera densità di Y_i , $g(\cdot)$, non è contenuta nel modello, allora θ_g , oggetto di inferenza, può non coincidere con l'oggetto di interesse $E(Y_i)$. Si riportano di seguito due esempi che mostrano le due possibili situazioni nel caso di errata specificazione del modello, sviluppando in dettaglio i risultati riportati in Royall & Tsou (2003, paragrafo 2).

Nel caso di una distribuzione normale $N(\theta, \sigma^2)$, con σ^2 fissata, l'identità $\theta_g = E(Y_i)$ è soddisfatta.

Verifica. Data $f(y; \theta) = 1/(\sqrt{2\pi}\sigma) \exp\{-(y - \theta)^2/2\sigma^2\}$, si ha

$$E_g[\log\{f(Y_i; \theta)\}] = E_g[-\log(\sqrt{2\pi}\sigma) - (Y_i - \theta)^2/2\sigma^2] = -E_g(Y_i - \theta)^2/2\sigma^2 + c,$$

con c costante non dipendente da θ . Per ogni valore finito di σ^2 , tale espressione è massimizzata in corrispondenza del valore di θ tale che $-\theta/\sigma^2 + E_g(Y_i)/\sigma^2 = 0$, cioè $\theta_g = E_g(Y_i)$. ■

Nel caso di una distribuzione log-normale $N(\mu, \sigma^2)$, con σ^2 fissata, invece, l'identità $\theta_g = E(Y_i)$ non è verificata.

Verifica. Posto $\theta = E(Y_i) = \exp\{\mu + \sigma^2/2\}$, si ha

$$f(y; \theta) = (2\pi\sigma^2 y^2)^{-1/2} \exp\{-[\log y - \log \theta + \sigma^2/2]^2/2\sigma^2\}.$$

Ora $E_g[\log\{f(Y_i; \theta)\}]$ è massimizzato quando $\log \theta = E_g\{\log(Y_i)\} + \sigma^2/2$, dunque in corrispondenza di $\theta_g = \exp\{E_g\{\log(Y_i)\} + \sigma^2/2\} \neq E_g(Y_i)$. ■

La seconda condizione per la robustezza della funzione di verosimiglianza viene soddisfatta quando l'approssimazione della probabilità di errata evidenza statistica in grandi campioni tramite la *bump function* continua a valere anche nel caso in cui la vera distribuzione non è inclusa nel modello statistico parametrico specificato.

In generale, la seconda condizione di robustezza della funzione di verosimiglianza non viene soddisfatta. Poniamoci nella condizione in cui il parametro che indicizza il modello sia scalare, $p = 1$. Per $\theta = \theta_g + c n^{-1/2}$, lo sviluppo del log-rapporto di verosimiglianza

tramite la formula di Taylor è

$$l(\theta) - l(\theta_g) = l_\theta(\theta_g)c/n^{1/2} + l_{\theta\theta}(\theta_g)c^2/2n + O_p(1/n^{1/2}). \quad (2.2)$$

Definite le quantità

$$a = E_g(-\partial^2[\log\{f(Y_i; \theta_g)\}]/\partial\theta^2) \quad (2.3)$$

e

$$b = E_g\{(\partial[\log\{f(Y_i; \theta_g)\}]/\partial\theta)^2\}, \quad (2.4)$$

si ha che il primo addendo del membro di destra di (2.2) converge in distribuzione a una normale con media nulla e varianza c^2b , mentre il secondo converge in probabilità a $-(c^2/2)a$. Dunque $l(\theta) - l(\theta_g)$ converge in distribuzione a una normale di media $-(c^2/2)a$ e varianza c^2b e $P_g\{L(\theta)/L(\theta_g) \geq k\}$ è approssimata dalla funzione $\Phi\{-ca/2b^{1/2} - \log k/cb^{1/2}\}$, che coincide con la *bump function* se $a = b$ ed è massimizzata in corrispondenza di $\Phi[-\{2 \log k a/b\}^{1/2}]$. Di conseguenza, la seconda condizione di robustezza è soddisfatta solo se $a = b$, dunque se lo è l'identità dell'informazione.

2.4 Trasformazione robusta della funzione di verosimiglianza

Dato il modello statistico \mathcal{F} , indicizzato dal parametro scalare θ , Royall & Tsou (2003) dimostrano che, nel caso in cui la funzione di verosimiglianza $L(\theta)$ soddisfi la prima condizione di robustezza, elevandola alla potenza data dal rapporto tra le quantità (2.3) e (2.4), a/b , si ottiene una funzione $L_{adj}(\theta) = L(\theta)^{a/b}$ che soddisfa entrambe le condizioni.

Le approssimazioni e dunque le conclusioni non vengono alterate sostituendo ad a e b i loro rispettivi stimatori consistenti, sotto campionamento casuale semplice, cioè

$$\hat{a} = \sum_{i=1}^n \left[-\partial^2 \log[f(Y_i; \theta)] / \partial\theta^2 \right] \Big|_{\theta=\hat{\theta}} / n \quad (2.5)$$

e

$$\hat{b} = \sum_{i=1}^n \left[\partial \log\{f(Y_i; \theta)\} / \partial\theta \right]^2 \Big|_{\theta=\hat{\theta}} / n. \quad (2.6)$$

Se il parametro θ è p -dimensionale, con $p > 1$, si possono ricavare le matrici A e B , con le loro rispettive stime empiriche \hat{A} e \hat{B} , di dimensione $p \times p$ in modo analogo, a partire dal vettore delle derivate prime e dalla matrice delle derivate seconde della

funzione di log-verosimiglianza. In questo caso, però, tali quantità non sono utili al fine di rendere robuste le funzioni di verosimiglianza. Si veda §2.6.

2.5 Trasformazione robusta della funzione di verosimiglianza profilo

La correzione vista nel paragrafo precedente può essere applicata, come mostrato per la prima volta da Stafford (1996), anche alla funzione di verosimiglianza profilo, in modo da ottenere un'alternativa robusta nei casi in cui il parametro del modello sia multidimensionale e si intenda ridurre la dimensione ponendo l'attenzione su un singolo parametro di interesse. Assumiamo che il parametro che identifica gli elementi del modello statistico parametrico (2.1) sia ora $\theta = (\psi, \lambda)$, p -dimensionale, dove ψ è il parametro di interesse unidimensionale e λ è il parametro di disturbo $(p - 1)$ -dimensionale. $L_P(\psi) = L(\psi, \hat{\lambda}_\psi)$ è dunque la funzione di verosimiglianza profilo associata a tale modello, dove $\hat{\lambda}_\psi$ è la stima di massima verosimiglianza di λ , fissato ψ .

Quando il modello è correttamente specificato, dunque la vera distribuzione delle Y_i , $f(y_i; \psi_0, \lambda_0)$, è contenuta in esso e ψ_0 è il vero valore di ψ , la funzione di verosimiglianza profilo $L_P(\psi)$ soddisfa entrambe le proprietà introdotte nel §2.1. Quando, invece, la vera densità $g(\cdot)$ non è inclusa nel modello, indichiamo con (ψ_g, λ_g) il valore del parametro tale che $\max_{(\psi, \lambda)} (E_g[\log\{f(Y; \psi, \lambda)\}]) = E_g[\log\{f(Y; \psi_g, \lambda_g)\}]$; si assume che la prima proprietà di robustezza (§2.1) continui ad essere soddisfatta, con ψ_g che gioca il ruolo di ψ_0 , mentre la seconda, in generale, non lo è.

Date le seguenti quantità:

$$I_{fg\psi\psi} = -E_g(\partial^2[\log\{f(Y_i; \psi, \lambda)\}]/\partial\psi^2)|_{(\psi_g, \lambda_g)},$$

$$I_{fg\psi\lambda} = -E_g(\partial^2[\log\{f(Y_i; \psi, \lambda)\}]/\partial\psi\partial\lambda^T)|_{(\psi_g, \lambda_g)},$$

$$I_{fg\lambda\psi} = -E_g(\partial^2[\log\{f(Y_i; \psi, \lambda)\}]/\partial\lambda\partial\psi)|_{(\psi_g, \lambda_g)},$$

$$I_{fg\lambda\lambda} = -E_g(\partial^2[\log\{f(Y_i; \psi, \lambda)\}]/\partial\lambda\partial\lambda^T)|_{(\psi_g, \lambda_g)}$$

e

$$\nu_{fg\psi\psi} = E_g\{(\partial[\log\{f(Y_i; \psi, \lambda)\}]/\partial\psi)^2\}|_{(\psi_g, \lambda_g)},$$

$$\nu_{fg\psi\lambda} = E_g\{(\partial[\log\{f(Y_i; \psi, \lambda)\}]/\partial\psi)(\partial[\log\{f(Y; \psi, \lambda)\}]/\partial\lambda^T)\}|_{(\psi_g, \lambda_g)},$$

$$\nu_{fg\lambda\psi} = E_g\{(\partial[\log\{f(Y_i; \psi, \lambda)\}]/\partial\lambda)(\partial[\log\{f(Y; \psi, \lambda)\}]/\partial\psi)\}|_{(\psi_g, \lambda_g)},$$

$$\nu_{fg\lambda\lambda} = E_g\{(\partial[\log\{f(Y_i; \psi, \lambda)\}]/\partial\lambda)(\partial[\log\{f(Y_i; \psi, \lambda)\}]/\partial\lambda^T)\}_{(\psi_g, \lambda_g)},$$

definiamo

$$a_{fg} = I_{fg\psi\psi} - I_{fg\psi\lambda}I_{fg\lambda\lambda}^{-1}I_{fg\lambda\psi} \quad (2.7)$$

e

$$b_{fg} = \nu_{fg\psi\psi} - 2\nu_{fg\psi\lambda}I_{fg\lambda\lambda}^{-1}I_{fg\lambda\psi} + I_{fg\psi\lambda}I_{fg\lambda\lambda}^{-1}\nu_{fg\lambda\lambda}I_{fg\lambda\lambda}^{-1}I_{fg\lambda\psi}. \quad (2.8)$$

Quando $a_{fg} = b_{fg}$, la seconda condizione di robustezza è soddisfatta. In generale, ciò si verifica solo quando il modello specificato contiene la vera densità $g(\cdot)$. Stimando i termini contenuti in a_{fg} e b_{fg} con i loro analoghi empirici, si ottengono le quantità \hat{a} e \hat{b} tali che, quando il modello basato su f soddisfa la prima delle condizioni di robustezza, la funzione $L_{adj_P}(\psi) = L_P(\psi)^{\hat{a}/\hat{b}}$ le soddisfa entrambe.

Se il parametro di interesse ψ è q -dimensionale, $1 \leq q < p$, estendendo le precedenti definizioni, si ottengono delle matrici di dimensione $q \times q$, A_{fg} e B_{fg} , costruite analogamente alle quantità (2.7) e (2.8), rispettivamente. Le matrici \hat{A} e \hat{B} si ottengono stimando i termini contenuti in A_{fg} e B_{fg} con le relative stime empiriche.

2.6 Stima sandwich della varianza asintotica

Consideriamo nuovamente il modello statistico parametrico (2.1), indicizzato dal parametro $\theta \in \Theta \subseteq \mathbb{R}^p$, e lo sviluppo di Taylor del secondo ordine della funzione di log-verosimiglianza ,

$$l_\theta(\hat{\theta}) \doteq l_\theta(\theta) + (\hat{\theta} - \theta)l_{\theta\theta}(\theta),$$

da cui si ottiene

$$\hat{\theta} - \theta \doteq -l_\theta(\theta)[l_{\theta\theta}(\theta)]^{-1} \doteq l_\theta(\theta)[j(\theta)]^{-1}.$$

Sotto condizioni di regolarità, la quantità $\hat{\theta} - \theta$ ha distribuzione approssimata normale con media nulla e matrice di covarianza asintotica definita dal prodotto matriciale

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

dove

$$H(\theta) = -E_\theta(l_{\theta\theta})$$

e

$$J(\theta) = E_\theta(l_\theta l_\theta^T) = Var_\theta(l_\theta).$$

Tale approssimazione asintotica della varianza è detta di tipo *sandwich*, per via della disposizione delle quantità che la determinano.

Se il modello è correttamente specificato, per l'identità dell'informazione (1.2), si ha $J(\theta) = H(\theta) = i(\theta)$; dunque la varianza asintotica sarebbe $i(\theta)^{-1}$, approssimabile tramite $i(\hat{\theta})$ o $j(\hat{\theta})$. Se il modello non è correttamente specificato, la varianza asintotica può essere stimata dal prodotto matriciale $\hat{H}(\theta)^{-1}\hat{J}(\theta)\hat{H}(\theta)^{-1}$, dove $\hat{J}(\theta)$ e $\hat{H}(\theta)$ sono delle stime empiriche ottenute, rispettivamente, come

$$\hat{H}(\theta) = \sum_{i=1}^n \left[\partial^2 \log\{f(Y_i; \theta)\} / \partial \theta \partial \theta^T \right] \Big|_{\theta=\hat{\theta}} / n = -j(\hat{\theta})/n$$

e

$$\hat{J}(\theta) = \sum_{i=1}^n \left[\partial \log\{f(Y_i; \theta)\} / \partial \theta \right]^2 \Big|_{\theta=\hat{\theta}} / n,$$

La stima asintotica della matrice di covarianza ottenuta è robusta, dunque valida anche in caso di errata specificazione del modello (Royall, 1986).

Notando la corrispondenza tra \hat{H} ed \hat{A} e tra \hat{J} e \hat{B} (§2.5), concludiamo che la quantità $\hat{A}^{-1}\hat{B}\hat{A}^{-1}$ è la varianza asintotica robusta di $\hat{\theta}$ e quindi di $\hat{\theta} - \theta$. Analogamente nel caso scalare per \hat{b}/\hat{a}^2 .

2.7 Estensioni della trasformazione robusta della log-verosimiglianza

Per quanto visto nel paragrafo precedente, nell'ambito di problemi di verifica d'ipotesi con $H_0 : \theta = \theta_0$, e anche nel caso in cui il modello statistico non sia correttamente specificato, la statistica test di Wald corretta è data da

$$W_{\text{adj}} = n(\hat{\theta} - \theta_0)^T \hat{A} \hat{B}^{-1} \hat{A} (\hat{\theta} - \theta_0) \quad (2.9)$$

ed ha distribuzione asintotica χ_p^2 , se θ è p -dimensionale (la distribuzione asintotica è χ_1^2 nel caso di parametro scalare). La statistica test *score* corretta si ottiene come $W_{\text{adj}} = l_{\theta}(\theta_0)^T \hat{B}^{-1} l_{\theta}(\theta_0)$ ed ha distribuzione asintotica nulla χ_p^2 .

La statistica test log-rapporto di verosimiglianza (1.3), utilizzabile nel caso di parametro scalare, diventa, nella sua forma corretta,

$$W_{\text{adj}} = 2 \frac{\hat{a}}{\hat{b}} [l(\hat{\theta}) - l(\theta_0)], \quad (2.10)$$

con distribuzione asintotica sotto H_0 pari un χ_1^2 , anche nel caso in cui il modello non sia correttamente specificato. Considerazioni analoghe valgono anche per test in cui si faccia riferimento alla funzione di verosimiglianza profilo con parametro di interesse scalare, dunque la statistica test

$$W_{adj_P} = 2 \frac{\hat{a}}{\hat{b}} [l(\hat{\psi}, \hat{\lambda}_{\hat{\psi}}) - l(\psi_0, \hat{\lambda}_{\psi_0})]$$

si distribuisce approssimativamente come un χ_1^2 , sotto $H_0 : \psi = \psi_0$, anche nelle medesime condizioni di cui sopra.

Capitolo 3

Applicazione ad alcuni modelli di campionamento casuale semplice

3.1 Introduzione

In questo capitolo si analizzano alcune applicazioni della correzione introdotta nel capitolo precedente nell'ambito di alcuni specifici modelli di campionamento casuale semplice, ovvero nel caso di campioni le cui osservazioni sono tra loro indipendenti ed identicamente distribuite. Si considerano, in particolare, dati di tipo conteggio e dati binari raggruppati. Vengono poi condotti degli studi di simulazione per esaminare l'effetto della correzione della funzione di verosimiglianza.

Per ciò che riguarda l'applicazione ai dati di conteggio, l'articolo di riferimento è Tsou (2006). Vengono verificati i risultati in esso riportati, sia nell'ambito del calcolo dei fattori di correzione, sia nell'ambito delle simulazioni. Vengono poi operate delle estensioni di queste ultime, tramite aggiunte e variazioni non presenti nell'articolo citato. L'applicazione ai dati binari viene infine condotta prendendo come riferimento quella per i dati di conteggio.

3.2 Applicazione a dati di conteggio

I dati di conteggio sono molto comuni nelle applicazioni statistiche. Si tratta di dati y_i derivanti, appunto, da conteggi il cui totale $\sum_{i=1}^n y_i$ non è prefissato. Per osservazioni di questo tipo, l'assunzione distributiva di riferimento è quello Poisson, caratterizzata dall'uguaglianza tra la media e la varianza.

3.2.1 Dati di conteggio indipendenti ed identicamente distribuiti

Date le osservazioni y_1, \dots, y_n , realizzazioni delle variabili casuali Y_1, \dots, Y_n indipendenti ed identicamente distribuite come $Y_i \sim Poi(\theta)$, dove $\theta > 0$ è un parametro unidimensionale tale che $\theta = E(Y_i) = Var(Y_i)$, $i = 1, \dots, n$, si ha $f(y_i; \theta) = \theta^{y_i} e^{-\theta} / y_i!$ e la funzione di verosimiglianza si ottiene come

$$L(\theta) = \prod_{i=1}^n f(y_i; \theta) = \prod_{i=1}^n \frac{\theta^{y_i} e^{-\theta}}{y_i!} \propto \theta^{\sum y_i} e^{-n\theta}. \quad (3.1)$$

Di conseguenza, la funzione di log-verosimiglianza è

$$l(\theta) = -n\theta + n\bar{y} \log \theta, \quad (3.2)$$

le cui prime due derivate sono $l_\theta = \partial l(\theta) / \partial \theta = -n + n\bar{y} / \theta$ e $l_{\theta\theta} = \partial^2 l(\theta) / \partial^2 \theta = -n\bar{y} / \theta^2$.

Se l'oggetto di interesse per l'inferenza è la media, $E_g(Y)$, la prima condizione di robustezza, secondo cui l'oggetto di inferenza rimane quello di interesse anche nel caso di errata specificazione del modello, è soddisfatta.

Verifica. In questo caso, l'oggetto di interesse θ coincide con la media della variabile di riferimento e si ha che

$$E_g[\log\{f(Y_i; \theta)\}] = E_g[Y_i \log(\theta) - \theta - \log(Y_i!)] = E_g(Y_i) \log(\theta) - \theta - E_g[\log(Y_i!).]$$

Massimizzando questa espressione, si ottiene: $\frac{dE_g[\log\{f(Y_i; \theta)\}]}{d\theta} = E_g(Y_i) \cdot 1/\theta - 1 = 0$, che implica $\theta_g = E_g(Y_i)$. ■

Poichè il parametro di interesse è unidimensionale, detto θ_0 il valore limite della stima di massima verosimiglianza $\hat{\theta}$, si ha che le quantità (2.5) e (2.6) sono, rispettivamente, del tipo $a = I_{\theta\theta}$ e $b = \nu_{\theta\theta}$, dove

$$I_{\theta\theta} = \lim_{n \rightarrow \infty} -E_g \left(\frac{l_{\theta\theta}}{n} \right) \Big|_{(\theta_0)} = \lim_{n \rightarrow \infty} -E_g \left(-\frac{n\bar{Y}}{n\theta^2} \right) \Big|_{(\theta_0)} = E_g \left(\frac{\bar{Y}}{\theta_0^2} \right) = \frac{1}{\theta_0}$$

e

$$\begin{aligned}
 \nu_{\theta\theta} &= \lim_{n \rightarrow \infty} E_g \left[\frac{(l_\theta)^2}{n} \right] \Big|_{(\theta_0)} = \lim_{n \rightarrow \infty} E_g \left[\left(-n + \frac{n\bar{Y}}{\theta_0} \right)^2 / n \right] \\
 &= \lim_{n \rightarrow \infty} E_g \left[\left(-n + \frac{\sum_{i=1}^n Y_i}{\theta_0} \right)^2 / n \right] = \lim_{n \rightarrow \infty} E_g \left[\left\{ \sum_{i=1}^n \left(\frac{Y_i}{\theta_0} - 1 \right) \right\}^2 / n \right] \\
 &= \lim_{n \rightarrow \infty} E_g \left[\left(-n + \frac{\sum_{i=1}^n Y_i}{\theta_0} \right)^2 / n \right] = \lim_{n \rightarrow \infty} E_g \left[\left\{ \sum_{i=1}^n \left(\frac{Y_i}{\theta_0} - 1 \right) \right\}^2 / n \right] \\
 &= \lim_{n \rightarrow \infty} E_g \left[\frac{1}{n\theta_0^2} \sum_{i=1}^n \left(Y_i - \theta_0 \right)^2 \right] = \frac{Var_g(Y)}{\theta_0^2}.
 \end{aligned}$$

I termini a e b possono essere stimati in modo consistente attraverso le loro relative quantità empiriche, $\hat{a} = 1/\bar{y}$ e $\hat{b} = \sum_{i=1}^n (y_i - \bar{y})^2 / n\bar{y}^2$, così da ottenere il fattore di correzione $\hat{a}/\hat{b} = n\bar{y} / \sum_{i=1}^n (y_i - \bar{y})^2$. La funzione di verosimiglianza robusta risulta quindi $L_{adj}(\theta) = \{\theta^{-n\bar{y}} e^{-n\theta}\}^{n\bar{y} / \sum_{i=1}^n (y_i - \bar{y})^2}$.

3.3 Applicazione a dati binari

I modelli per dati binari vengono impiegati nei casi in cui la risposta sia dicotomica, cioè caratterizzata da due modalità possibili. Osservazioni di questo tipo possono essere riportate nella struttura di dati non raggruppati o raggruppati. Nel primo caso, la risposta rappresenta l'esito di una singola prova e può dunque assumere valore 1 (successo) oppure 0 (insuccesso); nel secondo caso, la risposta indica il numero totale di successi per ogni combinazione delle variabili concomitanti.

3.3.1 Dati binari raggruppati indipendenti e identicamente distribuiti

Siano le osservazioni y_1, \dots, y_n realizzazioni delle variabili casuali indipendenti Y_1, \dots, Y_n , dove $Y_i = S_i/m_i$ e $S_i = m_i Y_i \sim Bi(m_i, \pi)$, secondo il modello (2.1). Le variabili sono essenzialmente identicamente distribuite, ad eccezione del fatto che gli indici m_1, \dots, m_n possono non essere uguali. Si ha $f(s_i; m_i, \pi) = \binom{m_i}{s_i} \pi^{s_i} (1 - \pi)^{m_i - s_i}$ e la funzione di

verosimiglianza è

$$\begin{aligned}
 L(\pi) &= \prod_{i=1}^n f(y_i; m_i, \pi) = \prod_{i=1}^n \binom{m_i}{s_i} \pi^{s_i} (1 - \pi)^{m_i - s_i} \\
 &= \prod_{i=1}^n \exp \left\{ \log \binom{m_i}{s_i} + s_i \log \pi + (m_i - s_i) \log(1 - \pi) \right\} \\
 &\propto \exp \left\{ \sum_{i=1}^n \left(s_i \log \left(\frac{\pi}{1 - \pi} \right) + m_i \log(1 - \pi) \right) \right\}.
 \end{aligned} \tag{3.3}$$

La funzione di log-verosimiglianza è

$$l(\pi) = \sum_{i=1}^n \left\{ s_i \log \left(\frac{\pi}{1 - \pi} \right) + m_i \log(1 - \pi) \right\}; \tag{3.4}$$

La derivata prima della funzione di verosimiglianza è

$$l_{\pi} = \frac{\partial l(\pi)}{\partial \pi} = \sum_{i=1}^n \left\{ \frac{s_i}{\pi(1 - \pi)} - \frac{m_i}{1 - \pi} \right\} = \sum_{i=1}^n \frac{s_i - m_i \pi}{\pi(1 - \pi)}$$

e la derivata seconda è

$$l_{\pi\pi} = \frac{\partial^2 l(\pi)}{\partial^2 \pi} = \sum_{i=1}^n \frac{2\pi s_i - m_i \pi^2 - s_i}{\pi^2(1 - \pi)^2} = \sum_{i=1}^n \frac{m_i(2\pi y_i - \pi^2 - y_i)}{\pi^2(1 - \pi)^2}.$$

Essendo l'oggetto di interesse il valore atteso $\pi = E_g(Y_i)$, la prima condizione di robustezza è verificata.

Verifica. Si ha che

$$\begin{aligned}
 E_g \left[\log \{ f(S_i; \pi) \} \right] &= E_g \left[\log \left\{ \binom{m_i}{S_i} \pi^{S_i} (1 - \pi)^{m_i - S_i} \right\} \right] \\
 &= E_g \left[\log \binom{m_i}{S_i} + S_i \log \pi + (m_i - S_i) \log(1 - \pi) \right] \\
 &= E_g \left[\log \binom{m_i}{S_i} \right] + m_i E(Y_i) \log \pi + (m_i - m_i E(Y_i)) \log(1 - \pi).
 \end{aligned}$$

Massimizzando questa espressione, si ottiene: $\frac{dE_g[\log\{f(S_i;\pi)\}]}{d\pi} = m_i E(Y_i)/\pi - m_i - m_i E(Y_i)/1 - \pi = 0$, che implica $\pi_g = E_g(Y_i)$. ■

Indicato con π_0 il valore limite della stima di massima verosimiglianza di π , le quantità

a e b si ottengono come

$$\begin{aligned} a = I_{\pi\pi} &= \lim_{n \rightarrow \infty} -E_g \left(\frac{l_{\pi\pi}}{n} \right) \Big|_{(\pi_0)} = \lim_{n \rightarrow \infty} \frac{1}{n} - E_g \left(\sum_{i=1}^n \frac{m_i(2\pi Y_i - \pi^2 - Y_i)}{\pi^2(1-\pi)^2} \right) \Big|_{(\pi_0)} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{m_i(2\pi^2 - \pi^2 - \pi)}{\pi^2(1-\pi)^2} \Big|_{(\pi_0)} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{m_i}{\pi(1-\pi)} \Big|_{(\pi_0)} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \frac{n}{\pi_0(1-\pi_0)} = \frac{1}{\pi_0(1-\pi_0)} \end{aligned}$$

e

$$\begin{aligned} b = \nu_{\pi\pi} &= \lim_{n \rightarrow \infty} E_g \left[\frac{(l_{\pi})^2}{n} \right] \Big|_{(\pi_0)} = \lim_{n \rightarrow \infty} \frac{1}{n} E_g \left[\sum_{i=1}^n \frac{S_i - m_i\pi}{\pi(1-\pi)} \right]^2 \Big|_{(\pi_0)} \\ &= \lim_{n \rightarrow \infty} \frac{1}{\pi_0^2(1-\pi_0)^2} E_g \left[\sum_{i=1}^n \frac{S_i - m_i\pi_0}{n} \right]^2 = \frac{Var_g(S)}{m_i^2 \pi_0^2 (1-\pi_0)^2} = \frac{Var_g(Y)}{\pi_0^2(1-\pi_0)^2}. \end{aligned}$$

Stimando consistentemente a e b , mediante la sostituzione dei valori π_0 e $Var_g(S)$ con le loro relative quantità empiriche, rispettivamente, \bar{y} (proporzione osservata di successi) e $\sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n}$, si ottiene il fattore di correzione empirico $\hat{a}/\hat{b} = n\bar{y}(1-\bar{y}) / \sum_{i=1}^n (y_i - \bar{y})^2$.

3.4 Studi di simulazione

Si conducono ora degli studi di simulazione per esaminare l'efficacia della stima della varianza ottenuta dall'utilizzo dei fattori di correzione e della trasformazione apportata alla funzione di verosimiglianza. Si considera, in particolare, il comportamento dei test log-rapporto di verosimiglianza e di Wald e si analizza la copertura dei corrispondenti intervalli di confidenza. Nell'Appendice A si riportano i comandi **R** per ottenere i risultati illustrati nei paragrafi §3.4.1 e §3.4.2.

3.4.1 Assunzione di distribuzione Poisson per dati generati da una binomiale negativa

Siano y_1, \dots, y_n realizzazioni delle variabili Y_1, \dots, Y_n , indipendenti ed identicamente distribuite. Le variabili Y_i sono tali che $Y_i + r \sim Bineg(r, \pi)$. I valori y_i indicano dunque il numero prove bernoulliane indipendenti con esito fallimentare, prima di osservare $r \in \{1, 2, \dots\}$ successi, con probabilità di successo π , $0 < \pi < 1$. Il supporto delle Y_i è

$\mathcal{S} = \{0, 1, \dots\}$ è la loro funzione di probabilità è

$$f_{Y_i}(y_i; \pi) = \binom{y_i + r - 1}{r - 1} \pi^r (1 - \pi)^{y_i} \quad (3.5)$$

Si ha inoltre che $E(Y_i) = r(1 - \pi)/\pi$ e $Var(Y_i) = r(1 - \pi)/\pi^2$.

Consideriamo come oggetto di inferenza, θ , il valore atteso $E(Y_i)$ e generiamo $B = 2000$ campioni di dati provenienti dalla distribuzione (3.5) per diversi valori della numerosità campionaria. Indicato con $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_B)^T$ il vettore contenente la media campionaria (stima di massima verosimiglianza per il valore atteso) calcolata in corrispondenza di ogni campione generato, calcoliamo le quantità $m(\hat{\theta})$, $s^2(\hat{\theta})$ e $med(\hat{\theta})$, cioè la media, varianza e mediana campionaria, rispettivamente.

Dopo aver individuato le quantità \hat{a} e \hat{b} secondo la metodologia illustrata nel §3.2.1, dunque sulla base dell'assunzione distributiva di tipo Poisson, calcoliamo $Var_{adj}(\hat{\theta}) = \hat{b}/n\hat{a}^2$ e $Var_{na}(\hat{\theta}) = 1/n\hat{a}$, al fine di individuare le differenze tra la stima robusta e non robusta della varianza di $\hat{\theta}$ (§2.6).

Specifichiamo, come in Tsou (2006), due particolari valori di θ , 6 e 7. Per diversi valori della numerosità campionaria (ai valori $n = 50$ e $n = 100$ riportati nell'articolo si aggiungono $n = 10$, $n = 20$ e $n = 500$), riportiamo nella Tabella 3.1 alcune quantità riguardanti la stima di massima verosimiglianza del valore atteso della distribuzione, $\hat{\theta}$: media, mediana e varianza campionarie, la stima robusta e non robusta della varianza dello stimatore. I dati vengono generati da una distribuzione binomiale negativa con $r = 2$.

La vicinanza tra i valori della media e della mediana di $\hat{\theta}$, per tutti i valori della numerosità campionaria conferma la simmetria della sua distribuzione, garantita asintoticamente dal teorema del limite centrale. Per quanto riguarda la stima della varianza, vediamo che la versione corretta approssima bene il valore della varianza campionaria, anche con una numerosità campionaria ridotta, al contrario di quanto si verifica con la stima non corretta. Poiché le quantità \hat{a} e \hat{b} sono individuate sulla base dell'assunzione della distribuzione Poisson, i valori osservati confermano la robustezza della stima *sandwich* della varianza.

TABELLA 3.1: Stima di alcune statistiche della SMV del valore atteso, θ , ottenute da dati generati da una distribuzione binomiale negativa con $r = 2$ e media θ .

	θ	$m(\hat{\theta})$	$med(\hat{\theta})$	$s^2(\hat{\theta})$	$Var_{adj}(\hat{\theta})$	$Var_{na}(\hat{\theta})$
$n = 10$	6	5.983	5.900	2.383	1.816	0.558
	7	7.000	6.900	3.026	2.363	0.655
$n = 20$	6	5.955	5.900	1.220	1.034	0.287
	7	6.986	6.900	1.615	1.365	0.338
$n = 50$	6	5.978	5.960	0.474	0.450	0.118
	7	6.963	6.920	0.634	0.592	0.137
$n = 100$	6	5.996	5.990	0.242	0.232	0.060
	7	7.004	7.000	0.316	0.306	0.070
$n = 500$	6	5.992	5.994	0.048	0.048	0.012
	7	6.999	7.004	0.063	0.063	0.014

Consideriamo poi il problema di verificare l'ipotesi $H_0 : \theta = \theta_0$ (Tsou (2006)) contro l'alternativa $H_1 : \theta \neq \theta_0$. Per verificare tale ipotesi, si conduce il test log-rapporto di verosimiglianza, confrontando i risultati ottenuti utilizzando la statistica test

$$W(\hat{\theta}) = 2[l(\hat{\theta}) - l(\theta_0)] \quad (3.6)$$

e quelli derivanti dall'utilizzo della sua versione trasformata,

$$W_{adj}(\hat{\theta}) = 2 \frac{\hat{a}}{\hat{b}} [l(\hat{\theta}) - l(\theta_0)]. \quad (3.7)$$

La stessa ipotesi viene inoltre verificata in questo studio utilizzando la statistica test di Wald,

$$W_e(\hat{\theta}) = n (\hat{\theta} - \theta_0)^2 j(\hat{\theta}) = \frac{(\hat{\theta} - \theta_0)^2}{Var_{na}(\hat{\theta})}, \quad (3.8)$$

e la sua trasformazione robusta

$$W_{e_{adj}}(\hat{\theta}) = n \frac{\hat{a}}{\hat{b}} (\hat{\theta} - \theta_0)^2 j(\hat{\theta}) = \frac{(\hat{\theta} - \theta_0)^2}{Var_{adj}(\hat{\theta})}. \quad (3.9)$$

Si osservano, in particolare, i valori empirici della probabilità di commettere un errore di primo tipo. Di seguito, indichiamo con $\hat{ep}I_{adj}$ la stima della probabilità di errore di primo tipo sulla base della statistica (3.7) e con $\hat{ep}I_{na}$ il medesimo valore ottenuto senza la correzione (3.6); analogamente, $\hat{ep}I_{e_{adj}}$ indica la stima della probabilità di errore di primo tipo utilizzando (3.9) e $\hat{ep}I_{e_{na}}$ quella ottenuta da (3.8). Poiché la distribuzione asintotica sotto l'ipotesi nulla delle suddette statistiche test è χ_1^2 , le stime di tali quantità sono ottenute come le proporzioni dei casi in cui le statistiche test sono maggiori del quantile di ordine $1 - \alpha$ di un χ_1^2 , dove α è un valore nominale fissato a priori.

Quando i dati vengono generati sotto l'ipotesi alternativa, specificando in particolare il valore $\theta = 7$, $\hat{ep}I_{adj}$, $\hat{ep}I_{na}$, $\hat{ep}I_{e_{adj}}$, $\hat{ep}I_{e_{na}}$ indicano i valori della potenza del test condotto. Si osservi che i valori della potenza non sono confrontabili poiché i valori effettivi degli α nei due test, robusto e non robusto, sono diversi. Non si può concludere, dunque, che il test sia più potente nel caso si utilizzi la statistica non trasformata rispetto al caso in cui essa viene utilizzata o viceversa. Considerazioni analoghe valgono anche nel caso in cui i dati vengono generati sotto H_0 . Le conclusioni non risultano però alterate anche prendendo sempre a riferimento il valore nominale 0.05. Nella Tabella 3.2 vengono riportati i valori $\hat{ep}I_{adj}$, $\hat{ep}I_{na}$, $\hat{ep}I_{e_{adj}}$, $\hat{ep}I_{e_{na}}$.

TABELLA 3.2: Test del log-rapporto di verosimiglianza e test di Wald per $H_0 : \theta = 6$ con $\alpha = 0.05$.

	Ipotesi	$\hat{ep}I_{adj}$	$\hat{ep}I_{na}$	$\hat{ep}I_{e_{adj}}$	$\hat{ep}I_{e_{na}}$
$n = 10$	$H_0 : \theta_0 = 6$	0.072	0.322	0.082	0.334
	$H_1 : \theta_0 = 7$	0.122	0.446	0.126	0.435
$n = 20$	$H_0 : \theta_0 = 6$	0.070	0.336	0.074	0.328
	$H_1 : \theta_0 = 7$	0.154	0.470	0.153	0.463
$n = 50$	$H_0 : \theta_0 = 6$	0.052	0.334	0.054	0.338
	$H_1 : \theta_0 = 7$	0.275	0.661	0.258	0.636
$n = 100$	$H_0 : \theta_0 = 6$	0.058	0.320	0.057	0.324
	$H_1 : \theta_0 = 7$	0.462	0.822	0.420	0.808
$n = 500$	$H_0 : \theta_0 = 6$	0.054	0.340	0.056	0.338
	$H_1 : \theta_0 = 7$	0.984	0.999	0.978	0.998

Le statistiche calcolate sulla base della trasformazione robusta sono più efficaci. In

particolare, quando i dati sono generati sotto l'ipotesi nulla, $\hat{e}pI_{adj}$ e $\hat{e}pI_{e_{adj}}$ approssimano meglio il valore nominale 0.05 rispetto a $\hat{e}pI_{na}$ e $\hat{e}pI_{e_{na}}$, rispettivamente.

Analizziamo le coperture empiriche degli intervalli di confidenza corrispondenti ai test condotti. Siano $\hat{c}p_{adj}$, $\hat{c}p_{na}$, $\hat{c}p_{e_{adj}}$ e $\hat{c}p_{e_{na}}$ le coperture empiriche degli intervalli di confidenza di livello nominale $1 - \alpha$ basati su $W_{adj}(\theta)$, $W(\theta)$, $W_{e_{adj}}(\theta)$ e W_e , rispettivamente. La Tabella 3.3 riporta tali valori, ottenuti specificando i medesimi valori di θ ed n di cui sopra.

Si osserva che i valori empirici della copertura garantita dagli intervalli di confidenza ottenuti servendosi della trasformazione robusta, $\hat{c}p_{adj}$ e $\hat{c}p_{e_{adj}}$, approssimano adeguatamente il valore nominale 0.95, al contrario di quanto si verifica per i valori $\hat{c}p_{na}$ e $\hat{c}p_{e_{na}}$. Risultati analoghi si ottengono specificando altri valori di θ ed r , come mostrato nelle Tabelle 3.4, 3.5 e 3.6, riportanti i risultati ottenuti specificando $r = 3$, $\theta_0 = 4$ come valore dell'ipotesi nulla e $\theta_0 = 5$ sotto l'ipotesi alternativa.

In tutte le procedure di stima illustrate, le approssimazioni migliorano al crescere della numerosità campionaria. A parità di condizioni, inoltre, non si registrano differenze significative tra le stime derivanti dall'utilizzo della statistica log-rapporto di verosimiglianza e di quella di Wald.

TABELLA 3.3: Intervalli di confidenza per θ di tipo log-rapporto di verosimiglianza e alla Wald con $1 - \alpha = 0.95$.

	θ	$\hat{c}p_{adj}$	$\hat{c}p_{na}$	$\hat{c}p_{e_{adj}}$	$\hat{c}p_{e_{na}}$
$n = 10$	6	0.928	0.678	0.918	0.666
	7	0.926	0.645	0.923	0.621
$n = 20$	6	0.930	0.664	0.926	0.672
	7	0.934	0.638	0.932	0.634
$n = 50$	6	0.948	0.666	0.946	0.662
	7	0.938	0.637	0.940	0.630
$n = 100$	6	0.942	0.680	0.943	0.676
	7	0.954	0.646	0.951	0.643
$n = 500$	6	0.946	0.686	0.944	0.688
	7	0.950	0.635	0.948	0.633

TABELLA 3.4: Stime puntuali per θ , ottenute da dati generati da una distribuzione binomiale negativa con $r = 3$ e media θ .

	θ	$m(\hat{\theta})$	$med(\hat{\theta})$	$s^2(\hat{\theta})$	$Var_{adj}(\hat{\theta})$	$Var_{na}(\hat{\theta})$
$n = 10$	4	4.007	3.900	0.973	0.744	0.376
	5	4.973	4.900	1.319	1.054	0.470
$n = 20$	4	3.996	4.000	0.479	0.417	0.194
	5	5.016	4.950	0.652	0.599	0.244
$n = 50$	4	3.984	3.980	0.177	0.178	0.079
	5	5.002	4.980	0.264	0.255	0.099
$n = 100$	4	4.006	4.000	0.095	0.091	0.040
	5	5.009	4.990	0.130	0.129	0.050
$n = 500$	4	4.001	4.000	0.019	0.019	0.008
	5	4.996	4.998	0.028	0.027	0.010

TABELLA 3.5: Test del log-rapporto di verosimiglianza e test di Wald per $H_0 : \theta = 4$ con $\alpha = 0.05$.

	Ipotesi	$\hat{e}pI_{adj}$	$\hat{e}pI_{na}$	$\hat{e}pI_{e_{adj}}$	$\hat{e}pI_{e_{na}}$
$n = 10$	$H_0 : \theta_0 = 4$	0.074	0.198	0.090	0.193
	$H_1 : \theta_0 = 5$	0.202	0.380	0.178	0.373
$n = 20$	$H_0 : \theta_0 = 4$	0.058	0.212	0.060	0.208
	$H_1 : \theta_0 = 5$	0.296	0.528	0.256	0.500
$n = 50$	$H_0 : \theta_0 = 4$	0.044	0.182	0.049	0.188
	$H_1 : \theta_0 = 5$	0.537	0.801	0.498	0.778
$n = 100$	$H_0 : \theta_0 = 4$	0.056	0.208	0.054	0.204
	$H_1 : \theta_0 = 5$	0.844	0.964	0.798	0.949
$n = 500$	$H_0 : \theta_0 = 4$	0.048	0.194	0.049	0.198
	$H_1 : \theta_0 = 5$	1.000	1.000	1.000	1.000

TABELLA 3.6: Intervalli di confidenza per θ di tipo log-rapporto di verosimiglianza e alla Wald con $1 - \alpha = 0.95$.

	θ	$\hat{c}p_{adj}$	$\hat{c}p_{na}$	$\hat{c}p_{e_{adj}}$	$\hat{c}p_{e_{na}}$
$n = 10$	4	0.926	0.802	0.910	0.807
	5	0.934	0.775	0.925	0.758
$n = 20$	4	0.942	0.788	0.940	0.702
	5	0.948	0.767	0.944	0.776
$n = 50$	4	0.956	0.818	0.951	0.812
	5	0.950	0.771	0.949	0.764
$n = 100$	4	0.944	0.792	0.946	0.796
	5	0.946	0.777	0.948	0.774
$n = 500$	4	0.952	0.806	0.951	0.802
	5	0.944	0.764	0.944	0.762

3.4.2 Assunzione di distribuzione binomiale per dati generati da una beta-binomiale

Per verificare l'efficacia della trasformazione robusta nei modelli di campionamento casuale semplice per dati binari raggruppati, introduciamo la distribuzione beta-binomiale. Si tratta di una distribuzione di probabilità discreta che estende quella binomiale, in cui la probabilità di successo segue a sua volta una distribuzione biparametrica. La distribuzione beta-binomiale permette di rappresentare dati binari raggruppati in presenza di sovradisersione, cioè quando la varianza delle variabili di riferimento è maggiore di quella prevista dal modello binomiale.

Siano y_1, \dots, y_n realizzazioni delle variabili Y_1, \dots, Y_n , indipendenti ed identicamente distribuite tali che $Y_i = S_i/m_i$, $S_i \sim Bi(m_i, \pi_i)$, $i = 1, \dots, n$. Si assume che $\pi_i \in (0, 1)$ segua una distribuzione Beta, $Be(\alpha_i, \beta_i)$ ed abbia dunque densità di probabilità

$$f_{\pi_i}(\pi_i; \alpha_i, \beta_i) = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} \pi_i^{\alpha_i-1} (1 - \pi_i)^{\beta_i-1},$$

media $E(\pi_i) = \alpha_i/(\alpha_i + \beta_i) = \mu_i$ e varianza $Var(\pi_i) = \alpha_i\beta_i/[(\alpha_i + \beta_i)^2(\alpha_i + \beta_i + 1)] = \mu_i(1 - \mu_i)\rho_i$, con $\rho_i \in (0, 1)$ parametro che regola la sovradisersione della distribuzione.

Si dimostra che la densità marginale di $S_i = m_i Y_i$ è

$$f_{S_i}(s_i; \mu_i, \rho_i) = \binom{m_i}{s_i} \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} \frac{\Gamma(\alpha_i + s_i)\Gamma(\beta_i + m_i - s_i)}{\Gamma(\alpha_i + \beta_i + m_i)} \quad (3.10)$$

ed è detta, appunto, beta-binomiale. Tra i parametri impiegati sussistono le seguenti relazioni: $\alpha_i = \mu_i(1 - \rho_i)/\rho_i$ e $\beta_i = (1 - \mu_i)(1 - \rho_i)/\rho_i$ e per $\rho_i \rightarrow 0$, la distribuzione converge ad una $Bi(m_i, \mu_i)$; tipicamente si assume $\rho_i = \rho$ per tutte le osservazioni. Per maggiori dettagli sulla distribuzione si rimanda a Salvan et al. (2020, capitolo 3).

Seguendo un procedimento analogo a quello illustrato nel paragrafo precedente, consideriamo $E(Y_i) = \pi$ come oggetto dell'inferenza e generiamo $B = 2000$ campioni di diversa numerosità campionaria dalla distribuzione (3.10), specificando i valori di π e assumendo $\rho_i = \rho$, $m_i = m$ per ogni valore di i . Per ciascuno dei B campioni generati, si calcola la media campionaria delle osservazioni (queste vengono divise per m volendo stimare il valore atteso della distribuzione delle Y_i) e se ne calcola media, $m(\hat{\pi})$, mediana, $med(\hat{\pi})$, e varianza campionarie, $s^2(\hat{\pi})$. Si individuano poi i fattori di correzione \hat{a} e \hat{b} , secondo l'assunzione di distribuzione binomiale (§3.3.1) e si calcolano le due stime della varianza: $Var_{adj}(\hat{\pi}) = \hat{b}/n\hat{a}$ e $Var_{na}(\hat{\pi}) = 1/n\hat{a}$. Nella Tabella 3.7 si riportano i suddetti valori, ottenuti ponendo $\pi = 0.3$, $\pi = 0.4$, $\rho = 0.6$, $m = 10$.

TABELLA 3.7: Stima di alcune statistiche della SMV del valore atteso, π , ottenute da dati generati da una distribuzione beta-binomiale con $m = 10$, $\rho = 0.6$ e probabilità di successo π .

	π	$m(\hat{\pi})$	$med(\hat{\pi})$	$s^2(\hat{\pi})$	$Var_{adj}(\hat{\pi})$	$Var_{na}(\hat{\pi})$
$n = 10$	0.3	0.297	0.290	0.013	0.009	0.017
	0.4	0.401	0.400	0.015	0.013	0.022
$n = 20$	0.3	0.303	0.300	0.007	0.006	0.010
	0.4	0.407	0.405	0.008	0.007	0.012
$n = 50$	0.3	0.300	0.298	0.003	0.003	0.004
	0.4	0.399	0.398	0.003	0.003	0.005
$n = 100$	0.3	0.301	0.300	0.001	0.001	0.002
	0.4	0.400	0.400	0.002	0.002	0.002
$n = 500$	0.3	0.300	0.300	0.000	0.000	0.000
	0.4	0.400	0.400	0.000	0.000	0.000

Anche in questo caso, i valori della media e della mediana sono molto vicini, confermando la simmetria della distribuzione dello stimatore. I vantaggi derivanti dall'utilizzo della stima robusta della varianza rispetto a quella non robusta non sono notevoli e sono coglibili solo per valori ridotti della numerosità campionaria.

Nella Tabella 3.8 sono riportati i valori di $e\hat{p}I_{adj}$, $e\hat{p}I_{na}$, $e\hat{p}I_{e_{adj}}$ e $e\hat{p}I_{e_{na}}$, ovvero delle stime della probabilità di errore di primo tipo e della potenza, per verificare $H_0 : \pi = 0.3$, ottenute utilizzando le statistiche test (3.7), (3.6), (3.9) e (3.8), rispettivamente, e specificando $H_1 : \pi = 0.4$.

TABELLA 3.8: Test del log-rapporto di verosimiglianza e test di Wald per $H_0 : \pi = 0.3$ con $\alpha = 0.05$.

	Ipotesi	$e\hat{p}I_{adj}$	$e\hat{p}I_{na}$	$e\hat{p}I_{e_{adj}}$	$e\hat{p}I_{e_{na}}$
$n = 10$	$H_0 : \pi = 0.3$	0.070	0.017	0.110	0.022
	$H_1 : \pi = 0.4$	0.150	0.051	0.162	0.067
$n = 20$	$H_0 : \pi = 0.3$	0.057	0.016	0.062	0.020
	$H_1 : \pi = 0.4$	0.240	0.128	0.238	0.128
$n = 50$	$H_0 : \pi = 0.3$	0.058	0.016	0.060	0.018
	$H_1 : \pi = 0.4$	0.447	0.268	0.436	0.245
$n = 100$	$H_0 : \pi = 0.3$	0.050	0.017	0.052	0.016
	$H_1 : \pi = 0.4$	0.754	0.582	0.726	0.550
$n = 500$	$H_0 : \pi = 0.3$	0.048	0.014	0.046	0.014
	$H_1 : \pi = 0.4$	1.000	1.000	1.000	1.000

Sotto H_0 , le statistiche robuste sono più efficaci di quelle non aggiustate, in quanto le stime della probabilità di errore di primo tipo approssimano meglio il valore nominale. Quando i dati vengono generati sotto l'ipotesi alternativa, specificando $\pi = 0.4$, i test robusti sembrano mostrare una potenza più elevata. Valgono, però, considerazioni analoghe a quelle fatte nel paragrafo precedente circa l'effettivo valore di α .

Nella Tabella 3.9 si riportano infine i valori delle coperture empiriche degli intervalli di confidenza corrispondenti ai test condotti. Per tutti i valori della numerosità campionaria, gli intervalli di confidenza basati sulle quantità corrette secondo la trasformazione robusta garantiscono una copertura che approssima meglio il valore nominale rispetto a quelli basati sulle quantità non trasformate. Anche in questo caso, le differenze tra i due tipo di intervalli di confidenza non sono marcate. Risultati pressoché analoghi si ottengono utilizzando valori alternativi di ρ .

TABELLA 3.9: Intervalli di confidenza per π di tipo log-rapporto di verosimiglianza e alla Wald con $1 - \alpha = 0.95$.

	π	$\hat{c}p_{adj}$	$\hat{c}p_{na}$	$\hat{c}p_{e_{adj}}$	$\hat{c}p_{e_{na}}$
$n = 10$	0.3	0.930	0.983	0.890	0.978
	0.4	0.932	0.982	0.928	0.980
$n = 20$	0.3	0.943	0.984	0.938	0.980
	0.4	0.952	0.986	0.948	0.986
$n = 50$	0.3	0.942	0.984	0.940	0.982
	0.4	0.951	0.986	0.948	0.986
$n = 100$	0.3	0.950	0.983	0.948	0.984
	0.4	0.944	0.986	0.946	0.985
$n = 500$	0.3	0.952	0.986	0.954	0.986
	0.4	0.946	0.985	0.946	0.985

Capitolo 4

Applicazione ad alcuni modelli lineari generalizzati

4.1 Introduzione

Dopo un breve richiamo alla teoria dei modelli lineari generalizzati, in questo capitolo si esamina come la correzione robusta per la funzione di verosimiglianza, proposta da Royall & Tsou (2003), si applichi ad essi. Si considerano, in particolare, il modello di regressione di Poisson con funzione di legame logaritmica e il modello di regressione logistica per dati binari raggruppati.

Per il modello di Poisson, si prende a riferimento Tsou (2006) e se ne verificano alcuni risultati, anche tramite gli studi di simulazione. Altri articoli consultati per condurre le applicazioni sono Blume et al. (2007) (contenente alcuni risultati circa la regressione logistica semplice) e Tsou & Chien (2014).

4.2 Modelli lineari generalizzati

Introdotti da Nelder & Wedderburn (1972), i modelli lineari generalizzati (GLM) per risposte univariate costituiscono una classe di modelli statistici che estende il modello di regressione lineare normale, al fine di trattare risposte con distribuzione non necessariamente normale, la cui media non è necessariamente la funzione identità del predittore lineare.

4.2.1 Famiglie di dispersione esponenziale

L'estensione da parte dei GLM a distribuzioni della variabile risposta Y_i diverse da quella normale deriva dall'assunzione secondo cui la densità delle Y_i appartenga alla classe delle famiglie di dispersione esponenziale.

La densità delle Y_i è dunque esprimibile nella forma

$$f(y_i; \omega_i, \phi) = \exp\left\{\frac{\omega_i y_i - b(\omega_i)}{a_i(\phi)} + c(y_i, \phi)\right\}, \quad (4.1)$$

dove $y_i \in \mathcal{Y}$, $\omega_i \in \Omega \subseteq \mathbb{R}$, $a_i(\phi) > 0$ e ω_i è detto parametro naturale mentre ϕ è detto parametro di dispersione. La funzione di log-verosimiglianza risulta essere

$$l(\omega_i, \phi) = \frac{\omega_i y_i - b(\omega_i)}{a_i(\phi)} + c(y_i, \phi),$$

da cui si ottengono $\partial(\omega_i, \phi)/\partial\omega_i = y_i - b'(\omega_i)/a_i(\phi)$ e $\partial^2(\omega_i, \phi)/\partial\omega_i^2 = -b''(\omega_i)/a_i(\phi)$.

Dall'identità (1.1) segue che

$$E(Y_i) = E_{\omega_i, \phi}(Y_i) = b'(\omega_i)$$

e da (1.2) si ottiene che

$$Var(Y_i) = Var_{\omega_i, \phi}(Y_i) = a_i(\phi)b''(\omega_i).$$

Posto $\mu_i = \mu(\omega_i) = E_{\omega_i, \phi}(Y_i)$, si ha

$$\mu(\omega_i) = b'(\omega_i) \implies Var_{\omega_i, \phi}(Y_i) = a_i(\phi)\mu'(\omega_i).$$

Indicando poi con $\omega(\mu_i)$ la funzione inversa di $\mu(\omega_i)$, si può scirvere

$$Var_{\omega_i, \phi}(Y_i) = a_i(\phi)b''(\omega_i)\Big|_{\omega_i=\omega(\mu_i)} = a_i(\phi)v(\mu_i),$$

dove $v(\mu_i) = b''(\omega_i)\Big|_{\omega_i=\omega(\mu_i)}$ è detta funzione di varianza. Secondo questa nuova parametrizzazione, si può indicare la distribuzione della variabile Y_i come

$$Y_i \sim DE_1(\mu_i, a_i(\phi)v(\mu_i)). \quad (4.2)$$

Per una trattazione più approfondita si rimanda a Salvani et al. (2020, capitolo 2).

4.2.2 Specificazione dei GLM

Date le osservazioni y_1, \dots, y_n , realizzazioni delle variabili casuali indipendenti Y_1, \dots, Y_n , il vettore riga delle variabili esplicative non stocastiche per l' i -esima osservazione, $\mathbf{x}_i = (x_{i0}, x_{i1}, \dots, x_{ip})$ per $i = 1, \dots, n$, e il vettore $\beta = (\beta_0, \dots, \beta_p)'$ dei coefficienti di regressione, i GLM permettono di studiare relazioni del tipo

$$g(E(Y_i)) = g(\mu_i) = \mathbf{x}_i\beta = \eta_i,$$

dove $Y_i \sim DE_1(\mu_i, a_i(\phi)v(\mu_i))$, η_i è noto come predittore lineare e $g(\cdot)$ è una funzione liscia invertibile detta di legame. Tra le possibili funzioni di legame, $g(\cdot)$ viene detta funzione di legame canonica se è tale che $g(\mu_i) = \omega(\mu_i) = \omega_i = \eta_i$.

4.3 Funzioni di verosimiglianza robusta nei GLM

L'estensione ai modelli lineari generalizzati della correzione delle funzioni di verosimiglianza al fine di renderle robuste appare naturale poiché, nella pratica, non è inusuale che la specificazione adattata possa fallire. Supponiamo che y_1, \dots, y_n siano realizzazioni delle variabili casuali indipendenti Y_1, \dots, Y_n . Sulla base di conoscenze preliminari, si assume un modello statistico per le distribuzioni di probabilità delle Y_i ,

$$\{f_{Y_i}(y_i; \beta) = f(\cdot; \eta_i(\beta)), i = 1, 2, \dots, n\}, \quad (4.3)$$

dove β è il vettore p -dimensionale dei parametri di regressione ignoti. In questo studio, assumiamo che il parametro di dispersione ϕ sia costante e noto.

Suddividiamo β come $\beta = (\psi, \lambda)$, dove ψ è il parametro di interesse unidimensionale e λ è il parametro di disturbo $(p-1)$ -dimensionale. Siano ψ_0 e λ_0 i valori limite delle stime di massima verosimiglianza, $\hat{\psi}$ e $\hat{\lambda}$, basate sul modello (4.3), quando il vero modello generatore delle Y_i è

$$\{g_i(\cdot; \theta) = g(\cdot; \eta_i(\psi, \lambda^*)), i = 1, 2, \dots, n\}, \quad (4.4)$$

dove λ^* è il vettore dei parametri di disturbo sotto $g(\cdot)$.

Supponiamo ora che la prima condizione di robustezza individuata da Royall & Tsou (2003) sia soddisfatta, cioè che il parametro di interesse ψ sotto il modello f rimanga il parametro di interesse sotto g , così che ψ_0 abbia la stessa interpretazione del vero valore del parametro di interesse. Indicate con $L(\psi, \lambda) = \prod_{i=1}^n f(y_i; \psi, \lambda)$ e $l(\psi, \lambda) = \log L(\psi, \lambda)$ le funzioni di verosimiglianza e log-verosimiglianza associate al modello f ,

individuare le quantità (2.7) e (2.8) e dette \hat{a} e \hat{b} le loro versioni empiriche, si ha che la funzione di log-verosimiglianza profilo robusta per ψ che soddisfa anche la seconda condizione di robustezza (§2.3) si ottiene come

$$\frac{\hat{a}}{\hat{b}} l(\psi, \hat{\lambda}_\psi). \quad (4.5)$$

Se il parametro di interesse ha dimensione $p > 1$, si ottengono delle matrici il cui uso è limitato alla stima robusta della matrice di covarianza. Notiamo infatti che non è ancora stata sviluppata una funzione di verosimiglianza aggiustata nel caso in cui il parametro di interesse sia multidimensionale. L'Appendice B contiene le quantità necessarie per ottenere le matrici A e B nell'ambito dei modelli di regressione sopra citati.

4.4 Regressione semplice di Poisson con legame logaritmico

Consideriamo le osservazioni y_1, \dots, y_n , realizzazioni delle variabili casuali indipendenti Y_1, \dots, Y_n , e il modello di regressione lineare semplice di Poisson con legame canonico, in cui si assume

$$\log E(Y_i) = \log \mu_i = \eta_i = \beta_0 + \beta_1 x_i, \quad (4.6)$$

da cui segue

$$\mu_i = \exp(\beta_0 + \beta_1 x_i).$$

La funzione di log-verosimiglianza associata a tale modello è

$$\begin{aligned} l(\beta_0, \beta_1) &= \sum_{i=1}^n y_i(\beta_0 + \beta_1 x_i) - \exp(\beta_0 + \beta_1 x_i) \\ &= \sum_{i=1}^n y_i(\beta_0 + \beta_1 x_i) - \mu_{i0}, \end{aligned} \quad (4.7)$$

dove μ_{i0} è il vero valore della media di Y_i sotto il vero modello g , (2.1). Assumiamo che il parametro β_1 sia di interesse. La funzione di verosimiglianza profilo per β_1 è data da

$$L_P(\beta_1) = \prod_{i=1}^n [\exp(\hat{\beta}_{0\beta_1} + \beta_1 x_i)]^{y_i} \exp[-\exp(\hat{\beta}_{0\beta_1} + \beta_1 x_i)],$$

dove $\hat{\beta}_{0\beta_1} = \log[\sum_{i=1}^n y_i / \sum_{i=1}^n \exp(\beta_1 x_i)]$ è la stima di massima verosimiglianza di β_0 con β_1 fissato.

Per individuare le quantità necessarie per la trasformazione robusta di tale funzione, ricaviamo le derivate prime e seconde della funzione di log-verosimiglianza (4.7) come segue:

$$\begin{aligned}
 l_{\beta_0} &= \sum_{i=1}^n y_i - \exp(\beta_0 + \beta_1 x_i) & l_{\beta_1} &= \sum_{i=1}^n y_i x_i - \exp(\beta_0 + \beta_1 x_i) x_i \\
 l_{\beta_0 \beta_0} &= \sum_{i=1}^n -\exp(\beta_0 + \beta_1 x_i) & l_{\beta_0 \beta_1} &= \sum_{i=1}^n -\exp(\beta_0 + \beta_1 x_i) x_i \\
 l_{\beta_1 \beta_0} &= \sum_{i=1}^n -\exp(\beta_0 + \beta_1 x_i) x_i & l_{\beta_1 \beta_1} &= \sum_{i=1}^n -\exp(\beta_0 + \beta_1 x_i) x_i^2.
 \end{aligned}$$

Calcoliamo poi

$$\begin{aligned}
 \nu_{\beta_0 \beta_0} &= E_g(l_{\beta_0}^2) = E_g\left(\sum_{i=1}^n [y_i - \exp(\beta_0 + \beta_1 x_i)]^2\right) = \sum_{i=1}^n \text{Var}_g(Y_i) \\
 \nu_{\beta_0 \beta_1} &= \nu_{\beta_1 \beta_0} = E_g(l_{\beta_0} l_{\beta_1}) \\
 &= E_g\left(\left[\sum_{i=1}^n y_i - \exp(\beta_0 + \beta_1 x_i)\right] \left[\sum_{i=1}^n y_i x_i - \exp(\beta_0 + \beta_1 x_i) x_i\right]\right) \\
 &= E_g\left(\sum_{i=1}^n [y_i x_i - \exp(\beta_0 + \beta_1 x_i)]^2 x_i\right) = \sum_{i=1}^n \text{Var}_g(Y_i) x_i \\
 \nu_{\beta_1 \beta_1} &= E_g(l_{\beta_1}^2) = E_g\left(\sum_{i=1}^n [y_i x_i - \exp(\beta_0 + \beta_1 x_i) x_i]^2\right) = \sum_{i=1}^n \text{Var}_g(Y_i) x_i^2
 \end{aligned}$$

Dalle (2.7) e (2.8), si ottengono

$$a = -l_{\beta_1 \beta_1} - \frac{(-l_{\beta_1 \beta_0})(-l_{\beta_0 \beta_1})}{-l_{\beta_0 \beta_0}} = \sum_{i=1}^n \exp(\beta_0 + \beta_1 x_i) x_i^2 - \frac{[\exp(\beta_0 + \beta_1 x_i) x_i]^2}{\exp(\beta_0 + \beta_1 x_i)} \quad (4.8)$$

e

$$\begin{aligned}
 b &= \sum_{i=1}^n \text{Var}_g(Y_i) x_i - 2 \frac{\sum_{i=1}^n \text{Var}_g(Y_i) x_i \sum_{i=1}^n \exp(\beta_0 + \beta_1 x_i) x_i}{\sum_{i=1}^n \exp(\beta_0 + \beta_1 x_i)} \\
 &+ \sum_{i=1}^n \text{Var}_g(Y_i) \left[\frac{\sum_{i=1}^n \exp(\beta_0 + \beta_1 x_i) x_i}{\sum_{i=1}^n \exp(\beta_0 + \beta_1 x_i)} \right]^2 \\
 &= \sum_{i=1}^n \text{Var}_g(Y_i) \left(x_i - \frac{\sum_{i=1}^n \exp(\beta_0 + \beta_1 x_i) x_i}{\sum_{i=1}^n \exp(\beta_0 + \beta_1 x_i)} \right)^2,
 \end{aligned} \quad (4.9)$$

dove $Var_g(Y_i)$ è la varianza di Y_i sotto il vero modello g . Le versioni empiriche di a e b , \hat{a} e \hat{b} , si ottengono sostituendo β_0 e β_1 con le loro stime di massima verosimiglianza, dunque stimando μ_{i0} con il valore stimato dal modello, $\hat{\mu}_i$, e $Var_g(Y_i)$ con $r_i^2 = (y_i - \hat{\mu}_i)^2$, ovvero con il quadrato dell' i -esimo residuo $r_i = y_i - \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)$.

4.5 Studio di simulazione per la regressione semplice di Poisson con legame logaritmico

Effettuiamo uno studio di simulazione per valutare le proprietà della trasformazione robusta nel caso del modello di regressione semplice di Poisson con legame canonico. I comandi utilizzati per ottenere i risultati illustrati nel seguito sono riportati nell'Appendice A.

Le osservazioni indipendenti, y_1, \dots, y_n , realizzazioni di Y_1, \dots, Y_n , vengono generate da una distribuzione binomiale negativa con indice r e media $E(Y_i) = \theta = \exp(\beta_0 + \beta_1 x_i)$. Fissato il valore di β_0 e supposto noto, si intende verificare l'ipotesi $H_0 : \beta_1 = \beta_{1_0}$ contro $H_1 : \beta_1 \neq \beta_{1_0}$.

Per $B = 2000$ volte, generiamo dei campioni di numerosità n e adattiamo un modello lineare generalizzato di tipo Poisson con funzione di legame logaritmica. Ricaviamo, per ciascuno degli adattamenti, il valore stimato per il coefficiente di regressione β_1 , ottenendo B replicazioni dello stesso. Calcoliamo poi le quantità $m(\hat{\beta}_1)$, $med(\hat{\beta}_1)$ e $s^2(\hat{\beta}_1)$, ovvero le stime della media, mediana e varianza campionarie.

Calcoliamo i valori di \hat{a} e \hat{b} , secondo (4.8) e (4.9) e, sulla base di questi, otteniamo le due stime della varianza di $\hat{\beta}_1$, $Var_{adj}(\hat{\beta}_1) = \hat{b}/n\hat{a}^2$ e $Var_{na}(\hat{\beta}_1) = 1/n\hat{a}$.

Per verificare l'ipotesi indicata per β_1 , si utilizzano i test log-rapporto di verosimiglianza e di Wald, nelle loro versioni robuste e non robuste. Le statistiche test di riferimento sono (3.6), (3.7), (3.8) e (3.9) e la notazione che seguirà è analoga a quella introdotta nel paragrafo §3.4.1, così come per le coperture empiriche degli intervalli di confidenza corrispondenti ai test condotti. Volendo verificare i risultati riportati in Tsou (2006), il valore dell'intercetta β_0 viene posto uguale a 1 e si specificano i valori 1 e 1.3 per β_1 . Le covariate x_i vengono generate da una distribuzione uniforme $U(0, 4)$. Le procedure vengono ripetute per $n = 50, 100, 500$ (riportate nell'articolo), con l'aggiunta di $n = 10, 20$. La Tabella 4.1 contiene i valori di $m(\hat{\beta}_1)$, $med(\hat{\beta}_1)$, $s^2(\hat{\beta}_1)$, $Var_{adj}(\hat{\beta}_1)$ e $Var_{na}(\hat{\beta}_1)$, ottenuti dalla specificazione descritta; la Tabella 4.2 contiene i valori $e\hat{p}I_{adj}$, $e\hat{p}I_{na}$, $e\hat{p}I_{e_{adj}}$ e $e\hat{p}I_{e_{na}}$ delle stime della probabilità di errore di primo tipo (sotto H_0) e

della stima della potenza dei test (sotto H_1); la Tabella 4.3 contiene i valori $\hat{c}p_{adj}$, $\hat{c}p_{na}$, $\hat{c}p_{e_{adj}}$ e $\hat{c}p_{e_{na}}$ delle coperture empiriche degli intervalli di confidenza corrispondenti.

TABELLA 4.1: Stima di alcune statistiche della SMV di β_1 , ottenute da dati generati da una distribuzione binomiale negativa con $r = 3$ e media $\exp(1 + \beta_1 x_i)$.

	β_1	$m(\hat{\beta}_1)$	$med(\hat{\beta}_1)$	$s^2(\hat{\beta}_1)$	$Var_{adj}(\hat{\beta}_1)$	$Var_{na}(\hat{\beta}_1)$
$n = 10$	1	0.989	0.973	0.108	0.062	0.011
	1.3	1.303	1.270	0.088	0.055	0.002
$n = 20$	1	0.997	0.982	0.046	0.031	0.003
	1.3	1.300	1.291	0.045	0.034	0.001
$n = 50$	1	0.996	0.989	0.017	0.016	0.001
	1.3	1.291	1.285	0.024	0.018	0.000
$n = 100$	1	0.998	0.996	0.008	0.008	0.000
	1.3	1.297	1.292	0.011	0.010	0.000
$n = 500$	1	0.998	0.997	0.002	0.002	0.000
	1.3	1.299	1.299	0.002	0.002	0.000

TABELLA 4.2: Test del log-rapporto di verosimiglianza e test di Wald per $H_0 : \beta_1 = 1$ con $\alpha = 0.05$.

	Ipotesi	$\hat{e}pI_{adj}$	$\hat{e}pI_{na}$	$\hat{e}pI_{e_{adj}}$	$\hat{e}pI_{e_{na}}$
$n = 10$	$H_0 : \beta_{1_0} = 1$	0.132	0.523	0.133	0.531
	$H_1 : \beta_{1_0} = 1.3$	0.294	0.856	0.288	0.843
$n = 20$	$H_0 : \beta_{1_0} = 1$	0.100	0.600	0.104	0.602
	$H_1 : \beta_{1_0} = 1.3$	0.393	0.912	0.370	0.905
$n = 50$	$H_0 : \beta_{1_0} = 1$	0.050	0.626	0.054	0.626
	$H_1 : \beta_{1_0} = 1.3$	0.570	0.968	0.548	0.968
$n = 100$	$H_0 : \beta_{1_0} = 1$	0.057	0.658	0.055	0.658
	$H_1 : \beta_{1_0} = 1.3$	0.857	0.997	0.839	0.997
$n = 500$	$H_0 : \beta_{1_0} = 1$	0.054	0.637	0.052	0.637
	$H_1 : \beta_{1_0} = 1.3$	1.000	1.000	1.000	1.000

TABELLA 4.3: Intervalli di confidenza per β_1 di tipo log-rapporto di verosimiglianza e alla Wald con $1 - \alpha = 0.95$.

	Ipotesi	$\hat{c}p_{adj}$	$\hat{c}p_{na}$	$\hat{c}p_{e_{adj}}$	$\hat{c}p_{e_{na}}$
$n = 10$	$H_0 : \beta_{1_0} = 1$	0.868	0.477	0.867	0.469
	$H_1 : \beta_{1_0} = 1.3$	0.903	0.207	0.900	0.204
$n = 20$	$H_0 : \beta_{1_0} = 1$	0.900	0.400	0.896	0.398
	$H_1 : \beta_{1_0} = 1.3$	0.916	0.243	0.916	0.244
$n = 50$	$H_0 : \beta_{1_0} = 1$	0.950	0.374	0.946	0.375
	$H_1 : \beta_{1_0} = 1.3$	0.927	0.213	0.922	0.212
$n = 100$	$H_0 : \beta_{1_0} = 1$	0.943	0.342	0.945	0.342
	$H_1 : \beta_{1_0} = 1.3$	0.943	0.220	0.943	0.217
$n = 500$	$H_0 : \beta_{1_0} = 1$	0.946	0.363	0.948	0.363
	$H_1 : \beta_{1_0} = 1.3$	0.960	0.233	0.959	0.232

Anche in questo caso, la vicinanza tra i valori delle medie campionarie e quelli delle mediane campionarie confermano la simmetria della distribuzione dello stimatore. La varianza robusta stima bene la varianza campionaria e la bontà della stima aumenta con l'aumentare della numerosità campionaria. Lo stesso comportamento non si registra per la stima non robusta della varianza.

Per entrambe le tipologie dei test statistici, le versioni robuste sono più efficaci di quelle che non lo sono: quando i dati sono generati sotto H_0 , il valore nominale di α viene approssimato meglio (con le dovute considerazioni già esposte nel capitolo precedente). Tale differenza si evidenzia soprattutto per valori elevati di n . Analogamente, le coperture empiriche degli intervalli di confidenza sono più vicine al valore nominale nel caso in cui vengono utilizzate le quantità trasformate.

4.6 Regressione logistica semplice

Siano le osservazioni y_1, \dots, y_n realizzazioni delle variabili casuali indipendenti Y_1, \dots, Y_n . Si assume che $m_i Y_i \sim Bi(m_i, \pi_i)$ sia dipendente dalla variabile esplicativa x_i , secondo la funzione di legame logit. Dunque il predittore lineare si può esprimere come

$$\eta_i = \mathbf{x}_i \beta = \beta_0 + \beta_1 x_i = \text{logit}[E(Y_i)] = \log\left(\frac{\pi_i}{1 - \pi_i}\right) \quad (4.10)$$

da cui segue che

$$\pi_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}}.$$

La funzione di log-verosimiglianza basata su tale modello di regressione è

$$\begin{aligned} l(\pi_i) &= \sum_{i=1}^n s_i \log \pi_i + (m_i - s_i) \log(1 - \pi_i) \\ &= \sum_{i=1}^n s_i \log\left(\frac{\pi_i}{1 - \pi_i}\right) + m_i \log(1 - \pi_i) \\ &= \sum_{i=1}^n s_i \eta_i - m_i \log(1 + e^{\eta_i}) \\ &= \sum_{i=1}^n s_i (\beta_0 + \beta_1 x_i) - m_i \log(1 + \exp\{\beta_0 + \beta_1 x_i\}). \end{aligned} \quad (4.11)$$

Ricaviamo le quantità necessarie per il calcolo del fattore di correzione, a/b , calcolando

$$l_{\beta_0} = \sum_{i=1}^n s_i - m_i \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \quad l_{\beta_1} = \sum_{i=1}^n s_i x_i - m_i \frac{\exp(\beta_0 + \beta_1 x_i) x_i}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

$$l_{\beta_0 \beta_0} = \sum_{i=1}^n -m_i \frac{\exp(\beta_0 + \beta_1 x_i)}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2} \quad l_{\beta_0 \beta_1} = \sum_{i=1}^n -m_i \frac{\exp(\beta_0 + \beta_1 x_i) x_i}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2}$$

$$l_{\beta_1 \beta_0} = \sum_{i=1}^n -m_i \frac{\exp(\beta_0 + \beta_1 x_i) x_i}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2} \quad l_{\beta_1 \beta_1} = \sum_{i=1}^n -m_i \frac{\exp(\beta_0 + \beta_1 x_i) x_i^2}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2}$$

$$\nu_{\beta_0 \beta_0} = E_g(l_{\beta_0}^2) = E_g\left(\sum_{i=1}^n \left[s_i - m_i \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}\right]^2\right) = \sum_{i=1}^n \text{Var}_g(Y_i)$$

$$\begin{aligned} \nu_{\beta_0 \beta_1} &= \nu_{\beta_1 \beta_0} = E_g(l_{\beta_0} l_{\beta_1}) \\ &= E_g\left(\left[\sum_{i=1}^n s_i - m_i \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}\right] \left[\sum_{i=1}^n s_i x_i - m_i \frac{\exp(\beta_0 + \beta_1 x_i) x_i}{1 + \exp(\beta_0 + \beta_1 x_i)}\right]\right) \\ &= E_g\left(\sum_{i=1}^n \left[s_i - m_i \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}\right]^2 x_i\right) = \sum_{i=1}^n \text{Var}_g(Y_i) x_i \end{aligned}$$

$$\nu_{\beta_1 \beta_1} = E_g(l_{\beta_1}^2) = E_g\left(\sum_{i=1}^n \left[s_i x_i - m_i \frac{\exp(\beta_0 + \beta_1 x_i) x_i}{1 + \exp(\beta_0 + \beta_1 x_i)}\right]^2\right) = \sum_{i=1}^n \text{Var}_g(Y_i) x_i^2,$$

da cui si ricavano

$$\begin{aligned} a &= -l_{\beta_1 \beta_1} - \frac{(-l_{\beta_1 \beta_0})(-l_{\beta_0 \beta_1})}{-l_{\beta_0 \beta_0}} = \sum_{i=1}^n \frac{m_i \exp(\beta_0 + \beta_1 x_i) x_i^2}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2} \\ &\quad - \left[\sum_{i=1}^n \frac{m_i \exp(\beta_0 + \beta_1 x_i) x_i}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2}\right]^2 / \sum_{i=1}^n \frac{m_i \exp(\beta_0 + \beta_1 x_i)}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2} \end{aligned} \quad (4.12)$$

e

$$\begin{aligned}
 b &= \sum_{i=1}^n Var_g(Y_i)x_i \\
 &- 2 \sum_{i=1}^n Var_g(Y_i)x_i \sum_{i=1}^n \frac{m_i \exp(\beta_0 + \beta_1 x_i)x_i}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2} / \sum_{i=1}^n \frac{m_i \exp(\beta_0 + \beta_1 x_i)}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2} \\
 &+ \sum_{i=1}^n Var_g(Y_i) \left\{ \sum_{i=1}^n \frac{m_i \exp(\beta_0 + \beta_1 x_i)x_i}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2} / \sum_{i=1}^n \frac{m_i \exp(\beta_0 + \beta_1 x_i)}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2} \right\}^2 \\
 &= \sum_{i=1}^n Var_g(Y_i) \left(x_i - \sum_{i=1}^n \frac{m_i \exp(\beta_0 + \beta_1 x_i)x_i}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2} / \sum_{i=1}^n \frac{m_i \exp(\beta_0 + \beta_1 x_i)}{[1 + \exp(\beta_0 + \beta_1 x_i)]^2} \right)^2,
 \end{aligned} \tag{4.13}$$

dove $Var_g(Y_i)$ è la varianza di Y_i sotto il vero modello g . La versione empirica di a e b , \hat{a} e \hat{b} , si ottengono sostituendo β_0 e β_1 con le loro stime di massima verosimiglianza e $Var_g(Y_i)$ con $r_i^2 = (y_i - \hat{\mu}_i)^2 = (y_i - m_i \hat{\pi}_i)^2$, dove $\hat{\pi}_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i) / [1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)]$.

Anche in questo caso, si potrebbero condurre degli studi di simulazione per verificare le proprietà della trasformazione proposta, generando i dati da una distribuzione beta-binomiale ed adattando un modello di regressione per dati binomiali, con funzione di legame logistica.

Conclusioni

In questa relazione si è affrontato il tema della robustezza della funzione di verosimiglianza, analizzando la trasformazione proposta da Royall & Tsou (2003). La medesima trasformazione, sulla base di quanto introdotto già da Stafford (1996), è stata analizzata anche nell'ambito della funzione di verosimiglianza profilo.

Si è proceduto all'applicazione dei metodi proposti ai modelli di campionamento casuale semplice, con particolare attenzione al modello Poisson e al modello binomiale. Sono state individuate le quantità necessarie per ottenere la trasformazione robusta nei suddetti casi e sono stati condotti degli studi di simulazione per verificare le proprietà della correzione. L'efficacia della trasformazione è stata confermata, in modo più evidente nel caso Poisson, per le procedure di stima intervallare e di verifica di ipotesi. Sono state confrontate le procedure di inferenza basate sulle versioni robuste della statistica log-rapporto di verosimiglianza e della statistica di Wald; il confronto tra le due statistiche non ha mostrato differenze notevoli.

La trasformazione oggetto dello studio è stata applicata poi ad alcuni modelli lineari generalizzati: il modello di regressione semplice di Poisson con funzione di legame logaritmica e il modello di regressione logistica semplice. Per il modello di regressione di Poisson si sono condotti degli studi di simulazione che hanno confermato l'efficacia della trasformazione robusta; anche in questo caso, non si sono registrate notevoli differenze tra i due tipi di test impiegati.

Molteplici sono le ulteriori possibili applicazioni da poter analizzare: il caso di diverse funzioni di legame per i modelli analizzati (in Tsou (2006) si analizza anche il caso del modello di Poisson con funzione di legame identità giungendo a risultati analoghi a quelli ottenuti con il legame canonico) o generare i dati da distribuzioni diverse negli studi di simulazione; si possono studiare modelli diversi da quelli considerati in questo studio (come fatto, ad esempio, con il modello normale e gamma in Tsou & Chien (2014), con risultati simili a quelli ottenuti in questo studio).

Appendice

Appendice A: Capitolo 3

Vengono di seguito riportati i comandi che permettono di ottenere i risultati delle simulazioni illustrate nel Capitolo 3.

Modello di campionamento casuale semplice Poisson

```
sim.nbin <- function(n, vero.theta, theta0, r, B = 2000){
  #generazione di B campioni n-dimensionali dalla distribuzione
  #binomiale negativa
  y <- matrix(rnbinom(n * B, size = r, mu = vero.theta),
              nrow = n, ncol = B)
  #media campionaria (SMV) di ogni campione
  y.bar <- apply(y, 2, mean)
  #varianza campionaria di ogni campione
  v <- apply(y, 2, var)
  #media delle medie campionarie
  m <- mean(y.bar)
  #mediana delle medie campionarie
  med <- median(y.bar)
  #varianza campionaria delle medie campionarie
  s2 <- var(y.bar)
  #fattori di correzione
  a.hat <- mean(1 / y.bar)
  b.hat <- mean(v / n / (y.bar^2) * (n-1))
  #varianza corretta = (b.hat/a.hat^2)/n
  adj.var <- (b.hat/a.hat^2)/n
  # varianza non corretta = 1/(n*a.hat)
  var <- 1 / (n * a.hat)
  #TRV corretto
  l.theta.hat <- -n * y.bar + n * y.bar * log(y.bar)
  l.theta.zero <- -n * theta0 + n * y.bar * log(theta0)
```

```

l.vero.theta <- -n * vero.theta + n * y.bar * log(vero.theta)
W <- 2 * a.hat / b.hat * (l.theta.hat - l.theta.zero)
W1 <- 2 * a.hat / b.hat * (l.theta.hat - l.vero.theta)
epI.adj <- mean(W >= qchisq(0.95, 1))
epI.adj.1 <- mean(W1 >= qchisq(0.95, 1))
# TRV non corretto
W.na <- 2 * (l.theta.hat - l.theta.zero)
W1.na <- 2 * (l.theta.hat - l.vero.theta)
epI.na <- mean(W.na >= qchisq(0.95, 1))
epI.na.1 <- mean(W1.na >= qchisq(0.95, 1))
# Wald corretto
We <- (y.bar - theta0)^2 / adj.var
We1 <- (y.bar - vero.theta)^2 / adj.var
epI.adj.e <- mean(We >= qchisq(0.95, 1))
epI.adj.e.1 <- mean(We1 >= qchisq(0.95, 1))
# Wald non corretto
We.na <- (y.bar - theta0)^2 / var
We1.na <- (y.bar - vero.theta)^2 / var
epI.na.e <- mean(We.na >= qchisq(0.95, 1))
epI.na.e.1 <- mean(We1.na >= qchisq(0.95, 1))
if(vero.theta == theta0){
  out <- round(c(m, med, s2, adj.var, var, epI.adj, epI.na,
                epI.adj.e, epI.na.e, 1-epI.adj, 1-epI.na,
                1-epI.adj.e, 1-epI.na.e), 3)
  names(out) <- c("m(y.bar)", "med(y.bar)", "s2(y.bar)",
                "adjvar(y.bar)", "var(y.bar)",
                "epI.adj.trv", "epI.na.trv", "epI.adj.We",
                "epI.na.We", "cp.adj.trv", "cp.na.trv",
                "cp.adj.We", "cp.na.We")
  return(out)
}
if(vero.theta != theta0){
  out <- round(c(m, med, s2, adj.var, var, epI.adj, epI.na,
                epI.adj.e, epI.na.e, 1-epI.adj.1,
                1-epI.na.1, 1-epI.adj.e.1, 1-epI.na.e.1), 3)
  names(out) <- c("m(y.bar)", "med(y.bar)", "s2(y.bar)",
                "adjvar(y.bar)", "var(y.bar)",
                "epI.adj.trv", "epI.na.trv", "epI.adj.We",
                "epI.na.We", "cp.adj.trv", "cp.na.trv",

```

```
                                "cp.adj.We", "cp.na.We")
    return(out)
  }
}

n1 <- 10
n2 <- 20
n3 <- 50
n4 <- 100
n5 <- 500

set.seed(123)
sim1.1 <- sim.nbin(n1, 6, 6, 2)
sim1.2 <- sim.nbin(n1, 7, 6, 2)
sim1.3 <- sim.nbin(n2, 6, 6, 2)
sim1.4 <- sim.nbin(n2, 7, 6, 2)
sim1.5 <- sim.nbin(n3, 6, 6, 2)
sim1.6 <- sim.nbin(n3, 7, 6, 2)
sim1.7 <- sim.nbin(n4, 6, 6, 2)
sim1.8 <- sim.nbin(n4, 7, 6, 2)
sim1.9 <- sim.nbin(n5, 6, 6, 2)
sim1.10 <- sim.nbin(n5, 7, 6, 2)

sim2.1 <- sim.nbin(n1, 4, 4, 3)
sim2.2 <- sim.nbin(n1, 5, 4, 3)
sim2.3 <- sim.nbin(n2, 4, 4, 3)
sim2.4 <- sim.nbin(n2, 5, 4, 3)
sim2.5 <- sim.nbin(n3, 4, 4, 3)
sim2.6 <- sim.nbin(n3, 5, 4, 3)
sim2.7 <- sim.nbin(n4, 4, 4, 3)
sim2.8 <- sim.nbin(n4, 5, 4, 3)
sim2.9 <- sim.nbin(n5, 4, 4, 3)
sim2.10 <- sim.nbin(n5, 5, 4, 3)
```

Modello di campionamento casuale semplice binomiale

```

library(VGAM)
simulazioni.betabin <- function(n, vero.pi, pi0, m, rho,
                                B = 2000){
  #generazione di B campioni n-dimensionali
  #dalla distribuzione binomiale negativa
  s <- matrix(rbetabinom(n * B, size = m, prob = vero.pi,
                        rho = rho), nrow = n, ncol = B)
  #media campionaria (SMV) di ogni campione
  y.bar <- apply(s, 2, function(x) mean(x/m))
  #varianza campionaria di ogni campione
  v <- apply(s, 2, function(x) var(x/m))
  #media delle medie campionarie
  m <- mean(y.bar)
  #mediana delle medie campionarie
  med <- median(y.bar)
  #varianza campionaria delle medie campionarie
  s <- var(y.bar)
  #fattori di correzione
  a <- mean(1 / (y.bar * (1-y.bar)))
  b <- mean(v / n / (y.bar^2) / ((1-y.bar)^2) * (n-1))
  #varianza corretta = (a^2/b)/n
  adj.var <- (b/a^2)/n
  # varianza non corretta = 1/(na)
  var <- 1 / (n * a)
  #TRV corretto
  l.pi.hat <- n*(y.bar)*log((y.bar)/(1-y.bar))+n*log(1-y.bar)
  l.pi.zero <- n*(y.bar)*log((pi0)/(1-pi0))+n*log(1-pi0)
  l.vero.pi <- n*(y.bar)*log((vero.pi)/(1-vero.pi))
  +n*log(1-vero.pi)
  W <- 2 * a / b * (l.pi.hat - l.pi.zero)
  W1 <- 2 * a / b * (l.pi.hat - l.vero.pi)
  epI.adj <- mean(W >= qchisq(0.95, 1))
  epI.adj.1 <- mean(W1 >= qchisq(0.95, 1))
  # TRV non corretto
  W.nc <- 2 * (l.pi.hat - l.pi.zero)
  W1.nc <- 2 * (l.pi.hat - l.vero.pi)
  epI.na <- mean(W.nc >= qchisq(0.95, 1))

```



```

epI.na.1 <- mean(W1.nc >= qchisq(0.95, 1))
# Wald corretto
We <- (y.bar - pi0)^2 / adj.var
We1 <- (y.bar - vero.pi)^2 / adj.var
epI.adj.e <- mean(We >= qchisq(0.95, 1))
epI.adj.e.1 <- mean(We1 >= qchisq(0.95, 1))
# Wald non corretto
We.nc <- (y.bar - pi0)^2 / var
We1.nc <- (y.bar - vero.pi)^2 / var
epI.na.e <- mean(We.nc >= qchisq(0.95, 1))
epI.na.e.1 <- mean(We1.nc >= qchisq(0.95, 1))
if(vero.pi == pi0){
  out <- round(c(m, med, s2, adj.var, var, epI.adj,
                epI.na, epI.adj.e, epI.na.e,
                1-epI.adj, 1-epI.na, 1-epI.adj.e,
                1-epI.na.e), 3)
  names(out) <- c("m(y.bar)", "med(y.bar)", "s2(y.bar)",
                "adjvar(y.bar)", "var(y.bar)",
                "epI.adj.trv", "epI.na.trv",
                "epI.adj.We", "epI.na.We",
                "cp.adj.trv", "cp.na.trv",
                "cp.adj.We", "cp.na.We")
  return(out)
}
if(vero.pi != pi0){
  out <- round(c(m, med, s2, adj.var, var, epI.adj,
                epI.na, epI.adj.e, epI.na.e,
                1-epI.adj.1, 1-epI.na.1, 1-epI.adj.e.1,
                1-epI.na.e.1), 3)
  names(out) <- c("m(y.bar)", "med(y.bar)", "s2(y.bar)",
                "adjvar(y.bar)", "var", "epI.adj.trv",
                "epI.na.trv", "epI.adj.We", "epI.na.We",
                "cp.adj.trv", "cp.na.trv",
                "cp.adj.We", "cp.na.We")
  return(out)
}
}
n1 <- 10

```

```
n2 <- 20
n3 <- 50
n4 <- 100
n5 <- 500

set.seed(459)
m1 <- rep(10, n1)
simulazioni.betabin(n1, 0.3, 0.3, m1, 0.6)
simulazioni.betabin(n1, 0.4, 0.3, m1, 0.6)
m2 <- rep(10, n2)
simulazioni.betabin(n2, 0.3, 0.3, m2, 0.6)
simulazioni.betabin(n2, 0.4, 0.3, m2, 0.6)
m3 <- rep(10, n3)
simulazioni.betabin(n3, 0.3, 0.3, m3, 0.6)
simulazioni.betabin(n3, 0.4, 0.3, m3, 0.6)
m4 <- rep(10, n4)
simulazioni.betabin(n4, 0.3, 0.3, m4, 0.6)
simulazioni.betabin(n4, 0.4, 0.3, m4, 0.6)
m5 <- rep(10, n5)
simulazioni.betabin(n5, 0.3, 0.3, m5, 0.6)
simulazioni.betabin(n5, 0.4, 0.3, m5, 0.6)
```

Modello di regressione semplice Poisson - legame logaritmico

```

sim.reg.log <- function(n, size, vero.beta1, beta1.0, B=2000){
  beta1.hat <- rep(NA, B)
  #variabile esplicativa generata dalla distribuzione U(1,4)
  x <- runif(n,1,4)
  y <- mu.hat <- res <- matrix(NA, nrow = n, ncol = B,
  byrow = FALSE)
  mu = exp(1 + vero.beta1*x)
  #generazione dei dati e adattamento del modello
  for(i in 1:B){
    y[,i] <- rnbinom(n, size = 3, mu = mu)
    poi.log <- glm(y[,i] ~ x, family = poisson)
    beta1.hat[i] <- coef(poi.log)[2]
    mu.hat[,i] <- fitted(poi.log)
  }
  #media campionaria di beta1.hat
  m <- mean(beta1.hat)
  #mediana campionaria di beta1.hat
  med <- median(beta1.hat)
  #varianza campionaria di beta1.hat
  s2 <- var(beta1.hat)
  #fattori di correzione
  x.tilde <- sum(x*exp(1 + vero.beta1*x))/
    sum(exp(1 + vero.beta1*x))
  a <- mean((x - x.tilde)^2*exp(1+vero.beta1*x))
  b <- mean((y-mu.hat)^2*(x - x.tilde)^2)
  #varianza corretta
  adj.var <- (a^2/b)^(-1)/n
  #varianza non corretta
  var <- 1 / a / n
  # TRV corretto
  #d: matrice n*B le cui colonne sono ottenute moltiplicando
  #l'i-esimo elemento di beta1.hat x
  d <- matrix(NA, n, B, byrow = FALSE)
  for(i in 1:B){
    d[,i] <- beta1.hat[i] * x
  }
  #SMV vincolata di beta0, fissato beta1.hat

```

```

beta0.p <- log(colSums(y)/colSums(exp(d)))
#Mhat: matrice n*B nelle cui colonne viene ripetuto n volte
#l'i-esimo valore di beta0.p
Mhat <- matrix(NA, n, B)
for(i in 1:B){
  Mhat[,i] <- beta0.p[i]
}
#d0: matrice n*B le cui colonne sono ottenute
#moltiplicando beta1.0 per x
d0 <- matrix(NA, n, B)
for(i in 1:B){
  d0[,i] <- beta1.0 * x
}
#SMV vincolata di beta0, fissato beta1.0
beta0.p0 <- log(colSums(y)/colSums(exp(d0)))
#M0: matrice n*B nelle cui colonne viene ripetuto n volte
#l'i-esimo valore di beta0.p0
M0 <- matrix(NA, n, B)
for(i in 1:B){
  M0[,i] <- beta0.p0[i]
}
#dv: matrice n*B le cui colonne sono ottenute
#moltiplicando vero.beta1 per x
dv <- matrix(NA, n, B)
for(i in 1:B){
  dv[,i] <- vero.beta1 * x
}
#SMV vincolata di beta0, fissato vero.beta1
beta0.pvero <- log(colSums(y)/colSums(exp(dv)))
#Mv: matrice n*B nelle cui colonne viene ripetuto n volte
#l'i-esimo valore di beta0.pvero
Mv <- matrix(NA, n, B)
for(i in 1:B){
  Mv[,i] <- beta0.pvero[i]
}
l.beta.hat <- colSums(y*(Mhat+d) - exp(Mhat+d))
l.beta.zero <- colSums(y*(M0+d0)-exp(M0+d0))
l.vero.beta <- colSums(y*(Mv+dv)-exp(Mv+dv))
W <- 2 * a / b * (l.beta.hat - l.beta.zero)

```

```

W1 <- 2 * a / b * (l.beta.hat - l.vero.beta)
epI.adj <- mean(W >= qchisq(0.95, 1))
epI.adj.1 <- mean(W1 >= qchisq(0.95, 1))
# TRV non corretto
W.nc <- 2 * (l.beta.hat - l.beta.zero)
W1.nc <- 2 * (l.beta.hat - l.vero.beta)
epI.na <- mean(W.nc >= qchisq(0.95, 1))
epI.na.1 <- mean(W1.nc >= qchisq(0.95, 1))
# Wald corretto
We <- (beta1.hat - beta1.0)^2 / adj.var
We1 <- (beta1.hat - vero.beta1)^2 / adj.var
epI.adj.e <- mean(We >= qchisq(0.95, 1))
epI.adj.e.1 <- mean(We1 >= qchisq(0.95, 1))
# Wald non corretto
We.na <- (beta1.hat - beta1.0)^2 / var
We1.na <- (beta1.hat - vero.beta1)^2 / var
epI.na.e <- mean(We.na >= qchisq(0.95, 1))
epI.na.e.1 <- mean(We1.na >= qchisq(0.95, 1))
if(vero.beta == beta0){
  out <- round(c(m, med, s2, adj.var, var, epI.adj, epI.na,
                epI.adj.e, epI.na.e, 1-epI.adj, 1-epI.na,
                1-epI.adj.e, 1-epI.na.e), 3)
  names(out) <- c("m(beta1.hat)", "med(beta1.hat)",
                 "s2(beta1.hat)", "adjvar(beta1.hat)",
                 "var(beta1.hat)", "epI.adj.trv",
                 "epI.na.trv", "epI.adj.We",
                 "epI.na.We", "cp.adj.trv", "cp.na.trv",
                 "cp.adj.We", "cp.na.We")
  return(out)
}
if(vero.beta != beta0){
  out <- round(c(m, med, s2, adj.var, var, epI.adj, epI.na,
                epI.adj.e, epI.na.e, 1-epI.adj.1,
                1-epI.na.1, 1-epI.adj.e.1, 1-epI.na.e.1), 3)
  names(out) <- c("m(beta1.hat)", "med(beta1.hat)",
                 "s2(beta1.hat)", "adjvar(beta1.hat)",
                 "var(beta1.hat)", "epI.adj.trv",
                 "epI.na.trv", "epI.adj.We",
                 "epI.na.We", "cp.adj.trv",

```

```
                                "cp.na.trv", "cp.adj.We", "cp.na.We")
  return(out)
}
}

n1 <- 10
n2 <- 20
n3 <- 50
n4 <- 100
n5 <- 500

set.seed(789)
sim.reg.log(n1, 3, 1, 1)
sim.reg.log(n1, 3, 1.3, 1)
sim.reg.log(n2, 3, 1, 1)
sim.reg.log(n2, 3, 1.3, 1)
sim.reg.log(n3, 3, 1, 1)
sim.reg.log(n3, 3, 1.3, 1)
sim.reg.log(n4, 3, 1, 1)
sim.reg.log(n4, 3, 1.3, 1)
sim.reg.log(n5, 3, 1, 1)
sim.reg.log(n5, 3, 1.3, 1)
```

Appendice B: Capitolo 4

Regressione di Poisson con legame logaritmico

Una generalizzazione per i modelli di regressione di Poisson con $p > 1$ parametri di regressione si può ottenere come segue. Consideriamo le osservazioni y_1, \dots, y_n , realizzazioni delle variabili casuali indipendenti Y_1, \dots, Y_n , e il modello di regressione lineare di poisson con legame canonico, in cui ogni Y_i dipende dal vettore delle covariate $\mathbf{x}_i = (x_{i0}, x_{i1}, \dots, x_{ip-1})^T$, con $x_{i0} = 1$. In particolare, si assume

$$\log E(Y_i) = \log \mu_i = \eta_i = \beta_0 + \dots + \beta_{p-2}x_{ip-2} + \beta_{p-1}x_{ip-1}. \quad (\text{A.1})$$

La funzione di log-verosimiglianza basata su tale modello di regressione è

$$\begin{aligned} l(\beta) &= \sum_{i=1}^n y_i(\beta_0 + \dots + \beta_{p-1}x_{ip-1}) - \exp(\beta_0 + \dots + \beta_{p-1}x_{ip-1}) \\ &= \sum_{i=1}^n y_i(\beta_0 + \dots + \beta_{p-2}x_{ip-2} + \beta_{p-1}x_{ip-1}) - \mu_{i0} \end{aligned} \quad (\text{A.2})$$

Assumiamo, senza perdita di generalità, che

$$\psi = (\beta_{p-1}, \beta_{p-2}, \dots, \beta_{p-q+1}, \beta_{p-q})^T = (\psi_1, \psi_2, \dots, \psi_{q-1}, \psi_q)^T$$

sia il vettore dei parametri di interesse di lunghezza q e che

$$\lambda = (\beta_0, \beta_1, \dots, \beta_{p-q+1})^T = (\lambda_1, \lambda_2, \dots, \lambda_{p-w})^T$$

sia il vettore $(p-q)$ -dimensionale dei parametri di disturbo. Siano μ_{i0} e μ'_{i0} i veri valori di μ_i e μ'_i , rispettivamente, dove μ'_i è la derivata prima della media rispetto a η_i .

Indicata con $Var_g(Y_i)$ la vera varianza delle Y_i , $i = 1, \dots, n$, e con $Z = (\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_{p-1})$ la matrice di disegno del modello dimensione $n \times p$, tale che $Z' = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{p-1})$, le quantità A e B sono delle matrici $q \times q$, i cui rispettivi elementi di posizione (u, v) , $u, v = 1, \dots, q$ sono

$$A_{(uv)} = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu_{i0} \left(x_{ip-u} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(u)}|}{|\Delta|} x_{ij-1} \right) \left(x_{ip-v} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(v)}|}{|\Delta|} x_{ij-1} \right) / n \quad (\text{A.3})$$

e

$$B_{(uv)} = \lim_{n \rightarrow \infty} \sum_{i=1}^n \text{Var}_g(Y_i) \left(x_{ip-u} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(u)}|}{|\Delta|} x_{ij-1} \right) \left(x_{ip-v} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(v)}|}{|\Delta|} x_{ij-1} \right) / n, \quad (\text{A.4})$$

dove $|\Delta|$ è il determinante della matrice $\Delta = W^T V^{-1} W$; la matrice W ha come colonne i vettori $(\mathbf{z}_0, \dots, \mathbf{z}_{j-2}, \mathbf{z}_{j-1}, \mathbf{z}_j, \dots, \mathbf{z}_{p-q-1})$ e $V = \text{diag}(1/\mu_{i0})$ è una matrice diagonale di ordine n . Le matrici $\Delta_{j(u)}$ e $\Delta_{j(v)}$ sono poi definite come $W^T V^{-1} W_{j(u)}$ e $W^T V^{-1} W_{j(v)}$, rispettivamente, dove $W_{j(u)} = (\mathbf{z}_0, \dots, \mathbf{z}_{j-2}, \mathbf{z}_{p-u}, \mathbf{z}_j, \dots, \mathbf{z}_{p-q-1})$ e $W_{j(v)} = (\mathbf{z}_0, \dots, \mathbf{z}_{j-2}, \mathbf{z}_{p-v}, \mathbf{z}_j, \dots, \mathbf{z}_{p-q-1})$ sono ottenute a partire dalla matrice W , sostituendo la j -esima colonna \mathbf{z}_{j-1} con \mathbf{z}_{p-u} e \mathbf{z}_{p-v} , rispettivamente. Nelle applicazioni, si possono ottenere delle stime consistenti di A e B sostituendo le quantità ignote tramite una loro stima consistente.

Al fine di calcolare gli elementi che costituiscono le quantità (A.3) e (A.4), definiamo alcune quantità, a partire dalla forma della funzione di log-verosimiglianza (A.2). Siano, per $j = 1, \dots, q$ e $t = 0, \dots, p - q - 1$

$$\begin{aligned} l_\psi &= \left[\sum_{i=1}^n y_i x_{ip-j} - \mu_{i0} x_{ip-j} \right] \\ &= \left[\sum_{i=1}^n x_{ip-1} (y_i - \mu_{i0}) \quad \dots \quad \sum_{i=1}^n x_{ip-w} (y_i - \mu_{i0}) \right] \end{aligned}$$

e

$$\begin{aligned} l_\lambda &= \left[\sum_{i=1}^n y_i x_{it} - \mu_{i0} x_{it} \right] \\ &= \left[\sum_{i=1}^n (y_i - \mu_{i0}) \quad \dots \quad \sum_{i=1}^n x_{ip-w-1} (y_i - \mu_{i0}) \right] \end{aligned}$$

i vettori di lunghezza, rispettivamente, q e $(p - q)$ delle derivate prime parziali di (A.2) rispetto al vettore dei parametri di interesse ψ e a quello dei parametri di disturbo λ (assumendo $x_{i0} = 1, \forall i$).

Da questi vettori si ricavano le seguenti matrici di derivate seconde

$$\begin{aligned} l_{\psi\psi} &= \left[\sum_{i=1}^n -\mu_{i0} x_{ip-j} x_{ip-k} \right] \\ &= - \begin{bmatrix} \sum_{i=1}^n \mu_{i0} x_{ip-1}^2 & \dots & \sum_{i=1}^n \mu_{i0} x_{ip-1} x_{ip-q} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n \mu_{i0} x_{ip-q} x_{ip-1} & \dots & \sum_{i=1}^n \mu_{i0} x_{ip-q}^2 \end{bmatrix}, \end{aligned}$$

di dimensione $q \times q$, con $j, k = 1, \dots, q$;

$$\begin{aligned} l_{\psi\lambda} &= \left[\sum_{i=1}^n -\mu_{i0} x_{ip-j} x_{it} \right] \\ &= - \begin{bmatrix} \sum_{i=1}^n \mu_{i0} x_{ip-1} x_{i0} & \cdots & \sum_{i=1}^n \mu_{i0} x_{ip-1} x_{ip-q-1} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n \mu_{i0} x_{ip-q} x_{i0} & \cdots & \sum_{i=1}^n \mu_{i0} x_{ip-q} x_{ip-q-1} \end{bmatrix}, \end{aligned}$$

di dimensione $q \times (p - q)$, con $j = 1, \dots, q$ e $t = 0, \dots, p - q - 1$;

$$\begin{aligned} l_{\lambda\lambda} &= \left[\sum_{i=1}^n -\mu_{i0} x_{is} x_{it} \right] \\ &= - \begin{bmatrix} \sum_{i=1}^n \mu_{i0} x_{i0}^2 & \cdots & \sum_{i=1}^n \mu_{i0} x_{i0} x_{ip-q-1} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n \mu_{i0} x_{i0} x_{ip-q-1} & \cdots & \sum_{i=1}^n \mu_{i0} x_{ip-q-1}^2 \end{bmatrix}, \end{aligned}$$

di dimensione $(p - q) \times (p - q)$, con $s, t = 0, \dots, p - q - 1$ e

$$\begin{aligned} l_{\lambda\psi} &= \left[\sum_{i=1}^n -\mu_{i0} x_{it} x_{ip-j} \right] \\ &= - \begin{bmatrix} \sum_{i=1}^n \mu_{i0} x_{i0} x_{ip-1} & \cdots & \sum_{i=1}^n \mu_{i0} x_{i0} x_{ip-q} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n \mu_{i0} x_{ip-q-1} x_{ip-1} & \cdots & \sum_{i=1}^n \mu_{i0} x_{ip-q-1} x_{ip-q} \end{bmatrix}, \end{aligned}$$

di dimensione $(p - q) \times q$, con $j = 1, \dots, w$ e $t = 0, \dots, p - q - 1$.

Sappiamo che $A = I_{\psi\psi} - I_{\psi\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\psi}$ e $B = \nu_{\psi\psi} - 2\nu_{\psi\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\psi} + I_{\psi\lambda} I_{\lambda\lambda}^{-1} \nu_{\lambda\lambda} I_{\lambda\lambda}^{-1} I_{\lambda\psi}$. Per quanto riguarda il calcolo delle quantità necessarie per ottenere la matrice A , si ha che $I_{\psi\psi} = \lim_{n \rightarrow \infty} -E(l_{\psi\psi}/n)$, $I_{\psi\lambda} = \lim_{n \rightarrow \infty} -E(l_{\psi\lambda}/n)$ e $I_{\lambda\psi} = \lim_{n \rightarrow \infty} -E(l_{\lambda\psi}/n)$. Inoltre, poiché $I_{\lambda\lambda} = \lim_{n \rightarrow \infty} -E(l_{\lambda\lambda}/n)$, detta Δ la matrice $l_{\lambda\lambda}$, si ha che

$$I_{\lambda\lambda}^{-1} = \frac{n}{|\Delta|} \begin{bmatrix} R_{1,1} & \cdots & R_{p-q,1} \\ \vdots & \ddots & \vdots \\ R_{1,p-q} & \cdots & R_{p-q,p-q} \end{bmatrix},$$

dove $R_{l,m} = (-1)^{l+m} |M_{(lm)}(\Delta)|$ e il (l, m) -esimo cofattore di Δ .

L'elemento in posizione (u, v) della matrice di dimensione $q \times q$, A_l , è

$$\begin{aligned}
A_{(uv)} &= I_{\psi_u \psi_v} - I_{\psi_u \lambda} I_{\lambda \lambda}^{-1} I_{\lambda \psi_v} - I_{\psi_v \lambda} I_{\lambda \lambda}^{-1} I_{\lambda \psi_u} + I_{\psi_v \lambda} I_{\lambda \lambda}^{-1} I_{\lambda \lambda} I_{\lambda \lambda}^{-1} I_{\lambda \psi_u} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \left\{ \sum_{i=1}^n \mu_{i0} x_{ip-u} x_{ip-v} - \frac{1}{|\Delta_l|} \sum_{i=1}^n \mu_{i0} \left(x_{ip-v} \sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(u)}| \right) \right. \\
&\quad \left. - \frac{1}{|\Delta|} \sum_{i=1}^n \mu_{i0} \left(x_{ip-u} \sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(v)}| \right) \right. \\
&\quad \left. + \frac{1}{|\Delta|^2} \sum_{i=1}^n \mu_{i0} \left(\sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(v)}| \right) \left(\sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(u)}| \right) \right\} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mu_{i0} \left(x_{ip-u} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(u)}|}{|\Delta|} x_{ij-1} \right) \left(x_{ip-v} - \sum_{j=1}^{p-q} x_{ij-1} \frac{|\Delta_{j(v)}|}{|\Delta|} x_{ij-1} \right),
\end{aligned}$$

dove $|\Delta_{j(u)}| = \sum_{m=1}^{p-q} (R_{j,m} \sum_{i=1}^n \mu_{i0} x_{im-1} x_{ip-u})$.

Con passaggi analoghi a quelli visti nel §3.4.2, si ottengono le quantità necessarie per ottenere la matrice di dimensione $q \times q$, B . In particolare,

$$\begin{aligned}
\nu_{\psi\psi} &= \left[\sum_{i=1}^n \text{Var}_g(Y_i) x_{ip-j} x_{ip-k} \right] \\
&= \begin{bmatrix} \sum_{i=1}^n \text{Var}_g(Y_i) x_{ip-1}^2 & \cdots & \sum_{i=1}^n \text{Var}_g(Y_i) x_{ip-1} x_{ip-q} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n \text{Var}_g(Y_i) x_{ip-q} x_{ip-1} & \cdots & \sum_{i=1}^n \text{Var}_g(Y_i) x_{ip-q}^2 \end{bmatrix}
\end{aligned}$$

è una matrice di dimensione $q \times q$, con $j, k = 1, \dots, q$;

$$\begin{aligned}
\nu_{\psi\lambda} &= \left[\sum_{i=1}^n \text{Var}_g(Y_i) x_{ip-j} x_{it} \right] \\
&= \begin{bmatrix} \sum_{i=1}^n \text{Var}_g(Y_i) x_{ip-1} x_{i0} & \cdots & \sum_{i=1}^n \text{Var}_g(Y_i) x_{ip-1} x_{ip-q-1} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n \text{Var}_g(Y_i) x_{ip-q} x_{i0} & \cdots & \sum_{i=1}^n \text{Var}_g(Y_i) x_{ip-q} x_{ip-q-1} \end{bmatrix}
\end{aligned}$$

è una matrice di dimensione $q \times (p-q)$, con $j = 1, \dots, q$ e $t = 0, \dots, p-q-1$ e

$$\begin{aligned}
\nu_{\lambda\lambda} &= \left[\sum_{i=1}^n \text{Var}_g(Y_i) x_{is} x_{it} \right] \\
&= \begin{bmatrix} \sum_{i=1}^n \text{Var}_g(Y_i) x_{i0}^2 & \cdots & \sum_{i=1}^n \text{Var}_g(Y_i) x_{i0} x_{ip-q-1} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n \text{Var}_g(Y_i) x_{i0} x_{ip-q-1} & \cdots & \sum_{i=1}^n \text{Var}_g(Y_i) x_{ip-q-1}^2 \end{bmatrix},
\end{aligned}$$

è una matrice di dimensione $(p-q) \times (p-q)$, con $s, t = 0, \dots, p-q-1$.

L'elemento in posizione (u, v) di B_l è

$$\begin{aligned}
 B_{(uv)} &= \nu_{\psi_u \psi_v} - I_{\psi_u \lambda} I_{\lambda \lambda}^{-1} \nu_{\lambda \psi_v} - \nu_{\lambda_u \lambda} I_{l \lambda \lambda}^{-1} I_{\lambda \psi_v} + I_{\psi_u \lambda} I_{\lambda \lambda}^{-1} \nu_{\lambda \lambda} I_{\lambda \lambda}^{-1} I_{\lambda \psi_v} \\
 &= \lim_{n \rightarrow \infty} \frac{1}{n} \left\{ \sum_{i=1}^n \text{Var}_g(Y_i) x_{ip-u} x_{ip-v} - \frac{1}{|\Delta|} \sum_{i=1}^n \text{Var}_g(Y_i) \left(x_{ip-v} \sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(u)}| \right) \right. \\
 &\quad \left. - \frac{1}{|\Delta|} \sum_{i=1}^n \text{Var}_g(Y_i) \left(x_{ip-u} \sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(v)}| \right) \right. \\
 &\quad \left. + \frac{1}{|\Delta|^2} \sum_{i=1}^n \text{Var}_g(Y_i) \left(\sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(u)}| \right) \left(\sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(v)}| \right) \right\} \\
 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \text{Var}_g(Y_i) \left(x_{ip-u} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(u)}|}{|\Delta|} x_{ij-1} \right) \left(x_{ip-v} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(v)}|}{|\Delta|} x_{ij-1} \right).
 \end{aligned}$$

Regressione logistica

Siano le osservazioni y_1, \dots, y_n realizzazioni delle variabili casuali indipendenti Y_1, \dots, Y_n . Nel modello di regressione logistica, si assume che $m_i Y_i \sim Bi(m_i, \pi_i)$ sia dipendente dal vettore delle covariate $\mathbf{x}_i = (x_{i0}, x_{i1}, \dots, x_{ip-1})'$, con $x_{i0} = 1$, secondo la funzione di legame *logit*. Ciò implica che il predittore lineare si può esprimere come

$$\eta_i = \mathbf{x}_i \beta = \beta_0 + \dots + \beta_{p-2} x_{ip-2} + \beta_{p-1} x_{ip-1} = \text{logit}[E(Y_i)] = \log\left(\frac{\pi_i}{1 - \pi_i}\right) \quad (\text{A.5})$$

da cui segue che

$$\pi_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}}.$$

La funzione di log-verosimiglianza basata su tale modello di regressione è

$$\begin{aligned}
 l(\pi_i) &= \sum_{i=1}^n s_i \log \pi_i + (m_i - s_i) \log(1 - \pi_i) \\
 &= \sum_{i=1}^n s_i \log\left(\frac{\pi_i}{1 - \pi_i}\right) + m_i \log(1 - \pi_i) \\
 &= \sum_{i=1}^n s_i \eta_i - m_i \log(1 + e^{\eta_i}).
 \end{aligned} \quad (\text{A.6})$$

Siano, senza perdita di generalità,

$$\psi = (\beta_{p-1}, \beta_{p-2}, \dots, \beta_{p-q+1}, \beta_{p-q})^T = (\psi_1, \psi_2, \dots, \psi_{q-1}, \psi_q)^T$$

il vettore dei parametri di interesse di lunghezza q e

$$\lambda = (\theta_0, \theta_1, \dots, \theta_{p-q+1})^T$$

il vettore $(p - q)$ -dimensionale dei parametri di disturbo. Siano μ_{i0} e μ'_{i0} i veri valori di μ_i e μ'_i , rispettivamente, con μ'_i derivata prima della media rispetto a η_i .

Detta $Var_g(Y_i)$ la vera varianza delle Y_i , $i = 1, \dots, n$, e $Z = (\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_{p-1})$ la matrice di disegno del modello dimensione $n \times p$, tale che $Z' = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{p-1})$, le quantità A e B sono delle matrici di dimensione $q \times q$ i cui rispettivi elementi di posizione (u, v) , $u, v = 1, \dots, q$ sono

$$A_{(uv)} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n m_i \mu'_{i0} \left(x_{ip-u} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(u)}|}{|\Delta|} x_{ij-1} \right) \left(x_{ip-v} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(v)}|}{|\Delta|} x_{ij-1} \right) \quad (\text{A.7})$$

e

$$B_{(uv)} = \lim_{n \rightarrow \infty} \sum_{i=1}^n m_i^2 Var_g(Y_i) \left(x_{ip-u} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(u)}|}{|\Delta|} x_{ij-1} \right) \left(x_{ip-v} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(v)}|}{|\Delta|} x_{ij-1} \right) / n, \quad (\text{A.8})$$

dove $|\Delta|$ è il determinante della matrice $\Delta = W^T V^{-1} W$, con la matrice W definita da $(\mathbf{z}_0, \dots, \mathbf{z}_{j-2}, \mathbf{z}_{j-1}, \mathbf{z}_j, \dots, \mathbf{z}_{p-q-1})$ e $V = diag(1/m_i \mu'_{i0})$, matrice diagonale di ordine n ; compaiono poi i determinanti $\Delta_{j(u)}$ e $\Delta_{j(v)}$ delle matrici, rispettivamente, $W^T V^{-1} W_{j(u)}$ e $W^T V^{-1} W_{j(v)}$, dove $W_{j(u)} = (\mathbf{z}_0, \dots, \mathbf{z}_{j-2}, \mathbf{z}_{j-v}, \mathbf{z}_j, \dots, \mathbf{z}_{p-q-1})$ e $W_{j(v)} = (\mathbf{z}_0, \dots, \mathbf{z}_{j-2}, \mathbf{z}_{j-u}, \mathbf{z}_j, \dots, \mathbf{z}_{p-q-1})$ sono ottenute a partire dalla matrice W , sostituendo la j -esima colonna x_{j-1} con x_{j-u} e x_{j-v} , rispettivamente. \hat{A} e \hat{B} , stime consistenti di A e B , si ottengono sostituendo le quantità ignote con loro stime consistenti.

Per calcolare (A.7) e (A.8), in riferimento alla funzione di log-verosimiglianza per il modello di regressione logistica (A.6), per $j = 1, \dots, q$ e $t = 0, \dots, p - q - 1$, se ne individuano i vettori delle derivate prime parziali, assumendo $x_{i0} = 1, \forall i$,

$$\begin{aligned} l_\psi &= \left[\sum_{i=1}^n s_i x_{ip-j} - m_i \mu_{i0} x_{ip-j} \right] \\ &= \left[\sum_{i=1}^n x_{ip-1} (s_i - m_i \mu_{i0}) \quad \dots \quad \sum_{i=1}^n x_{ip-q} (s_i - m_i \mu_{i0}) \right] \end{aligned}$$

e

$$\begin{aligned} l_\lambda &= \left[\sum_{i=1}^n s_i x_{it} - m_i \mu_{i0} x_{it} \right] \\ &= \left[\sum_{i=1}^n (s_i - m_i \mu_{i0}) \quad \dots \quad \sum_{i=1}^n x_{ip-q-1} (s_i - m_i \mu_{i0}) \right] \end{aligned}$$

ripettivamente, q e $(p - q)$ dimensionali. Posto $\mu'_{i0} = \partial\mu_{i0}/\partial\eta_i$, le matrici delle derivate seconde della funzione di log-verosimiglianza sono

$$l_{\psi\psi} = \left[\sum_{i=1}^n -m_i \mu'_{i0} x_{ip-j} x_{ip-k} \right]$$

$$= - \begin{bmatrix} \sum_{i=1}^n m_i \mu'_{i0} x_{ip-1}^2 & \cdots & \sum_{i=1}^n m_i \mu'_{i0} x_{ip-1} x_{ip-q} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n m_i \mu'_{i0} x_{ip-q} x_{ip-1} & \cdots & \sum_{i=1}^n m_i \mu'_{i0} x_{ip-q}^2 \end{bmatrix},$$

di dimensione $q \times q$, con $j, k = 1, \dots, w$;

$$l_{\psi\phi} = \left[\sum_{i=1}^n -m_i \mu'_{i0} x_{ip-j} x_{it} \right]$$

$$= - \begin{bmatrix} \sum_{i=1}^n m_i \mu'_{i0} x_{ip-1} x_{i0} & \cdots & \sum_{i=1}^n m_i \mu'_{i0} x_{ip-1} x_{ip-q-1} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n m_i \mu'_{i0} x_{ip-q} x_{i0} & \cdots & \sum_{i=1}^n m_i \mu'_{i0} x_{ip-q} x_{ip-q-1} \end{bmatrix},$$

di dimensione $q \times (p - q)$, con $j = 1, \dots, q$ e $t = 0, \dots, p - q - 1$;

$$l_{\lambda\lambda} = \left[\sum_{i=1}^n -m_i \mu'_{i0} x_{is} x_{it} \right]$$

$$= - \begin{bmatrix} \sum_{i=1}^n m_i \mu'_{i0} x_{i0}^2 & \cdots & \sum_{i=1}^n m_i \mu'_{i0} x_{i0} x_{ip-q-1} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n m_i \mu'_{i0} x_{i0} x_{ip-q-1} & \cdots & \sum_{i=1}^n m_i \mu'_{i0} x_{ip-q-1}^2 \end{bmatrix},$$

di dimensione $(p - q) \times (p - q)$, con $s, t = 0, \dots, p - q - 1$ e

$$l_{\lambda\psi} = \left[\sum_{i=1}^n -m_i \mu'_{i0} x_{it} x_{ip-j} \right]$$

$$= - \begin{bmatrix} \sum_{i=1}^n m_i \mu'_{i0} x_{i0} x_{ip-1} & \cdots & \sum_{i=1}^n m_i \mu'_{i0} x_{i0} x_{ip-q} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n m_i \mu'_{i0} x_{ip-q-1} x_{ip-1} & \cdots & \sum_{i=1}^n m_i \mu'_{i0} x_{ip-q-1} x_{ip-q} \end{bmatrix},$$

di dimensione $(p - q) \times q$, con $j = 1, \dots, q$ e $t = 0, \dots, p - q - 1$. Posto $I_{\psi\psi} = \lim_{n \rightarrow \infty} -E(l_{\psi\psi}/n)$, $I_{\psi\lambda} = \lim_{n \rightarrow \infty} -E(l_{\psi\lambda}/n)$, $I_{\lambda\psi} = \lim_{n \rightarrow \infty} -E(l_{\lambda\psi}/n)$ e $I_{\lambda\lambda} = \lim_{n \rightarrow \infty} -E(l_{\lambda\lambda}/n)$, detta Δ la matrice $l_{\lambda\lambda}$, si ha che

$$I_{\lambda\lambda}^{-1} = \frac{n}{|\Delta|} \begin{bmatrix} R_{1,1} & \cdots & R_{p-q,1} \\ \vdots & \ddots & \vdots \\ R_{1,p-q} & \cdots & R_{p-q,p-q} \end{bmatrix},$$

dove $R_{l,m} = (-1)^{l+m} |M_{(lm)}(\Delta)|$ e il (l, m) -esimo cofattore di Δ .

L'elemento in posizione (u, v) della matrice di dimensione $q \times q$, A , è

$$\begin{aligned}
A_{(uv)} &= I_{\psi_u \psi_v} - I_{\psi_u \lambda} I_{\lambda \lambda}^{-1} I_{\lambda \psi_v} - I_{\psi_v \lambda} I_{\lambda \lambda}^{-1} I_{\lambda \psi_u} + I_{\psi_v \lambda} I_{\lambda \lambda}^{-1} I_{\lambda \lambda} I_{\lambda \lambda}^{-1} I_{\lambda \psi_u} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \left\{ \sum_{i=1}^n m_i \mu'_{i0} x_{ip-u} x_{ip-v} - \frac{1}{|\Delta|} \sum_{i=1}^n m_i \mu'_{i0} \left(x_{ip-v} \sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(u)}| \right) \right. \\
&\quad \left. - \frac{1}{|\Delta|} \sum_{i=1}^n m_i \mu'_{i0} \left(x_{ip-u} \sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(v)}| \right) \right. \\
&\quad \left. + \frac{1}{|\Delta|^2} \sum_{i=1}^n m_i \mu'_{i0} \left(\sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(v)}| \right) \left(\sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(u)}| \right) \right\} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n m_i \mu'_{i0} \left(x_{ip-u} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(u)}|}{|\Delta|} x_{ij-1} \right) \left(x_{ip-v} - \sum_{j=1}^{p-q} x_{ij-1} \frac{|\Delta_{j(v)}|}{|\Delta|} x_{ij-1} \right),
\end{aligned}$$

dove $|\Delta_{j(u)}| = \sum_{m=1}^{p-w} (R_{j,m} \sum_{i=1}^n m_i \mu'_{i0} x_{im-1} x_{ip-u})$.

Per ottenere la matrice di dimensione $q \times q$, B , si individuano

$$\begin{aligned}
\nu_{\psi\psi} &= \left[\sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{ip-j} x_{ip-k} \right] \\
&= \begin{bmatrix} \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{ip-1}^2 & \cdots & \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{ip-1} x_{ip-q} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{ip-q} x_{ip-1} & \cdots & \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{ip-q}^2 \end{bmatrix}
\end{aligned}$$

di dimensione $q \times q$, con $j, k = 1, \dots, q$;

$$\begin{aligned}
\nu_{\psi\lambda} &= \left[\sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{ip-j} x_{it} \right] \\
&= \begin{bmatrix} \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{ip-1} x_{i0} & \cdots & \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{ip-1} x_{ip-q-1} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{ip-q} x_{i0} & \cdots & \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{ip-q} x_{ip-q-1} \end{bmatrix}
\end{aligned}$$

di dimensione $q \times (p - q)$, con $j = 1, \dots, q$ e $t = 0, \dots, p - q - 1$ e

$$\begin{aligned}
\nu_{\lambda\lambda} &= \left[\sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{is} x_{it} \right] \\
&= \begin{bmatrix} \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{i0}^2 & \cdots & \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{i0} x_{ip-q-1} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{i0} x_{ip-q-1} & \cdots & \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{ip-q-1}^2 \end{bmatrix},
\end{aligned}$$

di dimensione $(p - q) \times (p - q)$, con $s, t = 0, \dots, p - q - 1$.

L'elemento in posizione (u, v) di B è

$$\begin{aligned}
 B_{(uv)} &= \nu_{\psi_u \psi_v} - I_{\psi_u \lambda} I_{\lambda \lambda}^{-1} \nu_{\lambda \psi_v} - \nu_{\psi_u \lambda} I_{\lambda \lambda}^{-1} I_{\lambda \psi_v} + I_{\psi_u \lambda} I_{\lambda \lambda}^{-1} \nu_{\lambda \lambda} I_{\lambda \lambda}^{-1} I_{\lambda \psi_v} \\
 &= \lim_{n \rightarrow \infty} \frac{1}{n} \left\{ \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) x_{ip-u} x_{ip-v} \right. \\
 &\quad - \frac{1}{|\Delta|} \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) \left(x_{ip-v} \sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(u)}| \right) \\
 &\quad - \frac{1}{|\Delta|} \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) \left(x_{ip-u} \sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(v)}| \right) \\
 &\quad \left. + \frac{1}{|\Delta|^2} \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) \left(\sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(u)}| \right) \left(\sum_{j=1}^{p-q} x_{ij-1} |\Delta_{j(v)}| \right) \right\} \\
 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n m_i^2 \text{Var}_g(Y_i) \left(x_{ip-u} - \sum_{j=1}^{p-q} \frac{|\Delta_{j(u)}|}{|\Delta|} x_{ij-1} \right) \\
 &\quad \left(x_{ip-v} - \sum_{j=1}^{p-q} x_{ij-1} \frac{|\Delta_{j(v)}|}{|\Delta|} x_{ij-1} \right).
 \end{aligned}$$

Bibliografia

- BLUME, J. D., MCGARVEY, S. T., OLVEDA, R. M. & SU, L. (2007). Statistical evidence for GLM regression parameters: A robust likelihood approach. *Statistics in Medicine* **26**, 2919–2936.
- NELDER, J. A. & WEDDERBURN, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A* **135**, 370–384.
- PACE, L. & SALVAN, A. (2001). *Introduzione alla Statistica II Inferenza, Verosimiglianza, Modelli*. Cedam, Padova.
- ROYALL, R. (1986). Model robust confidence intervals using maximum likelihood estimators. *International Statistical Institute* **54**, 221–226.
- ROYALL, R. (2000). On the probability of observing misleading statistical evidence. *Journal of the American Statistical Association* **95**, 760–780.
- ROYALL, R. & TSOU, T. S. (2003). Interpreting statistical evidence by using imperfect models: robust adjusted likelihood functions. *Journal of the Royal Statistical Society B* **65**, 391–404.
- SALVAN, A., SARTORI, N. & PACE, L. (2020). *Modelli Lineari Generalizzati*. Springer, Milano.
- STAFFORD, J. E. (1996). A robust adjustment of the profile likelihood. *The Annals of Statistics* **24**, 336–352.
- TSOU, T. S. (2006). Robust Poisson regression. *Journal of Statistical Planning and Inference* **136**, 3173–3186.
- TSOU, T. S. & CHIEN, L. (2014). Parametric simultaneous robust inferences for regression coefficient under generalized linear models. *Journal of Statistical Computation and Simulation* **84**, 850–867.

