Università degli studi di Padova
Dipartimento di Scienze Statistiche

Corso di Laurea Magistrale in
Scienze Statistiche

Riduzione della distorsione in mediana in modelli di regressione per risposte ordinali

Relatore: Prof. Alessandra Salvan
Dipartimento di Scienze Statistiche

Correlatore: Prof. Euloge Clovis Kenne Pagui
Dipartimento di Scienze Statistiche

Laureando Vincenzo Gioia
Matricola N 1149570

Anno Accademico 2018/2019

Indice

Introduzione

La funzione di verosimiglianza ha un ruolo centrale nell'inferenza statistica. L'approccio di stima di massima verosimiglianza è l'approccio prediletto per la trattazione dei modelli di regressione per variabili risposta continue, categoriali su scala ordinale e su scala sconnessa o nominale. Tuttavia, in presenza di informazione campionaria limitata, dati sparsi o modelli complessi le proprietà asintotiche dello stimatore di massima verosimiglianza possono riflettere scarsamente l'esatta distribuzione campionaria, causando notevoli problemi nell'inferenza. Un ulteriore problema connesso con la stima di massima verosimiglianza si ha principalmente, ma non solo, nei modelli di regressione per variabili risposta discrete o categoriali dove vi è una probabilità positiva che le stime siano sulla frontiera dello spazio parametrico, o, detto in altri termini, il problema delle stime infinite.

La letteratura è ricca di metodi relativi alla riduzione della distorsione dello stimatore di massima verosimiglianza. Ne sono un esempio gli approcci basati sulla correzione asintotica dello distorsione e le sue varianti asintoticamente equivalenti, quali il *bootstrap* e il *jackknife*, che, tuttavia, dipendono dalla finitezza delle stime di massima verosimiglianza. L'approccio proposto da Firth (1993) è basato sulla correzione per la funzione di punteggio, piuttosto che per la stima di massima verosimiglianza stessa. Lo stimatore, soluzione dell'equazione di stima associata alla funzione di punteggio modificata, riduce la distorsione in media delle stime di massima verosimiglianza, senza dipendere dalla loro finitezza. Tale metodologia ha riscosso molta popolarità nei contesti applicativi e si è rivelata efficace sia nel ridurre la distorsione e sia nel prevenire la non finitezza delle stime in situazioni dove le stime di massima verosimiglianza risultano sulla frontiera dello spazio parametrico. La correzione del meccanismo di stima, secondo l'approccio di Firth (1993), utilizza come proprietà di centratura la non distorsione in media, la quale, essendo legata ad una specifica parametrizzazione, determina, per lo stimatore che riduce la distorsione in media, la mancanza della proprietà di equivarianza rispetto a trasformazioni biunivoche e regolari del parametro.

Un nuovo approccio di stima, sempre basato sulla correzione della funzione di punteggio, è stato proposto da Kenne Pagui *et al.* (2017). La differenza rispetto all'approccio di stima proposto da Firth (1993) risiede nell'adozione della non distorsione in mediana come indice di centratura. Lo stimatore risultante si dimostra efficace nel ridurre la distorsione in mediana per ciascuna componente del parametro e nel prevenire le stime sulla frontiera dello spazio parametrico. Il ricorso alla non distorsione in mediana, come proprietà di centratura, garantisce allo stimatore la proprietà di equivarianza per riparametrizzazioni monotone che trasformano ogni componente del vettore di parametri separatamente.

L'obiettivo del lavoro di tesi consiste nello studio dello stimatore risultante dall'approccio di stima proposto da Kenne Pagui *et al.* (2017) nei modelli di regressione per risposte ordinali che mettono in relazione le probabilità cumulate con i valori delle esplicative, ovvero nella formulazione base dei modelli link cumulati. Si ottengono le quantità necessarie dell'approccio di stima proposto da Kenne Pagui *et al.* (2017) per tale tipologia di modello, non ancora presente in letteratura, con conseguente traduzione su supporto informatico. Gli studi di simulazione effettuati hanno l'obiettivo di valutare le proprietà dello stimatore con distorsione ridotta in mediana in confronto ai suoi principali competitori, ovvero lo stimatore di massima verosiglianza e quello con distorsione ridotta in media.

Il Capitolo 1, introduttivo, riporta i principali concetti di base e definizioni dell'inferenza di verosimiglianza. Vengono esposti gli approcci per la riduzione della distorsione delle stime di massima verosimiglianza attraverso una modifica delle equazioni di stima basate sulla funzione di punteggio e in particolare l'approccio proposto da Firth (1993), risultante in stimatori con distorsione ridotta in media, e l'approccio proposto da Kenne Pagui *et al.* (2017), risultante in stimatori con distorsione ridotta in mediana componente per componente.

Nel Capitolo 2 viene presentato il modello di regressione per variabili risposta su scala ordinale che mette in relazione le probabilità cumulate con il predittore lineare (modello link cumulati), con particolare enfasi sulla versione con quote (*odds*) proporzionali. Verranno discusse le problematiche che possono insorgere nella stima dei parametri con il metodo della massima verosimiglianza e viene presentato l'approccio proposto da Firth (1993), formalizzato da Kosmidis (2014b), relativo al modello link cumulati.

Il nucleo del presente lavoro è esposto nel Capitolo 3 dove si calcolano le quantità necessarie all'implementazione dello stimatore con distorsione ridotta in mediana per i modelli link cumulati. Verranno, inoltre, affrontati gli aspetti computazionali connessi con l'implementazione di tale metodo.

Il Capitolo 4 è dedicato a un esempio di applicazione del metodo di stima risultante nello stimatore con distorsione ridotta in mediana a un insieme di dati reali, di numerosità campionaria limitata. È stato rivisitato uno studio, condotto con il metodo di stima di massima verosimiglianza, relativo all'effetto avverso del tamoxifene, farmaco utilizzato come adiuvante della terapia chirurgica in pazienti donne affette da cancro della mammella, sulla proliferazione del cancro all'endometrio.

Il Capitolo 5 riporta i risultati ottenuti tramite studi di simulazione. Si confronteranno, attraverso opportuni indici, i risultati ottenuti con il metodo di stima di massima verosimiglianza e con i metodi di stima basati sulle funzioni di punteggio modificate.

In Appendice si riporta il codice predisposto con il software di programmazione statistica R 3.5.0.

Capitolo 1

Teoria della verosimiglianza e metodi per la riduzione della distorsione

L'obiettivo di questo capitolo è definire il concetto di centratura in mediana delle stime di massima verosimiglianza secondo l'approccio proposto da Kenne Pagui *et al.* (2017). Attraverso una correzione della funzione di punteggio si ottiene uno stimatore non distorto in mediana componente per componente al terzo ordine ed equivariante rispetto a riparametrizzazioni che trasformano ogni componente del vettore di parametri separatamente.

Nel §1.1 vengono richiamati alcuni concetti fondamentali della teoria dell'inferenza statistica e in particolare di quella basata sulla verosimiglianza. Il §1.2 è dedicato ai metodi di stima basati sulla funzione di punteggio modificata. Nel §1.2.1. si esporranno brevemente i metodi di riduzione della distorsione in media delle stime di massima verosimiglianza presenti in letteratura e verrà discusso in dettaglio l'approccio proposto da Firth (1993). Il §1.2.2 è interamente dedicato alla metodologia proposta da Kenne Pagui *et al.* (2017).

I principali riferimenti bibliografici, relativi al §1.1, sono Pace e Salvan (1997, 2001). Il §1.2 ha come riferimenti bibliografici Firth (1993) e Kosmidis (2014a), per il metodo della riduzione della distorsione in media, e Kenne Pagui *et al.* (2017), per il metodo della riduzione della distorsione in mediana.

1.1 Inferenza di verosimiglianza

L'inferenza statistica è il procedimento mediante il quale si ottengono informazioni sulle caratteristiche di una popolazione attraverso l'osservazione di un campione, ovvero un aggregato di unità statistiche selezionato solitamente mediante un esperimento casuale. La teoria dell'inferenza statistica ha tra i suoi problemi quello di dedurre modelli appropriati alla luce di un insieme di osservazioni e lavora nella direzione opposta al Calcolo delle Probabilità che cerca di dedurre configurazioni osservabili in sequenze di eventi casuali. L'introduzione della funzione di verosimiglianza ha profondamente influenzato lo sviluppo di teorie e metodi statistici e ha permesso agevolmente di rovesciare il punto di vista del Calcolo delle Probabilità attraverso la specificazione del modello probabilistico in funzione di un insieme di parametri non noti. Essa ha un ruolo centrale nell'inferenza statistica parametrica e necessita per la sua definizione della nozione di modello statistico parametrico.

Sia $Y = (Y_1, \dots, Y_n)$ un vettore casuale di cui si osserva $y = (y_1, \dots, y_n)$. Si definisce modello statistico parametrico, $\mathcal{F}$, una famiglia di funzioni di densità di probabilità rispetto a una misura dominante (misura di Lebesgue nel caso assolutamente continuo e misura contatore in quello discreto) indicizzate tramite un vettore p-dimensionale di parametri θ,

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

Il modello statistico parametrico $\mathcal{F}$ è specificato attraverso la terna $(\mathcal{Y}, p_Y(y;\theta), \Theta)$, dove $\mathcal{Y}$ è lo spazio campionario, Θ è lo spazio parametrico e $p_Y(y;\theta)$ è la funzione di densità di probabilità o di massa di probabilità a seconda che ci si riferisca rispettivamente alle variabili casuali continue o a quelle discrete.

1.1.1 Stime, stimatori e proprietà collegate

Nella teoria della stima puntuale si vuole individuare il valore di θ che meglio di altri spieghi i dati osservati e pertanto si vuole individare una applicazione $\hat{\theta} : \mathcal{Y} \to \Theta$ che faccia corrispondere ad ogni elemento $y \in \mathcal{Y}$ un elemento in Θ, $\hat{\theta} = \hat{\theta}(y)$, detto stima puntuale di θ. La variabile casuale corrispondente

$\hat{\theta} = \hat{\theta}(Y)$ è detta stimatore. Uno stimatore è il risultato di un metodo di stima e quindi legato a una equazione di stima. Si è soliti caratterizzarlo per alcune proprietà desiderabili quali la non distorsione o la consistenza e per le proprietà distributive, siano esse esatte o asintotiche.

Uno stimatore $\hat{\theta}$ di θ è detto non distorto in media e il corrispondente valore campionario di $\hat{\theta}$ è una stima non distorta in media se vale

$$E_\theta[\hat{\theta}] = \theta, \qquad \forall\theta \in \Theta, \tag{1.1}$$

purchè il valor medio esista. Se tale relazione è soddisfatta al divergere della numerosità campionaria, $\lim_{n\to\infty} E_\theta[\hat{\theta}] = \theta, \forall\theta \in \Theta$, lo stimatore, $\hat{\theta}$, si definisce asintoticamente non distorto. Si definisce distorsione la quantità $b(\theta) = E_\theta[\hat{\theta}] - \theta$.

Se la distribuzione dello stimatore $\hat{\theta}$ di θ è continua e θ è scalare, lo stimatore $\hat{\theta}$ si definisce non distorto in mediana (Read, 1985) se ha la stessa probabilità di sottostimare o sovrastimare il vero valore del parametro, ovvero

$$P_\theta(\hat{\theta} \leq \theta) = 0.5, \qquad \theta \in \Theta. \tag{1.2}$$

Nel caso di un vettore di parametri, uno stimatore si definisce non distorto in mediana componente per componente se vale

$$P_\theta(\hat{\theta}_r < \theta_r) = 0.5, \qquad \theta_r \in \Theta, \tag{1.3}$$

dove θ_r, $r = 1, \ldots, p$, indica una generica componente di $\theta \in \Theta$. Se le relazioni (1.2) e (1.3) sono soddisfate al divergere della numerosità campionaria la non distorsione in mediana dello stimatore sarà in senso asintotico.

Si definisce errore quadratico medio (*Mean Squared Error*) di uno stimatore $\hat{\theta}$ la quantità $MSE_\theta(\hat{\theta}) = E_\theta[(\hat{\theta} - \theta)^2] = V_\theta[\hat{\theta}] + b(\theta)^2$, con $V_\theta[\hat{\theta}]$ varianza dello stimatore. Dati due stimatori del parametro θ, $\hat{\theta}_1$ e $\hat{\theta}_2$, se $\forall\theta$ vale $MSE_\theta(\hat{\theta}_1) < MSE_\theta(\hat{\theta}_2)$ allora lo stimatore $\hat{\theta}_1$ è più efficiente dello stimatore $\hat{\theta}_2$. Se al divergere della numerosità campionaria l'errore quadratico medio è pari a zero lo stimatore si dice consistente in media quadratica e se tale condizione è soddisfatta lo stimatore è consistente per θ, ovvero, $\forall \epsilon > 0$, vale

$$\lim_{n\to\infty} P_\theta(|\hat{\theta} - \theta| \geq \epsilon) = 0, \qquad \theta \in \Theta. \tag{1.4}$$

Le proprietà distributive dello stimatore $\hat{\theta}$ sotto θ, siano esse esatte o asintotiche, sono in grado di fornire informazioni sull'incertezza connessa con il procedimento di stima. Una possibile proprietà distributiva asintotica di uno stimatore di notevole utilità è la convergenza ad una distribuzione normale, che si ha, nel caso di parametro scalare, se al divergere della numerosità campionaria vale

$$\frac{\hat{\theta} - E_\theta[\hat{\theta}]}{\sqrt{V_\theta[\hat{\theta}]}} \xrightarrow{d} Z \sim \mathcal{N}(0,1), \qquad \theta \in \Theta.$$

1.1.2 Funzione di verosimiglianza e stima di massima verosimiglianza

Sia $y = (y_1, \ldots, y_n)$ un campione casuale da $Y = (Y_1, \ldots, Y_n)$ e sia $\mathcal{F}$ un modello statistico parametrico. Si definisce funzione di verosimiglianza per θ basata sul campione osservato y la funzione $\mathcal{L} : \Theta \to \mathbb{R}^+$ tale che

$$\mathcal{L}(\theta) = \mathcal{L}(\theta; y) = c(y) p_Y(y; \theta).$$

Essa è in grado di rappresentare l'informazione sul parametro ignoto θ, portata dal campione osservato y, dato il modello statistico parametrico $\mathcal{F}$, ovvero combina l'informazione pre-sperimentale contenuta nel modello statistico con quella sperimentale contenuta in y. La costante $c(y)$ è una costante di proporzionalità arbitraria che non dipende da θ ma solo dai dati e può essere rimossa senza alterare le conclusioni inferenziali su θ. In presenza di osservazioni indipendenti la funzione di verosimiglianza per θ è data dalla produttoria delle densità marginali, $p_{Y_i}(y_i; \theta)$, e, a meno di una costante moltiplicativa, risulta

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

Nel contesto dell'inferenza basata sulla funzione di verosimiglianza si è soliti lavorare con la trasformata logaritmica della funzione di verosimiglianza, detta log-verosimiglianza per θ, definita come

$$\ell(\theta) = \ell(\theta; y) = \log \mathcal{L}(\theta; y) = \log c(y) + \log p_Y(y; \theta),$$

con la convenzione che $\ell(\theta) = -\infty$ se $\mathcal{L}(\theta) = 0$. In presenza di osservazioni indipendenti risulta che la log-verosimiglianza per θ è, a meno di una costante additiva, pari a

$$\ell(\theta) = \sum_{i=1}^{n} \log p_{Y_i}(y_i; \theta).$$

Data una funzione di verosimiglianza $\mathcal{L}(\theta; y)$, con $\theta \in \Theta$, si definisce stima di massima verosimiglianza di θ un valore $\hat{\theta} \in \Theta$ che rende massima la funzione di verosimiglianza, ossia un valore $\hat{\theta}$ tale che $\mathcal{L}(\hat{\theta}; y) \geq \mathcal{L}(\theta; y)$, $\forall \theta \in \Theta$, esprimibile altrimenti come

$$\hat{\theta} = \arg \sup_{\theta \in \Theta} \mathcal{L}(\theta; y) = \arg \sup_{\theta \in \Theta} \ell(\theta; y).$$

Non è detto che tale stima esista o che sia unica. Se tuttavia $\hat{\theta} = \hat{\theta}(y)$ esiste unico con probabilità uno, o poichè molte proprietà dello stimatore sono asintotiche, tendente a uno al divergere della numerosità campionaria, la variabile casuale $\hat{\theta} = \hat{\theta}(Y)$ è detta stimatore di massima verosimiglianza. In modelli statistici parametrici che soddisfano tenui condizioni di regolarità (van der Vaart, 1998, §5.2) lo stimatore di massima verosimiglianza soddisfa la (1.4), ossia è consistente.

1.1.3 Quantità di verosimiglianza e momenti nulli

Con il termine quantità di verosimiglianza ci si riferisce alle quantità definite a partire dalla verosimiglianza, quali ad esempio le derivate parziali della funzione di log-verosimiglianza. Si definisce funzione di punteggio o funzione score il vettore delle derivate parziali prime della funzione di log-verosimiglianza

$$U(\theta) = [U_r(\theta)] = \left[\frac{\partial \ell(\theta; y)}{\partial \theta_r}\right], \quad r = 1, \ldots, p, \tag{1.5}$$

dove $[a_r]$ indica un vettore con r-esimo elemento a_r. La funzione di punteggio gioca un ruolo importante nel calcolo delle stime di massima verosimiglianza, $\hat{\theta}$, che in modelli regolari sono soluzione di

$$U(\theta) = 0, \tag{1.6}$$

detta equazione di verosimiglianza. La soluzione della (1.6) potrebbe non essere esprimibile in forma chiusa e, in tal caso, per ottenere le stime di massima verosimiglianza si necessita di algoritmi di calcolo numerici quali Newton-Raphson o Fisher-scoring (si veda ad esempio Wood, 2015, §5.1.1). La funzione di punteggio riveste una notevole importanza anche per gli approcci proposti da Firth (1993) e Kenne Pagui *et al.* (2017) in quanto una sua opportuna correzione all'interno di un procedimento di stima determina gli stimatori con distorsione ridotta in media e in mediana, discussi nel successivo §1.2.

La matrice hessiana della funzione di log-verosimiglianza, cambiata di segno, definisce la matrice di informazione osservata

$$j(\theta) = [j_{rs}(\theta)] = \left[-\frac{\partial^2 \ell(\theta; y)}{\partial \theta_r \partial \theta_s}\right], \quad r, s = 1, \ldots, p, \tag{1.7}$$

dove $[a_{rs}]$ indica una matrice $p \times p$, con generico elemento a_{rs}. La quantità $j(\hat{\theta})$ è un indice della caduta della log-verosimiglianza via via che ci si allontana da θ e quindi più è grande e più i valori distanti da $\hat{\theta}$ perdono sostegno empirico, ragione per cui rappresenta una misura dell'informazione che i dati forniscono sull'ignoto parametro θ.

Il valore atteso della matrice di informazione osservata è detto matrice di informazione di Fisher, definita come

$$i(\theta) = E_\theta[j(\theta)] = E_\theta\left[-\frac{\partial^2 \ell(\theta; Y)}{\partial \theta_r \partial \theta_s}\right], \quad r, s = 1, \ldots, p. \tag{1.8}$$

La quantità $i(\theta)$ risulta di notevole importanza per esprimere la distribuzione asintotica dello stimatore di massima verosimiglianza. Risultati di teoria asintotica del primo ordine (Pace e Salvan, 1997, §3.4.1) mostrano che, assumendo $i(\theta) = O(n)$, la distribuzione nulla della funzione di punteggio è, per il teorema del limite centrale, approssimabile con una distribuzione normale, ovvero $U(\theta) \dot{\sim} \mathcal{N}_p(0, i(\theta))$. Sotto la normalità asintotica della funzione di punteggio vale che

$$\hat{\theta} \dot{\sim} \mathcal{N}_p(\theta, i^{-1}(\theta)) \tag{1.9}$$

Lo stimatore di massima verosimiglianza è asintoticamente non distorto e la distorsione asintotica è di ordine $O(n^{-1})$. Lo stimatore di massima verosimiglianza è asintoticamente non distorto in mediana e si dimostra (Pfanzagl, 1970) che vale $P_\theta(\hat{\theta}_r < \theta_r) = 1/2 + O(n^{-1/2})$, per una componente θ_r di θ.

La notazione indiciale, utile per rappresentare vettori, matrici e tensori, ci permette di esprimere la generica derivata parziale di ordine m della funzione di log-verosimiglianza per θ come

$$U_{r_1 \dots r_m} = \frac{\partial^m \ell(\theta)}{\partial \theta_{r_1} \dots \partial \theta_{r_m}}. \tag{1.10}$$

Quindi U_r, $r = 1, \dots, p$, indica il generico elemento del vettore score, $U(\theta)$, mentre U_{rs}, U_{rst} indicano le derivate parziali di ordine superiore. Secondo tale notazione $i_{rs}(\theta)$ rappresenta l'elemento di posizione (r, s) della matrice di informazione attesa e $i^{rs}(\theta)$ l'elemento di posizione (r, s) della sua inversa.

Si definiscono momenti nulli i valori attesi delle derivate di $\ell(\theta)$ e dei loro prodotti valutati sotto la distribuzione nulla

$$\begin{aligned}
\nu_a &= E_\theta[U_a(\theta)], \\
\nu_{a,b} &= E_\theta[U_a(\theta)U_b(\theta)], \\
\nu_{a,b,c} &= E_\theta[U_a(\theta)U_b(\theta)U_c(\theta)], \\
\nu_{ab} &= E_\theta[U_{ab}(\theta)], \\
\nu_{a,bc} &= E_\theta[U_a(\theta)U_{bc}(\theta)], \\
\nu_{abc} &= E_\theta[U_{abc}(\theta)].
\end{aligned} \tag{1.11}$$

I momenti nulli (1.11) soddisfano le relazioni

$$\begin{aligned} &\nu_a = 0, \\ &\nu_{ab} + \nu_{a,b} = 0, \\ &\nu_{abc} + [3]\nu_{a,bc} + \nu_{a,b,c} = 0, \end{aligned} \tag{1.12}$$

dove $[3]\nu_{a,bc} = \nu_{a,bc} + \nu_{b,ac} + \nu_{c,ab}$. Tali relazioni, note come relazioni di Bartlett, sono coinvolte nello sviluppo dei metodi di riduzione della distorsione, discussi nel successivo §1.2, e saranno ampiamente utilizzate nei calcoli teorici presentati nel successivo §3.1. La relazione $\nu_a = 0$, ossia $E_\theta[U_a(\theta)] = 0$, stabilisce che l'equazione di stima (1.6) è non distorta.

1.1.4 Verosimiglianza profilo e quantità collegate

Nell'ambito di un modello statistico parametrico non necessariamente gli aspetti di interesse dell'inferenza coincidono con tutto il parametro $\theta \in \mathbb{R}^p$. Spesso possiamo identificare gli aspetti di interesse mediante un parametro $\tau \in \mathbb{R}^k$, detto parametro di interesse, e gli elementi accessori di variabilità con il parametro $\zeta \in \mathbb{R}^{p-k}$, detto parametro di disturbo. Inoltre i parametri $\tau \in T$ e $\zeta \in Z$ si suppongono a variazione indipendente, ossia $\Theta = T \times Z$. Sia $\mathcal{L}(\theta)$ la funzione di verosimiglianza per θ partizionato in due blocchi di componenti, $\theta = (\tau, \zeta)$, con τ parametro di interesse, si definisce funzione di verosimiglianza profilo per τ

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

dove $\hat{\zeta}_\tau$ è la stima di massima verosimiglianza di ζ per τ fissato. Si indica, dunque, con $\ell_p(\tau) = \ell(\tau, \hat{\zeta}_\tau)$ la funzione di log-verosimiglianza profilo per τ.

A seguito della partizione di θ in due blocchi risulta che il vettore score si partiziona come $U(\theta) = (U_\tau(\theta), U_\zeta(\theta))$ e il vettore k-dimensionale $U_\tau(\tau, \hat{\zeta}_\tau) = U_P(\tau)$ è detto score profilo per τ.

Analogamente, la matrice di informazione di Fisher e la sua inversa si partizionano nel seguente modo

$$i(\theta) = i(\tau, \zeta) = \begin{pmatrix} i_{\tau\tau} & i_{\tau\zeta} \\ i_{\zeta\tau} & i_{\zeta\zeta} \end{pmatrix},$$

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

dove $i^{\tau\tau} = (i_{\tau\tau} - i_{\tau\zeta} i_{\zeta\zeta}^{-1} i_{\zeta\tau})^{-1}$ e $i^{\tau\zeta} = -i^{\tau\tau} i_{\zeta\tau} i_{\zeta\zeta}^{-1}$.

1.1.5 Riparametrizzazioni e proprietà di invarianza rispetto alla parametrizzazione

Nella specificazione di un modello statistico parametrico, $\mathcal{F}$, il parametro θ funge da indicatore degli elementi di $\mathcal{F}$. Pertanto si può sostituire θ con una sua funzione biunivoca, continua e regolare, ossia infinitamente derivabile insieme con la sua inversa, e ottenere una riparametrizzazione del modello statistico parametrico $\mathcal{F}$. Sia $\psi = \psi(\theta)$ una trasformazione del parametro θ, con $\psi(\cdot)$ funzione biunivoca e regolare da $\Theta \subseteq \mathbb{R}^p$ in $\Psi \subseteq \mathbb{R}^p$. Il modello statistico parametrico può essere riscritto come

$$\mathcal{F} = \{p_Y(y; \psi) : \psi = \psi(\theta), \theta \in \Theta, y \in \mathcal{Y}\} = \{p_Y(y; \psi), \ \ \psi \in \Psi, y \in \mathcal{Y}\},$$

con $\Psi = \{\psi : \psi = \psi(\theta), \theta \in \Theta\}$. Trattandosi di formulazioni equivalenti, la scelta dell'una o dell'altra parametrizzazione è spesso una questione di convenienza ed è auspicabile che le conclusioni inferenziali cui si perviene siano invarianti rispetto alla parametrizzazione adottata (*principio di invarianza rispetto alla parametrizzazione*). Se una data procedura inferenziale produce una stima puntuale, il principio di invarianza rispetto alla parametrizzazione adottata richiede che lo stimatore goda della *proprietà di equivarianza.*

La funzione di verosimiglianza e il suo logaritmo non dipendono dalla parametrizzazione adottata. Dal momento che θ e $\psi = \psi(\theta)$ individuano lo stesso elemento di $\mathcal{F}$ valgono le seguenti relazioni

$$\mathcal{L}^{\Psi}(\psi) = \mathcal{L}^{\Theta}(\theta(\psi)),$$
$$\ell^{\Psi}(\psi) = \ell^{\Theta}(\theta(\psi)),$$

dove $\theta(\psi)$ indica l'applicazione inversa di $\psi(\theta)$. Tali relazioni permettono di definire la funzione di verosimiglianza come una funzione intrinseca, ossia non dipendente dal sistema di coordinate espresso dalla parametrizzazione. Ciò implica che lo stimatore di massima verosimiglianza gode della proprietà di equivarianza rispetto alla parametrizzazione, ovvero con ψ parametrizzazione alternativa di $\mathcal{F}$ si ha che $\hat{\psi} = \psi(\hat{\theta})$ e $\hat{\theta} = \theta(\hat{\psi})$. Tuttavia, solo alcuni dei criteri di valutazione di uno stimatore sono coerenti con il principio di invarianza rispetto alla parametrizzazione. Ad esempio il criterio di non distorsione non è invariante rispetto alla parametrizzazione e lo stimatore ricavato secondo tale criterio non gode della proprietà di equivarianza. Considerando, invece, come proprietà di centratura di uno stimatore la non distorsione in mediana allora, con θ scalare, se $\tilde{\theta}$ è uno stimatore non distorto in mediana di θ, si ha che lo stimatore di ψ, $\tilde{\psi} = \psi(\tilde{\theta})$, è non distorto in mediana.

Le derivate della log-verosimiglianza (1.10) e i loro valori attesi (1.11) dipendono dalla parametrizzazione scelta per il modello statistico parametrico $\mathcal{F}$ e si trasformano per effetto di riparametrizzazioni. Sia $\psi = \psi(\theta)$ una parametrizzazione alternativa del modello e siano θ^r, con $r = 1, \dots, p$, le generiche componenti di θ e ψ^a, con $a = 1, \dots, p$, le generiche componenti di ψ. Nella nuova parametrizzazione ψ, la funzione di punteggio e l'informazione attesa risultano, in notazione matriciale, legate alla funzione di punteggio e all'informazione attesa nella parametrizzazione θ dalle seguenti relazioni

$$U^{\Psi}(\psi) = [\theta^r_a] U^{\Theta}(\theta(\psi)),$$
$$i^{\Psi}(\psi) = [\theta^r_a]^T i^{\Theta}(\theta(\psi)) [\theta^r_a],$$

dove $[\theta^r_a]$ è la matrice $p \times p$ avente $\theta^r_a = \frac{\partial \theta^r(\psi)}{\partial \psi^a}$ come elemento di posto (r, a). La notazione matriciale non risulta conveniente per esprimere in forma compatta le leggi di trasformazione di quantità legate a derivate di ordine superiore al secondo. Si fa ricorso alla notazione indiciale per ottenere le relazioni (1.18) e (9.91) in Pace e Salvan (1997), relative alla derivata terza.

1.2 Metodi di stima basati su funzioni di punteggio modificate

Nei problemi regolari di stima (Severini, 2000, §3.4), lo stimatore di massima verosimiglianza ha distribuzione asintotica centrata nel vero valore del parametro. Tuttavia, in presenza di piccole numerosità campionarie, dati sparsi o modelli complessi, le proprietà asintotiche dello stimatore di massima verosimiglianza possono riflettere scarsamente l'esatta distribuzione campionaria e la distorsione delle stime può comportare un deterioramento dei risultati inferenziali. Un ulteriore problema connesso con l'approccio di stima di massima verosimiglianza è legato alla non finitezza delle stime, ovvero alle situazioni in cui le stime di massima verosimiglianza hanno probabilità positiva di essere sulla frontiera dello spazio parametrico.

L'approccio di stima basato sulle funzioni di punteggio modificate permette, introducendo una distorsione nella funzione di punteggio, di ottenere gli stimatori con distorsione ridotta in media, proposti da Firth (1993), e in mediana, proposti da Kenne Pagui *et al.* (2017). Tali approcci non richiedono il calcolo della stima di massima verosimiglianza e, oltre a ridurre la distorsione dello stimatore, rispettivamente in media e in mediana, possono garantire la finitezza delle stime in situazioni in cui le stime di massima verosimiglianza sono sulla frontiera dello spazio parametrico. Riprendendo quanto discusso nel §1.1.5, lo stimatore con distorsione ridotta in media secondo l'approccio di Firth (1993) non gode della proprietà di equivarianza per ogni trasformazione biettiva, ma risulta esattamente equivariante rispetto a trasformazioni lineari di θ. Lo stimatore con distorsione ridotta in mediana secondo l'approccio di Kenne Pagui *et al.* (2017), invece, gode della proprietà di equivarianza rispetto a riparametrizzazioni che trasformano ogni componente del parametro separatamente, ma non è equivariante rispetto a trasformazioni lineari di θ, quali i contrasti dei parametri. La combinazione dei due approcci mediante un aggiustamento "misto" (*mixed adjustment*) della funzione score è discussa in un recente lavoro di Kosmidis *et al.* (2019) e si dimostra efficace nel combinare le proprietà di equivarianza di entrambi i metodi rispetto a diverse componenti di θ.

Sia dato un modello statistico parametrico $\mathcal{F}$ indicizzato da un vettore di parametri $\theta \in \mathbb{R}^p$ e sia $U(\theta)$ la funzione di punteggio. L'idea base dell'approccio di tali metodi di stima è quella di sostituire le equazioni di verosimiglianza (1.6) con le nuove equazioni di stima

$$U(\theta) + B(\theta) = 0, \tag{1.13}$$

dove $B(\theta)$ è un termine vettoriale di correzione che si differenzia a secondo del tipo di approccio. Nel caso degli stimatori con distorsione ridotta in media, indicheremo $B(\theta)$ con $A^*(\theta)$, dove $A^*(\theta)$ è costruito in modo tale che lo stimatore risultante, $\hat{\theta}^*$, sia approssimativamente non distorto in media con distorsione di ordine $O(n^{-2})$, mentre nel caso degli stimatori con distorsione ridotta in mediana indicheremo $B(\theta)$ con $\tilde{A}(\theta)$, dove $\tilde{A}(\theta)$ è costruito in modo che lo stimatore risultante, $\tilde{\theta}$, sia approssimativamente non distorto in mediana con un errore di $O(n^{-3/2})$.

1.2.1 Riduzione della distorsione in media

Antecedenti alla proposta di Firth (1993), in letteratura diversi metodi sono stati proposti per la riduzione della distorsione dello stimatore di massima verosimiglianza. Tra i metodi più popolari vi è l'approccio basato sulla correzione asintotica della distorsione e le sue varianti asintoticamente equivalenti, quali il *bootstrap* e il *jackknife*. Tali metodi vengono definiti espliciti in quanto sono basati su una procedura che prevede la stima della distorsione $b(\theta)$ e la sua sottrazione alla stima di massima verosimiglianza al fine di ottenere una stima con distorsione inferiore. Questo modo di procedere è detto 'correttivo' in quanto la stima di massima verosimiglianza viene calcolata e, quindi, corretta. La dipendenza di tali metodi dalla stima di massima verosimiglianza risulta in una impossibilità di applicazione quando la stima risulta sulla frontiera dello spazio parametrico. Una rivisitazione dettagliata dei metodi di riduzione della distorsione presenti in letteratura è in Kosmidis (2014a).

L'approccio proposto da Firth (1993) si discosta dai precedenti in quanto è 'preventivo' piuttosto che 'correttivo', dal momento che non si basa sulla correzione della stima stessa ma sulla correzione del meccanismo di stima, ovvero ciò che viene modificato sono le funzioni di punteggio. Viene, inoltre, definito metodo implicito poichè lo stimatore risultante è soluzione di equazioni implicite, dove si approssima la distorsione dello stimatore obiettivo e si risolve il sistema di equazioni. L'approccio proposto da Firth (1993) non richiede il calcolo della stima di massima verosimiglianza. Ciò lo ha reso un metodo popolare in modelli in cui, con probabilità positiva, lo stimatore di massima verosimiglianza appartiene alla frontiera dello spazio parametrico. Tale probabilità può non essere trascurabile in situazioni di bassa numerosità campionaria e/o di sparsità dei dati.

Nei modelli parametrici regolari sotto campionamento casuale semplice con numerosità n, o, più in generale, se $i(\theta) = O(n)$, la distorsione della stimatore di massima verosimiglianza, $\hat{\theta}$, risulta essere pari a

$$E_{\theta}[\hat{\theta} - \theta] = b(\theta) + O(n^{-2}), \tag{1.14}$$

dove $b(\theta) = -i^{-1}(\theta)A^*(\theta) = O(n^{-1})$, e $A^*(\theta)$ ha componenti

$$A_r^*(\theta) = \frac{1}{2}\,\mathrm{Tr}\{i^{-1}(\theta)[P_r(\theta) + Q_r(\theta)]\}, \qquad r = 1, \ldots, p, \tag{1.15}$$

con $\mathrm{Tr}\{\cdot\}$ ad indicare la traccia della matrice. Le matrici $P_r(\theta)$ e $Q_r(\theta)$ contengono i momenti nulli definiti in (1.11) e in particolare

$$P_r = E_{\theta}[U(\theta)U(\theta)^T U_r(\theta)], \tag{1.16}$$

$$Q_r = E_{\theta}[-j(\theta)U_r(\theta)]. \tag{1.17}$$

Dalle equazioni sopra, l'elemento di posizione (s, t) di P_r è dato da $\nu_{r,s,t}$, mentre $\nu_{r,st}$ rappresenta l'elemento di posizione (s, t) di Q_r.

L'approccio proposto da Firth (1993) prevede una funzione punteggio modificata della forma

$$U^*(\theta) = U(\theta) + A^*(\theta), \tag{1.18}$$

dove il termine $A^*(\theta)$ è di ordine $O(1)$ ed è direttamente collegato a $b(\theta)$. La modificazione della funzione di punteggio pemette di rimuovere la distorsione di ordine $O(n^{-1})$ dallo stimatore di massima verosimiglianza. La soluzione di $U^*(\theta) = 0$ determina lo stimatore con distorsione ridotta in media, $\hat{\theta}^*$, e che risulta avere distorsione asintotica di ordine $O(n^{-2})$, ossia $E_\theta[\hat{\theta}^*] = \theta + O(n^{-2})$. Lo stimatore $\hat{\theta}^*$ non richiede che $\hat{\theta}$ sia finito. Si dimostra (Kosmidis, 2007, §6.3) che se lo stimatore di massima verosimiglianza $\hat{\theta}$ è consistente allora anche lo stimatore $\hat{\theta}^*$ è consistente e ha la stessa distribuzione asintotica dello stimatore di massima verosimiglianza, ossia $\hat{\theta}^* \dot{\sim} \mathcal{N}_p(\theta, i^{-1}(\theta))$. Le procedure inferenziali collegate alla normalità asintotica dello stimatore $\hat{\theta}^*$ sono analoghe a quelle in uso per lo stimatore di massima verosimiglianza.

La soluzione di $U^*(\theta) = 0$ può non esistere finita e la non reperebilità dell'espressione dello stimatore $\hat{\theta}^*$ in forma chiusa determina il ricorso ad algoritmi di calcolo numerico quali ad esempio l'algoritmo Fisher-scoring la cui $(k+1)$-esima iterazione è esprimibile come

$$\theta^{(k+1)} = \theta^{(k)} + i^{-1}(\theta^{(k)})U(\theta^{(k)}) + i^{-1}(\theta^{(k)})A^*(\theta^{(k)}). \tag{1.19}$$

La formulazione (1.19) differisce per la quantità $i^{-1}(\theta^{(k)})A^*(\theta^{(k)})$ dall'analoga espressione dell'algoritmo Fisher-scoring utilizzata per ottenere le stime di massima verosimiglianza.

Il prinicipale svantaggio dell'approccio di Firth (1993) è che lo stimatore risultante non è equivariante rispetto a ogni tipo di trasformazione biunivoca e ciò è legato al fatto che la distorsione di uno stimatore dipende dal modo in cui viene parametrizzato il modello. Tuttavia, lo stimatore che riduce la distorsione in media gode della proprietà di equivarianza per trasformazioni lineari del tipo $\psi(\theta) = L\theta$, dove L è una opportuna matrice di contrasti di parametri di dimensione $p \times p$ e non singolare. La mancanza di una proprietà desiderabile, quale l'equivarianza, è spesso considerata compensata dal vantaggio offerto nelle applicazioni pratiche avendo cura di scegliere una opportuna parametrizzazione.

1.2.2 Riduzione della distorsione in mediana

L'approccio proposto da Kenne Pagui *et al.* (2017) permette di ottenere la centratura in mediana delle stime di massima verosimiglianza mediante una modificazione della funzione di punteggio. Tale metodo, come quello proposto da Firth (1993), non richiede la finitezza delle stime di massima verosimiglianza e risulta efficace nel prevenire stime sulla frontiera dello spazio parametrico. Lo stimatore che si ottiene è non distorto in mediana al terzo ordine nel caso scalare e componente per componente nel caso di un vettore di parametri.

Nel caso di parametro scalare di interesse, viene modificata la funzione di punteggio e, adottando la mediana come indice di centratura di essa, si ottiene la funzione di punteggio modificata sottraendo alla funzione di punteggio la sua mediana approssimata. Nel caso di parametro scalare di interesse in presenza di parametri di disturbo viene modificata la funzione score profilo e si ottiene la funzione di punteggio modificata sottraendo alla funzione score profilo la sua mediana approssimata. La soluzione delle risultanti equazioni di stima, purchè aventi soluzione unica, restituiscono uno stimatore con distorsione ridotta in mediana. Tale stimatore gode della proprietà di equivarianza sotto riparametrizzazioni monotone nel caso scalare e, nel caso di parametro scalare in presenza di parametri di disturbo, la proprietà di equivarianza è soddisfata rispetto a riparametrizzazioni che non alterano l'interesse. Si rimanda a Kenne Pagui *et al.* (2017) per la trattazione esaustiva del caso scalare e del caso di parametro scalare di interesse in presenza di parametri di disturbo.

Nel caso di un vettore di parametri $\theta \in \mathbb{R}^p, p > 1$, l'approccio utilizzabile per il caso di parametro scalare di interesse, con e senza parametri di disturbo, non è percorribile data la scarsa maneggevolezza della definizione di mediana multivariata. Le definizioni disponibili proposte, quali considerare la mediana marginale o la mediana multivariata (Oja, 2013), non si adattano allo sviluppo della modifica in mediana del vettore score. L'approccio che si segue è quello di impostare un sistema di equazioni di stima che, per ogni θ_r, $r = 1, \ldots, p$, dia la stessa stima dell'equazione di stima basata sulla funzione score profilo modificata, equazione (8) in Kenne Pagui *et al.* (2017). Questo si ottiene definendo il vettore score modificato in mediana $\tilde{U}(\theta)$ con componenti

$$\tilde{U}_r(\theta) = U_r(\theta) - \gamma_{ra} U_a(\theta) + M_r, \qquad r = 1, \dots, p, \tag{1.20}$$

dove

$$\bar{U}_r(\theta) = U_r(\theta) - \gamma_{ra} U_a(\theta),$$

sono le componenti della score efficiente (*efficient score*), ovvero il termine dominante dello sviluppo della funzione score profilo, $U_P(\theta_r)$. Nella formula (1.20) e nelle quantità ad essa collegate gli indici $a, b, \dots$ prendono valori in $\{1, \dots, p\} \setminus \{r\}$. Qui e nel seguito si adotta la convenzione della somma di Einstein per cui quando un indice compare due o più volte si sottintende la somma rispetto a quell'indice sopra il campo di variazione. Il termine

$$M_r = -\kappa_{1r} + \frac{\kappa_{3r}}{6\kappa_{2r}}$$

rappresenta la mediana approssimata della funzione score profilo e i termini $\kappa_{jr}, j = 1, 2, 3$, sono i cumulanti approssimati di ordine j della funzione score profilo, $U_P(\theta_r)$, e risultano pari a

$$\begin{aligned}
\kappa_{1r} &= -\frac{1}{2}\nu^{ab}\{(\nu_{r,ab} - \gamma_{rc}\nu_{c,ab}) + (\nu_{r,a,b} - \gamma_{rc}\nu_{a,b,c}\},\\
\kappa_{2r} &= \nu_{r,r} - \gamma_{ra}\nu_{r,a},\\
\kappa_{3r} &= \nu_{r,r,r} - 3\gamma_{ra}\nu_{r,r,a} + 3\gamma_{ra}\gamma_{rb}\nu_{r,a,b} - \gamma_{ra}\gamma_{rb}\gamma_{rc}\nu_{a,b,c},
\end{aligned}$$

dove l'errore è di ordine $O(n^{-1})$ per κ_{1r} e $O(1)$ per κ_{2r} e κ_{3r}. Le quantità $\nu_{a,b}$, $\nu_{a,b,c}$, $\nu_{a,bc}$ sono i momenti nulli definiti in (1.11) e $\gamma_{ra} = \nu^{ab}\nu_{r,b}$ rappresenta il coefficiente di regressione di U_r su U_a, $a \in \{1, \dots, p\} \setminus \{r\}$. In linea con la notazione indiciale ν^{ab} rappresenta l'elemento (a, b) dell'inversa della matrice quadrata di dimensione $p - 1$ con elementi $\nu_{a,b}$. Inoltre tutte le quantità coinvolte vengono valutate in θ e non nelle stime vincolate.

La funzione di punteggio modificata risulta essere

$$\tilde{U}(\theta) = U(\theta) + i(\theta)M_1(\theta) \tag{1.21}$$

dove $M_1(\theta)$ ha elementi $M_{1r} = M_r/\kappa_{2r}$. La stima congiunta $\tilde{\theta}$ si ottiene come soluzione dell'equazione di stima $\tilde{U}(\theta) = 0$. Il sistema di equazioni $\tilde{U}(\theta) = 0$ può non avere soluzione e data l'analogia con il sistema di equazioni (1.18) per la riduzione della distorsione in media si può riproporre la stessa strategia di implementazione basata sull'algoritmo Fisher-scoring riscritto come

$$\theta^{(k+1)} = \theta^{(k)} + M_1(\theta^{(k)}) + i^{-1}(\theta^{(k)})U(\theta^{(k)}). \tag{1.22}$$

Quando disponibili, buoni valori iniziali per l'algoritmo sono le stime di massima verosimiglianza $\hat{\theta}$ o le stime non distorte in media $\hat{\theta}^*$ e la convergenza è legata alle particolari proprietà del modello specificato.

Nel caso continuo, la non distorsione in mediana delle componenti di $\tilde{\theta}$, con errore di ordine $O(n^{-3/2})$, segue dall'analoga proprietà di non distorsione in mediana dello stimatore ottenuto mediante l'equazione di stima basata sulla funzione score profilo modificata. Risulta dunque $P_\theta(\tilde{\theta}_r \leq \theta_r) = 1/2 + O(n^{-3/2})$, per cui la distorsione in mediana di $\tilde{\theta}$ ha ordine di errore inferiore rispetto a quella dello stimatore di massima verosimiglianza. Inoltre, lo stimatore con distorsione ridotta in mediana ha la stessa distribuzione asintotica dello stimatore di massima verosimiglianza e dello stimatore con distorsione ridotta in media, cioè $\tilde{\theta} \dot{\sim} \mathcal{N}_p(\theta, i^{-1}(\theta))$ e le procedure inferenziali collegate alla normalità asintotica dello stimatore $\tilde{\theta}$ sono analoghe a quelle in uso per lo stimatore di massima verosimiglianza e per lo stimatore con distorsione ridotta in media.

Un'importante proprietà di cui gode lo stimatore con distorsione ridotta in mediana, $\tilde{\theta}$, è l'equivarianza rispetto a riparametrizzazione che trasforma ogni componente di θ separatamente e ciò è dovuto al fatto che ogni componente di $\tilde{U}_r(\theta)$ si comporta tensorialmente rispetto a riparametrizzazioni di θ_r. Tuttavia, lo stimatore non risulta equivariante rispetto a trasformazioni lineari, quali i contrasti dei parametri, anche se in alcuni modelli, come il modello di regressione beta e il modello di regressione multinomiale, lo stimatore con distorsione ridotta in mediana continua a mantenere la sua proprietà di non distorsione in mediana molto accurata (si vedano Kosmidis *et al.*, 2019, e Kenne Pagui *et al.*, in elaborazione, 2019).

Semplificazioni algebriche per il metodo proposto da Kenne Pagui *et al.*

(2017) sono state ottenute dagli stessi autori (Kenne Pagui *et al.*, in elaborazione, 2019) e la nuova formulazione ha tra i suoi vantaggi quelli di impedire sommatorie multiple (si evita il calcolo dei cumulanti), di facilitare l'implementazione per modelli parametrici generali e, non ultimo, di ottenere notevoli vantaggi computazionali. Secondo la nuova formulazione, in fase di elaborazione, la funzione di punteggio modificata può essere riscritta come

$$\tilde{U}(\theta) = U(\theta) + \tilde{A}(\theta), \tag{1.23}$$

con

$$\tilde{A}(\theta) = A^*(\theta) - i(\theta)F(\theta). \tag{1.24}$$

Il termine $F(\theta)$ ha elementi $F_r(\theta) = [i^{-1}(\theta)]_r^T \tilde{F}_r(\theta)$, $r = 1, \ldots, p$, con $\tilde{F}_r(\theta)$ avente componenti

$$\tilde{F}_{r,t}(\theta) = \mathrm{Tr}\left\{ h_r(\theta) \left[\frac{1}{3} P_t(\theta) + \frac{1}{2} Q_t(\theta) \right] \right\}, \qquad t = 1, \ldots, p. \tag{1.25}$$

Le quantità $A^*(\theta)$, $P_t(\theta)$ e $Q_t(\theta)$ sono state definite rispettivamente in (1.15), (1.16) e (1.17). Il termine $h_r(\theta)$ è una matrice $p \times p$ tale che

$$h_r(\theta) = \frac{[i^{-1}(\theta)]_r [i^{-1}(\theta)]_r^T}{i^{rr}(\theta)}. \tag{1.26}$$

Secondo la formulazione (1.23) lo stimatore con distorsione ridotta in mediana, $\tilde{\theta}$, è soluzione di $\tilde{U}(\theta) = 0$ e la $(k+1)$-esima iterazione dell'algoritmo Fisher-scoring può essere riscritta come

$$\theta^{(k+1)} = \theta^{(k)} + i^{-1}(\theta^{(k)})\tilde{A}(\theta^{(k)}) + i^{-1}(\theta^{(k)})U(\theta^{(k)}), \tag{1.27}$$

Tale formulazione dell'algoritmo Fisher-scoring costituirà il nucleo centrale per l'implementazione dello stimatore con distorsione ridotta in mediana nel modello di regressione preso in esame nel prosieguo della trattazione.

Capitolo 2

Modelli per variabili risposta categoriali ordinali

L'obiettivo di questo capitolo è fornire i necessari richiami sui modelli link cumulati per dati ordinali, con particolare enfasi sulla tipologia con quote (*odds*) proporzionali, una delle più popolari applicazioni per la trattazione dei dati ordinali a partire dalla sua formalizzazione ad opera di McCullagh (1980). Si presenterà, inoltre, l'approccio ai modelli link cumulati secondo la metodologia proposta da Firth (1993) e formalizzata da Kosmidis (2014b).

Nel §2.1 si introduce la notazione usata e si motiva la scelta delle probabilità cumulate per la modellazione di variabili risposta su scala ordinale. Il §2.2 presenta i modelli link cumulati e, nello specifico, il §2.2.1 è dedicato alla versione con quote proporzionali (modelli logit cumulati). La specificazione di funzioni di legame diverse dal legame logit e l'interpretazione dei modelli link cumulati in termini di variabile latente continua sottostante vengono discusse rispettivamente nei §2.2.2 e §2.2.3. Nel §2.2.4 si riporta un'utile trasformazione monotona del parametro di regressione relativo a una variabile dicotomica in grado di fornire una misura di superiorità ordinale (Agresti e Kateri, 2017). Il §2.2.5 riporta due specificazioni alternative dei modelli logit cumulati che rilassano l'assunzione di proporzionalità. Il §2.3 è dedicato alla presentazione del lavoro di Kosmidis (2014b) relativo allo stimatore che riduce la distorsione in media nei modelli link cumulati. Nel §2.4 si tratta il problema della stime sulla frontiera dello spazio parametrico e nel §2.5 le proprietà di invarianza relative ai modelli link cumulati e di equivarianza relative allo stimatore di massima verosimiglianza e allo stimatore che riduce la distorsione in media.

I principali riferimenti bibliografici per questo capitolo sono Salvan *et al.* (2018, Capitolo 4), relativo alla notazione adottata, Agresti (2010, Capitoli 3 e 5), Agresti e Kateri (2017) e Kosmidis (2014b).

2.1 Variabili risposta su scala ordinale

In presenza di una variabile risposta qualitativa con $c > 2$ modalità (risposta politomica) è naturale modellarne la distribuzione mediante il modello multinomiale. Una variabile risposta di tipo categoriale può presentare o meno un ordinamento naturale tra le modalità e si avranno, rispettivamente, variabili risposta categoriali su scala ordinale e su scala sconnessa (o nominale). Variabili risposta categoriali su scala ordinale sono molto frequenti in diversi settori applicativi quali le scienze mediche, sociali e nei contesti aziendali.

La trattazione dei dati ordinali ha sofferto e soffre spesso dell'impropria utilizzazione di metodi statistici che trattano i dati categoriali di natura ordinale come se fossero di natura nominale, ignorando l'ordinamento delle modalità della variabile risposta, o di natura continua, assegnando valori numerici alle modalità della variabile risposta e usando metodi quali i minimi quadrati. Sebbene nuove tecniche di modellazione più flessibili e accurate a livello di previsione si adattino alla trattazione di dati ordinali (si veda ad esempio Azzalini e Scarpa, 2012, §5.10.3), emerge, comunque, l'importanza di usare appropriate tecniche di modellazione statistica in grado di sfruttare la natura ordinale della variabile risposta, specie quando l'obiettivo dell'analisi è di natura inferenziale.

Assumendo di disporre di dati non raggruppati, le osservazioni sulla variabile risposta sono costituite da n vettori $\mathbf{y_i} = (y_{i1}, \ldots, y_{ic})$, $i = 1, \ldots, n$, con $y_{ij} = 1$ se per il soggetto i-esimo si è osservata la modalità j-esima, $j = 1, \ldots, c$, e $y_{ij} = 0$ altrimenti. Si ha, dunque, che $\sum_{j=1}^{c} y_{ij} = 1$ e la variabile casuale c-dimensionale $\mathbf{Y_i} = (Y_{i1}, \ldots, Y_{ic})$, di cui $\mathbf{y_i}$ è l'osservazione campionaria, ha distribuzione multinomiale elementare con funzione di densità

$$p_{\mathbf{Y_i}}(\mathbf{y_i}, \pi_i) = \prod_{j=1}^{c} \pi_{ij}^{y_{ij}},$$

definita sul supporto $\left\{\mathbf{y_i} \in \{0,1\}^c : \sum_{j=1}^{c} y_{ij} = 1\right\}$. Il parametro è $\pi_i = (\pi_{i1}, \ldots, \pi_{ic})$, con la componente π_{ij}, tale che $\pi_{ij} \in (0,1)$ e soggetta al vincolo $\sum_{j=1}^{c} \pi_{ij} = 1$, che denota la probabilità di osservare la j-esima modalità per l'i-esima osservazione.

Indicate con j, $j = 1, \ldots, c$, le modalità della risposta, sia Y_i la variabile casuale che descrive la risposta, con modalità ordinate, dell'i-esimo soggetto e sia $F_{Y_i}(j) = P(Y_i \leq j)$ la funzione di ripartizione di tale variabile casuale (si noti la differenza di notazione rispetto al vettore casuale $\mathbf{Y_i}$ relativo alle variabili indicatrici). Lo studio della relazione tra la distribuzione della variabile risposta Y_i e il valore $\mathbf{x}_i$ delle variabili esplicative per l'i-esima unità statistica, nel caso di variabili risposta categoriali su scala ordinale, si traduce nel voler valutare l'impatto delle variabili esplicative in termini di ordinamento stocastico delle distribuzioni della variabile risposta. La nozione di ordinamento stocastico è particolarmente utile per il confronto tra due gruppi e si traduce nel confronto tra le funzioni di ripartizione della variabile risposta. Se per $k \neq i$ vale

- $F_{Y_i}(j) \leq F_{Y_k}(j), \forall j = 1, \ldots, c$, la distribuzione della variabile casuale Y_i è stocasticamente più grande della distribuzione della variabile casuale Y_k (per il soggetto i-esimo vi è una probabilità più elevata di essere nelle modalità più elevate della scala della variabile risposta);

- $F_{Y_i}(j) \geq F_{Y_k}(j), \forall j = 1, \ldots, c$, la distribuzione della variabile casuale Y_i è stocasticamente più piccola della distribuzione della variabile casuale Y_k;

con la disuguaglianza stretta valida per almeno un $j \in \{1, \ldots, c\}$.

Un'ovvia scelta per la modellazione di variabili risposta categoriali su scala ordinale è rappresentata da modelli che mettono in relazione le probabilità cumulate, $F_{Y_i}(j)$, con i valori delle variabili esplicative (modelli link cumulati).

Si ha

$$F_{Y_i}(j) = P(Y_i \leq j) = \sum_{h=1}^{j} \pi_{ih}, \qquad i = 1, \ldots, n, \ j = 1, \ldots, c\,,$$

dove $\pi_{ih} = P(Y_i = h) = P(Y_i \leq h) - P(Y_i \leq h-1)$.

2.2 Modelli link cumulati

Il modello link cumulati più popolare modella simultaneamete i $c-1$ link cumulati mettendo in relazione le probabilità cumulate $P(Y_i \leq j)$ con il predittore lineare mediante la formulazione

$$h[P(Y_i \leq j|\mathbf{x}_i)] = \alpha_j + \mathbf{x}_i\beta, \quad j = 1, \ldots, c-1, \quad i = 1, \ldots, n, \qquad (2.1)$$

dove $h[\cdot]$ è un' opportuna funzione di legame e $\mathbf{x}_i$ è il vettore riga contenente i valori delle p variabili esplicative. Il vettore dei parametri è dunque $\theta = (\alpha_1, \ldots, \alpha_{c-1}, \beta_1, \ldots, \beta_p)$ e tutte le componenti di tale vettore hanno come campo di esistenza quello dei numeri reali. I parametri $\beta_1, \ldots, \beta_p$ descrivono l'effetto delle variabili esplicative che risulta lo stesso per ogni link cumulato, mentre ogni link per la probabilità cumulata della modalità j ha una propria intercetta α_j. Risulta che gli α_j soddisfano $\alpha_1 \leq \ldots \leq \alpha_{c-1}$, poichè $P(Y_i \leq j)$ è crescente in j per ogni fissato $\mathbf{x}_i$.

In letteratura e in alcuni software di programmazione statistica si può trovare il modello parametrizzato come

$$h[P(Y_i \leq j|\mathbf{x}_i)] = \alpha_j - \mathbf{x}_i\beta, \quad j = 1, \ldots, c-1. \qquad (2.2)$$

Secondo tale parametrizzazione, l'interpretazione dell'effetto delle variabili esplicative è in linea con l'usuale interpretazione che si ha nei modelli di regressione lineare, ossia, ad esempio $\beta_r > 0$ indica che la distribuzione di Y_i diventa stocasticamente più grande all'aumetare di x_{ir}, a parità delle altre variabili esplicative.

2.2.1 Modelli logit cumulati con quote proporzionali

Il modello link cumulati che si ottiene con la funzione di legame logit

$$h[P(Y_i \leq j|\mathbf{x}_i)] = \text{logit}[P(Y_i \leq j|\mathbf{x}_i)] = \alpha_j + \mathbf{x}_i\beta, \quad j = 1, \dots, c-1, \quad (2.3)$$

è detto modello logit cumulato con quote proporzionali (*proportional odds model*, McCullagh, 1980). L'espressione equivalente del modello in termini di probabilità cumulate è

$$P(Y_i \leq j|\mathbf{x}_i) = \frac{e^{\alpha_j + \mathbf{x}_i\beta}}{1 + e^{\alpha_j + \mathbf{x}_i\beta}}, \quad (2.4)$$

da cui risulta che la probabilità che per l'individuo i-esimo si osservi la $j-$esima modalità, $j = 1, \dots, c-1$, è pari a

$$P(Y_i = j|\mathbf{x}_i) = \frac{e^{\alpha_j + \mathbf{x}_i\beta}}{1 + e^{\alpha_j + \mathbf{x}_i\beta}} - \frac{e^{\alpha_{j-1} + \mathbf{x}_i\beta}}{1 + e^{\alpha_{j-1} + \mathbf{x}_i\beta}},$$

con $\alpha_0 = -\infty$. Per tale modello vale la proprietà che il logaritmo del rapporto delle quote cumulate è proporzionale alla differenza tra due vettori di esplicative, con costante di proporzionalità indipendente dalla modalità j. Formalmente, siano $\mathbf{x}_i$ e $\mathbf{x}_{i'}$ due vettori p-dimensionali di variabili esplicative, si esprime la proprietà di quote cumulate proporzionali come

$$\log \frac{P(Y_i \leq j|\mathbf{x}_i)/P(Y_i > j|\mathbf{x}_i)}{P(Y_{i'} \leq j|\mathbf{x}_{i'})/P(Y_{i'} > j|\mathbf{x}_{i'})} = (\mathbf{x}_i - \mathbf{x}_{i'})\beta. \quad (2.5)$$

Risulta che per due differenti valori dell'$r-$esima variabile esplicativa, a parità delle altre variabili esplicative, la quota cumulata di avere risposta $Y_i \leq j$ per l'individuo $\mathbf{x}_i$ è pari a $e^{(x_{ir} - x_{i'r})\beta_r}$ volte la quota cumulata di avere risposta $Y_{i'} \leq j$ per l'individuo $\mathbf{x_{i'}}$, ovvero

$$\frac{P(Y_i \leq j|\mathbf{x}_{ir})}{P(Y_i > j|\mathbf{x}_{ir})} = e^{(x_{ir} - x_{i'r})\beta_r} \frac{P(Y_{i'} \leq j|\mathbf{x}_{i'r})}{P(Y_{i'} > j|\mathbf{x}_{i'r})}. \quad (2.6)$$

Il modello è detto con quote proporzionali in virtù della proprietà di effetto costante dei parametri di regressione su tutte le probabilità cumulate.

La stima dei parametri del modello logit cumulati, e in generale dei modelli link cumulati, predilige l'approccio di stima di massima verosimiglianza. La soluzione delle equazioni di verosimiglianza, $U(\theta) = 0$, può non esistere finita e non è esprimibile in forma chiusa. Si necessita, pertanto, di algoritmi di calcolo numerico per ottenere le stime di massima verosimiglianza. Nelle implementazioni presenti nei software di programmazione statistica sono di uso comune l'algoritmo Newton-Raphson, che sfrutta l'informazione osservata, $j(\theta)$, e l'algoritmo Fisher-scoring, equivalente all'algoritmo iterativo ai minimi quadrati pesati (ILSR) adottato per i modelli lineari generalizzati, che utilizza invece quella attesa, $i(\theta)$. Le stime che si ottengono con entrambi gli algoritmi sono identiche mentre gli standard error stimati differiscono e risultano, per una generica componente θ_r, rispettivamente $(j^{rr}(\hat{\theta}))^{1/2}$ e $(i^{rr}(\hat{\theta}))^{1/2}$, a seconda che si usi rispettivamente l'algoritmo Newton-Raphson o Fisher-scoring. L'algoritmo iterativo Fisher-scoring che porta alle stime di massima verosimiglianza dei parametri del modello prevede di aggiornare sequenzialmente fino a convergenza secondo

$$\theta^{(k+1)} = \theta^{(k)} + i(\theta^{(k)})^{-1} U(\theta^{(k)}). \tag{2.7}$$

Stabilito un criterio di arresto l'algoritmo dovrebbe convergere rapidamente alle stime di massima verosimiglianza $\hat{\theta}$ per una opportuna scelta dei valori iniziali $\theta^{(0)}$. Ottenute le stime di massima verosimiglianza, con i rispettivi standard error, si può procedere con le usuali procedure inferenziali basate sulla normalità asintotica dello stimatore di massima verosimiglianza e della funzione score.

2.2.2 Modelli link cumulati con diverse funzioni di legame

La scelta della funzione di legame logit è spesso la prima scelta quando si trattano i modelli link cumulati e permette di dare un'interpretazione dell'effetto mediante i rapporti delle quote (*odds ratio*). Specificazioni alternative alla funzione di legame logit sono disponibili e le più comunemente usate sono la funzione di legame probit e la funzione di legame log-log complementare (cloglog).

I modelli link cumulati con funzione di legame probit, l'inversa della funzione di distribuzione cumulata della normale standard, sono detti modelli probit cumulati. Denotando con $\Phi(\cdot)$ la funzione di ripartizione della normale standard, il modello probit cumulati ha la forma

$$\Phi^{-1}[P(Y_i \leq j|\mathbf{x_i})] = \alpha_j + \mathbf{x}_i\beta, \qquad j = 1, \ldots, c-1, \tag{2.8}$$

e l'espressione equivalente del modello in termini di probabilità cumulate risulta

$$P(Y_i \leq j|\mathbf{x_i}) = \Phi(\alpha_j + \mathbf{x}_i\beta), \qquad j = 1, \ldots, c-1.$$

A differenza del modello logit cumulati il modello probit cumulati non può essere definito con quote proporzionali poichè l'interpretazione risultante non si applica alle quote. Entrambi i modelli si adattano bene in situazioni simili dal momento che le distribuzioni logistica e normale hanno media uguale e deviazione standard simili. Le stime di massima verosimiglianza del modello logit cumulati risultano essere circa 1.6-1.8 volte quelle del modello probit cumulati.

Un'altra posssibile specificazione per la funzione di legame è la funzione log-log-complementare $h[P(Y_i \leq j)] = \log\{-\log[1 - P(Y_i \leq j)]\}$ che si differenzia dalla logit e dalla probit per il fatto che non risulta simmetrica. Il modello risultante è

$$\log\{-\log[1 - P(Y_i \leq j|\mathbf{x_i})]\} = \alpha_j + \mathbf{x}_i\beta \tag{2.9}$$

e ha trovato larga applicazione negli studi di analisi di sopravvivenza quando il tempo fino all'evento è misurato raggruppando i tempi in categorie ordinate. Il modello (2.9) è stato introdotto in letteratura (McCullagh, 1980) come modello a rischi proporzionali (*proportional hazards model*), in quanto adatta il modello a rischi proporzionali (Cox, 1972) per trattare tempi di sopravvivenza raggruppati.

2.2.3 Interpretazione dei modelli link cumulati in termini di variabile latente continua sottostante

La giustificazione dell'effetto comune del vettore β per ogni link cumulato nella formulazione (2.1) è legata all'ipotesi che Y_i sia ottenuta per discretizzazione di una variabile latente continua Y_i^* che soddisfa un modello di regressione $Y_i^* = -\mathbf{x}_i\beta + \epsilon_i, i = 1, \ldots, n$, dove le ϵ_i sono variabili casuali indipendenti ed identicamente distribuite con media zero e funzione di ripartizione $F(\cdot)$. Assegnando dei valori soglia α_j, $j = 1, \ldots, c$, tali che, con $-\infty = \alpha_0 \leq \alpha_1 \leq \ldots \leq \alpha_{c-1} \leq \alpha_c = +\infty$, si osserva

$$Y_i = j \qquad \text{se} \qquad \alpha_{j-1} \leq Y_i^* \leq \alpha_j,$$

risulta che

$$P(Y_i \leq j|\mathbf{x_i}) = P(Y_i^* \leq \alpha_j|\mathbf{x_i}) = P(\epsilon_i \leq \alpha_j + \mathbf{x}_i\beta) = F(\alpha_j + \mathbf{x}_i\beta). \tag{2.10}$$

Se $F(\cdot)$ è la funzione di ripartizione della logistica standard, $F(z) = \exp(z)/[1+\exp(z)]$, $z \in \mathbb{R}$, si ottiene il modello di regressione per logit cumulati con quote proporzionali (2.3). Con $F(\cdot)$ funzione di riparizione della normale standard, indicata con $\Phi(\cdot)$, si ottiene il modello di regressione probit cumulati (2.8). Con $F(\cdot)$ funzione di riparizione della distribuzione Gumbel (distribuzione dei valori estremi) standard, $F(z) = \exp[-\exp(-z)]$, $z \in \mathbb{R}$, si ottiene il modello di regressione link cumulato con funzione di legame log-log. Considerando la variabile casuale $Y = -Z$ dove Z è distribuita secondo una distribuzione Gumbel si ottiene che la funzione di ripartizione della variabile casuale Y è pari a $F_Y(z) = 1 - \exp[-\exp(z)]$, $z \in \mathbb{R}$, e si perviene, dunque, alla formulazione del modello di regressione link cumulato con funzione di legame log-log complementare (2.9).

Nel modello parametrizzato secondo (2.2) la variabile latente continua soddisfa un modello di regressione $Y_i^* = \mathbf{x}_i\beta + \epsilon_i$, $i = 1, \ldots, n$, dove gli ϵ_i sono definiti come sopra.

2.2.4 Misura di superiorità ordinale per confronti tra gruppi

Una interessante trasformazione monotona dei parametri relativi a variabili esplicative dicotomiche è stata proposta da Agresti e Kateri (2017) al fine di permetterne una interpretazione immediata in termini di confronto delle distribuzioni tra due gruppi. Tale trasformazione rappresenta dunque una misura di superiorità ordinale e fornisce la probabilità che una osservazione da una distribuzione assuma valori più piccoli di una osservazione indipendente da un'altra distribuzione, a parità di altre variabile esplicative. Tale misura risulta utile nel confronto tra due gruppi e si applica sia al modello di regressione lineare che ai modelli di regressione per variabili risposta su scala ordinale con variabile latente continua sottostante. Essa è una valida alternativa all'interpretazione dei rapporti delle quote cumulate e dei probit specie per la comunicazione dei risultati a persone con formazione non quantitativa.

Sia $\mathbf{x} = (x_1, \ldots, x_p)$ un vettore di variabili esplicative e sia x_r una variabile esplicativa dicotomica che indichi l'appartenenza di un individuo a un gruppo. Denotiamo con Y_{i1} la variabile risposta per una osservazione in uno dei due gruppi, ad esempio per $x_{ir} = 0$, e con Y_{i2} la variabile risposta indipendente per una osservazione dell'altro gruppo, ad esempio per $x_{ir} = 1$. Usando un modello basato sulla distribuzione condizionale di Y_i per i due gruppi, al valore delle esplicative $\mathbf{x_i}$, la misura di superiorità ordinale risulta essere

$$\Delta = P(Y_{i1} > Y_{i2} | \mathbf{x_i} \setminus \{x_{ir}\}) - P(Y_{i2} > Y_{i1} | \mathbf{x_i} \setminus \{x_{ir}\}),$$

con $\Delta \in [-1, 1]$ e se le distribuzioni Y_{i1} e Y_{i2} sono identicamente distribuite allora $\Delta = 0$. Per variabili risposta discrete, come pure per le variabili risposta categoriali su scala ordinale, una misura di superiorità ordinale risulta essere

$$\gamma = P(Y_{i1} > Y_{i2} | \mathbf{x_i} \setminus \{x_{ir}\}) + \frac{1}{2} P(Y_{i1} = Y_{i2} | \mathbf{x_i} \setminus \{x_{ir}\}), \qquad (2.11)$$

con $\gamma \in [0, 1]$ e se le distribuzioni Y_{i1} e Y_{i2} sono identicamente distribuite allora $\gamma = 1/2$. Le due misure di superiorità ordinale sono funzionalmente

legate dalle relazioni
$$\gamma = \frac{\Delta + 1}{2}, \quad \Delta = 2\gamma - 1.$$

Secondo la formulazione del modello (2.1) le espressioni di γ, in funzione delle diverse funzioni di legame, risultano

(i) logit: $\gamma(\beta_r) \approx \frac{\exp(-\beta_r/\sqrt{2})}{1+\exp(-\beta_r/\sqrt{2})}$,

(ii) probit: $\gamma(\beta_r) = \Phi(-\beta_r/\sqrt{2})$,

(iii) cloglog: $\gamma(\beta_r) = \frac{\exp(-\beta_r)}{1+\exp(-\beta_r)}$,

deve β_r è il parametro relativo alla variabile dicotomica, indicatrice dell'effetto del gruppo. Nel caso della funzione di legame logit il risultato è approssimato e non esatto per via del fatto che la differenza tra due distribuzioni logistiche non corrisponde a una distribuzione nota e una possibile opzione è quella di approssimare la distribuzione della differenza tra le due distribuzione logistiche con una distribuzione logistica con parametro β_r e parametro di scala $\sqrt{2}$. Per ulteriori dettagli si veda Agresti e Kateri (2017).

2.2.5 Modelli logit cumulati alternativi

La popolarità del modello logit cumulati con quote proporzionali è anche legata alla sua parsimonia poichè utilizza un solo parametro per ogni predittore e ciò ne garantisce una maggiore interpretabilità. Tale formulazione è, inoltre, legata all'interpretazione del modello in termini di una sottostante variabile latente continua, come discusso nel §2.2.3. Rilassando l'assunzione di proporzionalità delle quote nel modello logit cumulati si perviene a modelli che prevedono un effetto separato dei predittori per ogni logit e modelli che conservano l'assunzione di proporzionalità solo per una parte dei predittori.

Il modello che permette effetti differenti delle variabili esplicative per i differenti logit cumulati ha la forma

$$\text{logit}[P(Y_i \leq j|\mathbf{x_i})] = \alpha_j + \mathbf{x}_i\beta_j, \tag{2.12}$$

dove ogni predittore ha $c-1$ parametri e il vettore di parametri risultante è $\theta = (\alpha_1, \ldots, \alpha_{c-1}, \beta_{11}, \ldots, \beta_{1c-1}, \ldots, \beta_{p1}, \ldots, \beta_{pc-1})$. Tale formulazione non impone il parallelismo delle relazioni lineari relative a differenti logit cumulati e quindi le curve per differenti probabilità cumulate possono intersecarsi per alcuni valori delle variabili esplicative determinando una perdita dell'ordinamento stocastico delle distribuzioni della variabile risposta, a meno di non imporre opportuni vincoli. Conseguenza di ciò è che l'algoritmo di massimizzazione della verosimiglianza può non convergere e, qualora l'algoritmo convergesse, un miglioramento significativo dell'adattamento di tale modello ai dati, rispetto alla versione con quote proporzionali, può essere poco utile per via della perdita di interpretabilità e di parsimonia. Il modello che ammette vettori β dipendenti da j risulta preferibile quando si hanno pochi predittori qualitativi rispetto a situazioni con molti predittori, alcuni dei quali continui.

Il modello logit cumulati con quote proporzionali (2.3) è un modello annidato del modello (2.12) corrispondente al caso $\beta_{k1} = \ldots = \beta_{kc-1}$, per ogni $k = 1, \ldots, p$. Si può, dunque, effettuare un test al fine di verificare l'ipotesi che l'effetto sia lo stesso per ogni logit cumulato e, quindi, valga la struttura con quote proporzionali contro l'ipotesi alternativa di effetti separati. La versione con quote proporzionali ha un parametro di regressione per ogni predittore per un totale di p parametri, oltre ai $c - 1$ parametri di intercetta, mentre il modello più complesso ha $c-1$ parametri di regressione per ogni predittore per un totale di $p(c - 1)$ parametri, oltre ai $c - 1$ parametri di intercetta. Si può valutare, come in Peterson e Harrell (1990), l'allontanamento dall'assunzione di quote proporzionali mediante la statistica score

$$U(\hat{\theta}_0)^T i^{-1}(\hat{\theta}_0) U(\hat{\theta}_0), \tag{2.13}$$

dove $\hat{\theta}_0$ è la stima di massima verosimiglianza nel modello sotto l'ipotesi nulla $H_0 : \beta_1 = \ldots = \beta_{c-1}$. La statistica (2.13) ha distribuzione asintotica chi-quadrato con $p(c - 2)$ gradi di libertà. Tale statistica test risulta preferibile a quella basata sul test del rapporto di verosimiglianza e a quella di Wald che, usando la funzione di verosimiglianza del modello più generale, soffrono delle problematiche legate alla non convergenza dell'algoritmo testè menzionata. Tuttavia la statistica (2.13) risulta avere scarse prestazioni per dati sparsi e anche quando i dati non sono sparsi si osservano valori di significatività osservati troppo piccoli (Agresti, 2010, §3.5.5).

Una strada alternativa che ha riscosso un buon successo nella modellistica dei dati ordinali è stata considerata da Peterson e Harrel (1990), i quali hanno considerato un modello che si colloca tra il modello con quote proporzionali e il modello più generale con effetti separati per ogni logit. Tale modello è detto con quote proporzionali per un sottoinsieme di predittori (*partial proportional odds model*). Partizionando in blocchi il vettore delle variabili esplicative $\mathbf{x}_i = (\mathbf{u}_i, \mathbf{v}_i)$, il modello prevede una struttura con quote proporzionali per alcuni predittori, $\mathbf{u}_i$, mentre i predittori $\mathbf{v}_i$ non presentano tale struttura. Il modello ha la forma

$$\text{logit}[P(Y_i \leq j|\mathbf{x_i})] = \alpha_j + \mathbf{u}_i\beta + \mathbf{v}_i\gamma_j, \qquad j = 1, \ldots, c-1, \tag{2.14}$$

e per l'identificabilità si deve porre uno dei $\boldsymbol{\gamma_j}$ pari a $\mathbf{0}$, ad esempio $\boldsymbol{\gamma_1} = \mathbf{0}$. Il modello con quote proporzionali (2.3) è annidato in tale modello e corrisponde al caso $\boldsymbol{\gamma}_2 = \ldots = \boldsymbol{\gamma}_{c-1} = \mathbf{0}$ e considerando come insieme dei predittori $\mathbf{x}_i$ non partizionato. Pertanto si possono condurre test per modelli nidificati analoghi a quelli menzionati per il modello con effetti separati.

2.3 Riduzione della distorsione in media nei modelli link cumulati

Le equazioni di stima basate sulla funzione di punteggio modificata secondo l'approccio proposto da Firth (1993) sono state derivate da Kosmidis (2014b) per i modelli link cumulati. Lo stimatore risultante ha distorsione asintotica inferiore a quella dello stimatore di massima verosimiglianza, è equivariante rispetto a riparametrizzazioni lineari e soddisfa la proprietà di invarianza a seguito di inversione dell'ordine delle categorie. Inoltre lo stimatore risultante è quasi sempre finito, risulta avere buone proprietà frequentiste e ha comportamento asintotico analogo a quello dello stimatore di massima verosimiglianza.

Richiamando i concetti spiegati nel §1.2.1, si indica con $b(\theta)$ il termine dominante dello sviluppo asintotico della distorsione dello stimatore di massima verosimiglianza e siano $U(\theta)$ la funzione di punteggio e $i(\theta)$ la matrice di informazione attesa per $\theta = (\alpha_1, \ldots, \alpha_{c-1}, \beta_1, \ldots, \beta_p)$. Sia $A^*(\theta) = -i(\theta)b(\theta)$, le cui componente sono definite in (1.15). La funzione di punteggio modificata secondo il metodo di Firth (1993) ha componenti

$$U_r^*(\theta) = U_r(\theta) + A_r^*(\theta), \quad r = 1, \ldots, p + c - 1,$$

con A_r^* di ordine $O(1)$. La soluzione delle equazioni di stima basate sulla funzione di punteggio modificata risulta nello stimatore con distorsione ridotta in media, $\hat{\theta}^*$. Per i modelli link cumulati le quantità necessarie per ottenere lo stimatore con distorsione ridotta in media risultano analoghe a quelle utilizzate per lo stimatore con distorsione ridotta in mediana, esplicitate nei successivi §3.1 e §3.2. Tuttavia esse sono presentate in Kosmidis (2014b, §4.2) in forma compatta.

La soluzione delle equazioni $U_r^*(\theta) = 0$ può non esistere finita e non è esprimibile in forma chiusa. Come discusso nel §1.2.1, le stime $\hat{\theta}^*$ si possono ottenere numericamente mediante l'algoritmo Fisher-scoring (1.19). Kosmidis (2014b, §5.2) sottolinea che le iterazioni connesse all'algoritmo Fisher-scoring richiedono un notevole onere computazionale e si possono ottenere in modo alternativo mediante un approccio di stima basato sulla correzione iterativa delle frequenze multinomiali.

L'allontamento dall'assunzione di quote proporzionali, in analogia con quanto visto per l'approccio di stima di massima verosimiglianza nel §2.2.5 , può essere sottoposto a verica d'ipotesi (Kosmidis, 2014b, §12) mediante la statistica score modificata

$$U^*(\hat{\theta}_0^*)^T i^{-1}(\hat{\theta}_0^*) U^*(\hat{\theta}_0^*), \tag{2.15}$$

dove $\hat{\theta}_0^*$ denota le stime con distorsione ridotta in media nel modello soggetto a vincoli. La statistica score modificata ha distribuzione asintotica chi-quadrato con $p(c-2)$ gradi di libertà e ciò è dovuto alla relazione $U^*(\theta) = U(\theta) + A^*(\theta)$, con $A^*(\theta) = O(1)$.

2.4 Stime sulla frontiera dello spazio parametrico

Con numerosità campionarie sufficientemente grandi le stime di massima verosimiglianza esistono e sono uniche (McCullagh, 1980) mentre per numerosità campionarie finite e relativamente piccole le stime di massima verosimiglianza possono non esistere o essere "infinite" per certe configurazioni dei dati, ovvero vi è una probabilità positiva che le stime dei parametri siano sulla frontiera dello spazio parametrico. Il problema delle stime sulla frontiera dello spazio parametrico accade con frequenza per piccole numerosità campionarie, per un gran numero di parametri nel modello, dati altamente non bilanciati e dati sparsi nelle tabelle di contingenza. In un modello di regressione logistica per variabili risposta dicotomiche le stime di massima verosimiglianza non esistono o sono sulla frontiera dello spazio parametrico quando vi è separazione completa o quasi-completa (Albert e Anderson, 1984). Per un modello logit cumulati ciò accade se tale separazione avviene a seguito della collassabilità della variabile risposta ordinale in una binaria. Per una discussione approfondita di quali configurazioni dei dati determinino stime di massima verosimiglianza dei parametri di regressione sulla frontiera dello spazio parametrico si veda Agresti (2010, §3.4.5). Le stime sulla frontiera dello spazio parametrico possono:

(i) causare instabilità numeriche nella procedura di stima;

(ii) portare ad interpretazioni erronee quando la procedura di stima si basa su un algoritmo iterativo che viene fermato dopo un certo numero di iterazioni (fermando l'algoritmo ad un numero maggiore di iterazioni si può notare come le stime divergono gradualmente);

(iii) causare un deterioramento delle procedure inferenziali, in particolare quelle dipendenti dalle stime degli standard error.

Nei modelli link cumlati le stime di massima verosimiglianza dei parametri hanno probabilità positiva di essere sulla frontiera dello spazio parametrico. Le stime sulla frontiera dello spazio parametrico per tali modelli sono costituite da stime dei parametri di regressione $\beta = (\beta_1, \ldots, \beta_p)$ con valori infiniti e/o stime dei parametri di intercetta $-\infty = \alpha_0 \leq \alpha_1 \leq \ldots \leq \alpha_{c-1} \leq \alpha_c = +\infty$ per le quali $\hat{\alpha}_j = \hat{\alpha}_{j+1}$, per almeno un $j = 1, \ldots, c-2$. Pratt (1981) mostra che $\hat{\alpha}_j = \hat{\alpha}_{j+1}$ se e solo se le frequenze osservate per la j-esima modalità, $j = 2, \ldots, c-1$, sono zero. Inoltre, se la prima o l'ultima categoria hanno frequenze osservate pari a zero, allora le stime per α_1 e α_{c-1} saranno rispettivamente $-\infty$ e $+\infty$.

In letteratura sono presenti diverse soluzioni a tale problematica che non prevedono l'esclusione della variabile esplicativa per cui si ottiene una stima infinita del parametro corrispondente. Una possibilità nei modelli logit cumulati è il ricorso alla trasformazione logit empirica (McCullagh, 1980), la cui applicabilità è tuttavia limitata all'analisi delle tabelle di contingenza $2 \times k$ e non è generalizzabile per l'analisi di modelli link cumulati più generali. Kosmidis (2014b, §1) menziona tra le possibilità quella di ricorrere allo stimatore *ridge* la cui proprietà di *shrinkage* è in grado di garantire la finitezza delle stime. Tuttavia presenta una serie di svantaggi rispetto allo stimatore di massima verosimiglianza come la perdita dell'equivarianza rispetto a trasformazioni lineari dei parametri e la dipendenza dalla scelta del parametro di regolazione. Quando i dati sono sparsi una pratica usuale per impedire stime sulla frontiera e ridurre la distorsione è quella di aggiungere una piccola costante alle frequenze osservate. Tuttavia tale schema di aggiustamento per una costante non è raccomandato e perchè non è invariante rispetto a diverse rappresentazioni dei dati, ad esempio dati aggregati e disaggregati, e perchè non vi è una costante universale che porti a stime che sono ottimali in accordo a qualche criterio frequentista. L'approccio proposto da Firth (1993), come mostrato da Kosmidis (2014b), permette non solo di ridurre la distorsione dello stimatore ma può garantire la finitezza delle stime in molti casi dove le stime di massima verosimiglianza sono sulla frontiera dello spazio parametrico.

Riguardo alle stime sulla frontiera delle intercette $\alpha_1, \ldots, \alpha_{c-1}$, Kosmidis (2014b, §8.3, *Remark 1*) evidenzia come lo stimatore di massima verosimiglianza dei parametri $\beta_1, \ldots, \beta_p$ sia invariante per accorpamento di categorie non osservate con quelle adiacenti e può essere finito anche se le stime di alcuni dei parametri di intercetta sono sulla frontiera dello spazio parametrico. Quindi l'inferenza di massima verosimiglianza è possibile sui parametri β anche se una o più categorie non sono osservate. Anche lo stimatore con distorsione ridotta in media dei parametri β risulta invariante per accorpamento di categorie con frequenza osservata nulla. L'unica differenza rispetto allo stimatore di massima verosimiglianza è che se le categorie non osservate sono la prima e/o l'ultima le stime con distorsione ridotta in media, $\hat{\alpha}_1^*$ e/o $\hat{\alpha}_{c-1}^*$, sono comunque finite e pertanto non si necessita di accorpare tali categorie estreme.

Il problema sorge dal fatto che la maggior parte dei software non è in grado di riconoscere quando siamo in presenza di stime sulla frontiera dello spazio parametrico. Una strategia per individuare parametri con stime sulla frontiera nei modelli link cumulati si rifà alla diagnostica di Lesaffre e Albert (1989) per la regressione logistica multinomiale. Gli insiemi di dati che risultano in stime infinite dei β vengono individuati dall'osservazione di un valore assoluto delle stime e del corrispondente standard error stimato elevati (si possono usare ad esempio come soglie 20 per il valore assoluto delle stime e 100 per il corrispondente standard error stimato).

2.5 Proprietà di invarianza e di equivarianza nei modelli link cumulati

Sia $\psi = \psi(\theta)$, con $\psi(\cdot)$ funzione biettiva e regolare. Lo stimatore con distorsione ridotta in media, $\hat{\theta}^*$, non è equivariante per ogni $\psi(\cdot)$, proprietà che invece possiede lo stimatore di massima verosimiglianza. Tuttavia lo stimatore $\hat{\theta}^*$ è equivariante per trasformazioni lineari $\psi = \psi(\theta) = L\theta$ dove L è una opportuna matrice di contrasti di parametri, di dimensione $(p+c-1) \times (p+c-1)$ e non singolare. Deve, inoltre, essere verificato che nella nuova parametrizzazione ψ si abbia $\alpha_1' \leq \ldots \leq \alpha_{c-1}'$.

Con ragionamento analogo, Kosmidis (2014b, §7.2) mostra anche che lo stimatore con distorsione ridotta in media, $\hat{\theta}^*$, soddisfa la proprietà di invarianza a seguito dell'inversione dell'ordine delle categorie, nei casi un cui la distribuzione della sottostante variabile latente sia simmetrica (escludendo ad esempio la distribuzione Gumbel e quindi le formulazioni dei modelli con link log-log e log-log complementare). Si ottengono, dunque, le stesse probabilità per una determinata modalità invertendo l'ordine delle categorie e simultaneamente cambiando i segni dei parametri di regressione e il segno e l'ordine dei parametri di intercetta. Sia lo stimatore di massima verosimiglianza che lo stimatore con distorsione ridotta in media rispettano questa proprietà di equivarianza, dal momento che risultano equivarianti per trasformazioni lineari.

Una proprietà di invarianza del modello link cumulato è l'invarianza rispetto alla scelta delle modalità della risposta. Si perviene alle stesse conclusioni inferenziali sia che si consideri un modello link cumulato per una variabile risposta con c categorie sia che si conduca l'analisi accorpando una delle categorie con quella adiacente e quindi riducendo le modalità della variabile risposta di una unità, e sia in caso di raffinamento delle categorie. Tale proprietà ha una sua utilità pratica quando si vogliono confrontare le stime dei parametri che provengono da studi che usano scale differenti per la variabile risposta.

Capitolo 3

Riduzione della distorsione in mediana nei modelli link cumulati

L'obiettivo di questo capitolo è di ottenere lo stimatore con distorsione ridotta in mediana per i modelli link cumulati nella formulazione (2.1), secondo l'approccio proposto da Kenne Pagui *et al.* (2017) . Vengono inoltre fornite le quantità necessarie per l'implementazione relativa alle diverse funzioni di legame e, infine, sono affrontati gli aspetti di natura computazionale. Il presente capitolo è direttamente collegato all'Appendice A, dove è presente la traduzione in ambiente R delle formule matematiche e degli accorgimenti computazionali.

Nel §3.1 vengono esplicitate le principali quantità di verosimiglianza e i momenti nulli coinvolti nell'implementazione dello stimatore con distorsione ridotta in mediana. Il §3.2 presenta le quantità da sostituire nelle relazioni presentate al §3.1 in relazione alla funzione di legame scelta (logit, probit, log-log complementare). Il §3.3 analizza gli aspetti prettamente computazionali quali la scelta dei valori iniziali per l'algoritmo iterativo, l'utilizzo della tecnica dello *step halving* e alcuni problemi di instabilità numerica riscontrati.

I principali riferimenti bibliografici relativi al §3.3 sono Christensen (2019b), in merito alla discussione sulla scelta dei valori iniziali, Wood (2015), riferimento della tecnica dello *step-halving*, Christensen (2019a) e Kosmidis (2014b, Supplementary material), relativamente alle funzioni in ambiente R disponibili per il calcolo delle stime di massima verosimiglianza e delle stime con disorsione ridotta in media.

3.1 Quantità di verosimiglianza e momenti nulli

Si considera il modello link cumulati definito in (2.1), con vettore dei parametri $\theta = (\alpha_1, \ldots, \alpha_{c-1}, \beta_1, \ldots, \beta_p)$, con la condizione $-\infty = \alpha_0 \leq \alpha_1 \leq \ldots \leq \alpha_{c-1} \leq \alpha_c = +\infty$. Si indica il predittore lineare con $\eta_{ij} = \alpha_j + \mathbf{x}_i\beta$, $i = 1, \ldots, n$, $j = 1, \ldots, c-1$, e sia $\pi_{ij} = P(Y_i = j|\mathbf{x}_i) = F(\eta_{ij}) - F(\eta_{ij-1})$, con $F(\cdot)$ funzione di di ripartizione continua su $\mathbb{R}$. Con $y_{ij} = 1$ si indica se si è osservata la modalità j-esima, $j = 1, \ldots, c$, per l'i-esimo soggetto, e $y_{ij} = 0$ altrimenti. Si indicheranno con k, l, m gli indici che scorrono i parametri di intercetta e, salvo diversa specificazione, $k, l, m \in \{1, \ldots, c-1\}$. Con $r, s, t \in \{1, \ldots, p\}$ si denotano gli indici che scorrono i parametri di regressione β.

La funzione di verosimiglianza per θ e il suo logaritmo sono

$$\mathcal{L}(\theta) = \prod_{i=1}^{n}\prod_{j=1}^{c} P(Y_i = j|\mathbf{x}_i)^{y_{ij}} = \prod_{i=1}^{n}\prod_{j=1}^{c}[F(\eta_{ij}) - F(\eta_{ij-1})]^{y_{ij}},$$

$$\ell(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{c} y_{ij}\log[F(\eta_{ij}) - F(\eta_{ij-1})].$$

Gli elementi del vettore score, $U(\theta) = (U_{\alpha_1}, \ldots, U_{\alpha_{c-1}}, U_{\beta_1}, \ldots, U_{\beta_p})$, risultano

$$U_{\alpha_k} = \sum_{i=1}^{n} y_{ik}A(\eta_{ik}, \eta_{ik-1}) - y_{ik+1}B(\eta_{ik}, \eta_{ik+1}), \qquad k = 1, \ldots, c-1,$$

$$U_{\beta_r} = \sum_{i=1}^{n}\sum_{j=1}^{c} x_{ir}y_{ij}C(\eta_{ij}, \eta_{ij-1}), \qquad r = 1, \ldots, p,$$

dove

$$A(\eta_{ik}, \eta_{ik-1}) = \frac{f(\eta_{ik})}{\pi_{ik}},$$

$$B(\eta_{ik}, \eta_{ik+1}) = \frac{f(\eta_{ik})}{\pi_{ik+1}},$$
$$C(\eta_{ij}, \eta_{ij-1}) = \frac{f(\eta_{ij}) - f(\eta_{ij-1})}{\pi_{ij}},$$

avendo indicato con $f(\cdot)$ la funzione di densità corrispondente a $F(\cdot)$.

La matrice hessiana della funzione di log-verosimiglianza, simmetrica e di dimensione $p + c - 1$, è una matrice tridiagonale con elementi

$$U_{\alpha_k\alpha_k} = \sum_{i=1}^{n} y_{ik} D(\eta_{ik}, \eta_{ik-1}) - y_{ik+1} G(\eta_{ik}, \eta_{ik+1}),$$
$$U_{\alpha_k\alpha_{k+1}} = \sum_{i=1}^{n} y_{ik+1} H(\eta_{ik}, \eta_{ik+1}), \qquad k = 1, \ldots, c-2,$$
$$U_{\alpha_k\alpha_l} = 0, \qquad l \notin \{k-1, k, k+1\},$$
$$U_{\beta_r\alpha_k} = \sum_{i=1}^{n} x_{ir}[y_{ik} I(\eta_{ik}, \eta_{ik-1}) - y_{ik+1} J(\eta_{ik}, \eta_{ik+1})],$$
$$U_{\beta_r\beta_s} = \sum_{i=1}^{n} x_{ir} x_{is} y_{ij} K(\eta_{ij}, \eta_{ij-1}),$$

dove

$$D(\eta_{ik}, \eta_{ik-1}) = \frac{f'(\eta_{ik})}{\pi_{ik}} - \frac{f(\eta_{ik})^2}{\pi_{ik}^2},$$
$$G(\eta_{ik}, \eta_{ik+1}) = \frac{f'(\eta_{ik})}{\pi_{ik+1}} + \frac{f(\eta_{ik})^2}{\pi_{ik+1}^2},$$
$$H(\eta_{ik}, \eta_{ik+1}) = \frac{f(\eta_{ik}) f(\eta_{ik+1})}{\pi_{ik+1}^2},$$
$$I(\eta_{ik}, \eta_{ik-1}) = \frac{f'(\eta_{ik})}{\pi_{ik}} - \frac{f(\eta_{ik})[f(\eta_{ik}) - f(\eta_{ik-1})]}{\pi_{ik}^2},$$
$$J(\eta_{ik}, \eta_{ik+1}) = \frac{f'(\eta_{ik})}{\pi_{ik+1}} - \frac{f(\eta_{ik})[f(\eta_{ik+1}) - f(\eta_{ik})]}{\pi_{ik+1}^2},$$
$$K(\eta_{ij}.\eta_{ij-1}) = \frac{f'(\eta_{ij}) - f'(\eta_{ij-1})}{\pi_{ij}} - \frac{[f(\eta_{ij}) - f(\eta_{ij-1})]^2}{\pi_{ij}^2},$$

avendo indicato con $f'(\cdot)$ la derivata prima della funzione di densità $f(\cdot)$.

Dal momento che $E[Y_{ij}] = \pi_{ij}$ le componenti della matrice di informazione attesa di Fisher risultano

$$
\begin{aligned}
i_{\alpha_k\alpha_k} &= -\sum_{i=1}^{n} \pi_{ik} D(\eta_{ik}, \eta_{ik-1}) - \pi_{ik+1} G(\eta_{ik}, \eta_{ik+1}),\\
i_{\alpha_k\alpha_{k+1}} &= -\sum_{i=1}^{n} \pi_{ik+1} H(\eta_{ik}, \eta_{ik+1}), \qquad k = 1, \ldots, c-2,\\
i_{\alpha_k\alpha_l} &= 0, \qquad l \notin \{k-1, k, k+1\},\\
i_{\alpha_k\beta_r} &= -\sum_{i=1}^{n} x_{ir}[\pi_{ik} I(\eta_{ik}, \eta_{ik-1}) - \pi_{ik+1} J(\eta_{ik}, \eta_{ik+1})],\\
i_{\beta_r\beta_s} &= -\sum_{i=1}^{n}\sum_{j=1}^{c} x_{ir} x_{is} \pi_{ij} K(\eta_{ij}, \eta_{ij-1}).
\end{aligned}
$$

La relazione di Bartlett $\nu_{r,s} + \nu_{rs} = 0$ determina

$$
\begin{aligned}
\nu_{\alpha_k,\alpha_k} &= i_{\alpha_k\alpha_k},\\
\nu_{\alpha_k,\alpha_{k+1}} &= i_{\alpha_k\alpha_{k+1}}, \qquad k = 1, \ldots, c-2,\\
\nu_{\alpha_k,\beta_r} &= i_{\alpha_k\beta_r},\\
\nu_{\beta_r,\beta_s} &= i_{\beta_r\beta_s}.
\end{aligned}
$$

Per il calcolo dei momenti nulli della tipologia $\nu_{r,st}$ e $\nu_{r,s,t}$ si farà uso di

$$
E[Y_{ij}Y_{i'k}] = \begin{cases} 0 & i = i', j \neq k \\ \pi_{ik} & i = i', j = k \\ \pi_{ij}\pi_{i'k} & i \neq i'. \end{cases}
$$

I momenti nulli della tipologia $\nu_{r,st}$ risultano

$$\nu_{\alpha_k,\alpha_l\alpha_m} = E_\theta[U_{\alpha_k}U_{\alpha_l\alpha_m}] = \begin{cases} \nu_{\alpha_k,\alpha_k\alpha_k} & \\ \nu_{\alpha_k,\alpha_k\alpha_{k+1}} & \\ \nu_{\alpha_k,\alpha_{k+1}\alpha_{k+1}} & k = 1,\dots,c-2 \\ \nu_{\alpha_{k+1},\alpha_k\alpha_k} & k = 1,\dots,c-2 \\ \nu_{\alpha_{k+1},\alpha_k\alpha_{k+1}} & k = 1,\dots,c-2 \\ \nu_{\alpha_k,\alpha_l\alpha_l} & l \notin \{k-1,k,k+1\} \\ \nu_{\alpha_k,\alpha_l\alpha_{l+1}} & l \notin \{k-1,k\}, \end{cases}$$

$$\begin{aligned}
\nu_{\alpha_k,\alpha_k\alpha_k} &= \sum_{i=1}^n \pi_{ik}A(\eta_{ik},\eta_{ik-1})D(\eta_{ik},\eta_{ik-1}) + \pi_{ik+1}B(\eta_{ik},\eta_{ik+1})G(\eta_{ik},\eta_{ik+1}),\\
\nu_{\alpha_k,\alpha_k\alpha_{k+1}} &= -\sum_{i=1}^n \pi_{ik+1}B(\eta_{ik},\eta_{ik+1})H(\eta_{ik},\eta_{ik+1}),\\
\nu_{\alpha_k,\alpha_{k+1}\alpha_{k+1}} &= -\sum_{i=1}^n \pi_{ik+1}B(\eta_{ik},\eta_{ik+1})D(\eta_{ik+1},\eta_{ik}),\\
\nu_{\alpha_{k+1},\alpha_k\alpha_k} &= -\sum_{i=1}^n \pi_{ik+1}A(\eta_{ik+1},\eta_{ik})G(\eta_{ik},\eta_{ik+1}),\\
\nu_{\alpha_{k+1},\alpha_k\alpha_{k+1}} &= \sum_{i=1}^n \pi_{ik+1}A(\eta_{ik+1},\eta_{ik})H(\eta_{ik},\eta_{ik+1}),\\
\nu_{\alpha_k,\alpha_l\alpha_l} &= \nu_{\alpha_k,\alpha_l\alpha_l+1} = 0.
\end{aligned}$$

$$\nu_{\alpha_k,\alpha_l\beta_r} = E_\theta[U_{\alpha_k}U_{\alpha_l\beta_r}] = \begin{cases} \nu_{\alpha_k,\alpha_k\beta_r} & \\ \nu_{\alpha_k,\alpha_{k+1}\beta_r} & k = 1,\dots,c-2 \\ \nu_{\alpha_{k+1},\alpha_k\beta_r} & k = 1,\dots,c-2 \\ \nu_{\alpha_k,\alpha_l\beta_r} & l \notin \{k-1,k,k+1\}, \end{cases}$$

$$
\begin{aligned}
\nu_{\alpha_k,\alpha_k\beta_r} &= \sum_{i=1}^{n} x_{ir}[\pi_{ik}A(\eta_{ik},\eta_{ik-1})I(\eta_{ik},\eta_{ik-1}) + \pi_{ik+1}B(\eta_{ik},\eta_{ik+1})J(\eta_{ik},\eta_{ik+1})],\\
\nu_{\alpha_k,\alpha_{k+1}\beta_r} &= -\sum_{i=1}^{n} x_{ir}\pi_{ik+1}B(\eta_{ik},\eta_{ik+1})I(\eta_{ik+1},\eta_{ik}),\\
\nu_{\alpha_{k+1},\alpha_k\beta_r} &= -\sum_{i=1}^{n} x_{ir}\pi_{ik+1}A(\eta_{ik+1},\eta_{ik})J(\eta_{ik},\eta_{ik+1}),\\
\nu_{\alpha_k,\alpha_l\beta_r} &= 0.
\end{aligned}
$$

$$
\nu_{\beta_r,\alpha_k\alpha_l} = E_\theta[U_{\beta_r}U_{\alpha_k\alpha_l}] = \begin{cases} \nu_{\beta_r,\alpha_k\alpha_k} & \\ \nu_{\beta_r,\alpha_k\alpha_{k+1}} & k = 1,\dots,c-2 \\ \nu_{\beta_r,\alpha_k\alpha_l} & l \notin \{k-1,k,k+1\}, \end{cases}
$$

$$
\begin{aligned}
\nu_{\beta_r,\alpha_k\alpha_k} &= \sum_{i=1}^{n} x_{ir}[\pi_{ik}C(\eta_{ik},\eta_{ik-1})D(\eta_{ik},\eta_{ik-1}) - \pi_{ik+1}C(\eta_{ik+1},\eta_{ik})G(\eta_{ik},\eta_{ik+1})],\\
\nu_{\beta_r,\alpha_k\alpha_{k+1}} &= \sum_{i=1}^{n} x_{ir}\pi_{ik+1}C(\eta_{ik+1},\eta_{ik})H(\eta_{ik},\eta_{ik+1}),\\
\nu_{\beta_r,\alpha_k\alpha_l} &= 0.
\end{aligned}
$$

$$
\begin{aligned}
\nu_{\beta_r,\alpha_k\beta_s} &= E_\theta[U_{\beta_r}U_{\alpha_k\beta_s}] =\\
&= \sum_{i=1}^{n} x_{ir}x_{is}[\pi_{ik}C(\eta_{ik},\eta_{ik-1})I(\eta_{ik},\eta_{ik-1}) - \pi_{ik+1}C(\eta_{ik+1},\eta_{ik})J(\eta_{ik},\eta_{ik+1})].
\end{aligned}
$$

$$
\begin{aligned}
\nu_{\alpha_k,\beta_r\beta_s} &= E_\theta[U_{\alpha_k}U_{\beta_r\beta_s}] =\\
&= \sum_{i=1}^{n} x_{ir}x_{is}[\pi_{ik}A(\eta_{ik},\eta_{ik-1})K(\eta_{ik},\eta_{ik-1}) - \pi_{ik+1}B(\eta_{ik},\eta_{ik+1})K(\eta_{ik+1},\eta_{ik})].
\end{aligned}
$$

$$
\nu_{\beta_r,\beta_s\beta_t} = E_\theta[U_{\beta_r}U_{\beta_s\beta_t}] = \sum_{i=1}^{n}\sum_{j=1}^{c} x_{ir}x_{is}x_{it}\pi_{ij}C(\eta_{ij},\eta_{ij-1})K(\eta_{ij},\eta_{ij-1}).
$$

Sfruttando le relazioni di Bartlett $\nu_{rst}+\nu_{r,st}+\nu_{s,rt}+\nu_{t,rs}+\nu_{r,s,t}=0$ otteniamo i momenti del tipo $\nu_{r,s,t}$

$$\nu_{\alpha_k,\alpha_l,\alpha_m}=E_\theta[U_{\alpha_k}U_{\alpha_l}U_{\alpha_m}]=\begin{cases}\nu_{\alpha_k,\alpha_k,\alpha_k} & \\ \nu_{\alpha_k,\alpha_k,\alpha_{k+1}} & k=1,\dots,c-2\\ \nu_{\alpha_k,\alpha_{k+1},\alpha_{k+1}} & k=1,\dots,c-2\\ \nu_{\alpha_k,\alpha_l,\alpha_l} & l\notin\{k-1,k,k+1\}\\ \nu_{\alpha_k,\alpha_l,\alpha_{l+1}} & l\notin\{k-1,k\}\\ \nu_{\alpha_k,\alpha_l,\alpha_m} & l\notin\{k-1,k,k+1\}, m\notin\{l-1,l,l+1\},\end{cases}$$

$$\begin{aligned}
\nu_{\alpha_k,\alpha_k,\alpha_k} &= -\nu_{\alpha_k\alpha_k\alpha_k}-3\nu_{\alpha_k,\alpha_k\alpha_k},\\
\nu_{\alpha_k,\alpha_k,\alpha_{k+1}} &= -\nu_{\alpha_k\alpha_k\alpha_{k+1}}-2\nu_{\alpha_k,\alpha_k\alpha_{k+1}}-\nu_{\alpha_{k+1},\alpha_k\alpha_k},\\
\nu_{\alpha_k,\alpha_{k+1},\alpha_{k+1}} &= -\nu_{\alpha_k\alpha_{k+1}\alpha_{k+1}}-2\nu_{\alpha_{k+1},\alpha_k\alpha_{k+1}}-\nu_{\alpha_k,\alpha_{k+1}\alpha_{k+1}},\\
\nu_{\alpha_k,\alpha_l,\alpha_l} &= -\nu_{\alpha_k\alpha_l\alpha_l}-2\nu_{\alpha_l,\alpha_k\alpha_l}-\nu_{\alpha_k,\alpha_l\alpha_l},\\
\nu_{\alpha_k,\alpha_l,\alpha_{l+1}} &= -\nu_{\alpha_k\alpha_l\alpha_{l+1}}-\nu_{\alpha_{l+1},\alpha_k\alpha_l}-\nu_{\alpha_l,\alpha_k\alpha_{l+1}}-\nu_{\alpha_k,\alpha_l\alpha_{l+1}},\\
\nu_{\alpha_k,\alpha_l,\alpha_m} &= -\nu_{\alpha_k\alpha_l\alpha_m}-\nu_{\alpha_k,\alpha_l\alpha_m}-\nu_{\alpha_l,\alpha_k\alpha_m}-\nu_{\alpha_m,\alpha_k\alpha_l}.
\end{aligned}$$

I valori attesi delle derivate terze della log-verosimiglianza risultano

$$\begin{aligned}
\nu_{\alpha_k\alpha_k\alpha_k} &= \sum_{i=1}^n \pi_{ik}L(\eta_{ik},\eta_{ik-1})-\pi_{ik+1}M(\eta_{ik},\eta_{ik+1}),\\
\nu_{\alpha_k\alpha_k\alpha_{k+1}} &= \sum_{i=1}^n \pi_{ik+1}N(\eta_{ik},\eta_{ik+1}),\\
\nu_{\alpha_k\alpha_{k+1}\alpha_{k+1}} &= \sum_{i=1}^n \pi_{ik+1}O(\eta_{ik},\eta_{ik+1}),\\
\nu_{\alpha_k\alpha_l\alpha_l} &= \nu_{\alpha_k\alpha_l\alpha_{l+1}}=\nu_{\alpha_k\alpha_l\alpha_m}=0,
\end{aligned}$$

dove:

$$\begin{aligned}
L(\eta_{ik},\eta_{ik-1}) &= \frac{f''(\eta_{ik})}{\pi_{ik}} - 3\frac{f'(\eta_{ik})f(\eta_{ik})}{\pi_{ik}^2} + 2\frac{f(\eta_{ik})^3}{\pi_{ik}^3},\\
M(\eta_{ik},\eta_{ik+1}) &= \frac{f''(\eta_{ik})}{\pi_{ik+1}} + 3\frac{f'(\eta_{ik})f(\eta_{ik})}{\pi_{ik+1}^2} + 2\frac{f(\eta_{ik})^3}{\pi_{ik+1}^3},\\
N(\eta_{ik},\eta_{ik+1}) &= \frac{f'(\eta_{ik})f(\eta_{ik+1})}{\pi_{ik+1}^2} + 2\frac{f(\eta_{ik})^2 f(\eta_{ik+1})}{\pi_{ik+1}^3},\\
O(\eta_{ik},\eta_{ik+1}) &= \frac{f(\eta_{ik})f'(\eta_{ik+1})}{\pi_{ik+1}^2} - 2\frac{f(\eta_{ik})f(\eta_{ik+1})^2}{\pi_{ik+1}^3},
\end{aligned}$$

avendo indicato con $f''(\cdot)$ la derivata seconda della funzione di densità $f(\cdot)$.

$$\nu_{\alpha_k,\alpha_l,\beta_r} = E_\theta[U_{\alpha_k}U_{\alpha_l}U_{\beta_r}] = \begin{cases} \nu_{\alpha_k,\alpha_k,\beta_r} & \\ \nu_{\alpha_k,\alpha_{k+1},\beta_r} & k=1,\ldots,c-2 \\ \nu_{\alpha_k,\alpha_l,\beta_r} & l \notin \{k-1,k,k+1\}. \end{cases}$$

$$\begin{aligned}
\nu_{\alpha_k,\alpha_k,\beta_r} &= -\nu_{\alpha_k\alpha_k\beta_r} - 2\nu_{\alpha_k,\alpha_k\beta_r} - \nu_{\beta_r,\alpha_k\alpha_k},\\
\nu_{\alpha_k,\alpha_{k+1},\beta_r} &= -\nu_{\alpha_k\alpha_{k+1}\beta_r} - \nu_{\alpha_k,\alpha_{k+1}\beta_r} - \nu_{\beta_r,\alpha_k\alpha_{k+1}} - \nu_{\alpha_{k+1},\alpha_k\beta_r},\\
\nu_{\alpha_k,\alpha_l,\beta_r} &= -\nu_{\alpha_k\alpha_l\beta_r} - \nu_{\alpha_k,\alpha_l\beta_r} - \nu_{\beta_r,\alpha_k\alpha_l} - \nu_{\alpha_l,\alpha_k\beta_r}.
\end{aligned}$$

I valori attesi delle derivate terze della log-verosimiglianza risultano

$$\begin{aligned}
\nu_{\alpha_k\alpha_k\beta_r} &= \sum_{i=1}^n x_{ir}[\pi_{ik}P(\eta_{ik},\eta_{ik-1}) - \pi_{ik+1}Q(\eta_{ik},\eta_{ik+1})],\\
\nu_{\alpha_k\alpha_{k+1}\beta_r} &= \sum_{i=1}^n x_{ir}\pi_{ik+1}R(\eta_{ik},\eta_{ik+1}),\\
\nu_{\alpha_k\alpha_l\beta_r} &= 0,
\end{aligned}$$

dove:

$$P(\eta_{ik},\eta_{ik-1}) = \frac{f''(\eta_{ik})}{\pi_{ik}} - \frac{3f'(\eta_{ik})f(\eta_{ik}) - f'(\eta_{ik})f(\eta_{ik-1})}{\pi_{ik}^2} + 2\frac{f(\eta_{ik})^2[f(\eta_{ik}) - f(\eta_{ik-1})]}{\pi_{ik}^3},$$

$$Q(\eta_{ik},\eta_{ik+1}) = \frac{f''(\eta_{ik})}{\pi_{ik+1}} + \frac{3f'(\eta_{ik})f(\eta_{ik}) - f'(\eta_{ik})f(\eta_{ik+1})}{\pi_{ik+1}^2} - 2\frac{f(\eta_{ik})^2[f(\eta_{ik+1}) - f(\eta_{ik})]}{\pi_{ik+1}^3},$$

$$R(\eta_{ik},\eta_{ik+1}) = \frac{f'(\eta_{ik})f(\eta_{ik+1})}{\pi_{ik+1}^2} + \frac{f'(\eta_{ik+1})f(\eta_{ik})}{\pi_{ik+1}^2} - 2\frac{f(\eta_{ik})f(\eta_{ik+1})[f(\eta_{ik+1}) - f(\eta_{ik})]}{\pi_{ik+1}^3}.$$

$$\nu_{\alpha_k,\beta_r,\beta_s} = E_\theta[U_{\alpha_k}U_{\beta_r}U_{\beta_s}] = -\nu_{\alpha_k\beta_r\beta_s} - \nu_{\alpha_k,\beta_r\beta_s} - 2\nu_{\beta_r,\alpha_k\beta_s}.$$

Il valore atteso della derivata terza della log verosimiglianza risulta

$$\nu_{\alpha_k\beta_r\beta_s} = \sum_{i=1}^n x_{ir}x_{is}[\pi_{ik}S(\eta_{ik},\eta_{ik-1}) - \pi_{ik+1}T(\eta_{ik},\eta_{ik+1})],$$

dove

$$\begin{aligned} S(\eta_{ik},\eta_{ik-1}) &= \frac{f''(\eta_{ik})}{\pi_{ik}} - \frac{[f'(\eta_{ik}) - f'(\eta_{ik-1})]f(\eta_{ik})}{\pi_{ik}^2} \\ &- 2\frac{[f(\eta_{ik}) - f(\eta_{ik-1})]f'(\eta_{ik})}{\pi_{ik}^2} + 2\frac{[f(\eta_{ik}) - f(\eta_{ik-1})]^2 f(\eta_{ik})}{\pi_{ik}^3}, \end{aligned}$$

$$\begin{aligned} T(\eta_{ik},\eta_{ik+1}) &= \frac{f''(\eta_{ik})}{\pi_{ik+1}} - \frac{[f'(\eta_{ik+1}) - f'(\eta_{ik})]f(\eta_{ik})}{\pi_{ik+1}^2} \\ &- 2\frac{[f(\eta_{ik+1}) - f(\eta_{ik})]f'(\eta_{ik})}{\pi_{ik+1}^2} + 2\frac{[f(\eta_{ik+1}) - f(\eta_{ik})]^2 f(\eta_{ik})}{\pi_{ik+1}^3}. \end{aligned}$$

$$\nu_{\beta_r,\beta_s,\beta_t} = E_\theta[U_{\beta_r}U_{\beta_s}U_{\beta_t}] = -\nu_{\beta_r\beta_s\beta_t} - [3]\nu_{\beta_r,\beta_s\beta_t} = -\nu_{\beta_r\beta_s\beta_t} - 3\nu_{\beta_r,\beta_s\beta_t}.$$

Il valore atteso della derivata terza della log verosimiglianza risulta

$$\nu_{\beta_r\beta_s\beta_t} = \sum_{i=1}^n \sum_{j=1}^c x_{ir}x_{is}x_{it}\pi_{ij}U(\eta_{ij},\eta_{ij-1}),$$

dove

$$U(\eta_{ij}, \eta_{ij-1}) = \frac{f''(\eta_{ij}) - f''(\eta_{ij-1})}{\pi_{ij}} - 3\frac{[f'(\eta_{ij}) - f'(\eta_{ij-1})][f(\eta_{ij}) - f(\eta_{ij-1})]}{\pi_{ij}^2}$$
$$+ 2\frac{[f(\eta_{ij}) - f(\eta_{ij-1})]^3}{\pi_{ij}^3}.$$

3.2 Funzioni di legame

Nelle quantità calcolate nel §3.1 sono state lasciate non specificate la funzione di ripartizione $F(\cdot)$, la funzione di densità $f(\cdot)$ e le sue derivate prima, $f'(\cdot)$, e seconda $f''(\cdot)$. In virtù dell'interpretazione del modello link cumulato in termini di variabile latente continua sottostante, discusso in §2.2.3, a seconda della distribuzione del termine di errore sussistono le seguenti relazioni per le funzioni di legame:

- legame logit

$$\begin{aligned} F(\eta_{ij}) &= \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}}, \\ f(\eta_{ij}) &= \frac{e^{\eta_{ij}}}{(1 + e^{\eta_{ij}})^2} = F(\eta_{ij})[1 - F(\eta_{ij})], \\ f'(\eta_{ij}) &= f(\eta_{ij})[1 - 2F(\eta_{ij})], \\ f''(\eta_{ij}) &= f(\eta_{ij})[1 - 2F(\eta_{ij})]^2 - 2f(\eta_{ij})^2; \end{aligned}$$

- legame probit

$$\begin{aligned} F(\eta_{ij}) &= \Phi(\eta_{ij}), \\ f(\eta_{ij}) &= \phi(\eta_{ij}), \\ f'(\eta_{ij}) &= -\eta_{ij}\phi(\eta_{ij}), \\ f''(\eta_{ij}) &= \phi(\eta_{ij})(\eta_{ij}^2 - 1), \end{aligned}$$

  dove $\Phi(\eta_{ij})$ e $\phi(\eta_{ij})$ sono rispettivamente le funzioni di ripartizione e di densità della funzione normale standard;

- legame log-log complementare

$$F(\eta_{ij}) = 1 - e^{-e^{\eta_{ij}}},$$
$$f(\eta_{ij}) = [F(\eta_{ij}) - 1] \log[1 - F(\eta_{ij})],$$
$$f'(\eta_{ij}) = f(\eta_{ij})\{1 + \log[1 - F(\eta_{ij})]\},$$
$$f''(\eta_{ij}) = f'(\eta_{ij})\{1 + \log[1 - F(\eta_{ij})]\} - \frac{f(\eta_{ij})^2}{1 - F(\eta_{ij})}.$$

3.3 Aspetti computazionali

L'esistenza di una soluzione dell'equazione di verosimiglianza $U(\theta) = 0$, così come quella delle equazioni di stima basate su funzioni di punteggio modificate, $U^*(\theta) = 0$ e $\tilde{U}(\theta) = 0$, non è garantita. La mancanza di una soluzione in forma chiusa di tali equazioni di stima determina il ricorso ad un algoritmo di calcolo numerico iterativo e nello specifico è stato utlizzato l'algoritmo Fisher-scoring. Tale algoritmo, nella formulazione relativa allo stimatore che riduce la distorsione in mediana, risulta esprimibile come

$$\theta^{(k+1)} = \theta^{(k)} + i(\theta^{(k)})^{-1}\tilde{A}(\theta^{(k)}) + i(\theta^{(k)})^{-1}U(\theta^{(k)})$$

con $\tilde{A}(\theta)$ definito in (1.24). È stato adottato come criterio di arresto dell'algoritmo la condizione che il valore assoluto di tutte le componenti del vettore score, $|U_r(\theta^{(k)})|$, $r = 1, \ldots, p$, fosse inferiore ad un valore prefissato, e, come impostazione predefinita, è stata scelta come soglia 10^{-10}.

L'algoritmo necessita di un insieme di valori iniziali, $\theta^{(0)}$. Tale scelta risulta non banale e può comportare la mancata convergenza dell'algoritmo o problemi di instabilità numerica, derivanti dalla violazione della condizione $\alpha_1 \leq \ldots \leq \alpha_{c-1}$. Ottimi valori iniziali sono le stime di massima verisimiglianza, ove disponibili, e le stime ridotte in distorsione secondo il metodo di Firth (1993). La scelta dei valori iniziali per ottenere le stime di massima verosimiglianza proposta da Christensen (2019b, §3.1) prevede l'inizializzazione dei parametri di regressione β al valore zero e i parametri relativi alle intercette α ai quantili della distribuzione logistica, in modo che la differenza della funzione di ripartizione della distribuzione logistica tra due parametri di soglia sia la stessa. Nell'algoritmo riportato in Appendice tale scelta per i parametri relativi alle intercette è stata adottata soltanto per la funzione di legame logit mentre per le funzioni di legame probit e log-log complementare sono stati usati rispettivamente i quantili della distribuzione normale e i quantili della distribuzione della variabile casuale ottenuta dalla variabile casuale con distribuzione Gumbel cambiata di segno. Questa scelta dei valori iniziali si è resa necessaria dall'osservazione delle iterazioni dell'algoritmo che mostravano, in alcuni casi, un mancato rispetto della condizione $\alpha_1 \leq \ldots \leq \alpha_{c-1}$, ad esempio quando i quantili della distribuzione della distribuzione logistica venivano usati con la scelta della funzione di legame probit. Queste scelte dei valori iniziali sono risultate utili per ottenere le stime di massima verosimiglianza e le stime con distorsione ridotta in media, quando le stime di massima verosimiglianza risultavano sulla frontiera dello spazio parametrico. Per quanto discusso nel §1.2.2 al fine di ottenere le stime con distorsione ridotta in mediana sono stati usati come valori iniziali dell'algoritmo le stime con distorsione ridotta in media. Queste ultime considerazioni determinano anche un'accelerazione della convergenza, come evidenziato nella Tabella 3.1 per un generico insieme di dati e assumendo la funzione di legame logit, a seconda dei valori iniziali scelti e per tre diversi valori per il criterio di arresto: 10^{-6} (Iterazioni (1)), 10^{-8} (Iterazioni (2)) e 10^{-10} (Iterazioni (3)).

Stime	Valori iniziali	Iterazioni (1)	Iterazioni (2)	Iterazioni (3)
$\hat{\theta}$	$\theta^{(0)}$	8	10	12
$\hat{\theta}^*$	$\theta^{(0)}$	8	10	11
	$\hat{\theta}$	6	8	11
$\tilde{\theta}$	$\theta^{(0)}$	8	10	11
	$\hat{\theta}$	5	7	10
	$\hat{\theta}^*$	5	7	9

Tabella 3.1: Confronto numero iterazioni per diversi valori iniziali

Nell'implementazione dell'algoritmo Fisher-scoring è stata adottata una semplice modifica, presente in letteratura, che permette di avere maggior controllo sull'algoritmo stesso. Questa modifica ha come obiettivo quello di evitare che l'algoritmo compia passi di ottimizzazione troppo ampi che possono essere causa di problemi di instabilità numerica. Essa prende il nome di *step-halving* (si veda ad esempio Wood, 2015, §5.1.1). In caso di convergenza verso le stime deve verificarsi che $|\theta^{(k+1)} - \theta^{(k)}| < |\theta^{(k)} - \theta^{(k-1)}|$ e in caso contrario si dovrebbe riscontrare $|\theta^{(k+1)} - \theta^{(k)}| > |\theta^{(k)} - \theta^{(k-1)}|$. A tal proposito in caso di passo di aggiornamento troppo ampio si interviene proponendo un passo di aggiornamento meno ampio in grado di rispettare $|\theta^{(k+1)} - \theta^{(k)}| < |\tilde{\theta}^{(k)} - \tilde{\theta}^{(k-1)}|$ e ciò si ottiene mediante una penalizzazione esponenziale esprimibile come

$$\theta^{(k+1)} = \theta^{(k)} + \frac{1}{2^w}[i^{-1}(\theta^{(k)})\tilde{A}(\theta^{(k)}) + i^{-1}(\theta^{(k)})U(\theta^{(k)})],$$

dove $w \in \mathbb{N}$. Risulta che $w = 0$ se $|\theta^{(k+1)} - \theta^{(k)}| < |\ \theta^{(k)} - \theta^{(k-1)}|$ e la modifica ha effetto nullo. Nel caso in cui $|\theta^{(k+1)} - \theta^{(k)}| > |\theta^{(k)} - \theta^{(k-1)}|$ il passo viene premoltiplicato per $1/2, 1/4, \ldots$, fino al verificarsi della condizione $|\theta^{(k+1)} - \theta^{(k)}| < |\theta^{(k)} - \theta^{(k-1)}|$. Nonostante si sia posto $w \in \mathbb{N}$, è bene terminare la penalizzazione con $w = 10$, che rappresenta una notevole penalizzazione, e al suo raggiungimento si accetta l'ultimo valore e si procede con il passo successivo dell'algoritmo Fisher-scoring.

Come nell'algoritmo Newton-Raphson standard, la scelta dei valori iniziali per i parametri è cruciale affinchè l'algoritmo raggiunga convergenza in tempi ragionevoli. Qualche problema sorge ad esempio in caso di stime di massima verosimiglianza sulla frontiera dello spazio parametrico (§2.4, (i)) o in caso di approssimazioni numeriche dovute al limite di precisione del calcolatore. A tal proposito, nell'algoritmo in Appendice, si riporta una modifica che cerca di rimediare ai problemi di precisione del calcolatore. Negli studi di simulazione l'osservazione empirica di probabilità stimate pari a zero, legate ad approssimazioni numeriche, determinava la presenza di forme indeterminate nelle quantità utili per il calcolo delle stime e ciò risultava in una impossibilità di procedere con le successive iterazioni dell'algoritmo. La modifica effettuata prevede la sostituzione di tali valori pari a zero nelle probabilità stimate con un valore piccolo, dell'ordine di 10^{-16}. Questo artificio è riuscito a risolvere i problemi di instabilità numerica riscontrati in molte situazioni. Tale modifica, dove risultata efficace, ha permesso di superare i problemi legati all'invertibilità della matrice di informazione di Fisher e, nei casi di stime di massima verosimiglianza finite, esse risultano identiche a quelle ottenute con le funzioni `clm` (pacchetto R `ordinal`, Christensen, 2019a) e `bpolr` (Kosmidis, 2017b, Supplementary material). In alcune situazioni si sono riscontrati problemi di convergenza con la funzione `bpolr` non riscontrati con la funzione `clm`, e si sospetta che ciò sia dovuto a una non corretta scelta del vettore di valori iniziali. Nei casi in cui le funzioni `clm` e `bpolr` risultavano in una situazione di non inversione rispettivamente della matrice di informazione osservata e di quella attesa, la modifica effettuata ha permesso di aggirare il problema di inversione fornendo stime e/o standard error che per la diagnostica di Lesaffre e Albert (1989), menzionata in §2.4, risultano sulla frontiera dello spazio parametrico. Possibili soluzioni per evitare l'artificio usato potrebbero essere la trascrizione del codice generato in altri linguaggi di programmazione, che rendono disponibile la dichiarazione di valori con precisione maggiore (si può pensare alla traduzione di tale algoritmo in C++ con il pacchetto `Rcpp` di R), e/o la riformulazione algebrica delle quantità coinvolte nell'algoritmo.

Capitolo 4

Applicazione

Il presente capitolo ha come obiettivo quello di rivisitare, attraverso l'approccio di stima che risulta nello stimatore con distorsione ridotta in mediana, un'analisi statistica condotta con il metodo di stima di massima verosimiglianza e adottando il modello logit cumulati (2.3).

L'insieme di dati in esame è tratto, parzialmente, dai dati raccolti in uno studio svoltosi in Italia (Decensi *et al.*, 1996). L'obiettivo dello studio era di valutare l'effetto del tamoxifene, farmaco utilizzato negli anni '90 come adiuvante della terapia chirurgica in pazienti donne affette da cancro della mammella, sulla proliferazione del cancro all'endometrio. La somministrazione di 20mg di tamoxifene ha portato all'osservazione di una crescita nell'incidenza del cancro dell'endometrio e della mortalità ad esso collegata, ragione per cui furono condotti studi volti ad individuare marcatori del rischio di proliferazione del cancro all'endometrio.

In tale studio, condotto tra il 1992 e il 1994, 33 donne con cancro della mammella, nello stadio I o II, che avevano ricevuto il tamoxifene, sono state sottoposte a biopsia dell'endometrio. La difficoltà, anche per ragioni etiche, di trovare un gruppo di controllo formato da donne con cancro della mammella che non avessero ricevuto un trattamento ha determinato l'arruolamento di 37 donne in salute, purchè non riceventi nessuna terapia ormonale. Le donne appartenenenti al gruppo di controllo sono state sottoposte a biopsia dell'endometrio poichè presentavano segni quali il sanguinamento uterino anormale.

Le misurazioni della citometria a flusso del DNA hanno permesso di estrapolare la frazione iperdiploide (*hyperdiploid fraction*, HDF), frazione delle cellule in determinate fasi del ciclo cellulare, che viene considerata come indice di proliferazione del cancro all'endometrio. La frazione iperdiploide costituisce dunque, previa sua discretizzazione, la variabile risposta oggetto dell'analisi. Per ulteriori dettagli relativi agli aspetti di natura medico-biologica e del disegno dello studio si rimanda a Decensi *et al.* (1996).

L'obiettivo dell'analisi è valutare la relazione tra la frazione iperdiploide, HDF, e un insieme di variabili esplicative di interesse, quali la durata del trattamento a base di tamoxifene e lo stato di estensione temporale della menopausa, tenuto conto dell'età, dell'indice di massa corporea e del sanguimento uterino anormale. Nove delle settanta donne arruolate nello studio sono state escluse per una impossibilità di valutazione dell'HDF.

La distribuzione della variabile risposta, HDF, di natura continua, è stata resa di natura categoriale ordinale con quattro modalità, utilizzando i quartili della distribuzione empirica ($Q1 = 7.70$, $Q2 = 10$ e $Q3 = 16$). Si riportano nella Tabella 4.1 le frequenze osservate delle modalità della variabile risposta (proliferazione: 1=Bassa, 2=Medio-bassa, 3=Medio-alta, 4=Alta).

Modalità	1	2	3	4
Frequenze	15	17	14	15

Tabella 4.1: Frequenze osservate della variabile risposta

La codifica delle variabili esplicative qualitative nominali e delle variabili continue categorizzate è analoga a quella riportata nello studio originale (Decensi *et al.*, 1996). La durata di somministrazione del tamoxifene (DTM) è stata considerata una variabile con tre livelli (0=Nessuna somministrazione del farmaco, 1= durata della somministrazione inferiore a 36 mesi, 2=durata della somministrazione superiore a 36 mesi); l'estensione temporale della menopausa (DMN) è stata categorizzata in tre livelli (0=assenza di menopausa, 1= presenza di menopausa da meno di 5 anni, 2= presenza di menopausa da più di 5 anni); risultano dicotomizzate l'età (0= età inferiore a 53 anni, 1= età superiore a 53 anni) e l'indice di massa corporea (0=inferiore a 24.6 kg/m^2, 1=superiore a 24.6 kg/m^2); la presenza di sanguinamento uterino anormale è caratteristica dicotomica (0=assenza, 1=presenza).

I dati vengono modellati con il modello logit cumulati

$$\begin{aligned}\text{logit}[P(Y_i \leq j|\mathbf{x_i})] =& \alpha_j + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \\ & \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7}, \quad i = 1, \ldots, 61, \ j = 1, \ldots, 3,\end{aligned}$$

dove x_{i1} e x_{i2} sono le variabili indicatrici della durata di somministrazione del tamoxifene, con livello di riferimento costituito dalla non somministrazione del tamoxifene; x_{i3} e x_{i4} rappresentano le variabili indicatrici dello stato di estensione temporale della menopausa, con livello di riferimento la fase precedente alla menopausa; l'età, l'indice di massa corporea e la presenza di sanguinamento uterino anormale sono denotate rispettivamente con x_{i5}, x_{i6} e x_{i7} e ognuna di esse ha livello di riferimento rappresentato dall'etichetta 0.

Si riportano nella Tabella 4.2 le stime di massima verosimiglianza e nella Tabella 4.3 le stime con distorsione ridotta in mediana e i relativi standard error.

Parametro	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Stime	-2.80	-3.34	0.09	2.25	-0.68	-0.27	-1.58
Std. Error	1.11	1.22	0.66	0.87	0.61	0.50	1.05

Tabella 4.2: Stime di massima verosimiglianza dei coefficienti di regressione

Parametro	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Stime	-2.55	-3.04	0.06	2.04	-0.63	-0.25	-1.47
Std. Error	1.09	1.19	0.66	0.86	0.61	0.50	1.03

Tabella 4.3: Stime con distorsione ridotta in mediana dei coefficienti di regressione

Il modello con effetti separati per ogni logit cumulato è stato stimato con l'approccio di stima di massima verosimiglianza mediante le funzioni `clm` e `bpolr`, riscontrando la presenza di stime sulla frontiera dello spazio parametrico. Si possono condurre, tuttavia, test per verificare l'allontanamento dall'assunzione di quote proporzionali, quali il test del rapporto di verosimiglianza e lo score test. Tuttavia, riprendendo quanto discusso nel §2.2.5, data la presenza di un buon numero di variabili esplicative, si preferisce usare il modello più parsimonioso per ottenere una maggiore interpretabilità dell'effetto delle esplicative sulla variabile risposta.

I rapporti delle quote cumulate per le variabili di interesse, durata della somministrazione del tamoxifene ed estensione dello stato temporale della menopausa, aggiustate per l'età, l'indice di massa corporea e il sanguinamento uterino anormale sono riportate nella Tabella 4.4 con i rispettivi intervalli di confidenza di Wald al 95%. L'analisi delle Tabelle 4.2-4.4 evidenzia come la durata di somministrazione del tamoxifene abbia un effetto significativo nell'aumento della proliferazione delle cellule cancerogene endometriali e come l'estensione temporale dello stato di menopausa abbia invece un effetto protettivo contro il rischio di proliferazione. Le conclusioni che si ottengono stimando il modello con l'approccio di stima di massima verosimiglianza e con l'approccio di stima risultante nello stimatore con distorsione ridotta in mediana sono identiche da un punto di vista inferenziale. Tuttavia risulta evidente l'effetto di *shrinkage* delle stime parametri di regressione ottenuto con lo stimatore con distorsione ridotta in mediana e ciò determina, anche se non di molto, uno *shrinkage* delle stime dei rapporti delle quote cumulate verso 1.

Stima	Massima verosiglianza		Distorsione ridotta in mediana	
Variabile	*COR*	95% CI	*COR*	Wald CI 95%
DTM				
0	1.000	·	1.000	·
$\leq$ 36 mesi	0.061	(0.007-0.540)	0.082	(0.010-0.684)
$>$ 36 mesi	0.040	(0.003-0.390)	0.050	(0.005-0.516)
DMN				
assenza	1.000	·	1.000	·
$\leq$ 5 anni	1.092	(0.297-4.010)	1.060	(0.290-3.873)
$>$ 5 anni	9.503	(1.727-52.292)	7.543	(1.410-40.344)

Tabella 4.4: Stime dei rapporti delle quote cumulate (COR) aggiustate per età, indice di massa corporea e sanguinamento uterino anormale, con relativi intervalli di confidenza di Wald.

Capitolo 5

Studi di simulazione

In questo capitolo si presentano alcuni studi di simulazione per la verifica delle proprietà distributive dello stimatore di massima verosimiglianza, $\hat{\theta}$, e degli stimatori con distorsione ridotta in media, $\hat{\theta}^*$, e in mediana, $\tilde{\theta}$, relativi al modello link cumulati nella formulazione (2.1). I metodi di stima verranno valutati in funzione di quattro indici: il valore assoluto della distorsione stimata (|Bias|), la probabilità empirica di sottostima del vero valore del parametro (PU%), la radice dell'errore quadratico medio (RMSE) e la copertura empirica degli intervalli di Wald con livello di confidenza al 95% (Wald). I risultati sono stati ottenuti con 10000 simulazioni e si presentano i soli risultati relativi ai coefficienti di regressione del modello tralasciando la valutazione delle proprietà distributive dei parametri relativi alle intercette.

Il §4.1 presenta i risultati dello studio di simulazione ottenuto partendo da insiemi di dati frutto di generazione casuale e facendo variare la numerosità campionaria ($n = 30, 60, 90, 150$), il numero di variabili esplicative e le diverse funzioni di legame. Nel §4.2 si presentano i risultati dello studio di simulazione condotto partendo dall'insieme di dati (Jackman, 2004) e utilizzato nello studio di simulazione riportato da Kosmidis (2014b, §10). Relativamente a tale studio di simulazione, verranno riportati i risultati ottenuti mediante la specificazione delle funzioni di legame logit e probit. In entrambi gli studi di simulazione è stato valutato anche il comportamento della distorsione stimata, in valore assoluto, a seguito di trasformazione monotona del parametro relativo alle variabili esplicative dicotomiche, presentata nel §2.2.5.

5.1 Primo studio di simulazione

Gli studi di simulazione presentati in questo paragrafo sono stati effettuati considerando una variabile risposta su scala ordinale con tre modalità e, per le tre funzioni di legame considerate, facendo variare la numerosità campionaria n ($n = 30, 60, 90, 150$) e il numero di variabili esplicative considerate secondo le seguenti tre impostazioni:

(i) due variabili esplicative: una con distribuzione normale standard e una dicotomica con valori generati da una bernoulliana con probabilità 1/2;

(ii) tre variabili esplicative: una con distribuzione normale standard e due dicotomiche con valori generati da una bernoulliana con probabilità 1/2;

(iii) due variabili esplicative e la loro interazione: una con distribuzione normale standard e una dicotomica con valori generati da una bernoulliana con probabilità 1/2 .

Il meccanismo di generazione adottato per questo studio di simulazione è il seguente: si fissa la matrice del disegno, si genera la variabile risposta e si calcola la stima di massima verosimiglianza; successivamente, vengono generati 10000 campioni dalle stime di massima verosimiglianza mantenendo fissa la matrice del disegno.

Si presentano i risultati dello studio di simulazione condotto sotto le impostazioni (i) nella Tabella 5.1, (ii) nella Tabella 5.2 e nella Tabella 5.3 sotto l'impostazione (iii). Emergono con evidenza dall'analisi delle Tabelle 5.1-5.3 le buone proprietà dello stimatore con distorsione ridotta in mediana. Le differenze tra i tre stimatori risultano particolarmente marcate per basse numerosità campionarie e diventano via via meno evidenti al crescere della numerosità campionaria. Le stime ottenute con lo stimatore con distorsione ridotta in mediana si caratterizzano per valori della probabilità empirica di sottostima prossimi al 50%, una distorsione stimata confrontabile con quella ottenuta con lo stimatore con distorsione ridotta in media e nettamente inferiore a quella ottenuta con il metodo di stima di massima verosimiglianza. La copertura empirica degli intervalli di Wald è decisamente più vicina al livello nominale rispetto a quella ottenuta con lo stimatore di massima verosimiglianza e confrontabile con quella dello stimatore con distorsione ridotta in media. L'indicatore RMSE dello stimatore con distorsione ridotta in mediana si colloca tra quello dello stimatore di massima verosimiglianza e quello con distorsione ridotta in media, avente i valori più bassi. Come discusso nel §3.3, per basse numerosità campionarie si sono verificati alcuni problemi di instabilità numerica (nell'impostazione (i) con la funzione di legame cloglog il 0.02% dei campioni sia con lo stimatore $\hat{\theta}^*$ che con $\tilde{\theta}$; nell'impostazione (ii) con la funzione di legame cloglog 0.76% dei campioni con $\hat{\theta}^*$ e 0.95% con $\tilde{\theta}$; nell'impostazione (iii), rispettivamente con $\hat{\theta}^*$ e con $\tilde{\theta}$, 0.10% e 0.21% dei campioni con la funzione di legame logit, 0.19% e 0.22% con la funzione di legame probit, 0.49% e 0.53% con la funzione di legame cloglog).

Ai fini della valutazione del comportamento della distorsione stimata a seguito di riparametrizzazione monotona del parametro relativo alle variabili esplicative dicotomiche, utile per ottenere un confronto tra due gruppi, si presentano i risultati ottenuti per l'impostazione (i) nella Tabella 5.4, per l'impostazione (ii) nella Tabella 5.5 e per l'impostazione (iii) nella Tabella 5.6. Nella Tabella 5.5 si riporta anche la percentale di distorsione relativa stimata. Nella nuova parametrizzazione risulta evidente come lo stimatore con distorsione ridotta in mediana abbia valori della distorsione stimata inferiori a quelli dello stimatore con distorsione ridotta in media, che non gode della proprietà di equivarianza rispetto a riparametrizzazioni monotone come quelle presentate nel §2.2.4. Le differenze tra i due stimatori risultano marcate per numerosità campionarie limitate e diventano meno evidenti al crescere dell'informazione campionaria. Si può inoltre notare come lo stimatore di massima verosimiglianza, ad esclusione dello studio condotto sotto l'impostazione (i), mostri una distorsione stimata maggiore rispetto agli stimatori con distorsione ridotta in media e in mediana, e tale differenza svanisce al crescere dell'informazione campionaria.

Tabella 5.1: Primo studio di simulazione: impostazione (i). (Nota: Est.= Stimatore, Par.= Parametro)

			n=30				n=60				n=90				n=150			
Link	Est.	Par.	|Bias|	PU%	RMSE	Wald	|Bias|	PU%	RMSE	Wald	|Bias|	PU%	RMSE	Wald	|Bias|	PU%	RMSE	Wald
logit	$\hat{\theta}$	β_1	0.048	46.56	0.411	94.90	0.003	49.62	0.257	94.92	0.010	48.99	0.210	95.14	0.005	51.40	0.166	94.66
		β_2	0.021	49.88	0.825	94.61	0.035	52.01	0.516	94.67	0.020	51.52	0.407	95.27	0.004	49.62	0.312	94.68
	$\hat{\theta}^*$	β_1	<0.001	51.24	0.358	96.80	0.001	50.16	0.242	96.00	0.002	50.42	0.202	96.00	0.001	50.40	0.163	95.08
		β_2	0.009	51.41	0.740	96.48	0.003	49.84	0.489	95.87	0.001	49.53	0.393	95.80	0.001	50.23	0.306	95.12
	$\tilde{\theta}$	β_1	0.013	49.99	0.373	96.36	0.001	49.98	0.247	95.63	0.004	50.00	0.204	95.69	0.002	50.65	0.164	95.03
		β_2	0.002	50.89	0.764	96.01	0.010	50.30	0.496	95.58	0.004	49.86	0.397	95.66	<0.001	50.10	0.308	95.05
probit	$\hat{\theta}$	β_1	0.026	46.58	0.245	94.10	0.001	49.72	0.155	94.73	0.007	48.81	0.128	94.62	0.003	51.12	0.100	94.80
		β_2	0.016	49.76	0.497	93.83	0.016	51.71	0.309	94.47	0.010	51.34	0.246	94.65	0.001	50.29	0.189	94.64
	$\hat{\theta}^*$	β_1	<0.001	50.89	0.217	96.26	<0.001	50.11	0.147	95.70	0.002	49.96	0.123	95.55	0.001	50.37	0.098	95.12
		β_2	0.002	51.17	0.453	95.61	0.001	49.63	0.297	95.45	0.001	49.56	0.239	95.42	0.001	50.74	0.186	95.03
	$\tilde{\theta}$	β_1	0.006	49.89	0.224	95.75	<0.001	50.01	0.149	95.32	0.004	49.65	0.124	95.29	0.001	50.57	0.099	95.03
		β_2	0.003	50.69	0.464	95.20	0.003	49.98	0.299	95.18	0.001	49.85	0.240	95.26	0.001	50.67	0.186	94.96
cloglog	$\hat{\theta}$	β_1	0.036	46.65	0.301	93.84	0.001	49.38	0.174	93.81	0.009	47.10	0.135	94.42	0.001	49.99	0.111	94.29
		β_2	0.004	50.60	0.620	94.33	0.022	51.80	0.352	94.51	0.016	51.89	0.269	94.90	<0.001	50.33	0.205	94.57
	$\hat{\theta}^*$	β_1	0.002	50.57	0.259	96.39	0.001	50.04	0.163	95.29	0.004	48.84	0.129	95.42	<0.001	49.55	0.108	94.81
		β_2	0.009	50.83	0.561	96.05	0.002	49.42	0.336	95.53	0.002	49.99	0.260	95.50	0.002	50.60	0.201	94.96
	$\tilde{\theta}$	β_1	0.009	49.68	0.269	95.81	0.001	49.93	0.166	94.81	0.005	48.45	0.131	95.15	<0.001	49.66	0.109	94.64
		β_2	0.008	50.75	0.581	95.65	0.007	49.93	0.340	95.27	0.005	50.38	0.262	95.36	0.001	50.48	0.202	94.92

Tabella 5.2: Primo studio di simulazione: impostazione (ii). (Nota: Est.= Stimatore, Par.= Parametro)

			n=30[a]				n=60				n=90				n=150			
Link	Est.	Par.	|Bias|	PU%	RMSE	Wald	|Bias|	PU%	RMSE	Wald	|Bias|	PU%	RMSE	Wald	|Bias|	PU%	RMSE	Wald
logit	$\hat{\theta}$	β_1	0.103	55.55	0.509	94.56	0.015	47.82	0.261	94.25	<0.001	50.65	0.213	94.95	<0.001	50.37	0.165	94.99
		β_2	0.236	41.68	0.971	94.65	0.010	50.58	0.521	94.32	0.037	47.57	0.426	94.94	0.002	49.85	0.314	94.59
		β_3	0.212	43.53	1.020	94.50	0.010	50.17	0.525	94.19	0.021	48.16	0.421	94.54	0.002	49.92	0.313	94.59
	$\hat{\theta}^*$	β_1	0.006	47.43	0.415	96.47	0.001	50.30	0.243	95.79	0.003	51.40	0.203	95.84	0.001	50.03	0.161	95.58
		β_2	0.014	51.38	0.803	96.56	0.004	50.07	0.490	95.65	0.005	50.32	0.407	95.70	0.003	49.92	0.307	95.06
		β_3	0.013	51.68	0.840	96.52	0.001	49.42	0.494	95.49	0.003	49.91	0.403	95.31	0.003	49.77	0.306	95.11
	$\tilde{\theta}$	β_1	0.035	50.07	0.437	96.13	0.005	49.65	0.247	95.53	0.002	51.09	0.205	95.61	0.001	50.12	0.162	95.42
		β_2	0.063	49.28	0.848	96.31	0.005	50.13	0.496	95.35	0.011	49.80	0.410	95.56	0.003	49.90	0.308	94.99
		β_3	0.065	49.26	0.886	96.24	0.001	49.54	0.500	95.23	0.006	49.56	0.406	95.17	0.003	49.79	0.307	95.03
probit	$\hat{\theta}$	β_1	0.057	55.39	0.294	93.88	0.009	47.98	0.155	93.78	0.001	50.47	0.126	94.70	<0.001	50.06	0.101	94.81
		β_2	0.122	42.34	0.553	93.43	0.004	50.11	0.310	94.21	0.017	47.15	0.253	94.58	0.001	49.99	0.191	94.50
		β_3	0.119	43.87	0.595	93.39	0.005	50.07	0.313	94.01	0.013	48.35	0.251	94.29	0.001	49.81	0.190	94.54
	$\hat{\theta}^*$	β_1	0.003	47.63	0.242	96.27	0.001	50.21	0.146	95.43	0.002	50.82	0.121	95.64	0.001	49.85	0.099	95.25
		β_2	0.009	50.57	0.471	95.83	0.002	49.79	0.295	95.37	0.002	49.59	0.244	95.35	0.001	50.01	0.187	95.00
		β_3	0.006	51.44	0.496	96.25	<0.001	49.49	0.297	95.34	0.003	49.76	0.242	95.19	0.001	49.68	0.187	95.02
	$\tilde{\theta}$	β_1	0.017	50.10	0.253	95.91	0.003	49.67	0.148	95.01	0.002	50.65	0.122	95.47	<0.001	49.91	0.099	95.11
		β_2	0.030	49.06	0.489	95.50	0.002	49.79	0.297	95.17	0.005	49.21	0.245	95.24	0.001	50.01	0.188	94.96
		β_3	0.032	49.57	0.517	95.80	0.001	49.60	0.300	95.16	0.005	49.52	0.243	95.06	0.001	49.69	0.187	94.97
cloglog	$\hat{\theta}$	β_1	0.109	58.61	0.383	93.66	0.008	48.13	0.173	93.81	0.001	50.74	0.137	94.18	0.001	49.60	0.114	94.48
		β_2	0.175	41.94	0.684	93.84	0.021	52.38	0.351	93.55	0.014	48.48	0.259	94.44	0.003	49.31	0.210	94.75
		β_3	0.261	39.31	0.800	93.37	0.007	50.46	0.350	94.09	0.015	47.93	0.259	94.38	<0.001	50.26	0.210	94.71
	$\hat{\theta}^*$	β_1	0.002	46.72	0.288	96.06	<0.001	50.11	0.159	95.70	0.001	50.69	0.130	95.52	0.001	49.75	0.111	95.18
		β_2	0.002	52.44	0.536	96.25	0.005	50.51	0.329	94.99	0.001	50.26	0.247	95.57	0.001	49.59	0.206	95.35
		β_3	0.009	53.98	0.591	95.90	0.001	49.53	0.329	95.43	0.001	50.24	0.247	95.44	0.003	49.81	0.206	95.11
	$\tilde{\theta}$	β_1	0.024	50.07	0.306	95.84	0.002	49.73	0.162	95.38	0.001	50.77	0.131	95.27	<0.001	49.82	0.111	95.03
		β_2	0.037	50.20	0.580	95.99	0.007	50.73	0.333	94.74	0.003	49.94	0.249	95.45	0.002	49.49	0.206	95.30
		β_3	0.051	50.58	0.643	95.71	0.001	49.79	0.333	95.19	0.003	49.91	0.249	95.31	0.003	49.83	0.207	95.06

[a]% stime finite: logit link, $\hat{\theta}$=99.75%; probit link, $\hat{\theta}$=99.74%; cloglog link, $\hat{\theta}$=97.78%.

Tabella 5.3: Primo studio di simulazione: impostazione (iii). (Nota: Est.= Stimatore, Par.= Parametro)

			n=30 [a]				n=60				n=90				n=150			
Link	Est.	Par.	|Bias|	PU%	RMSE	Wald	|Bias|	PU%	RMSE	Wald	|Bias|	PU%	RMSE	Wald	|Bias|	PU%	RMSE	Wald
logit	$\hat{\theta}$	β_1	0.260	41.78	0.923	97.10	0.016	48.68	0.308	94.82	0.029	47.50	0.288	94.84	0.007	48.86	0.231	94.68
		β_2	0.095	51.30	1.166	96.10	0.066	53.22	0.559	94.64	0.038	52.77	0.443	94.60	<0.001	50.13	0.317	94.63
		β_3	0.414	61.26	1.292	95.53	0.043	52.25	0.597	94.74	0.043	52.99	0.463	94.50	0.019	52.18	0.341	94.53
	$\hat{\theta}^*$	β_1	0.017	54.45	0.576	97.00	<0.001	50.83	0.283	96.32	0.004	50.94	0.271	95.82	0.005	49.17	0.224	95.38
		β_2	0.006	49.06	0.885	97.79	0.010	49.82	0.518	96.01	0.002	49.20	0.420	95.56	0.003	50.46	0.309	95.20
		β_3	0.029	48.03	0.867	96.93	0.002	49.58	0.550	96.19	0.006	49.77	0.433	95.68	0.004	50.24	0.330	95.27
	$\tilde{\theta}$	β_1	0.101	49.72	0.681	97.05	0.006	50.05	0.292	95.90	0.013	49.75	0.277	95.47	0.005	49.10	0.226	95.08
		β_2	0.042	49.88	0.996	97.49	0.022	50.44	0.527	95.76	0.006	49.88	0.425	95.46	0.002	50.40	0.311	95.09
		β_3	0.114	51.03	0.966	96.85	0.012	50.30	0.561	95.82	0.014	50.44	0.440	95.43	0.007	50.77	0.332	95.16
probit	$\hat{\theta}$	β_1	0.165	41.18	0.608	97.17	0.008	49.23	0.185	94.33	0.016	47.55	0.171	94.32	<0.001	50.06	0.139	94.73
		β_2	0.050	51.10	0.704	94.54	0.036	53.05	0.335	94.43	0.019	52.25	0.263	94.72	0.001	49.99	0.191	94.50
		β_3	0.247	61.50	0.806	94.10	0.026	52.17	0.359	94.20	0.025	53.02	0.278	94.04	0.001	49.81	0.190	94.54
	$\hat{\theta}^*$	β_1	0.007	53.97	0.327	97.47	0.001	51.11	0.172	95.78	0.003	50.63	0.271	95.42	0.001	49.85	0.099	95.25
		β_2	0.003	49.25	0.506	96.85	0.007	49.98	0.314	95.74	0.001	49.07	0.420	95.59	0.001	50.01	0.187	95.00
		β_3	0.013	48.74	0.492	96.43	0.002	49.58	0.334	95.77	0.005	50.11	0.433	95.28	0.001	49.68	0.187	95.02
	$\tilde{\theta}$	β_1	0.057	49.81	0.389	95.50	0.002	50.60	0.176	95.16	0.007	49.58	0.165	95.13	<0.001	49.91	0.099	95.11
		β_2	0.012	49.71	0.540	96.41	0.012	50.52	0.318	95.50	0.002	49.62	0.254	95.45	0.001	50.01	0.188	94.96
		β_3	0.056	51.65	0.535	96.22	0.007	50.05	0.339	95.45	0.009	50.65	0.265	95.02	0.001	49.69	0.187	94.97
cloglog	$\hat{\theta}$	β_1	0.245	38.76	0.745	97.36	0.012	48.73	0.200	94.00	0.019	47.18	0.176	93.95	0.003	49.42	0.154	94.54
		β_2	0.078	51.74	0.492	97.18	0.041	53.13	0.385	94.77	0.029	52.74	0.291	94.57	0.004	49.71	0.206	94.94
		β_3	0.357	63.08	1.018	94.59	0.024	51.80	0.424	94.15	0.026	53.01	0.306	93.89	0.010	51.64	0.226	93.83
	$\hat{\theta}^*$	β_1	0.007	56.24	0.367	97.32	0.001	50.99	0.182	96.05	0.002	51.02	0.165	95.47	0.001	49.97	0.149	95.36
		β_2	0.011	48.69	0.703	98.08	0.006	49.86	0.359	95.92	0.004	49.41	0.276	95.56	0.004	49.62	0.202	95.41
		β_3	0.006	47.04	0.609	96.74	0.001	49.75	0.391	95.57	0.004	50.07	0.286	95.29	0.002	49.91	0.218	94.61
	$\tilde{\theta}$	β_1	0.062	50.57	0.446	97.40	0.004	50.33	0.188	95.42	0.007	49.77	0.168	95.03	0.001	49.88	0.150	95.17
		β_2	0.044	50.03	0.834	97.87	0.016	50.84	0.366	95.73	0.010	50.10	0.279	95.33	0.004	49.56	0.202	95.33
		β_3	0.076	50.93	0.696	96.92	0.006	50.21	0.399	95.25	0.009	50.93	0.290	95.00	0.004	50.33	0.220	94.49

[a] % stime finite: logit link $\hat{\theta}$=99.74%; probit link $\hat{\theta}$=99.73%; cloglog link $\hat{\theta}$=99.53%.

link	Stimatore	n=30	n=60	n=90	n=150
logit	$\gamma(\hat{\beta}_2)$	0.0008	0.0021	0.0011	0.0001
	$\gamma(\hat{\beta}_2^*)$	0.0049	0.0027	0.0022	0.0007
	$\gamma(\tilde{\beta}_2)$	0.0033	0.0016	0.0015	0.0005
probit	$\gamma(\hat{\beta}_2)$	0.0005	0.0017	0.0009	0.0001
	$\gamma(\hat{\beta}_2^*)$	0.0037	0.0022	0.0018	0.0006
	$\gamma(\tilde{\beta}_2)$	0.0023	0.0015	0.0013	0.0005
cloglog	$\gamma(\hat{\beta}_2)$	0.0030	0.0018	0.0017	0.0003
	$\gamma(\hat{\beta}_2^*)$	0.0037	0.0025	0.0013	0.0006
	$\gamma(\tilde{\beta}_2)$	0.0036	0.0015	0.0006	0.0005

Tabella 5.4: Distorsione stimata per la trasformazione monotona del parametro β_2, relativo alla variabile esplicativa dicotomica nell'impostazione (i).

		n=30		n=60		n=90		n=150	
link	Est.	|Bias|	%RB	|Bias|	%RB	|Bias|	%RB	|Bias|	%RB
logit	$\gamma(\hat{\beta}_2)$	0.0126	4.70	0.0011	0.21	0.0033	0.89	0.0004	0.08
	$\gamma(\hat{\beta}_3)$	0.0133	4.12	0.0006	0.11	0.0019	0.45	0.0004	0.08
	$\gamma(\hat{\beta}_2^*)$	0.0116	4.33	0.0001	0.02	0.0017	0.46	0.0005	0.10
	$\gamma(\hat{\beta}_3^*)$	0.0106	3.28	0.0010	0.19	0.0011	0.26	0.0006	0.12
	$\gamma(\tilde{\beta}_2)$	0.0063	2.35	0.0002	0.04	0.0006	0.16	0.0005	0.10
	$\gamma(\tilde{\beta}_3)$	0.0042	1.30	0.0007	0.13	0.0005	0.12	0.0006	0.12
probit	$\gamma(\hat{\beta}_2)$	0.0128	1.83	0.0008	0.16	0.0028	0.73	0.0004	0.08
	$\gamma(\hat{\beta}_3)$	0.0142	2.17	0.0007	0.11	0.0022	0.52	0.0002	0.02
	$\gamma(\hat{\beta}_2^*)$	0.0069	0.99	0.0001	0.02	0.0010	0.26	0.0004	0.08
	$\gamma(\hat{\beta}_3)$	0.0068	1.04	0.0006	0.11	0.0002	0.05	0.0004	0.08
	$\gamma(\tilde{\beta}_2)$	0.0033	0.47	0.0002	0.04	0.0004	0.10	0.0004	0.08
	$\gamma(\tilde{\beta}_3)$	0.0019	0.29	0.0004	0.08	0.0001	0.02	0.0004	0.08
cloglog	$\gamma(\hat{\beta}_2)$	0.0160	5.65	0.0033	0.59	0.0022	0.51	0.0005	0.10
	$\gamma(\hat{\beta}_3)$	0.0210	8.72	0.0006	0.11	0.0023	0.55	0.0003	0.06
	$\gamma(\hat{\beta}_2^*)$	0.0121	4.28	0.0005	0.09	0.0007	0.16	0.0002	0.02
	$\gamma(\hat{\beta}_3^*)$	0.0167	6.94	0.0012	0.22	0.0011	0.27	0.0010	0.18
	$\gamma(\tilde{\beta}_2)$	0.0062	2.19	<0.0001	<0.01	0.0003	0.07	0.0003	0.03
	$\gamma(\tilde{\beta}_3)$	0.0083	3.45	0.0007	0.13	0.0005	0.12	0.0009	0.16

Tabella 5.5: Distorsione e distorsione relativa stimate per le trasformazioni monotone dei parametri β_2 e β_3, relativi alle variabili esplicative dicotomiche nell'impostazione (ii). (Nota: Est.= Stimatore, %RB=%|Bias relativo|)

link	Stimatore	n=30	n=60	n=90	n=150
logit	$\gamma(\hat{\beta}_2)$	0.0081	0.0057	0.0030	0.0004
	$\gamma(\hat{\beta}_2^*)$	0.0021	0.0025	0.0031	0.0009
	$\gamma(\tilde{\beta}_2)$	0.0021	0.0007	0.0019	0.0008
probit	$\gamma(\hat{\beta}_2)$	0.0079	0.0056	0.0028	0.0004
	$\gamma(\hat{\beta}_2^*)$	0.0008	0.0014	0.0024	0.0004
	$\gamma(\tilde{\beta}_2)$	0.0021	0.0001	0.0016	0.0004
cloglog	$\gamma(\hat{\beta}_2)$	0.0098	0.0049	0.0041	0.0008
	$\gamma(\hat{\beta}_2^*)$	0.0015	0.0024	0.0013	0.0008
	$\gamma(\tilde{\beta}_2)$	0.0038	0.0004	0.0001	0.0009

Tabella 5.6: Distorsione stimata per la trasformazione monotona del parametro β_2, relativo alla variabile esplicativa dicotomica nell'impostazione (iii).

5.2 Secondo studio di simulazione

Il secondo studio di simulazione è stato condotto sull'insieme di dati *admit*, disponibile nel paccchetto R `plsc` (Jackman, 2017). I dati in esame riportano il punteggio assegnato a $n = 106$ candidati al dottorato in Scienze Politiche di Stanford. Sono disponibili alcune caratteristiche dei candidati quali il genere, i punteggi nelle parti quantitative e verbali, il grado di interessamento per la politica americana e per la teoria politica in generale. Il punteggio assegnato è su scala ordinale con cinque modalità (1=più basso e 5=più alto) con frequenze assolute osservate, nella Tabella 5.7.

Modalità	1	2	3	4	5
Frequenze	23	24	2	37	20

Tabella 5.7: Frequenze assolute osservate della variabile risposta

I dati vengono modellati secondo il modello link cumulati nella formulazione (2.1) e, in particolare,

$$h[P(Y_i \leq j)] = \alpha_j+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3}+\beta_4 x_{i4}+\beta_5 x_{i5}, \;\; i = 1, \ldots, 106, \; j = 1, \ldots, 4,$$

con $h[\cdot]$ opportuna funzione di legame, che nella specificazione logit conduce ai risultati nella Tabella 5.8 e nella specificazione probit quelli nella Tabella 5.9. I punteggi nelle parti quantitative e verbali, x_{i1} e x_{i2}, di natura continua, sono stati standardizzati. Gli interessi per la politica americana, x_{i3}, e per la teoria politica, x_{i4}, hanno natura dicotomica (1=interesse e 0=poco interesse) così come il genere, x_{i5} (1=donne, 0=uomini). Le stime di massima verosimiglianza, arrotondate alla terza cifra decimale, da cui sono stati generati i valori della variabile risposta risultano essere per le funzioni di legame

(i) logit: $\hat{\theta} = (-1.406, 0.525, 0.658, 3.341, -1.993, -0.892, -2.816, -0.009, -1.215)$

(ii) probit: $\hat{\theta} = (-0.840, 0.271, 0.349, 1.873, -1.169, -0.492, -1.627, 0.002, -0.640)$

I risultati presentati nella Tabella 5.8, in relazione al modello con funzione di legame logit, e nella Tabella 5.9, per la funzione di legame probit, mostrano con evidenza quanto ci si attendeva. Lo stimatore di massima verosimiglianza ha su tutti gli indici i valori peggiori: una maggiore distorsione stimata, un'alta stima della probabilità di sottostimare il vero valore del parametro, una copertura un pò inferiore a quella nominale. Lo stimatore con distorsione ridotta in media ha come atteso i più bassi valori in termini di distorsione stimata e una stima della probabilità di sottostima migliore rispetto a quella dello stimatore di massima verosimiglianza ma comunque per la maggior parte dei parametri distante dal 50%. Lo stimatore con distorsione ridotta in mediana mostra una proprietà di centratura notevole, ossia valori delle stime della probabilità di sottostima prossimi al 50%, e risulta confrontabile con lo stimatore con distorsione ridotta in media in termini e di distorsione stimata e di copertura. L'analisi del RMSE mostra che i valori sono molto simili ma in generale lo stimatore $\hat{\theta}$ ha i valori più elevati, $\hat{\theta}^*$ i più bassi e $\tilde{\theta}$ intermedi.

Nelle Tabelle 5.8-5.9 si riportano, inoltre, la stima della distorsione e della distorsione relativa in percentuale ottenuta nelle parametrizzazioni discusse in §2.2.5. Risulta evidente come lo stimatore che riduce la distorsione in mediana ha valori inferiori rispetto allo stimatore di massima verosimiglianza e a quello che riduce la distorsione in media per i parametri $\gamma(\beta_3)$ e $\gamma(\beta_5)$ e confrontabili per $\gamma(\beta_4)$.

Est.	Par.	|Bias|	PU	RMSE	Wald	Par.	|Bias|	%RB
	β_1	0.136	62.41	0.379	94.20			
	β_2	0.055	56.83	0.252	93.98			
$\hat{\theta}$	β_3	0.220	58.16	0.862	94.74	$\gamma(\beta_3)$	0.0034	0.39
	β_4	0.002	50.11	0.788	94.54	$\gamma(\beta_4)$	0.0003	0.06
	β_5	0.073	55.52	0.488	94.57	$\gamma(\beta_5)$	0.0057	0.81
	β_1	0.006	47.60	0.327	94.86			
	β_2	0.002	48.24	0.230	95.14			
$\hat{\theta}^*$	β_3	0.014	47.76	0.769	95.50	$\gamma(\beta_3)$	0.0101	1.15
	β_4	0.002	49.81	0.738	95.81	$\gamma(\beta_4)$	0.0004	0.08
	β_5	0.001	49.41	0.454	95.33	$\gamma(\beta_5)$	0.0045	0.64
	β_1	0.029	50.16	0.333	94.97			
	β_2	0.011	49.85	0.233	95.11			
$\tilde{\theta}$	β_3	0.061	50.11	0.786	95.48	$\gamma(\beta_3)$	0.0069	0.78
	β_4	0.005	49.85	0.750	95.52	$\gamma(\beta_4)$	0.0009	0.18
	β_5	0.011	50.53	0.458	95.22	$\gamma(\beta_5)$	0.0027	0.38

Tabella 5.8: Secondo studio di simulazione: funzione di legame logit. (Nota: %RB= %|Bias relativo|, Est.= Stimatore, Par.=Parametro)

Est.	Par.	|Bias|	PU	RMSE	Wald	Par.	|Bias|	%RB
	β_1	0.077	63.39	0.207	93.66			
	β_2	0.027	57.01	0.138	93.98			
$\hat{\theta}$	β_3	0.120	58.10	0.478	94.29	$\gamma(\beta_3)$	0.0050	0.57
	β_4	0.008	49.62	0.461	93.91	$\gamma(\beta_4)$	0.0020	0.40
	β_5	0.039	54.47	0.274	94.49	$\gamma(\beta_5)$	0.0066	0.98
	β_1	0.005	48.31	0.178	94.76			
	β_2	$<$0.001	48.74	0.127	95.03			
$\hat{\theta}^*$	β_3	0.005	48.08	0.426	95.44	$\gamma(\beta_3)$	0.0095	1.09
	β_4	0.001	50.32	0.431	95.30	$\gamma(\beta_4)$	0.0001	0.02
	β_5	0.002	48.78	0.256	95.47	$\gamma(\beta_5)$	0.0022	0.33
	β_1	0.017	50.83	0.181	94.96			
	β_2	0.004	50.19	0.128	94.90			
$\tilde{\theta}$	β_3	0.030	50.15	0.436	95.36	$\gamma(\beta_3)$	0.0063	0.72
	β_4	0.006	49.79	0.438	95.00	$\gamma(\beta_4)$	0.0014	0.28
	β_5	0.008	49.82	0.259	95.36	$\gamma(\beta_5)$	0.0007	0.10

Tabella 5.9: Secondo studio di simulazione: funzione di legame probit. (Nota: %RB= %|Bias relativo|, Est.= Stimatore, Par.=Parametro)

Conclusioni

L'obiettivo del presente lavoro è stato lo studio dello stimatore con distorsione ridotta in mediana, secondo l'approccio proposto da Kenne Pagui *et al.* (2017), nei modelli di regressione per variabili risposta ordinali e in particolare nei modelli link cumulati nella loro formulazione base (McCullagh, 1980).

L'approccio di stima di massima verosimiglianza è stato ed è tutt'ora l'approccio più popolare per la trattazione dei modelli link cumulati. Tuttavia, in presenza di piccole numerosità campionarie, dati sparsi o modelli complessi, le proprietà asintotiche dello stimatore di massima verosimiglianza possono riflettere scarsamente l'esatta distribuzione campionaria e la distorsione delle stime ha come conseguenza il deterioramento delle procedure inferenziali. Un ulteriore problema connesso con l'approccio di stima di massima verosimiglianza è legato alle stime sulla frontiera dello spazio parametrico. Lo sviluppo dei metodi di riduzione della distorsione in media, secondo l'approccio di Firth (1993), è stato formalizzato (Kosmidis, 2014b) per i modelli link cumulati. Lo stimatore che si ottiene dall'equazione di stima basata sulla funzione punteggio modificata è efficace nel ridurre la distorsione delle stime di massima verosimiglianza e nel prevenire in alcune situazioni le stime sulla frontiera dello spazio parametrico. Tuttavia tale stimatore ha come svantaggio quello di non essere equivariante rispetto a trasformazioni monotone delle componenti del parametro.

L'approccio di stima proposto da Kenne Pagui *et al.* (2017) si basa anch'esso sulla correzione della funzione di punteggio. Tuttavia si differenzia dall'approccio di Firth (1993) poichè utilizza come proprietà di centratura la non distorsione in mediana. Lo stimatore risultante dalla soluzione dell'equazione di stima basata sulla funzione punteggio modificata risulta non distorto in mediana al terzo ordine componente per componente ed equivariante rispetto a riparametrizzazioni monotone che trasformano ogni componente del vettore di parametri separatamente, proprietà desiderabile per uno stimatore. Il calcolo delle quantità necessarie allo sviluppo dello stimatore che riduce la distorsione in mediana nei modelli link cumulati e la loro traduzione su supporto informatico rappresentano il nucleo del presente lavoro.

È stata mostrata un'applicazione dello stimatore con distorsione ridotta in mediana a un insieme di dati reali. Le conclusioni sono state qualitativamente concordi con l'analisi ottenuta con il metodo di stima di massima verosimiglianza (Decensi *et al.*, 1996). L'effetto di *shrinkage* delle stime dei parametri di regressione non porta a conclusioni inferenziali differenti da quelle presenti nell'articolo originale. Questo è principalmente dovuto al forte effetto che hanno le variabili di interesse sulla variabile risposta.

Gli studi di simulazione condotti hanno permesso di verificare le ottime proprietà dello stimatore con distorsione ridotta in mediana. È risultata evidente la centratura in mediana delle stime ottenute, valutata mediante la probabilità empirica di sottostima, e una distorsione stimata confrontabile con quella ottenibile con l'approccio di stima che determina lo stimatore con distorsione ridotta in media. Gli altri indici, quali la radice quadrata dell'errore quadratico medio e la copertura empirica degli intervalli di confidenza, sono sostanzilmente simili a quelli ottenibili con lo stimatore con distorsione ridotta in media. Inoltre si è mostrato come, a seguito di trasformazione monotona del parametro relativo a una variabile esplicativa dicotomica, corrispondente a una misura utile per il confronto tra due gruppi (Agresti e Kateri, 2017), lo stimatore con distorsione ridotta in mediana ha una distorsione stimata, in valore assoluto, inferiore a quella dello stimatore con distorsione ridotta in media. Tale proprietà, unita con la centratura in mediana, rendono l'approccio di stima proposto da Kenne Pagui *et al.* (2017) una valida alternativa ai suoi principali competitori, ovvero lo stimatore di massima verosimiglianza e lo stimatore con distorsione ridotta in media, specie in situazioni di numerosità campionaria limitata.

Per i calcoli numerici si è usato il software di programmazione statistica R 3.5.0 e il codice relativo all'implementazione è disponibile in Appendice.

Appendice: codice R

Funzioni e quantità utili

```
quantile_func<-function(z){
 log(-log(1-z))
}

is.integer0 <- function(x){
  is.integer(x) && length(x) == 0L
}

quantita_chiave<-function(par,x,q,
                         type=c("logit","probit","cloglog")){
 n<-nrow(x)
 p<-ncol(x)
 alpha<-par[1:(q-1)]
 alpha<-c(-Inf,alpha,Inf)
 beta<-par[q:(q+p-1)]

 mat1<-matrix(rep(alpha[-1],n),ncol=q,byrow=TRUE)
 mat2<-matrix(rep(alpha[1:q],n),ncol=q,byrow=TRUE)
 mat3<-matrix(rep(alpha[3:(q+1)],n),ncol=q-1,byrow=TRUE)

 eta_beta <-x%*%beta
 eta_ij<- apply(mat1,2,function(t) t+eta_beta)
 eta_ijm1<-apply(mat2,2,function(t) t+eta_beta)
```

```
eta_ijp1<-apply(mat3,2,function(t) t+eta_beta)

type<-match.arg(type)
pfun<-switch(type,
           logit = make.link(type)$linkinv,
           probit = make.link(type)$linkinv,
           cloglog = make.link(type)$linkinv)
dfun<-switch(type,
           logit = make.link(type)$mu.eta,
           probit = make.link(type)$mu.eta,
           cloglog = make.link(type)$mu.eta)
d1fun<-switch(type,
           logit = function(eta) dfun(eta)*(1-2*pfun(eta)),
           probit = function(eta) -eta*dfun(eta),
           cloglog = function(eta) dfun(eta)*(1-exp(eta)))
d2fun<-switch(type,
           logit = function(eta) dfun(eta)*(1-2*pfun(eta))^2-
                                 2*dfun(eta)^2,
           probit = function(eta) (eta^2-1)*dfun(eta),
           cloglog = function(eta) d1fun(eta)*(1+log(1-pfun(eta)))-
                                   dfun(eta)^2/(1-pfun(eta)))

rip_ij<-pfun(eta_ij)
rip_ijm1<-pfun(eta_ijm1)
rip_ijp1<-pfun(eta_ijp1)

den_ij<-dfun(eta_ij)
den_ijm1<-dfun(eta_ijm1)
den_ijp1<-dfun(eta_ijp1)
if(type=="probit" | type=="cloglog")
{
 den_ij[,q]<-rep(1e-12,n)
 den_ijp1[,q-1]<-rep(1e-12,n)
```

```
 }
 pi_ij<-rip_ij-rip_ijm1
 pi_ijp1<-rip_ijp1-rip_ij[,-q]

#modifica per risolvere problemi di instabilià numerica
 for(j in 1:q)
 {
  for(i in 1:n){
   if(pi_ij[i,j]==0) pi_ij[i,j]<-1e-16
    if(j<q){
   if(pi_ijp1[i,j]==0) pi_ijp1[i,j]<-1e-16
    }
   }
 }

 der_den_ij<-d1fun(eta_ij)
 der_den_ijm1<-d1fun(eta_ijm1)
 der_den_ijp1<-d1fun(eta_ijp1)

 if(type=="probit" | type=="cloglog")
 {
  der_den_ij[,q]<-rep(1e-12,n)
  der_den_ijp1[,q-1]<-rep(1e-12,n)
  der_den_ijm1[,1]<-rep(1e-12,n)
 }

 der2_den_ij<-d2fun(eta_ij)
 der2_den_ijm1<-d2fun(eta_ijm1)
 if(type=="probit" | type=="cloglog"){
  der2_den_ij[,q]<-rep(1e-12,n)
  der2_den_ijm1[,1]<-rep(1e-12,n)
 }
```

```
A<-den_ij/pi_ij
B<-den_ij[,-q]/pi_ijp1
C<-(den_ij-den_ijm1)/pi_ij
D<-der_den_ij/pi_ij-(den_ij/pi_ij)^2
G<-der_den_ij[,-q]/pi_ijp1+(den_ij[,-q]/pi_ijp1)^2
H<-den_ij[,-q]*den_ijp1/pi_ijp1^2
I<-der_den_ij/pi_ij-den_ij*(den_ij-den_ijm1)/pi_ij^2
J<-der_den_ij[,-q]/pi_ijp1-
    den_ij[,-q]*(den_ijp1-den_ij[,-q])/pi_ijp1^2
K<-(der_den_ij-der_den_ijm1)/pi_ij-(den_ij-den_ijm1)^2/ pi_ij^2
L<-der2_den_ij/pi_ij-
  3*der_den_ij*den_ij/pi_ij^2+2*den_ij^3/pi_ij^3
M<-der2_den_ij[,-q]/pi_ijp1+
   3*der_den_ij[,-q]*den_ij[,-q]/pi_ijp1^2+
   2*den_ij[,-q]^3/pi_ijp1^3
N<-der_den_ij[,-q]*den_ijp1/pi_ijp1^2+
   2*den_ij[,-q]^2*den_ijp1/pi_ijp1^3
O<-den_ij[,-q]*der_den_ijp1/pi_ijp1^2-
   2*den_ij[,-q]*den_ijp1^2/pi_ijp1^3
P<-der2_den_ij/pi_ij-
   der_den_ij*(3*den_ij-den_ijm1)/pi_ij^2+
   2*den_ij^2*(den_ij-den_ijm1)/pi_ij^3
Q<-der2_den_ij[,-q]/pi_ijp1+
   der_den_ij[,-q]*(3*den_ij[,-q]-den_ijp1)/pi_ijp1^2-
   2*den_ij[,-q]^2*(den_ijp1-den_ij[,-q])/pi_ijp1^3
R<-der_den_ij[,-q]*den_ijp1/pi_ijp1^2+
   der_den_ijp1*den_ij[,-q]/pi_ijp1^2-
   2*den_ij[,-q]*den_ijp1*(den_ijp1-den_ij[,-q])/pi_ijp1^3
S<-der2_den_ij/pi_ij-
   den_ij*(der_den_ij-der_den_ijm1)/pi_ij^2-
   2*der_den_ij*(den_ij-den_ijm1)/pi_ij^2+
   2*den_ij*(den_ij-den_ijm1)^2/pi_ij^3
```

```
 T<-der2_den_ij[,-q]/pi_ijp1-
    den_ij[,-q]*(der_den_ijp1-der_den_ij[,-q])/pi_ijp1^2-
    2*der_den_ij[,-q]*(den_ijp1-den_ij[,-q])/pi_ijp1^2+
    2*den_ij[,-q]*(den_ijp1-den_ij[,-q])^2/pi_ijp1^3
 U<-(der2_den_ij-der2_den_ijm1)/pi_ij-
    3*(der_den_ij-der_den_ijm1)*(den_ij-den_ijm1)/pi_ij^2+
    2*(den_ij-den_ijm1)^3/pi_ij^3
 return(list(A=A[,-q],B=B,C=C,D=D[,-q],G=G,H=H,
             I=I[,-q],J=J,K=K,L=L[,-q], M=M,N=N,O=O,
             P=P[,-q],Q=Q,R=R,S=S[,-q],T=T,U=U,
            pi_ij=pi_ij,pi_ijp1=pi_ijp1))
}
```

Funzione score

```
score_fun<-function(par,x,y,type="logit"){
 q<-ncol(y)
 res<-rep(0,length(par))
 fit<-quantita_chiave(par,x,q,type)
 res[1:(q-1)]<-with(fit,{colSums(y[,-q]*A-y[,-1]*B)})
 for(i in 1:ncol(x)){
  X<-matrix(rep(x[,i],q),ncol=q,byrow=FALSE)
  res[i+q-1]<- with(fit,{sum(colSums(X*y*C))})
 }
 return(res)
}
```

Informazione attesa di Fisher

```
info_fun<-function(par,x,y,type="logit"){
 q<-ncol(y)
 p<-ncol(x)
 infoFisher<-matrix(0,nrow=q+p-1,ncol=q+p-1)
 fit<-quantita_chiave(par,x,q,type)
```

```
#i_aj_aj
if(q==2)
{
 infoFisher[1,1]<-with(fit,{
                  -colSums(pi_ij[,-q]*D-pi_ijp1*G)})
} else {
diag(infoFisher[1:(q-1),1:(q-1)])<-with(fit,{
                  -colSums(pi_ij[,-q]*D-pi_ijp1*G)})
#i_aj_aj+1
for(i in 1:q-1){
 infoFisher[i,i+1]<-with(fit,{ -colSums(pi_ijp1*H)})[i]
 infoFisher[i+1,i]<- infoFisher[i,i+1]
 }
}
#i_aj_br
for(i in 1:p){
 X<-matrix(rep(x[,i],q-1),ncol=q-1,byrow=FALSE)
 if(q==2){
  infoFisher[q+i-1,1:(q-1)] <-with(fit,{
           -sum(X*(pi_ij[,-q]*I-pi_ijp1*J))})
 infoFisher[1:(q-1),q+i-1]<- infoFisher[q+i-1,1:(q-1)]
 } else {
 infoFisher[q+i-1,1:(q-1)] <-with(fit,{
      -colSums(X*(fit$pi_ij[,-q]*fit$I[,-q]-fit$pi_ijp1*fit$J))})
 infoFisher[1:(q-1),q+i-1]<- infoFisher[q+i-1,1:(q-1)]
 }
#i_br_bs
 for(j in 1:p){
  if(j<=i){
  X1<-matrix(rep(x[,i],q),ncol=q,byrow=FALSE)
  X2<-matrix(rep(x[,j],q),ncol=q,byrow=FALSE)
  infoFisher[i+q-1,j+q-1]<-with(fit,{
```

```
                      -sum(colSums(X1*X2*pi_ij*K))})
   }
  infoFisher[j+q-1,i+q-1]<-infoFisher[i+q-1,j+q-1]
  }
 }
 return(infoFisher)
}
```

Momenti nulli

```
nu_quantities<-function(par,x,q,type="logit")
{
 p<-ncol(x)
 fit<-quantita_chiave(par,x,q,type)
 #nu_r,st
 nu_a_aa<-array(0,dim=c(q-1))
 if(q>2){
  nu_a_aap1<-array(0,dim=c(q-2))
  nu_a_ap1ap1<-array(0,dim=c(q-2))
  nu_ap1_aa<-array(0,dim=c(q-2))
  nu_ap1_aap1<-array(0,dim=c(q-2))
  nu_a_ap1b<-array(0,dim=c(p,q-2))
  nu_ap1_ab<-array(0,dim=c(p,q-2))
  nu_b_aap1<-array(0,dim=c(p,q-2))
 }
 nu_a_ab<-array(0,dim=c(p,q-1))
 nu_b_aa<-array(0,dim=c(p,q-1))
 nu_b_ab<-array(0,dim=c(p,q-1,p))
 nu_a_bb<-array(0,dim=c(p,q-1,p))
 nu_b_bb<-array(0,dim=c(p,p,p))

 #nu_r,s,t
 nu_a_a_a<-array(0,dim=c(q-1))
```

```
nu_a_a_ap1<-array(0,dim=c(q-2))
nu_a_ap1_ap1<-array(0,dim=c(q-2))
nu_a_a_b<-array(0,dim=c(p,q-1))
nu_a_ap1_b<-array(0,dim=c(p,q-2))
nu_a_b_b<-array(0,dim=c(p,q-1,p))
nu_b_b_b<-array(0,dim=c(p,p,p))

nu_a_aa<-with(fit,{
                   colSums(pi_ij[,-q]*A*D+pi_ijp1*B*G)})
nu_a_a_a<-with(fit,{-colSums(pi_ij[,-q]*L-pi_ijp1*M)-
                    3*nu_a_aa})
if(p!=1 & q>3){
 nu_a_aap1<-with(fit,{
                   -colSums(pi_ijp1[,-(q-1)]*B[,-(q-1)]*H[,-(q-1)])})
 nu_a_ap1ap1<-with(fit,{
                    -colSums(pi_ijp1[,-(q-1)]*B[,-(q-1)]*D[,-1])})
 nu_ap1_aa<-with(fit,{
                   -colSums(pi_ijp1[,-(q-1)]*A[,-1]*G[,-(q-1)])})
 nu_ap1_aap1<-with(fit,{
                    colSums(pi_ijp1[,-(q-1)]*A[,-1]*H[,-(q-1)])})
 nu_a_a_ap1<-with(fit,{-colSums(pi_ijp1[,-(q-1)]*N[,-(q-1)])-
                 2*nu_a_aap1-nu_ap1_aa})
 nu_a_ap1_ap1<-with(fit,{-colSums(pi_ijp1[,-(q-1)]*O[,-(q-1)])-
                 2*nu_ap1_aap1-nu_a_ap1ap1})
 }

if(p==1 | q==3){
 nu_a_aap1<-with(fit,{
                   -sum(pi_ijp1[,-(q-1)]*B[,-(q-1)]*H[,-(q-1)])})
 nu_a_ap1ap1<-with(fit,{
                   -sum(pi_ijp1[,-(q-1)]*B[,-(q-1)]*D[,-1])})
 nu_ap1_aa<-with(fit,{
                   -sum(pi_ijp1[,-(q-1)]*A[,-1]*G[,-(q-1)])})
```

```
 nu_ap1_aap1<-with(fit,{
                   sum(pi_ijp1[,-(q-1)]*A[,-1]*H[,-(q-1)])})
 nu_a_a_ap1<-with(fit,{-sum(pi_ijp1[,-(q-1)]*N[,-(q-1)])-
                2*nu_a_aap1-nu_ap1_aa})
 nu_a_ap1_ap1<-with(fit,{-sum(pi_ijp1[,-(q-1)]*O[,-(q-1)])-
                2*nu_ap1_aap1-nu_a_ap1ap1})
 }

for(i in 1:p){
  X1<-matrix(rep(x[,i],q-1),ncol=q-1,byrow=FALSE)
  if(q>2){
   X2<-matrix(rep(x[,i],q-2),ncol=q-2,byrow=FALSE)
   }
  nu_a_ab[i,]<-with(fit,{
                  colSums(X1*(pi_ij[,-q]*A*I+pi_ijp1*B*J))})
  if(q>2){
   nu_a_ap1b[i,]<- with(fit,{
             -colSums(X2*pi_ijp1[,-(q-1)]*B[,-(q-1)]*I[,-1])})
   nu_ap1_ab[i,]<- with(fit,{
             -colSums(X2*pi_ijp1[,-(q-1)]*A[,-1]*J[,-(q-1)])})
   }
 nu_b_aa[i,]<-with(fit,{
        colSums(X1*(pi_ij[,-q]*C[,-q]*D-pi_ijp1*C[,-1]*G))})
 if(q>2){
  nu_b_aap1[i,]<-with(fit,{
        colSums(X2*pi_ijp1[,-(q-1)]*C[,-c(1,q)]*H[,-(q-1)])})
  }
 nu_a_a_b[i,]<-with(fit,{-colSums(X1*(pi_ij[,-q]*P-pi_ijp1*Q[,-q]))-
                 2*nu_a_ab[i,]-nu_b_aa[i,]})
 if(q>2){
  nu_a_ap1_b[i,]<-with(fit,{
         -colSums(X2*pi_ijp1[,-(q-1)]*R[,-(q-1)])-
          nu_b_aap1[i,]-nu_a_ap1b[i,]-nu_ap1_ab[i,]})
```

```
  }
 for(j in 1:p){
  X1<-matrix(rep(x[,i],q-1),ncol=q-1,byrow=FALSE)
  X2<-matrix(rep(x[,j],q-1),ncol=q-1,byrow=FALSE)
  nu_b_ab[j,,i]<- with(fit,{
        colSums(X1*X2*(pi_ij[,-q]*C[,-q]*I-pi_ijp1*C[,-1]*J))})
  nu_a_bb[j,,i]<-with(fit,{
        colSums(X1*X2*(pi_ij[,-q]*A*K[,-q]-pi_ijp1*B*K[,-1]))})

   nu_a_b_b[j,,i]<-with(fit,{
        -colSums(X1*X2*(pi_ij[,-q]*S-pi_ijp1*T))-
         2*nu_b_ab[j,,i]-nu_a_bb[j,,i]})
   for(k in 1:p){
   X1<-matrix(rep(x[,i],q),ncol=q,byrow=FALSE)
   X2<-matrix(rep(x[,j],q),ncol=q,byrow=FALSE)
   X3<-matrix(rep(x[,k],q),ncol=q,byrow=FALSE)
   nu_b_bb[i,j,k]<- with(fit,{
                   sum(colSums(X1*X2*X3*pi_ij*C*K))})
   nu_b_b_b[i,j,k]<-with(fit,
                    {-sum(colSums(X1*X2*X3*pi_ij*U))-
                    3*nu_b_bb[i,j,k]})
   }
  }
 }

if(q>2){
 return(list(nu_a_aa=nu_a_aa,
           nu_a_aap1=nu_a_aap1,
           nu_a_ap1ap1=nu_a_ap1ap1,
           nu_ap1_aa= nu_ap1_aa,
           nu_ap1_aap1=nu_ap1_aap1,
           nu_a_ab=nu_a_ab,
           nu_a_ap1b=nu_a_ap1b,
```

```
             nu_ap1_ab=nu_ap1_ab,
             nu_b_aa=nu_b_aa,
             nu_b_aap1=nu_b_aap1,
             nu_b_ab=nu_b_ab,
             nu_a_bb=nu_a_bb,
             nu_b_bb=nu_b_bb,
             nu_a_a_a=nu_a_a_a,
             nu_a_a_ap1=nu_a_a_ap1,
             nu_a_ap1_ap1= nu_a_ap1_ap1,
             nu_a_a_b=nu_a_a_b,
             nu_a_ap1_b=nu_a_ap1_b,
             nu_a_b_b=nu_a_b_b,
             nu_b_b_b=nu_b_b_b))
 }
 if(q==2){
  return(list(nu_a_aa=nu_a_aa,
             nu_a_ab=nu_a_ab,
             nu_b_aa=nu_b_aa,
             nu_b_ab=nu_b_ab,
             nu_a_bb=nu_a_bb,
             nu_b_bb=nu_b_bb,
             nu_a_a_a=nu_a_a_a,
             nu_a_a_b=nu_a_a_b,
             nu_a_b_b=nu_a_b_b,
             nu_b_b_b=nu_b_b_b))
 }
}
```

Algoritmo Fisher-scoring

```
fisher.scoring<- function(x,y, start = NULL,
       type=c("logit","probit","cloglog"),maxiter=100,
       tolerance=10^(-10),criterion="ML"){
```

```
q<-ncol(y)
p<-ncol(x)
type<-match.arg(type)
if(is.null(start)) {
 start<-switch(type,
              logit =c(qlogis((1:(q-1))/q),rep(0,p)) ,
              probit =c(qnorm((1:(q-1))/q),rep(0,p)) ,
              cloglog =c(quantile_func((1:(q-1))/q),rep(0,p)))
}
step<-.Machine$integer.max
par<-start
for (k in 1:maxiter){
 par0<-par
 inv<-try(solve(info_fun(par,x,y,type)),TRUE)
 if(!is.character(inv)){
 info_inv<-inv
 }
 score<-score_fun(par,x,y,type)
 if(criterion=="BR" | criterion=="MBR"){
 nu<-nu_quantities(par,x,q,type)
 Q<-P<-array(0,dim=c((p+q-1),(p+q-1),(p+q-1)))

 for(i in 1:(q-1)){
  Q[i,i,i]<-nu$nu_a_aa[i]
  P[i,i,i]<-nu$nu_a_a_a[i]
  Q[i,q:(q+p-1),i]<- Q[q:(q+p-1),i,i]<-nu$nu_a_ab[,i]
  P[i,q:(q+p-1),i]<- P[q:(q+p-1),i,i]<-nu$nu_a_a_b[,i]

  for(j in 1:p){
   Q[q+j-1,q:(q+p-1),i]<-nu$nu_a_bb[,i,j]
   P[q+j-1,q:(q+p-1),i]<-nu$nu_a_b_b[,i,j]
   }
  }
```

```
if(q>2){
for(i in 1:(q-2)){
  Q[i,i+1,i]<-Q[i+1,i,i]<-nu$nu_a_aap1[i]
  P[i,i+1,i]<-P[i+1,i,i]<-P[i,i,i+1]<-nu$nu_a_a_ap1[i]
  Q[i+1,i+1,i]<-nu$nu_a_ap1ap1[i]
  P[i+1,i+1,i]<-nu$nu_a_ap1_ap1[i]
  Q[i,i,i+1]<-nu$nu_ap1_aa[i]
  Q[i,i+1,i+1]<-Q[i+1,i,i+1]<-nu$nu_ap1_aap1[i]
  P[i,i+1,i+1]<-P[i+1,i,i+1]<-nu$nu_a_ap1_ap1[i]
  Q[i+1,q:(q+p-1),i]<-Q[q:(q+p-1),i+1,i]<-nu$nu_a_ap1b[,i]
  Q[i,q:(q+p-1),i+1]<-Q[q:(q+p-1),i,i+1]<-nu$nu_ap1_ab[,i]
  P[i,q:(q+p-1),i+1]<-P[q:(q+p-1),i,i+1]<-nu$nu_a_ap1_b[,i]
  P[i+1,q:(q+p-1),i]<-P[q:(q+p-1),i+1,i]<-nu$nu_a_ap1_b[,i]
 }
}
for(i in 1:p){
 diag(Q[,,i+q-1])[1:q-1]<-nu$nu_b_aa[i,]
 diag(P[,,i+q-1])[1:q-1]<-nu$nu_a_a_b[i,]
 Q[1:(q-1),q:(p+q-1),i+q-1]<-t(nu$nu_b_ab[,,i])
 Q[q:(p+q-1),1:(q-1),i+q-1]<-nu$nu_b_ab[,,i]
 P[1:(q-1),q:(p+q-1),i+q-1]<-t(nu$nu_a_b_b[,,i])
 P[q:(p+q-1),1:(q-1),i+q-1]<-nu$nu_a_b_b[,,i]
 Q[q:(q+p-1),q:(q+p-1),i+q-1]<-nu$nu_b_bb[,,i]
 P[q:(q+p-1),q:(q+p-1),i+q-1]<-nu$nu_b_b_b[,,i]
 if(q>2){
  for(j in 1:(q-2)){
   Q[j,j+1,i+q-1]<-Q[j+1,j,i+q-1]<-nu$nu_b_aap1[i,j]
   P[j,j+1,i+q-1]<-P[j+1,j,i+q-1]<-nu$nu_a_ap1_b[i,j]
  }
 }
}

Astar1<-rep(0,q+p-1)
```

```
 for(i in 1:(q+p-1)){
  Astar1[i]<-0.5*sum(diag(info_inv%*%(Q[,,i]+P[,,i])))
  }
 }

 if(criterion=="MBR"){
  h<-array(0,dim=c((p+q-1),(p+q-1),(p+q-1)))
  for(i in 1:(q+p-1)){
   h[,,i]<-info_inv[,i]%*%t(info_inv[,i])/diag(info_inv)[i]
   }

  F_theta<-rep(0,q+p-1)
  for(i in 1:(q+p-1)){
    F_tilde<-rep(0,q+p-1)
    for(j in 1:(q+p-1)){
     F_tilde[j]<- sum(diag(h[,,i]%*%(P[,,j]/3+Q[,,j]/2)))
    }
    F_theta[i]<-t(info_inv[,i])%*%F_tilde
   }
  }

 stepPrev<-step
 stepFactor<-0
 testhalf<-TRUE
 if(criterion=="ML"){
  while(testhalf & stepFactor<11){
   par <- par+2^(-stepFactor)*(step<-info_inv%*% score)
  testhalf<-drop(crossprod(stepPrev)<crossprod(step))
  stepFactor<-stepFactor+1
  #print(stepFactor)
 }
}
if(criterion=="BR"){
```

```
  while(testhalf & stepFactor<11){
   par <- par+
       2^(-stepFactor)*(step<-info_inv%*%score+info_inv%*%Astar1)
   testhalf<-drop(crossprod(stepPrev)<crossprod(step))
   stepFactor<-stepFactor+1
  }
 }
 if(criterion=="MBR"){
 while(testhalf & stepFactor<11){
  par <- par+
       2^(-stepFactor)*(step<-info_inv%*%score+
                         info_inv%*%Astar1-F_theta)
  testhalf<-drop(crossprod(stepPrev)<crossprod(step))
  stepFactor<-stepFactor+1
  }
 }

 if(criterion=="ML"){
  if(all(abs(score_fun(par,x,y,type))<=tolerance)) break
 }
 if(criterion=="BR"){
  if(all(abs((score_fun(par,x,y,type)+Astar1))<=tolerance)) break
 }
 if(criterion=="MBR") {
  if(all(abs((score_fun(par,x,y,type)+Astar1-
     info_fun(par,x,y,type)%*%F_theta))<=tolerance)) break
  }
 }
  return(list(est=par,std=sqrt(diag(solve(info_fun(par,x,y,type)))),
             iterations=k,start=start))
}
```

Funzioni utili per gli studi di simulazione

```
#Generatore delle risposte
library(plyr)
get_simulate_data<-function(x,y,type="logit",
                   nsimu=10000,seed=1234){
 fit_ml<-fisher.scoring(y=t(y),
         x=x,criterion="ML",type=type)
 true_coefs <-fit_ml$est
 fitted_probabilities<-quantita_chiave(true_coefs,x,q,type)
 simulate <- function(fitted){
             freq <- sapply(1:nrow(x),
                     function(j) rmultinom(1, 1,
                     fitted_probabilities$pi_ij[j, ]))}
 set.seed(seed)
 seeds <- sample(seq.int(nsimu*1000), nsimu, replace = FALSE)
 datasets <- llply(seeds, function(seed) {
              set.seed(seed)
              dat <- simulate(fitted_probabilities$pi_ij[j, ])})
 return(list(ml_est=true_coefs,x=x,datasets=datasets))
}

#Simulazioni stime di massima verosimiglianza
simul_results_ML<-function(data,nsimu,type="logit",
                  criterion="ML",maxiter=500){
 est<-matrix(NA,nrow=nsimu,ncol=p+q-1)
 se<-matrix(NA,nrow=nsimu,ncol=p+q-1)
 conv<-rep(NA,nsimu)
 for(i in 1:nsimu){
  if(i%%100==0) print(i)
  current_data <- t(data$datasets[[i]])
  ncol0<-which(apply(current_data,2,sum)==0)
  if(is.integer0(ncol0)){
```

```
   ml_fit <- try(fisher.scoring(y=current_data,
                 x=data$x,criterion=criterion,
                 type=type,maxiter=maxiter),TRUE)
   if(!is.character(ml_fit)){
    est[i,]<-ml_fit$est
    se[i,]<-ml_fit$std
    conv[i]<-ml_fit$iterations
   }
  } else {
   ml_fit <- try(fisher.scoring(y=current_data[,-ncol0],
                 x=data$x,criterion=criterion,
                 maxiter=maxiter,type=type),TRUE)
   if(!is.character(ml_fit)){
    n.cell0<-length(ncol0)
    est[i,]<-c(ml_fit$est[1:(q-n.cell0-1)],rep(-9999, n.cell0),
                    ml_fit$est[(q-1):(q+p-n.cell0-1)])
    se[i,]<-c(ml_fit$std[1:(q- n.cell0-1)],rep(-9999, n.cell0),
                   ml_fit$std[(q-1):(q+p-n.cell0-1)])
    conv[i]<-ml_fit$iterations
    print(c("Warning on dataset",i,"Deleted",n.cell0,
            "columns of the response matrix due to zero counts"))
   }
  }
 }
 return(list(est=est,se=se,conv=conv,maxiter=maxiter))
}

#Simulationi stime con distorsione ridotta in media
simul_results_BR<-function(data,nsimu=10000,type="logit",
                   criterion="BR",mle.est=NULL){
 est<-matrix(NA,nrow=nsimu,ncol=p+q-1)
 se<-matrix(NA,nrow=nsimu,ncol=p+q-1)
 conv<-rep(NA,nsimu)
```

```
for(i in 1:nsimu){
 if(i%%100==0) print(i)
 current_data <- t(data$datasets[[i]])
 ncol0<-which(apply(current_data,2,sum)==0)
if(i %in% inf.est & !(is.integer0(inf.est))){
  if(is.integer0(ncol0)){
  br_fit <- try(fisher.scoring(y=current_data,
                x=data$x,criterion=criterion,
                type=type),TRUE)
  if(!is.character(br_fit)){
   est[i,]<-br_fit$est
   se[i,]<-br_fit$std
   conv[i]<-br_fit$iterations
  }
  next
 }else{
   if( (ncol0==1 |ncol0==q)){
    br_fit <- try(fisher.scoring(y=current_data,
                  x=data$x,criterion=criterion,
                 type=type),TRUE)
   if(!is.character(br_fit)){
    est[i,]<-br_fit$est
    se[i,]<-br_fit$std
    conv[i]<-br_fit$iterations
   }
   next
  } else {
   br_fit <- try(fisher.scoring(y=current_data[,-ncol0],
                 x=data$x,criterion=criterion,type=type),TRUE)
     if(!is.character(br_fit)){
     n.cell0<-length(ncol0)
     est[i,]<-c(br_fit$est[1:(q-n.cell0-1)], rep(-9999,n.cell0),
               br_fit$est[(q-1):(q+p-n.cell0-1)])
```

```
     se[i,]<-c(br_fit$std[1:(q-n.cell0-1)],rep(-9999,n.cell0),
              br_fit$std[(q-1):(q+p-n.cell0-1)])
     conv[i]<-br_fit$iterations
     print(c("Warning on dataset",i,"Deleted",n.cell0,
           "columns of the response matrix due to zero counts"))
    }
   next
   }
  }
 }
 if( (is.integer0(ncol0)) & !(i %in% inf.est)){
  if(!is.null(mle.est)){
   start<-mle.est$est[i,]
  } else {
   start<-NULL
  }
  br_fit <- try(fisher.scoring(y=current_data, x=data$x,
               criterion=criterion,start=start,
               type=type),TRUE)
  if(!is.character(br_fit)){
   est[i,]<-br_fit$est
   se[i,]<-br_fit$std
   conv[i]<-br_fit$iterations
  }
 } else {
  if(is.logical(ncol0==1 |ncol0==q)& (ncol0==1 |ncol0==q)){
   br_fit <- try(fisher.scoring(y=current_data,
                x=data$x,criterion=criterion,type=type),TRUE)
   if(!is.character(br_fit)){
    est[i,]<-br_fit$est
    se[i,]<-br_fit$std
    conv[i]<-br_fit$iterations
   }
```

```
   } else {
    br_fit <- try(fisher.scoring(y=current_data[,-ncol0],
                  x=data$x,criterion=criterion,type=type),TRUE)
    if(!is.character(br_fit)){
     n.cell0<-length(ncol0)
     est[i,]<-c(br_fit$est[1:(q-n.cell0-1)],rep(-9999,n.cell0),
                     br_fit$est[(q-1):(q+p-n.cell0-1)])
     se[i,]<-c(br_fit$std[1:(q-n.cell0-1)],rep(-9999,n.cell0),
                    br_fit$std[(q-1):(q+p-n.cell0-1)])
     conv[i]<-br_fit$iterations
     print(c("Warning on dataset",i,"Deleted",n.cell0,
            "columns of the response matrix due to zero counts"))
    }
   }
  }
 }
 return(list(est=est,se=se,conv=conv))
}

#Simulationi stime con distorsione ridotta in mediana
simul_results_MBR<-function(data,nsimu=10000,type="logit",
                   criterion="MBR",br.est=NULL){
 est<-matrix(NA,nrow=nsimu,ncol=p+q-1)
 se<-matrix(NA,nrow=nsimu,ncol=p+q-1)
 conv<-rep(NA,nsimu)
 for(i in 1:nsimu){
  if(i%%100==0) print(i)
  current_data <- t(data$datasets[[i]])
  ncol0<-which(apply(current_data,2,sum)==0)
  if(i %in% inf.est & !(is.integer0(inf.est))){
   if(is.integer0(ncol0)){
   mbr_fit <- try(fisher.scoring(y=current_data,
                 x=data$x,criterion=criterion,
```

```
                type=type),TRUE)
   if(!is.character(mbr_fit)){
    est[i,]<-mbr_fit$est
    se[i,]<-mbr_fit$std
    conv[i]<-mbr_fit$iterations
   }
   next
  }else{
    if( (ncol0==1 |ncol0==q)){
     mbr_fit <- try(fisher.scoring(y=current_data,
                   x=data$x,criterion=criterion,
                  type=type),TRUE)
    if(!is.character(mbr_fit)){
     est[i,]<-mbr_fit$est
     se[i,]<-mbr_fit$std
     conv[i]<-mbr_fit$iterations
    }
    next
   }else{
    mbr_fit <- try(fisher.scoring(y=current_data[,-ncol0],
                  x=data$x,criterion=criterion,type=type),TRUE)
      if(!is.character(mbr_fit)){
      n.cell0<-length(ncol0)
      est[i,]<-c(mbr_fit$est[1:(q-n.cell0-1)],rep(-9999,n.cell0),
               mbr_fit$est[(q-1):(q+p-n.cell0-1)])
      se[i,]<-c(mbr_fit$std[1:(q-n.cell0-1)],rep(-9999,n.cell0),
              mbr_fit$std[(q-1):(q+p-n.cell0-1)])
      conv[i]<-mbr_fit$iterations
      print(c("Warning on dataset",i,"Deleted",n.cell0,
              "columns of the response matrix due to zero counts"))
     }
   next
   }
```

```
 }
 }
 if( (is.integer0(ncol0)) & !(i %in% inf.est)){
  if(!is.null(br.est)){
   start<-br.est$est[i,]
  } else {
   start<-NULL
  }
  mbr_fit <- try(fisher.scoring(y=current_data,x=data$x,
                 criterion=criterion,start=start,
                 type=type),TRUE)
  if(!is.character(mbr_fit)){
   est[i,]<-mbr_fit$est
   se[i,]<-mbr_fit$std
   conv[i]<-mbr_fit$iterations
  }
 } else {
  if(is.logical((ncol0==1 |ncol0==q)) &(ncol0==1|ncol0==q)){
   if(!is.null(br.est)){
    start<-br.est$est[i,]
   } else {
    start<-NULL
   }
   mbr_fit <- try(fisher.scoring(y=current_data,
                 x=data$x,criterion=criterion,
                 type=type),TRUE)
   if(!is.character(mbr_fit)){
    est[i,]<-mbr_fit$est
    se[i,]<-mbr_fit$std
    conv[i]<-mbr_fit$iterations
   }
  } else {
   mbr_fit <- try(fisher.scoring(y=current_data[,-ncol0],
```

```
                   x=data$x,criterion=criterion,
                   type=type),TRUE)
    if(!is.character(mbr_fit)){
     n.cell0<-length(ncol0)
     est[i,]<-c(mbr_fit$est[1:(q-n.cell0-1)],rep(-9999,n.cell0),
                 mbr_fit$est[(q-1):(q+p-n.cell0-1)])
     se[i,]<-c(mbr_fit$std[1:(q-n.cell0-1)],rep(-9999,n.cell0),
                    mbr_fit$std[(q-1):(q+p-n.cell0-1)])
     conv[i]<-mbr_fit$iterations
     print(c("Warning on dataset",i,"Deleted",n.cell0,
          "columns of the response matrix due to zero counts"))
    }
   }
  }
 }
 return(list(est=est,se=se,conv=conv))
}

#Diagnostica per valori infiniti o problemi numerici
diagnostics<-function(results,maxiter=500){
 ind_inf<-integer(0)
 index_NA<- which(is.na(results$est[,1]))
 index_not_conv<-which(results$conv==maxiter)
 index_NA_not_conv<-c(index_NA,index_not_conv)
 deleted_column<- which(apply(results$est==-9999,2,sum)!=0)
 if(is.integer0(index_NA_not_conv)) {
  infiniteEstimates<-(apply(results$est,2,function(x) any(abs(x)>20))|
                      apply(results$se,2,function(x) any(abs(x)>100)))
  if(sum(infiniteEstimates)>0){
   ind_inf<-which(abs(results$est[,1])>20 | results$se[,1]>100 )
   if(is.integer0(deleted_column)){
    for(i in 2:(p+q-1))
    ind_inf<-sort(union(ind_inf,which(abs(results$est[,i])>20|
```

```
                                             results$se[,i]>100)))
   } else {
    for(i in 2:(p+q-1)){
     if(i!=deleted_column){
      ind_inf<-sort(union(ind_inf,which(abs(results$est[,i])>20|
                                              results$se[,i]>100)))
     }
    }
   }
  }
 inf.est<-sort(ind_inf)
 } else {
  infiniteEstimates <- apply(results$est[-index_NA_not_conv,], 2,
                      function(x) any(abs(x) > 20)) |
                      apply(results$se[-index_NA_not_conv,], 2,
                      function(x) any(abs(x) > 100))
  if(sum(infiniteEstimates)>0 ){
   ind_inf<-which(abs(results$est[,1])>20 |
                     results$se[,1]>100 )
   for(i in 2:(p+q-1))
   ind_inf<-sort(union(ind_inf,which(abs(results$est[,i])>20 |
                                         results$se[,i]>100 )))
  }
  if(sum(infiniteEstimates)>0 ){
   inf.est<-sort(union(index_NA_not_conv,ind_inf))
   } else {
   inf.est<-sort(index_NA_not_conv)
  }
  inf.est
 }
}

#Risultati simulazione
```

```
estimator_results<-function(data,results,inf.est,
                            nsimu=10000,levels=0.05){
 trueparmat<-matrix(rep(data$ml_est,nsimu),nsimu,
                    p+q-1,byrow=TRUE)
 fin.est<-100*(nsimu-length(inf.est))/nsimu
 if(is.integer0(inf.est)){
  bias<-round(colMeans(results$est[,q:(q+p-1)]-
                       trueparmat[,q:(q+p-1)]),3)
  PU<-100*colMeans(results$est[,q:(q+p-1)]<=
                   trueparmat[,q:(q+p-1)])
  RMSE<-round(sqrt(colMeans((results$est[,q:(q+p-1)]-
                             trueparmat[,q:(q+p-1)])^2)),3)
  z0.975<-matrix(qnorm(1-levels/2),nsimu,p+q-1)
  coverage<-colMeans((abs((results$est[,q:(q+p-1)]-
                           trueparmat[,q:(q+p-1)])/
                           results$se[,q:(q+p-1)])<=
                           z0.975[,q:(q+p-1)]))
  } else {
  bias<-round(colMeans(results$est[-inf.est,q:(q+p-1)]-
                       trueparmat[-inf.est,q:(q+p-1)]),3)
  PU<-100*colMeans(results$est[-inf.est,q:(q+p-1)]<=
                   trueparmat[-inf.est,q:(q+p-1)])
  RMSE<-round(sqrt(colMeans((results$est[-inf.est,q:(q+p-1)]-
                             trueparmat[-inf.est,q:(q+p-1)])^2)),3)
  z0.975<-matrix(qnorm(1-levels/2),nsimu,p+q-1)
  coverage<-colMeans((abs((results$est[-inf.est,q:(q+p-1)]-
                           trueparmat[-inf.est,q:(q+p-1)])/
                           results$se[-inf.est,q:(q+p-1)])<=
                           z0.975[-inf.est,q:(q+p-1)]))
 }
 return(list(fin.est=fin.est,bias=bias,PU=PU,
             RMSE=RMSE,coverage=coverage))
}
```

```
#Distorsione e distorsione relativa nella nuova parametrizzazione
gamma_bias<-function(beta_true,res,inf.est,
                     type=c("logit","probit","cloglog")){
 type<-match.arg(type)
 ripar<-switch(type,
        logit =function(x) exp(-x/sqrt(2))/(1+exp(-x/sqrt(2))) ,
        probit =function(x) pnorm(-x/sqrt(2)),
        cloglog =function(x) exp(-x)/(1+exp(-x)))
 true_gamma<-ripar(beta_true)
 if(is.integer0(inf.est)){
  bias_gamma<-round(abs(mean(ripar(res)- true_gamma)),4)
 } else {
  bias_gamma<-round(abs(mean(ripar(res[-inf.est])- true_gamma)),4)
 }
 relative_bias<-100*bias_gamma/true_gamma
 return(list(bias_gamma=bias_gamma,relative_bias=relative_bias))
}
```

Esempio di studio di simulazione

```
set.seed(25)
W<-10000
n<-60
x1<-rnorm(n)
x2<-sample(0:1,n,replace=TRUE)
x<-model.matrix(~x1+x2)[,-1]
y<-rmultinom(n,1,prob=c(1/3,1/3,1/3))
q<-ncol(t(y))
p<-ncol(x)

data<-get_simulate_data(x,y,nsimu=W,type="logit")
results_ML<-simul_results_ML(data,nsimu=W,type="logit",
```

```
                             criterion="ML",maxiter=500)
inf.est<-diagnostics(results_ML)
res<-estimator_results(data,results_ML,inf.est,
                       nsimu=W,levels=0.05)
bias_new_par<-gamma_bias(as.vector(data$ml_est)[4],
                         results_ML$est[,4],
                         inf.est,type="logit")

results_BR<-simul_results_BR(data,nsimu=W,type="logit",
                             criterion="BR",
                             mle.est=results_ML)
inf.est_BR<-diagnostics(results_BR)
res_br<-estimator_results(data,results_BR,inf.est_BR,
                          nsimu=10000,levels=0.05)
bias_new_par_br<-gamma_bias(as.vector(data$ml_est)[4],
                            results_BR$est[,4],
                            inf.est_BR,type="logit")

results_MBR<-simul_results_BR(data,nsimu=W,type="logit",
                              criterion="MBR",
                              br.est=results_BR)
inf.est_MBR<-diagnostics(results_MBR)
res_mbr<-estimator_results(data,results_MBR,inf.est_MBR,
                           nsimu=W,levels=0.05)
bias_new_par_mbr<-gamma_bias(as.vector(data$ml_est)[4],
                             results_MBR$est[,4],
                             inf.est_MBR,type="logit")
```

Bibliografia

[1] Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. 2nd ed. Wiley.

[2] Agresti, A. e Kateri, M. (2017). Ordinal probability effect measures for group comparisons in multinomial cumulative link models. *Biometrika*, **73**, 214-219.

[3] Albert, A. e Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, **71**, 1-10.

[4] Azzalini, A. (1996). *Statistical Inference Based on the Likelihood*. Chapman & Hall.

[5] Azzalini, A. e Scarpa, B. (2012). *Data Analysis and Data Mining*. Oxford University Press.

[6] Christensen, R. H. B. (2019a). ordinal - regression models for ordinal data. *R package version 2019.4-25*. (disponibile a http://www.cran.r-project.org/package=ordinal/).

[7] Christensen, R. H. B. (2019b). *Cumulative link models for ordinal regression with the R package Ordinal*. (disponibile a http://cran.uni-muenster.de/web/packages/ordinal/vignettes/clm_article.pdf).

[8] Cox, D. R. (1972). Regression models and life tables *Journal of the Rojal Statistical Society Series B*, **34**, 187-220.

[9] Decensi, A., Fontana, V., Bruno, S., Gustavino, C., Gatteschi, B. e Costa, A. (1996). Effect of tamoxifen on endometrial proliferation. *Journal of Clinical Oncology*, **14**, 434-440.

[10] Firth, D. (1993). Bias reduction of maximum likelihhod estimates. *Biometrika*,**80**, 27-38.

[11] Jackman, S. (2004). What do we learn from graduate admission committes?: a multiple rater, latent variable model, with incomplete discrete and continuous indicators. *Political Analysis.* **12**, 400-424.

[12] Jackman, S. (2017). pscl: Classes and methods for R developed in the Political Science Computational Laboratory. United States Studies Centre, University of Sydney. Sydney, New South Wales, Australia. *R package version 1.5.2.* (disponibile a https://github.com/atahk/pscl/).

[13] Kenne Pagui, E. C. , Salvan, A. e Sartori, N. (2017). Median bias reduction of maximum likelihhod estimates. *Biometrika*, **104**, 4, 923-938.

[14] Kenne Pagui, E. C. , Salvan, A. e Sartori, N. (in fase di elaborazione, 2019). Efficient implementation of median bias reduction.

[15] Kosmidis, I. (2007). *Bias Reduction in Exponential Family Nonlinear Models.* PhD Thesis. Department of Statistics, University of Warwick, Coventry. (Disponibile a http://www.ikosmidis.com/files/ikosmidis_thesis.pdf).

[16] Kosmidis, I. (2014a). Bias in parametric estimation: reduction and useful side-effects. *Wiley Interdisciplinary Reviews: Computational Statistics*, **6**, 185–196.

[17] Kosmidis, I. (2014b). Improved estimation in cumulative link models. *Journal of the Royal Statistical Society Series B*, **76**, 169–196.

[18] Lesaffre, E. e Albert, A. (1989). Partial separation in logistic discrimination. *Journal of the Royal Statistical Society Series B*, **51**, 109–116.

[19] McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society Series B*, **42** , 109–142.

[20] Oja, H. (2013). Multivariate median, In *Robustness and Complex Data Structures*, C. Becker, R. Fried e S. Kuhnt, eds. Berlin: Springer, pp. 3-15.

[21] Pace, L. e Salvan, A. (1997). *Principles of Statistical Inference from a Neo-fisherian Perspective*, vol. 4. World Scientific Pub Co Inc.

[22] Pace, L. e Salvan, A. (2001). *Introduzione alla Statistica II - Inferenza, Verosimiglianza, Modelli.* Cedam, Padova.

[23] Pfanzagl, J. (1970). On the asymptotic efficiency of median unbiased estimates. *Annals Mathematical Statistics.* **41**, 1500-1509.

[24] Peterson, B. e Harrell, Jr. F. E. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics*, **39**, 205-217.

[25] Pratt, J. W. (1981). Concavity of the log likelihood. *Journal of Americal Statistical Association*, **76**, 103-106; correzioni, **77** (1982), 954.

[26] Read, C. B. (1985). Median unbiased estimators. In *Encyclopedia of Statistical Sciences*, S. Kots, N. Johnson and C. Read, eds., Vol. 5. New York: Wiley, pp. 424-426.

[27] Salvan, A., Sartori, N. e Pace, L. (2018). *Modelli Statistici 2.* Materiale didattico per l'insegnamento di Modelli Statistici 2, Corsi di Laurea in Scienze Statistiche, Università di Padova.

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

[29] van der Vaart, W.A. (1998). *Asymptotic Statistics.* Cambridge University Press.

[30] Wood, S. N. (2015). *Core Statistics.* Cambridge University Press, Institute of Mathematical Statistics Textbooks.

Ringraziamenti

All'alba del trentesimo anno di età posso dire di aver ricevuto tanto dalle persone. Si presenta qui l'occasione per poter dedicare loro qualche piccolo pensiero. Il primo pensiero va ai miei genitori che hanno saputo trasmettermi l'amore per ogni cosa intrapresa e scolastica e lavorativa. C'è tutto il loro desidero nella vita in questo mio percorso.
Un pensiero va a mio fratello da cui ho imparato tanto. Mi ha insegnato a nutrire di pazienza e di fiducia i miei sogni anche quando la delusione e la tristezza reclamavano il loro spazio, ad essere inclusivi piuttosto che esclusivi nei rapporti con le persone e che il singolo individuo è più importante delle moltitudini. Sembrerà strano detto da uno studente di Statistica. Non dimentico mia cognata e i miei bellissimi nipotini, capaci di trasportarmi nel mondo della fanciullezza. Un grazie di cuore ai miei cugini Tina e Domenico e alle loro famiglie. Se ho intrapreso questo percorso di studi a Genova è anche grazie a loro. Un pensiero va anche al resto della famiglia, agli zii e ai cugini.
Un pensiero particolare va al mio relatore, la Professoressa Salvan, e al mio correlatore, il Professore Kenne Pagui, a cui ho cercato di rubare il mestiere dello statistico con gli occhi. Ricorderò la loro passione e dedizione per l'insegnamento, la loro precisione e il loro interesse per la ricerca. Spero di aver fatto tesoro di tutti gli insegnamenti ricevuti e che mi siano di aiuto nei futuri percorsi lavorativi. La scelta di seguire un insegnamento della Professoressa Salvan e di svolgere il lavoro di tesi con lei si sono rivelate tra le più belle del lungo percorso universitario che mi porto sulle spalle. Del Professore Kenne Pagui ho avvertito la fiducia sin dai primi giorni che abbiamo iniziato a discutere del lavoro di tesi. È stato un onore aver lavorato a questo argomento di ricerca e spero di aver ripagato la fiducia con questo mio piccolo contributo.

Ringrazio, inoltre, tutti i docenti del Dipartimento di Statistica che ho avuto la fortuna di conoscere. Di ognuno di loro ho cercato di imparare il meglio che potessero darmi e mi accompagnerà il loro ricordo. Non dimentico anche tutte le persone che lavorano all'interno del Dipartimento: i tecnici informatici, i responsabili della biblioteca e tutto il personale. Li ricorderò per la loro cordialità e disponibilità.
Mi preme ringraziare, inoltre, il Professor Vincenzo Fontana, mio relatore della tesi di laurea triennale e responsabile dell'Ufficio Statistico dell'IRCCS AOU San Martino-IST, e per la sua disponibilità nel reperirmi i dati che ho analizzato nell'esempio di applicazione presentato nella tesi e per tutti i consigli ricevuti da lui. Desidero ringraziare i professori di SMID e in modo particolare le Professoresse Rogantin e Riccomagno che in questi anni di studio a Padova si sono interessate al mio percorso.
Non dimentico gli amici. Un grazie di cuore va a tutti gli amici castelluccesi, lucani, italiani e di ogni nazionalità che ho avuto la fortuna di incontrare durante il mio percorso scolastico ed extra-scolastico e che hanno saputo rendere felice il tempo libero.
Ringrazio, infine, il mio spirito per avermi sostenuto nel raggiungere il traguardo della laurea.