

Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Magistrale in  
Scienze Statistiche



**METODI ALTERNATIVI  
ALLA REGRESSIONE LOG-BINOMIALE  
PER L'INFERENZA SUL RISCHIO RELATIVO**

Relatore Prof. Alessandra Salvan  
Dipartimento di Scienze Statistiche

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

Laureando Francesco Pozza  
Matricola N 1206982

Anno Accademico 2019/2020



# Indice

<b>Introduzione</b>	<b>ix</b>
<b>1 Teoria della verosimiglianza e metodi per la riduzione della distorsione</b>	<b>1</b>
1.1 Inferenza statistica e verosimiglianza . . . . .	2
1.1.1 Modello statistico parametrico . . . . .	2
1.1.2 Stime, stimatori e relative proprietà . . . . .	2
1.1.3 Funzione di verosimiglianza e stima di massima verosimiglianza . . . . .	4
1.1.4 Informazione osservata e risultati distributivi ad essa collegati . . . . .	5
1.1.5 Relazioni di Bartlett . . . . .	6
1.2 Metodi di riduzione della distorsione . . . . .	7
1.2.1 Riduzione della distorsione in media . . . . .	8
1.2.2 Riduzione della distorsione in mediana . . . . .	10
<b>2 Modelli di regressione per il rischio relativo</b>	<b>13</b>
2.1 Notazione . . . . .	14
2.2 Modelli lineari generalizzati e stima di $\log(\text{RR}_i)$ . . . . .	14
2.2.1 Modello logistico . . . . .	15
2.2.2 Regressione log-binomiale . . . . .	18
2.2.3 Modello di Poisson e quasi-verosimiglianza . . . . .	18
2.3 Il modello proposto da Richardson <i>et al.</i> (2017) . . . . .	20

2.3.1	Esempio: confronto tra due gruppi . . . . .	21
2.3.2	Verifica della proprietà di variazione indipendente tra $\eta_{i1}$ e $\eta_{i2}$ . . . . .	23
2.3.3	Interpretazione dei parametri e relazione con il modelli log-binomiale e logistico . . . . .	23
2.3.4	Parametrizzazione alternativa . . . . .	25
<b>3</b>	<b>Quantità per la riduzione della distorsione in media e media- na</b>	<b>29</b>
3.1	Quantità di verosimiglianza . . . . .	29
3.1.1	Derivate della funzione di log-verosimiglianza . . . . .	29
3.1.2	Prolungamento per continuità delle derivate di $\pi_{i0}$ . . .	32
3.1.3	Informazione attesa di Fisher . . . . .	33
3.2	Riduzione della distorsione in media e mediana . . . . .	33
3.2.1	Definizione di $P_{\gamma_k}(\theta)$ , $P_{\beta_k}(\theta)$ , $Q_{\gamma_k}(\theta)$ e $Q_{\beta_k}(\theta)$ . . . .	33
3.2.2	Definizione di $F_1$ , $F_{2\gamma_j}$ e $F_{2\beta_j}$ . . . . .	36
3.2.3	Equazioni <i>score</i> modificate . . . . .	38
3.3	Parametrizzazione alternativa per la componente di disturbo .	39
<b>4</b>	<b>Studi di simulazione ed applicazione dei modelli ad un caso reale</b>	<b>41</b>
4.1	Scelta di $\gamma$ e $\beta$ . . . . .	42
4.2	Simulazioni con uguale numero di trattati e non trattati . . .	43
4.3	Simulazioni con numero di trattati e non trattati sbilanciato .	47
4.4	Caso studio . . . . .	51
	<b>Conclusioni</b>	<b>56</b>
<b>A</b>	<b>Dettagli tecnici</b>	<b>57</b>
A.1	Legame tra $\pi_{i0}$ , $\pi_{i1}$ e i parametri $\eta_{i1}$ e $\eta_{i2}$ . . . . .	57
A.2	Legame tra $\pi_{i0}$ , $\pi_{i1}$ e i parametri $\eta_{i1}$ e $\eta_{i3}$ . . . . .	60

<b>B Codice R utilizzato</b>	<b>63</b>
B.1 Funzione per la stima dei modelli sviluppati nella tesi . . . . .	63
B.2 Funzione per gli studi di simulazione . . . . .	81
<b>Bibliografia</b>	<b>97</b>



# Introduzione

L'obiettivo di molti studi epidemiologici è determinare in che modo un insieme di variabili esplicative influenzi una variabile binaria di interesse. In tale contesto la regressione logistica è di gran lunga il modello statistico più utilizzato. Ciò implica che l'effetto delle variabili esplicative venga parametrizzato in termini di log-rapporto delle quote. Questa scelta è in qualche modo obbligata in alcuni disegni sperimentali, come negli studi caso-controllo. In numerose altre situazioni è tuttavia spesso desiderabile valutare l'effetto di un'esplicativa, al netto delle altre variabili esplicative disponibili, in termini di rischio relativo in quanto più facilmente interpretabile (si veda per esempio McNutt *et al.* 2003). Sfortunatamente, la regressione log-binomiale, i cui coefficienti parametrizzano direttamente il logaritmo del rischio relativo, è spesso non utilizzabile nella pratica in quanto afflitta da gravi problemi di instabilità numerica (Barros e Hirakata, 2003). Per questa ragione, negli ultimi decenni, vi è stato un crescente interesse riguardo lo sviluppo di metodi per la stima del rischio relativo. Una soluzione particolarmente interessante, nel caso in cui le procedure inferenziali siano focalizzate sull'effetto di un fattore di esposizione binario e sulle sue interazioni con le altre variabili esplicative, è stato sviluppato da Richardson *et al.* (2017). Questo metodo si contraddistingue dagli altri proposti per via di una diversa parametrizzazione del modello, dove la parte di interesse e quella di disturbo sono tra loro con variazione indipendente. In questo modo è possibile ottenere facilmente la stime di massima verosimiglianza e i relativi *standard error* senza porre vincoli sui parametri.

L'obiettivo di questa tesi è quello di migliorare le proprietà dello stimatore proposto da Richardson *et al.* (2017) sotto vari punti di vista. In primo luogo, dato che i modelli per l'inferenza sul rischio relativo sono spesso utilizzati in presenza di numerosità campionarie limitate e di un numero elevato di variabili esplicative, lo stimatore di massima verosimiglianza potrebbe essere soggetto ad una distorsione non trascurabile. Per un'inferenza più accurata sono stati, quindi, implementati i metodi di riduzione della distorsione in media e mediana proposti da Firth (1993) e Kenne Pagui *et al.* (2017). Queste metodologie permettono, infatti, di migliorare le proprietà distributive dello stimatore di massima verosimiglianza focalizzandosi sulla riduzione della distorsione in media e mediana. Inoltre, tali metodi non dipendono dall'esistenza dello stimatore di massima verosimiglianza e, anzi, sono spesso efficaci nel prevenire stime infinite.

Un altro aspetto al quale è stata dedicata molta attenzione è lo sviluppo di una parametrizzazione alternativa per la componente di disturbo, più facile da trattare matematicamente rispetto a quella specificata da Richardson *et al.* (2017). Anche in questo frangente sono stati derivati gli stimatori con distorsione in media e mediana ridotti.

Il Capitolo 1 è dedicato ad una schematica rassegna delle nozioni di base dell'inferenza statistica. Vengono, inoltre, introdotti i metodi per la riduzione della distorsione in media e mediana dello stimatore di massima verosimiglianza, con particolare riguardo al metodo proposto da Firth (1993) e a quello sviluppato da Kenne Pagui *et al.* (2017).

Nel Capitolo 2 sono descritti alcuni modelli per l'inferenza sul rischio relativo. Tra questi particolare attenzione è stata riservata al modello proposto da Richardson *et al.* (2017) del quale si analizzano aspetti interpretativi e proprietà matematiche. Al termine del capitolo viene anche proposta ed analizzata nel dettaglio una diversa specificazione del parametro di disturbo.

Nel Capitolo 3 vengono derivate le quantità necessarie per implementare, nel modello di Richardson *et al.* (2017) e nel modello con parametro di disturbo alternativo, i metodi di riduzione della distorsione in media e mediana.



Il Capitolo 4 è dedicato ad un confronto empirico, basato su vari studi di simulazione, delle proprietà distributive degli stimatori ottenuti.

Infine, nell'Appendice A vengono riportati alcuni risultati matematici richiamati nel Capitolo 2. Il codice, scritto con l'ausilio del programma R 3.6.2 ed utilizzato per le simulazioni del Capitolo 4 è riportato nell'Appendice B.



# Capitolo 1

## Teoria della verosimiglianza e metodi per la riduzione della distorsione

In questo capitolo sono richiamate alcune nozioni di base dell'inferenza statistica, le quali risulteranno necessarie nei capitoli successivi della relazione. In particolare, dopo aver definito il concetto di modello statistico parametrico, vengono delineate le nozioni di stima e stimatore. Nei paragrafi successivi viene invece introdotta la teoria della verosimiglianza.

Infine, la parte conclusiva del capitolo è una schematica introduzione ai metodi di riduzione della distorsione in media e mediana dello stimatore di massima verosimiglianza. Particolare attenzione in questo frangente viene dedicata al metodo di Firth (1993) e al metodo di riduzione della distorsione in mediana proposto da Kenne Pagui *et al.* (2017).

I riferimenti principali per il paragrafo 1.1 sono Pace e Salvani (1996, 2001). Il paragrafo 1.2 ha invece come riferimenti principali Kosmidis (2014), Firth (1993) e Kenne Pagui *et al.* (2017).

## 1.1 Inferenza statistica e verosimiglianza

### 1.1.1 Modello statistico parametrico

L'intera teoria dell'inferenza statistica si basa sull'assunzione fondamentale che i dati osservati  $y = y^{oss}$ , relativi ad un determinato fenomeno di interesse, possano essere considerati come la realizzazione di un vettore casuale  $Y$ . Ad  $Y$  è associato uno spazio campionario  $\mathcal{Y}$  e una funzione di probabilità (o di densità nel caso continuo) indicata con  $p_Y^0(y)$ . Dato che, nella maggior parte dei casi la forma di  $p_Y^0(y)$  non è nota a priori, essa deve essere in qualche modo dedotta a partire dai dati stessi. Si definisce quindi **modello statistico**, in seguito indicato con  $\mathcal{F}$ , una famiglia di funzioni di probabilità compatibile con la generazione di  $y^{oss}$ . Se  $p_Y^0(y) \in \mathcal{F}$  il modello statistico viene detto **correttamente specificato**. Inoltre, se gli elementi di  $\mathcal{F}$  possono essere individuati attraverso un parametro  $\theta \in \Theta$ , ossia

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

e  $\Theta \subseteq \mathbb{R}^p$ , per qualche  $p \in \mathbb{N}^+$ , il modello statistico viene detto **parametrico**. Nel caso contrario si parla di **modello statistico non parametrico**. Se esiste una corrispondenza biunivoca tra  $\Theta$  e gli elementi di  $\mathcal{F}$ , il parametro  $\theta$  viene detto **identificabile**, nel caso contrario il modello è **sovraparametrizzato**. In modelli parametrici con parametro identificabile si ha perciò che  $p_Y^0(y) = p_Y(y; \theta_0)$  per un unico  $\theta_0 \in \Theta$  detto **vero valore del parametro**. In questo caso, è quindi equivalente fare inferenza su  $p_Y^0(y)$  o su  $\theta_0$ .

### 1.1.2 Stime, stimatori e relative proprietà

In seguito alla specificazione di un modello statistico  $\mathcal{F}$ , con spazio parametrico  $\Theta$ , risulta naturale voler identificare un valore per il parametro  $\theta$  che più di altri sia compatibile con i dati osservati. Nella teoria della stima puntuale, tale obiettivo viene raggiunto mediante l'utilizzo di una opportuna applicazione  $\hat{\theta} : \mathcal{Y} \rightarrow \Theta$ , la quale assegna a ciascun punto dello spazio cam-

pionario  $\mathcal{Y}$  un unico punto dello spazio parametrico  $\Theta$ . Il valore assunto da questa applicazione sui dati osservati, ossia  $\hat{\theta}(y)$ , viene detto **stima** di  $\theta_0$ . La variabile casuale  $\hat{\theta}(Y)$  viene invece detta **stimatore** di  $\theta$ . Dato che di norma la scelta di uno stimatore per  $\theta$  non è univoca, può essere utile effettuare un confronto tra stimatori diversi valutando la loro distribuzione sotto  $\theta_0$ . Si definisce quindi **errore sistematico di stima** la quantità

$$b(\theta) = E_{\theta}(\hat{\theta} - \theta),$$

dove si intende con  $\theta$  un generico valore di  $\theta_0$ .

Uno stimatore viene detto **non distorto** in media se è soddisfatta la condizione  $b(\theta) = 0$ , per ogni  $\theta \in \Theta$ . Il concetto di non distorsione può essere reso meno stringente introducendo la nozione di **non distorsione asintotica**. Uno stimatore viene detto asintoticamente non distorto se

$$\lim_{n \rightarrow +\infty} E_{\theta}(\hat{\theta} - \theta) = 0, \quad \theta \in \Theta.$$

Il concetto di non distorsione di uno stimatore può essere esteso anche a quantità diverse dalla media. In particolare, se il parametro  $\theta$  è scalare e la distribuzione di  $\hat{\theta}(Y)$  è continua, lo stimatore viene definito **non distorto in mediana** se

$$P_{\theta}(\hat{\theta} < \theta) = 1/2, \quad \theta \in \Theta. \quad (1.1)$$

Nel caso in cui  $\theta$  sia un parametro  $p$ -dimensionale, uno stimatore si dice non distorto in mediana componente per componente se l'uguaglianza

$$P_{\theta}(\hat{\theta}_j < \theta_j) = 1/2,$$

dove  $\theta_j$ ,  $j = 1, \dots, p$ , rappresenta la  $j$ -esima componente di  $\theta$ , vale per ogni  $\theta \in \Theta$ .

Un'altra quantità, utile per valutare le proprietà campionarie di uno stimatore, è l'**errore quadratico medio** (*Mean Squared error*), definito come

$$\text{MSE}(\theta) = E_{\theta} [(\hat{\theta} - \theta)^2] = b(\theta)^2 + \text{Var}_{\theta}(\hat{\theta}). \quad (1.2)$$

Infine, uno stimatore  $\hat{\theta}$  è detto **consistente** per  $\theta$  se

$$\lim_{n \rightarrow +\infty} P_{\theta} \left( |\hat{\theta} - \theta| \geq \varepsilon \right) = 0,$$

quale sia  $\theta \in \Theta$ . Una condizione sufficiente, affinché uno stimatore sia consistente, è che l'errore quadratico medio tenda a zero, al crescere della numerosità campionaria  $n$ .

### 1.1.3 Funzione di verosimiglianza e stima di massima verosimiglianza

Sia  $\mathcal{F}$  un modello statistico parametrico per i dati  $y$ , avente funzione del modello  $p_Y(y; \theta)$  con  $\theta \in \Theta \subseteq \mathbb{R}^p$ . Se si considera  $p_Y(y; \theta)$  funzione di  $\theta$  con  $y$  fissato, la funzione  $L : \Theta \rightarrow [0, +\infty)$  definita come

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

è detta **funzione di verosimiglianza** (*likelihood function*). Attraverso la funzione di verosimiglianza è possibile identificare quelle configurazioni del parametro  $\theta$  maggiormente supportate dall'evidenza empirica fornita da  $y$ . In particolare, dati due possibili valori parametrici  $\theta_1, \theta_2 \in \Theta$ , un valore di  $L(\theta_1)/L(\theta_2)$  superiore ad 1 esprime un maggiore evidenza sperimentale a supporto di  $\theta_1$  come parametro del modello rispetto a  $\theta_2$ . L'interpretazione è invertita per valori di  $L(\theta_1)/L(\theta_2)$  minori di 1. Il confronto tra  $L(\theta_1)$  e  $L(\theta_2)$ , basato sul rapporto  $L(\theta_1)/L(\theta_2)$  implica che nella funzione di verosimiglianza possano essere eliminati tutti quei fattori che non dipendono da  $\theta$ .

Nel caso in cui il modello statistico ipotizzato per i dati  $y = (y_1, \dots, y_n)$  assuma che questi siano la realizzazione di  $n$  variabili casuali indipendenti  $Y_i, i = 1, \dots, n$ , con distribuzione marginale  $p_{Y_i}(y_i; \theta)$ , la funzione di verosimiglianza assumerà la forma

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

Spesso, per quanto riguarda le procedure inferenziali su  $\theta$ , risulta più facile trattare la **funzione di log-verosimiglianza**, data dal logaritmo naturale della funzione di verosimiglianza,  $\ell(\theta) = \log L(\theta)$ , con  $\ell(\theta) = -\infty$  se  $L(\theta) = 0$ . Infatti, in quanto funzione monotona crescente, l'applicazione del logaritmo naturale garantisce che le regioni dello spazio parametrico con verosimiglianza elevata mantengono anche una log-verosimiglianza elevata, nel contempo, però, la derivazione di molti risultati teorici risulta semplificata. Sotto l'ipotesi di osservazioni indipendenti la log-verosimiglianza assume la forma

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

Da quanto esposto finora, risulta chiaro come la massimizzazione della funzione di log-verosimiglianza (o di verosimiglianza), rappresenti un metodo generale per l'ottenimento di una stima di  $\theta$  basata sui dati osservati  $y$ . Si definisce **stima di massima verosimiglianza** di  $\theta$  un valore  $\hat{\theta} \in \Theta$  tale che  $\ell(\hat{\theta}) \geq \ell(\theta)$  per ogni  $\theta \in \Theta$ . Si noti però che, in generale, l'esistenza e unicità di  $\hat{\theta}$  non è garantita. Nel caso in cui  $\hat{\theta}$  esista unico con probabilità uno, viene definito **stimatore di massima verosimiglianza** la variabile casuale  $\hat{\theta}(Y)$ . Se il modello statistico ipotizzato è con **verosimiglianza regolare**, ossia  $\Theta$  è un sottoinsieme aperto di  $\mathbb{R}^p$  e  $\ell(\theta)$  risulta essere differenziabile almeno tre volte con derivate parziali continue in  $\Theta$ , la stima di massima verosimiglianza viene ottenuta ponendo ogni componente della **funzione score**

$$U(\theta) = \frac{\partial \ell(\theta)}{\partial \theta},$$

pari a zero, risolvendo il sistema in  $p$  equazioni che ne deriva e verificando che i valori ottenuti corrispondano ad un punto di massimo assoluto.

#### 1.1.4 Informazione osservata e risultati distributivi ad essa collegati

Nell'ambito dell'inferenza di verosimiglianza si definisce **informazione osservata**,  $j(\theta)$ , l'opposto della matrice hessiana, rispetto a  $\theta$ , della funzione

di log-verosimiglianza, ossia

$$j(\theta) = -\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^T}.$$

Il valore atteso di quest'ultima

$$i(\theta) = E_\theta \{j(\theta)\} = -E_\theta \left\{ \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^T} \right\},$$

prende il nome di **informazione attesa** ed è strettamente legata alla distribuzione asintotica di alcune quantità connesse alla funzione di verosimiglianza. In particolare, si può dimostrare come, in modelli che soddisfano ulteriori condizioni di regolarità, valgano sotto  $\theta$ , le seguenti proprietà

$$U(\theta) \xrightarrow{d} N_p(0, i(\theta)),$$

$$\hat{\theta} \xrightarrow{d} N_p(\theta, i(\theta)^{-1}).$$

Questi risultati risultano particolarmente utili nel determinare la distribuzione asintotica delle statistiche **log-rapporto di verosimiglianza** ( $W$ ), **score** ( $W_u$ ) e di **Wald** ( $W_e$ ), definite come

$$W = 2 \left[ \ell(\hat{\theta}) - \ell(\theta) \right],$$

$$W_u = U(\theta)^T i(\theta)^{-1} U(\theta),$$

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

ed utilizzate per la costruzione di regioni di confidenza e test di ipotesi. In particolare, sotto  $\theta$ , le tre statistiche hanno la stessa distribuzione asintotica  $\chi_p^2$ .

### 1.1.5 Relazioni di Bartlett

Nello sviluppo dei metodi di riduzione della distorsione, trattati nella seconda parte di questo capitolo, ricoprono un ruolo fondamentale alcune relazioni di



annullamento. Nell'introdurle, è opportuno ricondursi alla notazione utilizzata da McCullagh (1987) per le derivate della funzione di log-verosimiglianza. In particolare, queste vengono indicate con

$$U_r = U_r(\theta) = \frac{\partial \ell(\theta)}{\partial \theta_r}, \quad U_{rs} = U_{rs}(\theta) = \frac{\partial^2 \ell(\theta)}{\partial \theta_r \partial \theta_s},$$

dove  $\theta_r$  e  $\theta_s$  rappresentano due generiche componenti del parametro  $\theta$ . I valori attesi delle derivate della funzione di log-verosimiglianza sono invece espressi secondo la struttura

$$\begin{aligned} \nu_r &= E_\theta(U_r), & \nu_{rs} &= E_\theta(U_{rs}), & \nu_{r,s} &= E_\theta(U_r U_s), \\ \nu_{rs,t} &= E_\theta(U_{rs} U_t), & \nu_{r,s,t} &= E_\theta(U_r U_s U_t). \end{aligned}$$

Similmente con  $\nu^{s,t}$  viene indicato il generico elemento dell'inversa della matrice con elementi  $\nu_{s,t}$ .

In modelli statistici parametrici regolari, si può mostrare come valgano le seguenti relazioni di annullamento

$$\begin{aligned} \nu_r &= 0, \\ \nu_{rs} + \nu_{r,s} &= 0, \\ \nu_{rst} + \nu_{rs,t} + \nu_{rt,s} + \nu_{r,st} + \nu_{r,s,t} &= 0, \end{aligned}$$

note anche come **relazioni di Bartlett**.

## 1.2 Metodi di riduzione della distorsione

In problemi regolari di stima, i metodi di verosimiglianza permettono di derivare, all'interno di una teoria unificata, una vasta gamma di procedure inferenziali le quali soddisfano asintoticamente vari criteri di ottimalità. In particolare, lo stimatore di massima verosimiglianza risulta sempre asintoticamente non distorto e normalmente distribuito. In generale, però, non è garantita la non distorsione per campioni finiti. Tale problematica può portare ad una significativa perdita di efficacia delle procedure inferenziali, come

osservato, per esempio, in Kosmidis (2014). Per questo motivo, un crescente interesse è stato riservato allo sviluppo di procedure in grado di ridurre la distorsione dello stimatore di massima verosimiglianza in presenza di numerosità campionarie limitate. Tali metodi si basano sull'osservazione che, sotto opportune condizioni di regolarità, la distorsione di  $\hat{\theta}$  può essere scritta come

$$b(\theta) = \frac{b_1(\theta)}{n} + \frac{b_2(\theta)}{n^2} + O(n^{-3}), \quad (1.3)$$

e che quindi si possa ottenere uno stimatore con distorsione ridotta eliminando i termini di ordine minore presenti nella (1.3).

Si definiscono metodi per la riduzione della distorsione **espliciti** quelle procedure che ottengono una diminuzione della distorsione dello stimatore di massima verosimiglianza sottraendo a  $\hat{\theta}$  una stima di  $b(\theta)$ . Tra i metodi espliciti più utilizzati è possibile trovare il metodo *jackknife*, proposto da Quenouille (1956), il metodo *bootstrap* ed alcune tecniche basate su approssimazioni del termine di distorsione attraverso sviluppi asintotici di  $b(\theta)$  (Kosmidis, 2014). Sebbene in generale di facile attuazione, le procedure esplicite presentano alcune problematiche, prima fra tutte il fatto che necessitano della finitezza della stima di massima verosimiglianza per essere implementate.

Un'alternativa, ai metodi espliciti di riduzione della distorsione, è data dai metodi **impliciti**. Questi ultimi permettono di ridurre la distorsione di uno stimatore incorporandola direttamente nell'equazione di stima.

### 1.2.1 Riduzione della distorsione in media

Tra i metodi impliciti, particolarmente rivelante risulta la proposta di Firth (1993), la quale permette di eliminare il termine di ordine  $O(n^{-1})$  dalla (1.3) attraverso un'opportuna penalizzazione della funzione *score*. Per introdurre questo metodo, si consideri innanzitutto il caso in cui il modello ipotizzato per i dati appartenga alla famiglia esponenziale monoparametrica, con funzione di densità

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

dove  $s(y)$  rappresenta la statistica sufficiente e  $\theta \in \Theta \subseteq \mathbb{R}$ . In questo caso, la funzione *score* risulta pari a

$$U(\theta) = s(y) - K'(\theta),$$

con  $K'(\theta) = (\partial/\partial\theta)K(\theta)$ , e l'eventuale distorsione dello stimatore di massima verosimiglianza,  $\hat{\theta}$ , è dovuta all'effetto congiunto di due elementi

- La non distorsione della funzione *score*, ossia  $E_\theta \{U(\theta)\} = 0$ ,
- Una curvatura della funzione *score* non nulla, ossia  $U''(\theta) \neq 0$ .

L'idea fondamentale, sulla quale si basa l'intero lavoro di Firth (1993), è quindi quella di ridurre la distorsione presente nello stimatore di massima verosimiglianza introducendo un termine di penalizzazione nella funzione *score*, così da bilanciare gli effetti di una curvatura non nulla.

Questa soluzione risulta efficace anche in modelli ben più complessi, con parametro  $\theta \in \Theta \subseteq \mathbb{R}^p$ , ed è esprimibile in forma generale attraverso l'equazione di stima

$$U^*(\theta) = U(\theta) + A^*(\theta) = 0,$$

dove  $A^*(\theta)$  vettore di dimensione  $p$ . Firth (1993) mostra come si possano derivare due diverse espressioni di  $A^*(\theta)$ , basate rispettivamente sull'informazione attesa  $i(\theta)$  e su quella osservata  $j(\theta)$ , ossia

$$A^{*(E)}(\theta) = -i(\theta)b_1(\theta)/n, \quad A^{*(O)}(\theta) = -j(\theta)b_1(\theta)/n.$$

Nel caso in cui venga utilizzata la forma basata sull'informazione attesa, un generico elemento di  $A^*(\theta)$  risulta pari a

$$A_r^{*(E)} = \nu^{u,v} (\nu_{r,u,v} + \nu_{r,uv}) / 2,$$

mentre con l'informazione osservata assume la forma

$$A_r^{*(O)} = -U_{rs} \nu^{s,t} \nu^{u,v} (\nu_{t,u,v} + \nu_{t,uv}) / 2,$$

Si noti che in entrambi i casi è stata utilizzata la convezione per la somma di Einstein e perciò la somma è implicata su ogni indice ripetuto  $r, s, \dots$

Infine, un'espressione equivalente di  $A^{*(E)}(\theta)$ , maggiormente efficiente dal punto di vista computazione, è reperibile in Kosmidis e Firth (2010). In particolare, definendo le matrici  $P_r(\theta) = E_\theta \{U(\theta)U(\theta)^T U_r(\theta)\}$  e  $Q_r(\theta) = -E_\theta \{j(\theta)U_r(\theta)\}$ , per  $r = 1, \dots, p$ , si può dimostrare che

$$A^{*(E)}(\theta) = F_1/2,$$

dove  $F_1$  è un vettore  $p$ -dimensionale avente come elementi

$$F_{1r} = \text{tr} [i(\theta)^{-1} \{P_r(\theta) + Q_r(\theta)\}] \quad (r = 1, \dots, p).$$

Per quanto riguarda la distribuzione dello stimatore ottenuto, questa risulta asintoticamente equivalente a quella dello stimatore di massima verosimiglianza.

## 1.2.2 Riduzione della distorsione in mediana

Una delle problematiche principali, riguardanti i metodi di riduzione della distorsione dello stimatore di massima verosimiglianza, è legata al fatto che la distorsione di uno stimatore è inevitabilmente legata alla parametrizzazione utilizzata. Questo implica che gli stimatori così ottenuti non risultano equivarianti rispetto a trasformazioni biunivoche del parametro. Per questa ragione, Kenne Pagui *et al.* (2017) hanno sviluppato un metodo per la riduzione della distorsione basato sulla mediana, e non sulla media, come indice di centratura. In particolare, la loro proposta si basa, come per il metodo di Firth (1993), sulla derivazione di una nuova equazione di stima, in questo caso ottenuta sottraendo sistematicamente alla funzione *score* un'approssimazione della sua mediana. La mediana della distribuzione dello stimatore così ottenuto sarà quindi approssimativamente il vero valore del parametro. A differenza degli altri metodi di riduzione della distorsione in mediana, presenti in letteratura, il metodo proposto da Kenne Pagui *et al.* (2017) non necessita di valori finiti della stima di massima verosimiglianza, ed è anzi efficace nel prevenire stime infinite. Nel caso in cui venga considerato un modello statistico parametrico regolare con parametro multidimensionale  $\theta = (\theta_1, \dots, \theta_p)$ , per  $\theta \in \Theta \subseteq \mathbb{R}^p$ ,

Kenne Pagui *et al.* (2017) mostrano come si possa ottenere uno stimatore  $\tilde{\theta}$  approssimativamente non distorto in mediana, componente per componente, con errore di ordine  $O(n^{-3/2})$ . Tale risultato viene raggiunto definendo un sistema di equazioni di stima basato sul vettore *score* modificato in mediana  $\tilde{U}(\theta)$  le cui componenti sono esprimibili, utilizzando la convenzione per la somma di Einstein, come

$$\tilde{U}_r = U_r - \rho_{ra}U_a - \kappa_{1r} + \kappa_{3r}/(6\kappa_{2r}), \quad (1.4)$$

dove  $\rho_{ra} = \nu^{ab}\nu_{r,b}$ , con  $\nu^{ab}$  generico elemento dell'inversa della matrice ottenuta eliminando da  $i(\theta)$  la riga e la colonna  $r$ -esima, mentre  $\kappa_{jr}, j = 1, 2, 3$ , rappresenta una approssimazione del  $j$ -esimo cumulante della *score* profilo relativa alla componente  $\theta_r$ . Nello specifico, le espressioni dei tre cumulanti sono

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

ed il termine di errore è di ordine  $O(n^{-1})$  per  $\kappa_{1r}$  mentre per  $\kappa_{2r}$  e  $\kappa_{3r}$  è di ordine  $O(1)$ .

Lo stimatore  $\tilde{\theta}$ , ottenuto risolvendo  $\tilde{U}(\theta) = 0$ , risulta inoltre equivariante rispetto a trasformazioni biunivoche che agiscono separatamente sulle componenti del parametro  $\theta$ .

Infine, l'espressione (1.4), basata sulla somma ripetuta di un gran numero di termini, può risultare sconveniente dal punto di vista applicativo. Per questa ragione, in Kenne Pagui *et al.* (2020) è sviluppata una forma alternativa di  $\tilde{U}(\theta)$  basata su prodotti matriciali, la quale permette anche di evidenziare la connessione tra il metodo di correzione in mediana e quello proposto da Firth (1993).

Nello specifico, richiamando la notazione utilizzata nel paragrafo 1.2.1 per

le quantità  $P_s(\theta)$ ,  $Q_s(\theta)$  e  $F_1$ , si può mostrare come

$$\begin{aligned}\tilde{U}(\theta) &= U(\theta) + \tilde{A}(\theta) \\ &= U(\theta) + A^*(\theta) - i(\theta)\tilde{F}_2,\end{aligned}\tag{1.5}$$

dove  $A^*(\theta)$  rappresenta la penalizzazione apportata alla *score* dal metodo di Firth (1993), nella forma descritta in Kosmidis e Firth (2010), mentre  $\tilde{F}_2$  è un vettore di dimensione  $p$  con elementi

$$\tilde{F}_{2r} = [i(\theta)^{-1}]_r^T F_{2,r} \quad (r = 1, \dots, p).\tag{1.6}$$

Nell'espressione (1.6), la quantità  $[i(\theta)^{-1}]_r$  rappresenta la colonna  $r$ -esima di  $i(\theta)^{-1}$  mentre  $F_{2,r}$  è un vettore di dimensione  $p$  avente elementi pari a

$$F_{2s,r} = \text{tr} \left[ h_r(\theta) \left\{ \frac{1}{3} P_s(\theta) + \frac{1}{2} Q_s(\theta) \right\} \right] \quad (s = 1, \dots, p),$$

dove  $h_r(\theta) = [i(\theta)^{-1}]_r [i(\theta)^{-1}]_r^T / i^{rr}(\theta)$  è una matrice di dimensione  $p \times p$ .

## Capitolo 2

# Modelli di regressione per il rischio relativo

L'obiettivo di questo capitolo è descrivere alcuni dei metodi utilizzati per l'inferenza sul rischio relativo. In particolare, nel paragrafo 2.2 viene analizzata sotto vari aspetti la modellazione del rischio relativo attraverso i modelli lineari generalizzati. Il paragrafo 2.3 è invece dedicato ad una analisi approfondita del modello proposto da Richardson *et al.* (2017), per la stima del rischio relativo nel caso in cui la variabile esplicativa di primario interesse sia di tipo dicotomico. Particolarmente rilevante è il paragrafo 2.3.3 nel quale viene trattata l'interpretazione dei parametri del modello. Quest'ultima, infatti, non sembrava del tutto chiara nell'articolo originale

Infine, il paragrafo 2.3.4 rappresenta uno dei contributi originali di questa tesi in quanto viene proposta una diversa specificazione del parametro di disturbo, rispetto a quella presente in Richardson *et al.* (2017), la quale risulta più facilmente trattabile dal punto di vista matematico e presenta meno problemi computazionali. Ulteriori approfondimenti del modello di Richardson *et al.* (2017) sono riportati nell'Appendice A.

## 2.1 Notazione

Sia  $y = (y_1, \dots, y_n)$  il vettore di osservazioni su una risposta binaria per  $n$  unità statistiche. Per ciascuna osservazione sono considerate fissate  $p$  variabili esplicative,  $(x_i, t_i) = (x_{i0}, x_{i1}, \dots, x_{ip-2}, t_i)$ , con  $x_{i0} = 1$ ,  $i = 1, \dots, n$ , mentre  $t_i$  è una variabile dicotomica che indica se il soggetto  $i$ -esimo è stato esposto ad un generico fattore di esposizione binario  $t$ . Si assume inoltre che  $y_i$  sia la realizzazione di una variabile casuale bernoulliana  $Y_i \sim Be(\pi_i)$  con valore atteso  $E(Y_i; x_i, t_i) = P(Y_i = 1; x_i, t_i) = \pi_i(x_i, t_i) = \pi_i$ .

Nel caso in cui il contesto lo renda necessario, viene indicata con  $\pi_{i1} = \pi_i(x_i, t_i = 1)$  la probabilità di osservare il fenomeno di interesse dato il vettore di esplicative  $(x_i, t_i = 1)$ . Similmente,  $\pi_{i0}$  rappresenta la probabilità di osservare il fenomeno di interesse dato il vettore di esplicative  $(x_i, t_i = 0)$ . Infine, si indica con il termine logaritmo del rischio relativo,  $\log(RR_i)$ , il logaritmo del rapporto tra  $\pi_{i1}$  e  $\pi_{i0}$ , ossia  $\log(RR_i) = \log\left(\frac{\pi_{i1}}{\pi_{i0}}\right)$ . Il simbolo  $\simeq$  viene utilizzato nel caso in cui due quantità differiscano tra loro in modo trascurabile.

## 2.2 Modelli lineari generalizzati e stima di $\log(RR_i)$

I modelli lineari generalizzati rappresentano probabilmente la via concettualmente più immediata per l'inferenza sul rischio relativo. Nel caso oggetto di studio, il modello assume, per ciascuna delle  $n$  osservazioni presenti nel campione, che

$$g(\pi_i) = \eta_i,$$

dove  $g(\cdot)$  rappresenta un'opportuna funzione di legame e  $\eta_i = \eta_i(x_i, t_i)$  un predittore lineare relativo al vettore delle variabili esplicative  $(x_i, t_i)$ . Quest'ultima quantità, nelle pagine seguenti potrà essere definita in modo diverso a seconda che vengano comprese o meno le interazioni tra variabili confondenti  $x_i$  e il fattore di esposizione binario  $t_i$ . Nel caso in cui non vi siano



interazioni viene usata la notazione

$$\begin{aligned}\eta_i &= x_i\beta + \gamma_0 t_i \\ &= \beta_0 x_{i0} + \beta_1 x_{i1} + \dots + \beta_{p-2} x_{ip-2} + \gamma_0 t_i,\end{aligned}$$

dove  $x_{i0} = 1$ ,  $i = 1, \dots, n$  e  $\beta_0$  rappresenta l'intercetta del modello. Il coefficiente  $\gamma_0$  misura l'effetto del fattore di esposizione binario  $t_i$  mentre il vettore  $\beta = (\beta_0, \beta_1, \dots, \beta_{p-2})$ , contiene i coefficienti relativi all'intercetta e alle rimanenti  $p-2$  variabili esplicative. In presenza di interazioni, il predittore lineare assume la forma

$$\begin{aligned}\eta_i &= x_i\beta + x_i t_i \gamma \\ &= \beta_0 x_{i0} + \beta_1 x_{i1} + \dots + \beta_{p-2} x_{ip-2} + \gamma_0 t_i + \gamma_1 t_i x_{i1} + \dots + \gamma_{p-2} t_i x_{ip-2},\end{aligned}$$

con  $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_{p-2})$  vettore dei coefficienti relativo alle interazioni tra il fattore di esposizione binario  $t_i$  e le altre esplicative disponibili. Infine, il predittore lineare può essere scritto in forma matriciale come  $\eta = Z\theta$ , dove  $z_i = (x_i, x_i t_i)$  rappresenta la riga  $i$ -esima della matrice di disegno  $Z$  e  $\theta = (\beta_0, \beta_1, \dots, \beta_{p-2}, \gamma_0, \dots, \gamma_{p-2})$ . In assenza di interazione abbiamo invece  $z_i = (x_i, t_i)$  e  $\theta = (\beta_0, \beta_1, \dots, \beta_{p-2}, \gamma_0)$ .

## Scelta della funzione di legame

In letteratura, per la stima di  $\log(\text{RR}_i)$  tramite i modelli lineari generalizzati, sono stati proposti in prevalenza due metodi: la regressione logistica e la regressione log-binomiale. Analizziamo ora vantaggi e svantaggi di entrambi.

### 2.2.1 Modello logistico

Il modello logistico assume come funzione di legame

$$g(\pi_i) = \text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \eta_i,$$

ed è quindi facile osservare che, in assenza di interazioni tra  $t_i$  e  $x_i$ , il coefficiente  $\gamma_0$  relativo al fattore di esposizione binario risulta

$$\gamma_0 = \log(\text{OR}_i) = \log\left(\frac{\pi_{i1}(1 - \pi_{i0})}{\pi_{i0}(1 - \pi_{i1})}\right), \quad (2.1)$$

dove  $\log(\text{OR}_i)$  è il log-rapporto delle quote per l'osservazione  $i$ -esima. Dalla (2.1) si ottiene perciò che, se  $(1 - \pi_{i0}) \simeq (1 - \pi_{i1}) \simeq 1$ , allora  $\gamma_0 = \log(\text{OR}_i) \simeq \log(\text{RR}_i)$ . L'espressione esatta di  $\log(\text{RR}_i)$  è invece

$$\log(\text{RR}_i) = \log\left(\frac{1 + e^{-x_i\beta}}{1 + e^{-x_i\beta - \gamma_0}}\right), \quad (2.2)$$

in questo caso però non è più possibile interpretare in modo diretto il coefficiente  $\gamma_0$ . In presenza di interazioni tra  $t_i$  e  $x_i$  si ottiene un risultato analogo osservando che

$$\text{logit}(\pi_{i1}) - \text{logit}(\pi_{i0}) = x_i\beta + x_i\gamma - x_i\beta = x_i\gamma,$$

ossia

$$\log\left(\frac{\pi_{i1}(1 - \pi_{i0})}{\pi_{i0}(1 - \pi_{i1})}\right) = \log(\text{OR}_i) = x_i\gamma = \gamma_0 + \sum_{j=1}^{p-2} x_{ij}\gamma_j, \quad (2.3)$$

e quindi, se  $(1 - \pi_{i0}) \simeq (1 - \pi_{i1}) \simeq 1$ , si ha che  $x_i\gamma \simeq \log(\text{RR}_i)$ . L'espressione esatta di  $\log(\text{RR}_i)$  in questo caso è

$$\log(\text{RR}_i) = \log\left(\frac{1 + e^{-x_i\beta}}{1 + e^{-x_i\beta - x_i\gamma}}\right). \quad (2.4)$$

## Problematiche legate all'impiego della regressione logistica per l'inferenza sul rischio relativo

L'inferenza su  $\log(\text{RR}_i)$  basata sull'approssimazione  $\log(\text{OR}_i) \simeq \log(\text{RR}_i)$  risulta necessaria negli studi caso-controllo in quanto, in questo contesto, una stima di  $\log(\text{RR}_i)$  non è ottenibile per via del disegno sperimentale (Zou, 2004). In letteratura però non sono rari i casi in cui si è scelto di utilizzare la regressione logistica quando, almeno teoricamente, era direttamente ottenibile una stima del rischio relativo (McNutt *et al.*, 2003). Come visto nel paragrafo precedente, questa pratica è in linea di principio accettabile se il fenomeno di interesse è raro in ogni strato della popolazione oggetto di studio, ossia se  $\pi_i \simeq 0$ ,  $i = 1, \dots, n$ . Se tale condizione non risulta soddisfatta,

l'impiego dei coefficienti derivanti dalla regressione logistica come surrogati del logaritmo del rischio relativo può portare a stime fortemente distorte di quest'ultimo (Barros, 2003).

### Approssimazione di Zang e Yu (1998)

Come visto nelle espressioni (2.1) e (2.3), i coefficienti derivanti dalla regressione logistica possono essere interpretati in termini di log-rapporto delle quote. Sfruttando la relazione che intercorre tra  $\text{OR}_i$  e  $\text{RR}_i$ , Zang e Yu (1998) hanno proposto un metodo alternativo alle (2.2) e (2.4) per ottenere una stima di  $\log(\text{RR}_i)$  direttamente dalla stima dei coefficienti della regressione logistica. Questa si basa sull'osservazione che

$$\begin{aligned}
 \log(\text{RR}_i) &= \log\left(\frac{\pi_{i1}}{\pi_{i0}}\right) \\
 &= \log\left(\frac{\pi_{i1}(1-\pi_{i0})}{\pi_{i0}(1-\pi_{i1})} \frac{(1-\pi_{i0})(1-\pi_{i1})}{(1-\pi_{i0})(1-\pi_{i1}) + \pi_{i1}(1-\pi_{i0})}\right) \\
 &= \log\left(\frac{\frac{\pi_{i1}(1-\pi_{i0})}{\pi_{i0}(1-\pi_{i1})}}{(1-\pi_{i0}) + \pi_{i0} \frac{\pi_{i1}(1-\pi_{i0})}{\pi_{i0}(1-\pi_{i1})}}\right) \\
 &= \log\left(\frac{\text{OR}_i}{(1-\pi_{i0}) + \pi_{i0}\text{OR}_i}\right), \tag{2.5}
 \end{aligned}$$

ed anche per via della sua semplicità la (2.5) è stata estensivamente utilizzata all'inizio degli anni 2000. Va tuttavia osservato che spesso il metodo non è stato applicato correttamente. Un errore comune era quello di considerare la quantità  $\log\left(\frac{\text{OR}_i}{(1-\pi_{i0}) + \pi_{i0}\text{OR}_i}\right)$ , dove  $\pi_{i0}$  rappresenta l'incidenza del fenomeno di interesse nel gruppo dei non esposti, come una stima  $\log(\text{RR}_i)$  aggiustata per l'effetto delle altre variabili esplicative (McNutt *et al.*, 2003). Questo porta ovviamente ad una stima distorta, in quanto non viene considerato l'effetto delle variabili confondenti. Per evitare di incorrere in questo problema, McNutt *et al.* (2003) raccomandano, nel caso in cui si vogliono impiegare i coefficienti ottenuti dalla regressione logistica, di utilizzare le espressioni (2.2) e (2.4). Queste danno risultati equivalenti alla corretta applicazione della (2.5) ma si basano direttamente su  $\pi_{i0}$  e  $\pi_{i1}$  invece che su  $\log(\text{OR}_i)$ .

Sebbene entrambi i metodi risultino metodologicamente corretti, in tutti e due i casi viene persa la struttura lineare imposta sul predittore del modello logistico, rendendo quindi l'interpretazione dei risultati più complessa.

### 2.2.2 Regressione log-binomiale

La via forse più intuitiva per modellare il rischio relativo è adottare un modello lineare generalizzato binomiale con funzione di legame logaritmica. In questo caso il modello assume

$$\begin{aligned}\log(\pi_i) &= x_i\beta + \gamma t_i x_i \\ &= \beta_0 x_{i0} + \beta_1 x_{i1} + \dots + \beta_{p-2} x_{ip-2} + \gamma_0 t_i x_{i0} + \gamma_1 t_i x_{i1} + \dots + \gamma_{p-2} t_i x_{ip-2},\end{aligned}$$

e dunque,

$$\log(\text{RR}_i) = x_i\gamma.$$

La principale criticità di questo metodo risiede nel fatto che, a differenza di quanto accade per la regressione logistica, utilizzando il logaritmo come funzione di legame, è necessario porre dei vincoli in fase di stima dei parametri per far sì che le probabilità stimate rimangano comprese tra 0 e 1 (Lumley *et al.*, 2006). Ciò si traduce in pratica imponendo che, per ciascun valore osservato del vettore delle variabili esplicative  $x_i$ , il predittore lineare  $\eta_i$  assuma valori compresi tra  $-\infty$  e 0. Dal punto di vista computazionale tale restrizione rende il modello instabile. Infatti, gli algoritmi utilizzati per la risoluzione delle equazioni di stima spesso non convergono, in particolare quando le stime si trovano ai limiti dello spazio parametrico (Barros, 2003), rendendo la regressione log-binomiale difficilmente utilizzabile.

### 2.2.3 Modello di Poisson e quasi-verosimiglianza

Una stima, asintoticamente non distorta, di  $\log(\text{RR}_i)$  si può ottenere anche attraverso l'impiego di un modello di Poisson con funzione di legame logaritmica. In questo caso si assume impropriamente che  $y_i$ , con  $i = 1, \dots, n$ , sia

la realizzazione di una variabile casuale di Poisson  $Y_i \sim Po(\pi_i)$  con media  $\pi_i = e^{x_i\beta + \gamma t_i x_i}$ . Si assume perciò che

$$\log(\pi_i) = x_i\beta + \gamma t_i x_i,$$

e la stima dei parametri viene ottenuta agevolmente massimizzando la log-verosimiglianza di Poisson,

$$\ell(\beta, \gamma) = \sum_{i=1}^n [y_i(x_i\beta + \gamma t_i x_i) - e^{x_i\beta + \gamma t_i x_i}].$$

Sebbene l'equazione di stima rimanga non distorta, ipotizzando per i dati una distribuzione di Poisson invece che una distribuzione binomiale, il modello risulta non correttamente specificato. Questo si traduce in pratica in stime troppo grandi della varianza dei coefficienti. Una soluzione a questo problema è reperibile in Zou (2004) dove si utilizza, per la stima degli errori standard dei coefficienti, uno stimatore sandwich.

Risultati del tutto equivalenti a quelli ottenuti ipotizzando per i dati un modello di Poisson sono stati raggiunti da Carter *et al.* (2005) utilizzando un approccio di quasi verosimiglianza. In particolare, gli Autori osservano che, se  $Y_i$  segue una distribuzione bernoulliana con  $\log(\pi_i) = x_i\beta + \gamma t_i x_i$ , le equazioni di verosimiglianza risultano

$$\begin{cases} \frac{\partial \ell(\beta, \gamma)}{\partial \beta} = \sum_{i=1}^n \frac{y_i - \pi_i}{1 - \pi_i} \frac{\partial \eta_i}{\partial \beta} = 0 \\ \frac{\partial \ell(\beta, \gamma)}{\partial \gamma} = \sum_{i=1}^n \frac{y_i - \pi_i}{1 - \pi_i} \frac{\partial \eta_i}{\partial \gamma} = 0 \end{cases}$$

e gran parte dei problemi di convergenza che si riscontrano in fase di stima del modello, sono dovuti alla quantità  $(1 - \pi_i)$  nel denominatore di ciascuna delle equazioni precedenti. Viene perciò proposto di stimare  $\theta = (\beta, \gamma)$  tramite il seguente sistema di equazioni non distorte

$$\begin{cases} Q_\beta(\theta) = 0 \\ Q_\gamma(\theta) = 0 \end{cases}$$

con

$$Q_\beta(\theta) = \sum_{i=1}^n Q_\beta(\theta)_i = \sum_{i=1}^n (y_i - \pi_i) \frac{\partial \eta_i}{\partial \beta}, \quad (2.6)$$

$$Q_\beta(\theta) = \sum_{i=1}^n Q_\beta(\theta)_i = \sum_{i=1}^n (y_i - \pi_i) \frac{\partial \eta_i}{\partial \gamma}. \quad (2.7)$$

Lo stimatore che ne deriva,  $\hat{\theta}^Q = [\hat{\beta}^Q, \hat{\gamma}^Q]$ , risulta avere varianza asintotica pari a

$$Var(\hat{\theta}^Q) = \left[ \sum_{i=1}^n E[J_i(\theta)] \right]^{-1} \left[ \sum_{i=1}^n E[Q_i(\theta) Q_i(\theta)^T] \right]^{-1} \left[ \sum_{i=1}^n E[J_i(\theta)] \right]^{-1} \quad (2.8)$$

dove, in questo caso,  $Q_i(\theta) = [Q_\beta(\beta)_i, Q_\gamma(\gamma)_i]$ , con queste ultime definite nelle espressioni (2.6) e (2.7), mentre  $J_i(\theta) = \frac{-\partial Q_i(\beta)}{\partial \beta \partial \beta^T}$  rappresenta l'informazione osservata.

### 2.3 Il modello proposto da Richardson *et al.* (2017)

Richardson *et al.* (2017) propongono un nuovo metodo per l'inferenza sul rischio relativo, quando è di interesse valutare l'effetto di un fattore di esposizione binario al netto delle altre variabili esplicative. In particolare viene proposta una riparametrizzazione  $(\eta_{i1}, \eta_{i2})$  di  $(\pi_{i0}, \pi_{i1})$  dove

$$\eta_{i1} = \log(\text{RR}_i),$$

e

$$\eta_{i2} = \log\left(\frac{\pi_{i1}\pi_{i0}}{(1-\pi_{i1})(1-\pi_{i0})}\right),$$

sono parametri a variazione indipendente, come verrà mostrato nel paragrafo 2.3.2. Le probabilità  $\pi_{i1}$  e  $\pi_{i0}$  sono legate a  $\eta_{i1}$  e  $\eta_{i2}$  da un sistema di due equazioni in due incognite dal quale si ottiene che

$$\pi_{i0} = \pi_{i0}(\eta_{i1}, \eta_{i2}) = \frac{-(e^{\eta_{i1}} + 1)e^{\eta_{i2}} + \sqrt{e^{2\eta_{i2}}(e^{\eta_{i1}} + 1)^2 + 4e^{\eta_{i1} + \eta_{i2}}(1 - e^{\eta_{i2}})}}{2e^{\eta_{i1}}(1 - e^{\eta_{i2}})}, \quad (2.9)$$

$$\pi_{i1} = \pi_{i1}(\eta_{i1}, \eta_{i2}) = \pi_{i0}e^{\eta_{i1}}.$$

Infine entrambi i parametri sono legati alle esplicative disponibili attraverso un predittore lineare

$$\eta_{i1} = x_i\gamma = \gamma_0x_{i0} + \gamma_1x_{i1} + \dots + \gamma_{p-2}x_{ip-2},$$

$$\eta_{i2} = x_i\beta = \beta_0x_{i0} + \beta_1x_{i1} + \dots + \beta_{p-2}x_{ip-2},$$

dove  $x_i = (x_{i0}, x_{i1}, x_{i2}, \dots, x_{ip-2})$  rappresenta come al solito il vettore delle variabili confondenti con  $x_{i0} = 1$  per  $i = 1, \dots, n$ .

La log-verosimiglianza risulta

$$\ell(\gamma, \beta) = \sum_{i=1}^n [y_i \log(\pi_{i0}e^{\eta_{i1}t_i}) + (1 - y_i) \log(1 - \pi_{i0}e^{\eta_{i1}t_i})]. \quad (2.10)$$

Dato che  $\eta_{i1}$  e  $\eta_{i2}$  sono con variazione indipendente e sono entrambi definiti su  $\mathbb{R}$ , la massimizzazione della log-verosimiglianza può essere effettuata senza nessun vincolo sui valori di  $\gamma$  e  $\beta$ . In questo modo, viene così risolto il problema dell'instabilità numerica dei coefficienti nella regressione log-binomiale. Inoltre, le componenti di  $\gamma$  sono direttamente interpretabili in termini di 'effetti' sul logaritmo del rischio relativo.

### 2.3.1 Esempio: confronto tra due gruppi

L'esempio più semplice in cui può essere applicato il modello del paragrafo precedente è quello in cui l'unica esplicativa è il fattore di esposizione binaria  $t$ . Si tratta quindi del problema del confronto tra due binomiali. In questo caso i predittori lineari assumono la forma

$$\log(\text{RR}_i) = \eta_{i1} = \gamma_0x_{i0} = \gamma_0,$$

$$\log \frac{\pi_{i0}\pi_{i1}}{(1 - \pi_{i0})(1 - \pi_{i1})} = \eta_{i2} = \beta_0x_{i0} = \beta_0.$$

Per quanto riguarda  $\pi_{i0}$  si ha

$$\pi_{i0} = \frac{-(e^{\gamma_0} + 1)e^{\beta_0} + \sqrt{e^{2\beta_0}(e^{\gamma_0} + 1)^2 + 4e^{\gamma_0 + \beta_0}(1 - e^{\beta_0})}}{2e^{\gamma_0}(1 - e^{\beta_0})}$$

mentre  $\pi_{i1} = \pi_{i0}e^{\gamma_0}$ . Per valutare, in questa situazione semplice, le proprietà dello stimatore di massima verosimiglianza è stato condotto un studio di simulazione. In particolare sono state effettuate 10000 simulazioni per vari valori di  $\pi_{i0}$ ,  $\pi_{i1}$  e per diverse numerosità nei campioni di soggetti esposti e non esposti. In ciascun caso sono stati calcolati quattro indici: la distorsione relativa percentuale stimata (Relative Bias), lo *standard error* dello stimatore (SE), ossia il valore medio, nel corso delle 10000 simulazioni, della radice quadrata degli elementi diagonali dell'inversa della matrice di informazione attesa stimata, la percentuale empirica di sottostima (PU%) e la copertura empirica degli intervalli di Wald con livello di confidenza nominale 95% (Wald).

Dalla Tabella 2.1 è possibile notare come lo stimatore presenti una percentuale di sottostima prossima a 50%, ma ciò è dovuto alla semplicità dei casi analizzati. Inoltre la copertura degli intervalli di Wald risulta abbastanza distante dal valore nominale, specialmente nell'ultima simulazione. In linea di massima anche la distorsione in media è presente, spesso in modo importante.

**Tabella 2.1:** Risultati delle simulazioni riguardanti la stima di  $\gamma_0$  per varie combinazioni di  $\pi_{i0}$ ,  $\pi_{i1}$ ,  $n_{i0}$  e  $n_{i1}$ . Gli standard error di simulazione per SE, Relative Bias e Wald sono riportati tra parentesi. Gli standard error di simulazione per PU% risultano pari a 0.5.

$n_1$	$n_0$	$\pi_{i1}$	$\pi_{i0}$	PU%	Relative Bias	SE	Wald
15	15	0.7	0.3	49.1	6.9 (0.6)	0.464 (0.001)	96.11 (0.19)
15	15	0.3	0.7	50.8	-7.1 (0.6)	0.464 (0.001)	96.33 (0.19)
15	20	0.5	0.8	49.5	-6.5 (0.7)	0.293 (0.001)	94.32 (0.23)
15	15	0.8	0.1	52.7	-5.1 (0.3)	0.774 (0.002)	93.32 (0.25)
15	10	0.4	0.45	49.4	-24.4 (4.7)	0.528 (0.002)	97.12 (0.18)



### 2.3.2 Verifica della proprietà di variazione indipendente tra $\eta_{i1}$ e $\eta_{i2}$

Per dimostrare che  $\eta_{i1}$  e  $\eta_{i2}$  sono parametri con variazione indipendente può essere utile scrivere  $\eta_{i2}$  in funzione di  $\eta_{i1}$  e  $\pi_{i0}$ . In particolare si può notare che

$$\begin{aligned}\eta_{i2} &= \log \frac{\pi_{i1}\pi_{i0}}{(1-\pi_{i1})(1-\pi_{i0})} = \log \frac{\pi_{i1}\pi_{i0}^2}{\pi_{i0}(1-\pi_{i1})(1-\pi_{i0})} \\ &= \eta_{i1} + \log \frac{\pi_{i0}^2}{(1-\pi_{i0}e^{\eta_{i1}})(1-\pi_{i0})}.\end{aligned}\quad (2.11)$$

Dalla (2.11) risulta quindi chiaro come la componente che rende  $\eta_{i2}$  con variazione indipendente rispetto a  $\eta_{i1}$  sia  $\log \frac{\pi_{i0}^2}{(1-\pi_{i0}e^{\eta_{i1}})(1-\pi_{i0})}$ .

Infatti se  $\eta_{i1} > 0$  si ottiene che

$$\log \frac{\pi_{i0}^2}{(1-\pi_{i0}e^{\eta_{i1}})(1-\pi_{i0})} \in \mathbb{R}, \quad \pi_{i0} \in (0, e^{-\eta_{i1}}).$$

Mentre per  $\eta_{i1} < 0$

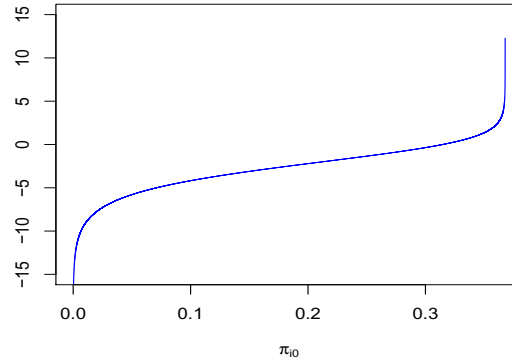
$$\log \frac{\pi_{i0}^2}{(1-\pi_{i0}e^{\eta_{i1}})(1-\pi_{i0})} \in \mathbb{R}, \quad \pi_{i0} \in (0, 1).$$

Nelle Figure 2.1 e 2.2 si può osservare graficamente come il valore di  $\log \frac{\pi_{i0}^2}{(1-\pi_{i0}e^{\eta_{i1}})(1-\pi_{i0})}$  non venga influenzato da quello assunto da  $\eta_{i1}$ . Dalla relazione (2.11) segue che anche  $\eta_{i2}$  risulta essere un parametro con variazione indipendente rispetto a  $\eta_{i1}$ .

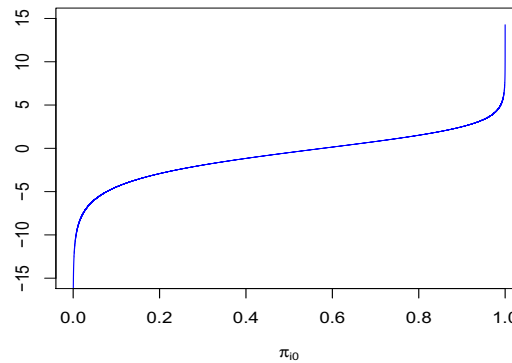
### 2.3.3 Interpretazione dei parametri e relazione con il modelli log-binomiale e logistico

Per quanto riguarda l'interpretazione dei parametri presenti in  $\eta_{i1}$ , si vede facilmente che  $\gamma_0$  è equivalente al parametro relativo a  $t_i$  nella regressione log-binomiale mentre  $\gamma_1, \dots, \gamma_{p-2}$  rappresentano, sempre nel modello log-binomiale, i parametri associati alle interazioni tra il fattore di esposizione binario  $t$  e le altre covariate disponibili. I parametri in  $\eta_{i2}$  sono invece una

**Figura 2.1:** Andamento di  $\log \frac{\pi_{i0}^2}{(1-\pi_{i0}e^{\eta_{i1}})(1-\pi_{i0})}$  al variare di  $\pi_{i0}$  con  $\eta_{i1} = 1$



**Figura 2.2:** Andamento di  $\log \frac{\pi_{i0}^2}{(1-\pi_{i0}e^{\eta_{i1}})(1-\pi_{i0})}$  al variare di  $\pi_{i0}$  con  $\eta_{i1} = -1$



combinazione lineare dei coefficienti della regressione logistica. Infatti, un generico  $\beta_j$ ,  $j = 1, \dots, p - 2$ , nel modello di Richardson *et al.* (2017) è ottenibile moltiplicando per 2 il coefficiente associato alla variabile  $x_{ij}$  e sommando il coefficiente relativo all'interazione tra  $x_{ij}$  e  $t_i$  nel modello logistico.

A titolo esemplificativo, ipotizziamo di disporre di un campione di  $n$  unità statistiche e di voler valutare attraverso il logaritmo del rischio relativo  $\log(\text{RR}_i)$  l'effetto del fattore di esposizione binario  $t_i$ , al netto di un'ulteriore

variabile esplicativa  $x_{i1}$ , sulla probabilità di osservare una determinata condizione di interesse  $y_i$  nell'unità  $i$ -esima. Se assumiamo che vi sia un effetto di interazione tra  $t_i$  e  $x_{i1}$ , il modello log-binomiale corrispondente ha il seguente predittore lineare

$$\log(\pi_i) = \beta_0 x_{i0} + \beta_1 x_{i1} + \gamma_0 t_i + \gamma_1 t_i x_{i1}.$$

Effettuando la differenza  $\log(\pi_{i1}) - \log(\pi_{i0})$ , a parità di  $x_{i1}$ , si ottiene che

$$\begin{aligned} \log(\text{RR}_i) &= \beta_0 + \beta_1 x_{i1} + \gamma_0 + \gamma_1 x_{i1} - \beta_0 - \beta_1 x_{i1} \\ &= \gamma_0 + \gamma_1 x_{i1}. \end{aligned}$$

Dato che in questo caso il modello di Richardson *et al.* (2017) assume per  $\eta_{i1}$  la forma

$$\eta_{i1} = \gamma_0 + \gamma_1 x_{i1},$$

i coefficienti relativi a  $\eta_{i1}$  coincidono con il vettore  $\gamma = (\gamma_0, \gamma_1)$  della regressione log-binomiale. Similmente, se si assume per i dati il modello di regressione logistica, si ha che

$$\log \frac{\pi_i}{1 - \pi_i} = \beta_0 + \beta_1 x_{i1} + \gamma_0 t_i + \gamma_1 t_i x_{i1}.$$

Calcolando il log-prodotto delle quote,  $\log \frac{\pi_{i0}\pi_{i1}}{(1-\pi_{i0})(1-\pi_{i1})}$ , è immediato notare che

$$\begin{aligned} \log \frac{\pi_{i0}\pi_{i1}}{(1-\pi_{i0})(1-\pi_{i1})} &= \beta_0 + \beta_1 x_{i1} + \beta_0 + \beta_1 x_{i1} + \gamma_0 t_i + \gamma_1 t_i x_{i1} \\ &= 2\beta_0 + 2\beta_1 x_{i1} + \gamma_0 + \gamma_1 x_{i1}, \end{aligned}$$

e che quindi i coefficienti relativi a  $\eta_{i2}$ , nel modello di Richardson *et al.* (2017), coincidono con il vettore  $(2\beta_0 + \gamma_0, 2\beta_1 + \gamma_1)$  dove  $\gamma_0, \gamma_1, \beta_0$  e  $\beta_1$  sono i parametri dalla regressione logistica.

### 2.3.4 Parametrizzazione alternativa

Come mostrato in dettaglio nel paragrafo A.1 dell'Appendice A, la parametrizzazione descritta nel paragrafo 2.3 permette di ottenere, per qualsiasi

combinazione di  $(\eta_{i1}, \eta_{i2}) \in \mathbb{R}^2$ , dei valori di  $(\pi_{i0}, \pi_{i1})$  entrambi compresi tra 0 e 1. Nonostante questo, nel caso in cui  $\eta_{i2}$  risulti pari a 0 la (2.9) non è definita ed è necessario prolungarla per continuità.

Sfruttando quanto visto in precedenza per il parametro  $\eta_{i2}$ , in questo paragrafo verrà introdotto un nuovo parametro di disturbo  $\eta_{i3}$ , sempre con variazione indipendente rispetto a  $\eta_{i1}$ , il quale mappa direttamente  $\pi_{i0}$  in  $(0, 1)$  senza bisogno di ulteriori estensioni per continuità. In particolare se si definisce

$$\eta_{i3} = \log \frac{\pi_{i0}}{(1 - \pi_{i0}e^{\eta_{i1}})(1 - \pi_{i0})}, \quad (2.12)$$

si ha che  $\eta_{i3}$  risulta con variazione indipendente rispetto a  $\eta_{i1}$ . Per dimostrare quest'ultima affermazione, si consideri innanzi tutto il caso in cui  $\eta_{i1} > 0$ . In questo frangente si ha che  $\eta_{i3}$  è una funzione continua e monotona crescente di  $\pi_{i0}$ , definita per  $\pi_{i0} \in (0, e^{-\eta_{i1}})$ . Inoltre,

$$\begin{aligned} \lim_{\pi_{i0} \rightarrow 0} \log \frac{\pi_{i0}}{(1 - \pi_{i0}e^{\eta_{i1}})(1 - \pi_{i0})} &= -\infty, & (\eta_{i1} \geq 0), \\ \lim_{\pi_{i0} \rightarrow e^{-\eta_{i1}}} \log \frac{\pi_{i0}}{(1 - \pi_{i0}e^{\eta_{i1}})(1 - \pi_{i0})} &= \infty, & (\eta_{i1} \geq 0). \end{aligned}$$

Similmente, nel caso in cui  $\eta_{i1} < 0$ , si ottiene che  $\eta_{i3}$  è una funzione continua e monotona crescente per  $\pi_{i0} \in (0, 1)$  e che in aggiunta

$$\begin{aligned} \lim_{\pi_{i0} \rightarrow 0} \log \frac{\pi_{i0}}{(1 - \pi_{i0}e^{\eta_{i1}})(1 - \pi_{i0})} &= -\infty, & (\eta_{i1} < 0), \\ \lim_{\pi_{i0} \rightarrow e^{-\eta_{i1}}} \log \frac{\pi_{i0}}{(1 - \pi_{i0}e^{\eta_{i1}})(1 - \pi_{i0})} &= \infty, & (\eta_{i1} < 0). \end{aligned}$$

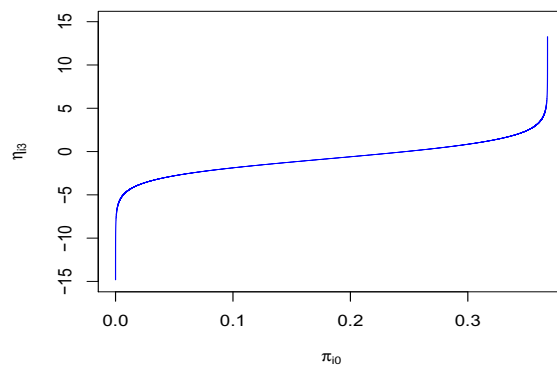
Dato che in entrambi i casi considerati il valore di  $\eta_{i3}$  non è in alcun modo influenzato dal valore assunto da  $\eta_{i1}$ , i parametri  $\eta_{i1}$  e  $\eta_{i3}$  sono quindi con variazione indipendente. Tale affermazione è verificabile anche graficamente nelle Figure 2.3 e 2.4.

A questo punto è possibile derivare le espressioni di  $\pi_{i0}$  e  $\pi_{i1}$  in funzione di  $\eta_{i1}$  e  $\eta_{i3}$ , le quali risultano pari a

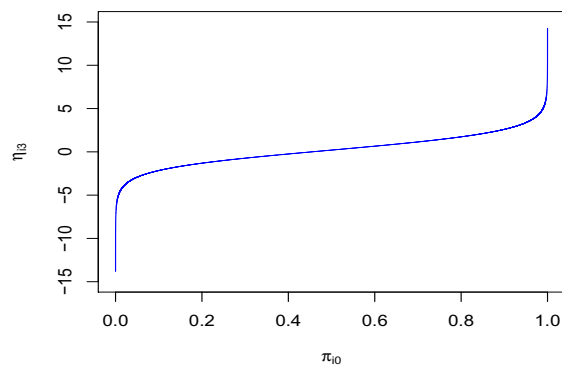
$$\begin{aligned} \pi_{i1} &= \pi_{i0}e^{\eta_{i1}} \\ \pi_{i0} &= \frac{[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})] - \sqrt{[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2 - 4e^{\eta_{i1} + 2\eta_{i3}}}}{2e^{\eta_{i1} + \eta_{i3}}}. \end{aligned}$$

In questo caso, i valori di  $\pi_{i0}$  e  $\pi_{i1}$  sono compresi tra 0 e 1 per ogni  $(\eta_{i1}, \eta_{i3}) \in \mathbb{R}^2$ . La dimostrazione di questo risultato è riportata all'interno del paragrafo A.2 in Appendice A.

**Figura 2.3:** Andamento di  $\eta_{i3}$  al variare di  $\pi_{i0}$  con  $\eta_{i1} = 1$



**Figura 2.4:** Andamento di  $\eta_{i3}$  al variare di  $\pi_{i0}$  con  $\eta_{i1} = -1$





# Capitolo 3

## Quantità necessarie per i metodi di riduzione della distorsione in media e mediana

In questo capitolo vengono riportate le quantità necessarie per implementare i metodi di riduzione della distorsione in media e mediana sia per il modello di Richardson *et al.* (2017) che per il modello con componente di disturbo alternativa (2.12). Nel calcolo di queste quantità è impiegata una notazione parzialmente discordante rispetto a quella utilizzata nel Capitolo 1. In particolare, per quanto riguarda le quantità di verosimiglianza, viene resa esplicita la dipendenza di queste ultime dalle componenti di  $\beta$  e  $\gamma$ .

### 3.1 Quantità di verosimiglianza per il modello di Richardson *et al.* (2017)

#### 3.1.1 Derivate della funzione di log-verosimiglianza

A partire dalla log-verosimiglianza (2.10) ed indicando con  $U_{\gamma_r} = (\partial/\partial\gamma_r) \ell(\gamma, \beta)$ ,  $U_{\beta_r} = (\partial/\partial\beta_r) \ell(\gamma, \beta)$ ,  $U_{\gamma_r\gamma_s} = (\partial^2/(\partial\gamma_r\partial\gamma_s)) \ell(\gamma, \beta)$ ,  $U_{\gamma_r\beta_s} = (\partial^2/(\partial\gamma_r\partial\beta_s))$

$\ell(\gamma, \beta)$ ,  $U_{\beta_r \beta_s} = (\partial^2 / (\partial \beta_r \partial \beta_s)) \ell(\gamma, \beta)$  le derivate prime e seconde di  $\ell(\gamma, \beta)$  si ottiene, per il modello proposto da Richardson *et al.* (2017), che

$$\begin{aligned} U_{\gamma_r} &= \sum_{i=1}^n \left[ \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \frac{\partial \pi_i}{\partial \eta_{i1}} \frac{\partial \eta_{i1}}{\partial \gamma_r} \right], \\ U_{\beta_r} &= \sum_{i=1}^n \left[ \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \frac{\partial \pi_i}{\partial \eta_{i2}} \frac{\partial \eta_{i2}}{\partial \beta_r} \right], \\ U_{\gamma_r \gamma_s} &= \sum_{i=1}^n \left[ -\frac{(y_i - y_i \pi_i + \pi_i^2)}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right)^2 + \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \frac{\partial^2 \pi_i}{\partial \eta_{i1}^2} \right] \frac{\partial \eta_{i1}}{\partial \gamma_r} \frac{\partial \eta_{i1}}{\partial \gamma_s}, \\ U_{\beta_r \beta_s} &= \sum_{i=1}^n \left[ -\frac{(y_i - y_i \pi_i + \pi_i^2)}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right)^2 + \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \frac{\partial^2 \pi_i}{\partial \eta_{i2}^2} \right] \frac{\partial \eta_{i2}}{\partial \beta_r} \frac{\partial \eta_{i2}}{\partial \beta_s}, \\ U_{\gamma_r \beta_s} &= \sum_{i=1}^n \left[ -\frac{(y_i - y_i \pi_i + \pi_i^2)}{[\pi_i (1 - \pi_i)]^2} \frac{\partial \pi_i}{\partial \eta_{i1}} \frac{\partial \pi_i}{\partial \eta_{i2}} + \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \frac{\partial^2 \pi_i}{\partial \eta_{i1} \partial \eta_{i2}} \right] \frac{\partial \eta_{i1}}{\partial \gamma_r} \frac{\partial \eta_{i2}}{\partial \beta_s}, \end{aligned}$$

dove  $\frac{\partial \eta_{i1}}{\partial \gamma_s} = \frac{\partial \eta_{i2}}{\partial \beta_s} = x_{is}$ , per  $s = 0, \dots, p-2$ .

Inoltre, dato che vale la relazione  $\pi_i = \pi_{i0} e^{t_i \eta_{i1}}$ , si ricava che

$$\frac{\partial \pi_i}{\partial \eta_{i1}} = \left( \frac{\partial \pi_{i0}}{\partial \eta_{i1}} \right) e^{t_i \eta_{i1}} + \pi_{i0} e^{t_i \eta_{i1}} t_i,$$

$$\frac{\partial \pi_i}{\partial \eta_{i2}} = \left( \frac{\partial \pi_{i0}}{\partial \eta_{i2}} \right) e^{t_i \eta_{i1}},$$

con

$$\frac{\partial \pi_{i0}}{\partial \eta_{i1}} = -\frac{e^{\eta_{i2} - \eta_{i1}}}{2(e^{\eta_{i2}} - 1)} + \frac{2e^{\eta_{i2}} + e^{-\eta_{i1}}(1 - e^{\eta_{i1}})e^{2\eta_{i2}}}{2(e^{\eta_{i2}} - 1)\sqrt{4e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}}}$$

$$\begin{aligned} \frac{\partial \pi_{i0}}{\partial \eta_{i2}} &= -\frac{e^{-\eta_{i1}}(e^{\eta_{i1}} + 1)e^{\eta_{i2}}}{2(e^{\eta_{i2}} - 1)^2} \\ &+ \frac{2e^{\eta_{i2}} + e^{-\eta_{i1}}(e^{2\eta_{i1}} + 1)e^{2\eta_{i2}}}{2(e^{\eta_{i2}} - 1)^2 \sqrt{4e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}}}. \end{aligned}$$



Le derivate seconde di  $\pi_i$  rispetto a  $\eta_{i1}$  e  $\eta_{i2}$  assumono invece la forma

$$\begin{aligned}\frac{\partial^2 \pi_i}{\partial^2 \eta_{i1}} &= \frac{\partial^2 \pi_{i0}}{\partial \eta_{i1}^2} e^{t_i \eta_{i1}} + 2 \frac{\partial \pi_{i0}}{\eta_{i1}} e^{t_i \eta_{i1}} t_i + \pi_{i0} e^{t_i \eta_{i1}} t_i^2 \\ &= e^{t_i \eta_{i1}} \left( \frac{\partial^2 \pi_{i0}}{\partial^2 \eta_{i1}} + 2 \frac{\partial \pi_{i0}}{\partial \eta_{i1}} + \pi_{i0} t_i \right), \\ \frac{\partial^2 \pi_i}{\partial^2 \eta_{i2}} &= e^{t_i \eta_{i1}} \frac{\partial^2 \pi_{i0}}{\partial^2 \eta_{i2}}, \\ \frac{\partial^2 \pi_i}{\partial \eta_{i1} \partial \eta_{i2}} &= e^{t_i \eta_{i1}} \frac{\partial^2 \pi_{i0}}{\partial \eta_{i1} \partial \eta_{i2}} + t_i e^{t_i \eta_{i1}} \frac{\partial \pi_{i0}}{\partial \eta_{i2}},\end{aligned}$$

con

$$\begin{aligned}\frac{\partial^2 \pi_i}{\partial^2 \eta_{i1}} &= \frac{e^{\eta_{i2} - \eta_{i1}}}{2(e^{\eta_{i2}} - 1)} - \frac{(1 - e^{\eta_{i1}}) e^{2\eta_{i2} - \eta_{i1}} + e^{2\eta_{i2}}}{2(e^{\eta_{i2}} - 1) \sqrt{4e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}}} \\ &+ \frac{e^{2\eta_{i2}} (e^{\eta_{i2} + \eta_{i1}} - e^{\eta_{i2}} + 2) (e^{\eta_{i2} + \eta_{i1}} - e^{\eta_{i2}} - 2e^{\eta_{i1}})}{2(e^{\eta_{i2}} - 1) [4e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}]^{\frac{3}{2}}},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \pi_{i0}}{\partial^2 \eta_{i2}} &= \frac{(e^{\eta_{i1}} + 1) e^{2\eta_{i2} - \eta_{i1}}}{(e^{\eta_{i2}} - 1)^3} - \frac{(e^{\eta_{i1}} + 1) e^{\eta_{i2} - \eta_{i1}}}{2(e^{\eta_{i2}} - 1)^2} \\ &- \frac{[e^{\eta_{i2}} (e^{\eta_{i2}} + 1) + (e^{2\eta_{i1}} + 1) e^{2\eta_{i2} - \eta_{i1}}]}{(e^{\eta_{i2}} - 1)^3 \sqrt{4e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}}} \\ &- \frac{(e^{2\eta_{i1}} + 1) (2e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}) e^{2\eta_{i2} - \eta_{i1}}}{2(e^{\eta_{i2}} - 1)^2 [4e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}]^{\frac{3}{2}}} \\ &- \frac{e^{\eta_{i2}} (2e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}})}{(e^{\eta_{i2}} - 1)^2 [4e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}]^{\frac{3}{2}}},\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \pi_{i0}}{\partial \eta_{i1} \partial \eta_{i2}} &= \frac{e^{2\eta_{i2} - \eta_{i1}}}{2(e^{\eta_{i2}} - 1)^2} - \frac{e^{\eta_{i2} - \eta_{i1}}}{2(e^{\eta_{i2}} - 1)} \\
&+ \frac{(e^{\eta_{i1}} - 1) e^{3\eta_{i2} - \eta_{i1}} - 2e^{\eta_{i2}}}{2(e^{\eta_{i2}} - 1)^2 \sqrt{4e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}}} \\
&+ \frac{(1 - e^{\eta_{i1}}) e^{2\eta_{i2} - \eta_{i1}}}{(e^{\eta_{i2}} - 1) \sqrt{4e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}}} \\
&- \frac{e^{\eta_{i2}} (4e^{\eta_{i2} + \eta_{i1}} + 2(e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}})}{2(e^{\eta_{i2}} - 1) [4e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}]^{\frac{3}{2}}} \\
&- \frac{(1 - e^{\eta_{i1}}) (4e^{\eta_{i2} + \eta_{i1}} + 2(e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}) e^{2\eta_{i2} - \eta_{i1}}}{4(e^{\eta_{i2}} - 1) [4e^{\eta_{i2} + \eta_{i1}} + (e^{\eta_{i1}} - 1)^2 e^{2\eta_{i2}}]^{\frac{3}{2}}}.
\end{aligned}$$

### 3.1.2 Prolungamento per continuità delle derivate di $\pi_{i0}$

Come visto per  $\pi_{i0}$  nella (A.7), anche le sue derivate non sono definite per  $\eta_{i2} = 0$  ma è possibile prolungare per continuità. Infatti, utilizzando lo sviluppo di Taylor di  $e^{\eta_{i2}}$  attorno a  $\eta_{i2} = 0$ , ossia

$$e^{\eta_{i2}} = 1 + \eta_{i2} + o(\eta_{i2}),$$

si ottiene che

$$\begin{aligned}
\lim_{\eta_{i2} \rightarrow 0} \frac{\partial \pi_{i0}}{\partial \eta_{i1}} &= - \frac{e^{\eta_{i1}}}{(e^{\eta_{i1}} + 1)^2}, \\
\lim_{\eta_{i2} \rightarrow 0} \frac{\partial \pi_{i0}}{\partial \eta_{i2}} &= \frac{e^{\eta_{i1}}}{(e^{\eta_{i1}} + 1)^3}, \\
\lim_{\eta_{i2} \rightarrow 0} \frac{\partial^2 \pi_{i0}}{\partial \eta_{i1}^2} &= \frac{(e^{\eta_{i1}} - 1) e^{\eta_{i1}}}{(e^{\eta_{i1}} + 1)^3}, \\
\lim_{\eta_{i2} \rightarrow 0} \frac{\partial \pi_{i0}^2}{\partial \eta_{i2}^2} &= - \frac{(e^{\eta_{i1}} - 1)^2 e^{\eta_{i1}}}{(e^{\eta_{i1}} + 1)^5}, \\
\lim_{\eta_{i2} \rightarrow 0} \frac{\partial \pi_{i0}^2}{\partial \eta_{i1} \partial \eta_{i2}} &= - \frac{e^{\eta_{i1}} (2e^{\eta_{i1}} - 1)}{(e^{\eta_{i1}} + 1)^4}.
\end{aligned}$$

### 3.1.3 Informazione attesa di Fisher

Utilizzando i risultati del paragrafo 3.1.1 è agevole ottenere l'espressione dell'informazione attesa di Fisher,  $i(\theta)$ , con  $\theta = (\gamma, \beta)$ . In particolare si ha che

$$\begin{aligned} i(\theta)_{\gamma_r, \gamma_s} &= \sum_{i=1}^n \frac{x_{ir}x_{is}}{\pi_i(1-\pi_i)} \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right)^2, \\ i(\theta)_{\beta_r, \beta_s} &= \sum_{i=1}^n \frac{x_{ir}x_{is}}{\pi_i(1-\pi_i)} \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right)^2, \\ i(\theta)_{\gamma_r, \beta_s} &= \sum_{i=1}^n \frac{x_{ir}x_{is}}{\pi_i(1-\pi_i)} \frac{\partial \pi_i}{\partial \eta_{i1}} \frac{\partial \pi_i}{\partial \eta_{i2}}, \end{aligned}$$

con  $r, s = 0, \dots, p-2$ .

## 3.2 Riduzione della distorsione in media e mediana per il modello di Richardson *et al.* (2017)

### 3.2.1 Definizione di $P_{\gamma_k}(\theta)$ , $P_{\beta_k}(\theta)$ , $Q_{\gamma_k}(\theta)$ e $Q_{\beta_k}(\theta)$

Per implementare i metodi di riduzione della distorsione in media e mediana è necessario innanzitutto definire, nel caso del modello di Richardson *et al.* (2017), le matrici  $P_{\gamma_k}(\theta)$ ,  $P_{\beta_k}(\theta)$ ,  $Q_{\gamma_k}(\theta)$  e  $Q_{\beta_k}(\theta)$  per  $k = 0, \dots, p-2$ , descritte nel Capitolo 1. In particolare, utilizzando i risultati del paragrafo 3.1.1, si ottiene

$$P_{\gamma_k}(\theta) = \left[ \begin{array}{ccc|ccc} P_{\gamma_k}(\theta)_{\gamma_0\gamma_0} & \cdots & P_{\gamma_k}(\theta)_{\gamma_0\gamma_{p-2}} & P_{\gamma_k}(\theta)_{\gamma_0\beta_0} & \cdots & P_{\gamma_k}(\theta)_{\gamma_0\beta_{p-2}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ P_{\gamma_k}(\theta)_{\gamma_{p-2}\gamma_0} & \cdots & P_{\gamma_k}(\theta)_{\gamma_{p-2}\gamma_{p-2}} & P_{\gamma_k}(\theta)_{\gamma_{p-2}\beta_0} & \cdots & P_{\gamma_k}(\theta)_{\gamma_{p-2}\beta_{p-2}} \\ \hline P_{\gamma_k}(\theta)_{\beta_0\gamma_0} & \cdots & P_{\gamma_k}(\theta)_{\beta_0\gamma_{p-2}} & P_{\gamma_k}(\theta)_{\beta_0\beta_0} & \cdots & P_{\gamma_k}(\theta)_{\beta_0\beta_{p-2}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ P_{\gamma_k}(\theta)_{\beta_{p-2}\gamma_0} & \cdots & P_{\gamma_k}(\theta)_{\beta_{p-2}\gamma_{p-2}} & P_{\gamma_k}(\theta)_{\beta_{p-2}\beta_0} & \cdots & P_{\gamma_k}(\theta)_{\beta_{p-2}\beta_{p-2}} \end{array} \right],$$

con

$$\begin{aligned} P_{\gamma_k}(\theta)_{\gamma_s \gamma_r} &= \sum_{i=1}^n \mathbb{E}_\theta \left[ \left( \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \right)^3 \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right)^3 \frac{\partial \eta_{i1}}{\partial \gamma_k} \frac{\partial \eta_{i1}}{\partial \gamma_s} \frac{\partial \eta_{i1}}{\partial \gamma_r} \right] \\ &= \sum_{i=1}^n \left[ \frac{(1 - 2\pi_i)}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right)^3 x_{ik} x_{is} x_{ir} \right], \end{aligned}$$

$$\begin{aligned} P_{\gamma_k}(\theta)_{\beta_s \beta_r} &= \sum_{i=1}^n \mathbb{E}_\theta \left[ \left( \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \right)^3 \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right) \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right)^2 \frac{\partial \eta_{i1}}{\partial \gamma_k} \frac{\partial \eta_{i2}}{\partial \beta_s} \frac{\partial \eta_{i2}}{\partial \beta_r} \right] \\ &= \sum_{i=1}^n \left[ \frac{(1 - 2\pi_i)}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right) \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right)^2 x_{ik} x_{is} x_{ir} \right], \end{aligned}$$

$$\begin{aligned} P_{\gamma_k}(\theta)_{\gamma_s \beta_r} &= \sum_{i=1}^n \mathbb{E}_\theta \left[ \left( \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \right)^3 \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right)^2 \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right) \frac{\partial \eta_{i1}}{\partial \gamma_k} \frac{\partial \eta_{i1}}{\partial \gamma_s} \frac{\partial \eta_{i2}}{\partial \beta_r} \right] \\ &= \sum_{i=1}^n \left[ \frac{(1 - 2\pi_i)}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right)^2 \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right) x_{ik} x_{is} x_{ir} \right], \end{aligned}$$

per  $k, s, r = 0, \dots, p-2$ .

Similmente  $P_{\beta_k}(\theta)$  assume la forma

$$P_{\beta_k}(\theta) = \left[ \begin{array}{ccc|ccc} P_{\beta_k}(\theta)_{\gamma_0 \gamma_0} & \cdots & P_{\beta_k}(\theta)_{\gamma_0 \gamma_{p-2}} & P_{\beta_k}(\theta)_{\gamma_0 \beta_0} & \cdots & P_{\beta_k}(\theta)_{\gamma_0 \beta_{p-2}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ P_{\beta_k}(\theta)_{\gamma_{p-2} \gamma_0} & \cdots & P_{\beta_k}(\theta)_{\gamma_{p-2} \gamma_{p-2}} & P_{\beta_k}(\theta)_{\gamma_{p-2} \beta_0} & \cdots & P_{\beta_k}(\theta)_{\gamma_{p-2} \beta_{p-2}} \\ P_{\beta_k}(\theta)_{\beta_0 \gamma_0} & \cdots & P_{\beta_k}(\theta)_{\beta_0 \gamma_{p-2}} & P_{\beta_k}(\theta)_{\beta_0 \beta_0} & \cdots & P_{\beta_k}(\theta)_{\beta_0 \beta_{p-2}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ P_{\beta_k}(\theta)_{\beta_{p-2} \gamma_0} & \cdots & P_{\beta_k}(\theta)_{\beta_{p-2} \gamma_{p-2}} & P_{\beta_k}(\theta)_{\beta_{p-2} \beta_0} & \cdots & P_{\beta_k}(\theta)_{\beta_{p-2} \beta_{p-2}} \end{array} \right],$$

con

$$\begin{aligned} P_{\beta_k}(\theta)_{\gamma_s \gamma_r} &= \sum_{i=1}^n \mathbb{E}_\theta \left[ \left( \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \right)^3 \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right) \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right)^2 \frac{\partial \eta_{i2}}{\partial \beta_k} \frac{\partial \eta_{i1}}{\partial \gamma_s} \frac{\partial \eta_{i1}}{\partial \gamma_r} \right] \\ &= \sum_{i=1}^n \left[ \frac{(1 - 2\pi_i)}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right) \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right)^2 x_{ik} x_{is} x_{ir} \right], \end{aligned}$$

$$\begin{aligned} P_{\beta_k}(\theta)_{\beta_s \beta_r} &= \sum_{i=1}^n \mathbb{E}_\theta \left[ \left( \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \right)^3 \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right)^3 \frac{\partial \eta_{i2}}{\partial \beta_k} \frac{\partial \eta_{i2}}{\partial \beta_s} \frac{\partial \eta_{i2}}{\partial \beta_r} \right] \\ &= \sum_{i=1}^n \left[ \frac{(1 - 2\pi_i)}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right)^3 x_{ik} x_{is} x_{ir} \right], \end{aligned}$$

$$\begin{aligned}
P_{\beta_k}(\theta)_{\gamma_s \beta_r} &= \sum_{i=1}^n \mathbb{E}_\theta \left[ \left( \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \right)^3 \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right) \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right)^2 \frac{\partial \eta_{i2}}{\partial \beta_k} \frac{\partial \eta_{i1}}{\partial \gamma_s} \frac{\partial \eta_{i2}}{\partial \beta_r} \right] \\
&= \sum_{i=1}^n \left[ \frac{(1 - 2\pi_i)}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right) \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right)^2 x_{ik} x_{is} x_{ir} \right],
\end{aligned}$$

per  $k, s, r = 0, \dots, p-2$ .

La matrice  $Q_{\gamma_k}(\theta)$  risulta invece pari a

$$Q_{\gamma_k}(\theta) = \left[ \begin{array}{ccc|ccc}
Q_{\gamma_k}(\theta)_{\gamma_0 \gamma_0} & \cdots & Q_{\gamma_k}(\theta)_{\gamma_0 \gamma_{p-2}} & Q_{\gamma_k}(\theta)_{\gamma_0 \beta_0} & \cdots & Q_{\gamma_k}(\theta)_{\gamma_0 \beta_{p-2}} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
Q_{\gamma_k}(\theta)_{\gamma_{p-2} \gamma_0} & \cdots & Q_{\gamma_k}(\theta)_{\gamma_{p-2} \gamma_{p-2}} & Q_{\gamma_k}(\theta)_{\gamma_{p-2} \beta_0} & \cdots & Q_{\gamma_k}(\theta)_{\gamma_{p-2} \beta_{p-2}} \\
\hline
Q_{\gamma_k}(\theta)_{\beta_0 \gamma_0} & \cdots & Q_{\gamma_k}(\theta)_{\beta_0 \gamma_{p-2}} & Q_{\gamma_k}(\theta)_{\beta_0 \beta_0} & \cdots & Q_{\gamma_k}(\theta)_{\beta_0 \beta_{p-2}} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
Q_{\gamma_k}(\theta)_{\beta_{p-2} \gamma_0} & \cdots & Q_{\gamma_k}(\theta)_{\beta_{p-2} \gamma_{p-2}} & Q_{\gamma_k}(\theta)_{\beta_{p-2} \beta_0} & \cdots & Q_{\gamma_k}(\theta)_{\beta_{p-2} \beta_{p-2}}
\end{array} \right],$$

con

$$\begin{aligned}
Q_{\gamma_k}(\theta)_{\gamma_s \gamma_r} &= \mathbb{E}_\theta \left[ \sum_{i=1}^n \left[ \left( -\frac{y_i - y_i \pi_i + \pi_i^2}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right)^2 + \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \frac{\partial^2 \pi_i}{\partial^2 \eta_{i1}} \right) \frac{\partial \eta_{i1}}{\partial \gamma_s} \frac{\partial \eta_{i1}}{\partial \gamma_r} \right] \right. \\
&\quad \times \left. \sum_{i=1}^n \left[ \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \frac{\partial \pi_i}{\partial \eta_{i1}} \frac{\partial \eta_{i1}}{\partial \gamma_k} \right] \right] \\
&= \sum_{i=1}^n \left[ -\frac{(1 - 2\pi_i)}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right)^3 + \frac{1}{\pi_i (1 - \pi_i)} \frac{\partial^2 \pi_i}{\partial^2 \eta_{i1}} \frac{\partial \pi_i}{\partial \eta_{i1}} \right] x_{ik} x_{is} x_{ir},
\end{aligned}$$

$$\begin{aligned}
Q_{\gamma_k}(\theta)_{\gamma_s \beta_r} &= \mathbb{E}_\theta \left[ \sum_{i=1}^n \left[ \left( -\frac{y_i - y_i \pi_i + \pi_i^2}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right) \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right) + \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \frac{\partial^2 \pi_i}{\partial \eta_{i1} \partial \eta_{i2}} \right) \right. \right. \\
&\quad \times \left. \frac{\partial \eta_{i1}}{\partial \gamma_s} \frac{\eta_{i2}}{\partial \beta_r} \right] \times \sum_{i=1}^n \left[ \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \frac{\partial \pi_i}{\partial \eta_{i1}} \frac{\partial \eta_{i1}}{\partial \gamma_k} \right] \right] \\
&= \sum_{i=1}^n \left[ -\frac{(1 - 2\pi_i)}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right)^2 \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right) + \frac{1}{\pi_i (1 - \pi_i)} \frac{\partial^2 \pi_i}{\partial \eta_{i1} \partial \eta_{i2}} \frac{\partial \pi_i}{\partial \eta_{i1}} \right] x_{ik} x_{is} x_{ir},
\end{aligned}$$

$$\begin{aligned}
Q_{\gamma_k}(\theta)_{\beta_s \beta_r} &= \mathbb{E}_\theta \left[ \sum_{i=1}^n \left[ \left( -\frac{y_i - y_i \pi_i + \pi_i^2}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right)^2 + \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \frac{\partial^2 \pi_i}{\partial^2 \eta_{i2}} \right) \frac{\partial \eta_{i2}}{\partial \beta_s} \frac{\eta_{i2}}{\partial \beta_r} \right] \right. \\
&\quad \times \left. \sum_{i=1}^n \left[ \frac{y_i - \pi_i}{\pi_i (1 - \pi_i)} \frac{\partial \pi_i}{\partial \eta_{i1}} \frac{\partial \eta_{i1}}{\partial \gamma_k} \right] \right] \\
&= \sum_{i=1}^n \left[ -\frac{(1 - 2\pi_i)}{[\pi_i (1 - \pi_i)]^2} \left( \frac{\partial \pi_i}{\partial \eta_{i1}} \right) \left( \frac{\partial \pi_i}{\partial \eta_{i2}} \right)^2 + \frac{1}{\pi_i (1 - \pi_i)} \frac{\partial^2 \pi_i}{\partial^2 \eta_{i2}} \frac{\partial \pi_i}{\partial \eta_{i1}} \right] x_{ik} x_{is} x_{ir},
\end{aligned}$$

per  $k, s, r = 0, \dots, p-2$ .

In modo analogo è possibile ottenere l'espressione per  $Q_{\beta_k}(\theta)$  ossia

$$Q_{\beta_k}(\theta) = \left[ \begin{array}{ccc|ccc} Q_{\beta_k}(\theta)_{\gamma_0\gamma_0} & \cdots & Q_{\beta_k}(\theta)_{\gamma_0\gamma_{p-2}} & Q_{\beta_k}(\theta)_{\gamma_0\beta_0} & \cdots & Q_{\beta_k}(\theta)_{\gamma_0\beta_{p-2}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ Q_{\beta_k}(\theta)_{\gamma_{p-2}\gamma_0} & \cdots & Q_{\beta_k}(\theta)_{\gamma_{p-2}\gamma_{p-2}} & Q_{\beta_k}(\theta)_{\gamma_{p-2}\beta_0} & \cdots & Q_{\beta_k}(\theta)_{\gamma_{p-2}\beta_{p-2}} \\ \hline Q_{\beta_k}(\theta)_{\beta_0\gamma_0} & \cdots & Q_{\beta_k}(\theta)_{\beta_0\gamma_{p-2}} & Q_{\beta_k}(\theta)_{\beta_0\beta_0} & \cdots & Q_{\beta_k}(\theta)_{\beta_0\beta_{p-2}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ Q_{\beta_k}(\theta)_{\beta_{p-2}\gamma_0} & \cdots & Q_{\beta_k}(\theta)_{\beta_{p-2}\gamma_{p-2}} & Q_{\beta_k}(\theta)_{\beta_{p-2}\beta_0} & \cdots & Q_{\beta_k}(\theta)_{\beta_{p-2}\beta_{p-2}} \end{array} \right],$$

con

$$\begin{aligned} Q_{\beta_k}(\theta)_{\gamma_s\gamma_r} &= E_{\theta} \left[ \sum_{i=1}^n \left[ \left( -\frac{y_i - y_i\pi_i + \pi_i^2}{[\pi_i(1-\pi_i)]^2} \left( \frac{\partial\pi_i}{\partial\eta_{i1}} \right)^2 + \frac{y_i - \pi_i}{\pi_i(1-\pi_i)} \right) \frac{\partial^2\pi_i}{\partial^2\eta_{i1}} \right] \frac{\partial\eta_{i1}}{\partial\gamma_s} \frac{\partial\eta_{i1}}{\partial\gamma_r} \right. \\ &\quad \left. \times \sum_{i=1}^n \left[ \frac{y_i - \pi_i}{\pi_i(1-\pi_i)} \frac{\partial\pi_i}{\partial\eta_{i2}} \frac{\partial\eta_{i2}}{\partial\beta_k} \right] \right] \\ &= \sum_{i=1}^n \left[ -\frac{(1-2\pi_i)}{[\pi_i(1-\pi_i)]^2} \left( \frac{\partial\pi_i}{\partial\eta_{i1}} \right)^2 \left( \frac{\partial\pi_i}{\partial\eta_{i2}} \right) + \frac{1}{\pi_i(1-\pi_i)} \frac{\partial^2\pi_i}{\partial^2\eta_{i1}} \frac{\partial\pi_i}{\partial\eta_{i2}} \right] x_{ik}x_{is}x_{ir}, \end{aligned}$$

$$\begin{aligned} Q_{\beta_k}(\theta)_{\gamma_s\beta_r} &= E_{\theta} \left[ \sum_{i=1}^n \left[ \left( -\frac{y_i - y_i\pi_i + \pi_i^2}{[\pi_i(1-\pi_i)]^2} \frac{\partial\pi_i}{\partial\eta_{i1}} \frac{\partial\pi_i}{\partial\eta_{i2}} + \frac{y_i - \pi_i}{\pi_i(1-\pi_i)} \frac{\partial^2\pi_i}{\partial\eta_{i1}\partial\eta_{i2}} \right) \frac{\partial\eta_{i1}}{\partial\gamma_s} \frac{\partial\eta_{i2}}{\partial\beta_r} \right] \right. \\ &\quad \left. \times \sum_{i=1}^n \left[ \frac{y_i - \pi_i}{\pi_i(1-\pi_i)} \frac{\partial\pi_i}{\partial\eta_{i2}} \frac{\partial\eta_{i2}}{\partial\beta_k} \right] \right] \\ &= \sum_{i=1}^n \left[ -\frac{(1-2\pi_i)}{[\pi_i(1-\pi_i)]^2} \left( \frac{\partial\pi_i}{\partial\eta_{i1}} \right) \left( \frac{\partial\pi_i}{\partial\eta_{i2}} \right)^2 + \frac{1}{\pi_i(1-\pi_i)} \frac{\partial^2\pi_i}{\partial\eta_{i1}\partial\eta_{i2}} \frac{\partial\pi_i}{\partial\eta_{i2}} \right] x_{ik}x_{is}x_{ir}, \end{aligned}$$

$$\begin{aligned} Q_{\beta_k}(\theta)_{\beta_s\beta_r} &= E_{\theta} \left[ \sum_{i=1}^n \left[ \left( -\frac{y_i - y_i\pi_i + \pi_i^2}{[\pi_i(1-\pi_i)]^2} \frac{\partial\pi_i}{\partial\eta_{i2}} + \frac{y_i - \pi_i}{\pi_i(1-\pi_i)} \frac{\partial^2\pi_i}{\partial^2\eta_{i2}} \right) \frac{\partial\eta_{i2}}{\partial\beta_s} \frac{\partial\eta_{i2}}{\partial\beta_r} \right] \right. \\ &\quad \left. \times \sum_{i=1}^n \left[ \frac{y_i - \pi_i}{\pi_i(1-\pi_i)} \frac{\partial\pi_i}{\partial\eta_{i2}} \frac{\partial\eta_{i2}}{\partial\beta_k} \right] \right] \\ &= \sum_{i=1}^n \left[ -\frac{(1-2\pi_i)}{[\pi_i(1-\pi_i)]^2} \left( \frac{\partial\pi_i}{\partial\eta_{i2}} \right)^3 + \frac{1}{\pi_i(1-\pi_i)} \frac{\partial^2\pi_i}{\partial^2\eta_{i2}} \frac{\partial\pi_i}{\partial\eta_{i2}} \right] x_{ik}x_{is}x_{ir}. \end{aligned}$$

### 3.2.2 Definizione di $F_1$ , $F_{2\gamma_j}$ e $F_{2\beta_j}$

Una volta ottenute  $P_{\gamma_k}(\theta)$ ,  $P_{\beta_k}(\theta)$ ,  $Q_{\gamma_k}(\theta)$  e  $Q_{\beta_k}(\theta)$  risulta necessario definire i vettori  $F_1$ ,  $F_{2,\gamma_j}$  e  $F_{2,\beta_j}$ , di dimensione  $2(p-1)$ . Nel caso del modello

proposto da Richardson *et al.* (2017)  $F_1$  assume la forma

$$F_1 = \begin{pmatrix} F_{1\gamma_0} \\ \vdots \\ F_{1\gamma_{p-2}} \\ F_{1\beta_0} \\ \vdots \\ F_{1\beta_{p-2}} \end{pmatrix},$$

con

$$\begin{aligned} F_{1\gamma_j} &= \text{tr} [i(\theta)^{-1} \{P_{\gamma_j}(\theta) + Q_{\gamma_j}(\theta)\}] & (j = 0, \dots, p-2), \\ F_{1\beta_j} &= \text{tr} [i(\theta)^{-1} \{P_{\beta_j}(\theta) + Q_{\beta_j}(\theta)\}] & (j = 0, \dots, p-2). \end{aligned}$$

$F_{2,\gamma_j}$  risulta invece pari a

$$F_{2,\gamma_j} = \begin{pmatrix} F_{2\gamma_0,\gamma_j} \\ \vdots \\ F_{2\gamma_{p-2},\gamma_j} \\ F_{2\beta_0,\gamma_j} \\ \vdots \\ F_{2\beta_{p-2},\gamma_j} \end{pmatrix} \quad (j = 0, \dots, p-2),$$

con

$$\begin{aligned} F_{2\gamma_k,\gamma_j} &= \text{tr} \left[ h_{\gamma_j}(\theta) \left\{ \frac{1}{3}P_{\gamma_k}(\theta) + \frac{1}{2}Q_{\gamma_k}(\theta) \right\} \right] & (k = 0, \dots, p-2), \\ F_{2\beta_k,\gamma_j} &= \text{tr} \left[ h_{\gamma_j}(\theta) \left\{ \frac{1}{3}P_{\beta_k}(\theta) + \frac{1}{2}Q_{\beta_k}(\theta) \right\} \right] & (k = 0, \dots, p-2), \end{aligned}$$

dove  $h_{\gamma_j}(\theta) = [i(\theta)^{-1}]_{\gamma_j} [i(\theta)^{-1}]_{\gamma_j}^T / i(\theta)_{\gamma_j\gamma_j}^{-1}$ .

Infine  $F_{2,\beta_j}$  assume la forma

$$F_{2\beta_j} = \begin{pmatrix} F_{2\gamma_0,\beta_j} \\ \vdots \\ F_{2\gamma_{p-2},\beta_j} \\ F_{2\beta_0,\beta_j} \\ \vdots \\ F_{2\beta_{p-2},\beta_j} \end{pmatrix} \quad (j = 0, \dots, p-2),$$

con

$$F_{2\gamma_k, \beta_j} = \text{tr} \left[ h_{\beta_j}(\theta) \left\{ \frac{1}{3} P_{\gamma_k}(\theta) + \frac{1}{2} Q_{\gamma_k}(\theta) \right\} \right] \quad (k = 0, \dots, p-2),$$

$$F_{2\beta_k, \beta_j} = \text{tr} \left[ h_{\beta_j}(\theta) \left\{ \frac{1}{3} P_{\beta_k}(\theta) + \frac{1}{2} Q_{\beta_k}(\theta) \right\} \right] \quad (k = 0, \dots, p-2),$$

dove  $h_{\beta_j}(\theta) = [i(\theta)^{-1}]_{\beta_j} [i(\theta)^{-1}]_{\beta_j}^T / i(\theta)_{\beta_j \beta_j}^{-1}$ .

### 3.2.3 Equazioni *score* modificate

A questo punto è possibile ottenere l'equazione *score* modificata, per la riduzione della distorsione in mediana, nella forma proposta da Kenne Pagui *et al.* (2020) ossia

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

dove  $U(\theta) = (U_{\gamma_0}, \dots, U_{\gamma_{p-2}}, U_{\beta_0}, \dots, U_{\beta_{p-2}})$  è un vettore di dimensione  $2(p-1)$  mentre il termine di aggiustamento  $\tilde{A}(\theta)$  può essere scritto come

$$\tilde{A}(\theta) = \frac{1}{2} F_1 - i(\theta) \tilde{F}_2, \quad (3.1)$$

dove

$$\tilde{F}_2 = \begin{pmatrix} \tilde{F}_{2\gamma_0} \\ \vdots \\ \tilde{F}_{2\gamma_{p-2}} \\ \tilde{F}_{2\beta_0} \\ \vdots \\ \tilde{F}_{2\beta_{p-2}} \end{pmatrix},$$

è un vettore di dimensione  $2(p-1)$  con elementi  $\tilde{F}_{2\gamma_j} = [i(\theta)^{-1}]_{\gamma_j}^T F_{2,\gamma_j}$ ,  $\tilde{F}_{2\beta_j} = [i(\theta)^{-1}]_{\beta_j}^T F_{2,\beta_j}$  per  $j = 1, \dots, p-2$ . Utilizzando i risultati esposti da Kenne Pagui *et al.* (2020) è possibile, eliminando il secondo termine della (3.1), ottenere l'equazione *score* modificata per la riduzione della distorsione in media proposta da Firth (1993). Questa è quindi pari a

$$U(\theta)^* = U(\theta) + \frac{1}{2} F_1.$$



### 3.3 Parametrizzazione alternativa per la componente di disturbo

L'implementazione dei metodi di riduzione della distorsione, nel modello avete (2.12) come parametro di disturbo, avviene ricalcando quanto esposto nel paragrafo 3.2. È infatti sufficiente modificare l'espressione di  $\pi_{i0}$ , che in questo caso è funzione dei predittori lineari  $\eta_{i1}$  e  $\eta_{i3}$ . Rispetto a questi ultimi, le derivate parziali di  $\pi_{i0}$  risultano pari a

$$\frac{\partial \pi_{i0}}{\partial \eta_{i1}} = -\frac{e^{-\eta_{i3}-\eta_{i1}} (e^{\eta_{i3}} + 1)}{2} - \frac{e^{-\eta_{i3}-\eta_{i1}} [(e^{\eta_{i1}} - 1) e^{2\eta_{i3}} - (e^{\eta_{i1}} + 2) e^{\eta_{i3}} - 1]}{2\sqrt{(e^{\eta_{i1}} - 1)^2 e^{2\eta_{i3}} + 2(e^{\eta_{i1}} + 1) e^{\eta_{i3}} + 1}},$$

$$\frac{\partial \pi_{i0}}{\partial \eta_{i3}} = -\frac{e^{-\eta_{i3}-\eta_{i1}}}{2} + \frac{e^{-\eta_{i3}-\eta_{i1}} [(e^{\eta_{i1}} + 1) e^{\eta_{i3}} + 1]}{2\sqrt{(e^{\eta_{i1}} - 1)^2 e^{2\eta_{i3}} + 2(e^{\eta_{i1}} + 1) e^{\eta_{i3}} + 1}},$$

$$\begin{aligned} \frac{\partial^2 \pi_{i0}}{\partial^2 \eta_{i1}} &= \frac{e^{-\eta_{i3}-\eta_{i1}} (e^{\eta_{i3}} + 1)}{2} - \frac{e^{-\eta_{i3}-\eta_{i1}} (e^{\eta_{i3}} + 1)^2}{2\sqrt{(e^{\eta_{i1}} - 1)^2 e^{2\eta_{i3}} + 2(e^{\eta_{i1}} + 1) e^{\eta_{i3}} + 1}} \\ &\quad + \frac{(e^{\eta_{i3}+\eta_{i1}} - e^{\eta_{i3}} + 1) (e^{2\eta_{i3}+\eta_{i1}} - e^{\eta_{i3}+\eta_{i1}} - e^{2\eta_{i3}} - 2e^{\eta_{i3}} - 1)}{2[(e^{\eta_{i1}} - 1)^2 e^{2\eta_{i3}} + 2(e^{\eta_{i1}} + 1) e^{\eta_{i3}} + 1]^{\frac{3}{2}}}, \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \pi_{i0}}{\partial \eta_{i1} \partial \eta_{i3}} &= \frac{e^{-\eta_{i3}-\eta_{i1}}}{2} - \frac{e^{-\eta_{i3}-\eta_{i1}} (e^{2\eta_{i3}+\eta_{i1}} - e^{2\eta_{i3}} + 1)}{2\sqrt{(e^{\eta_{i1}} - 1)^2 e^{2\eta_{i3}} + 2(e^{\eta_{i1}} + 1) e^{\eta_{i3}} + 1}} \\ &\quad + \frac{e^{-\eta_{i1}} (e^{\eta_{i3}+2\eta_{i1}} - 2e^{\eta_{i3}+\eta_{i1}} + e^{\eta_{i3}} + e^{\eta_{i1}} + 1)}{2[(e^{\eta_{i1}} - 1)^2 e^{2\eta_{i3}} + 2(e^{\eta_{i1}} + 1) e^{\eta_{i3}} + 1]^{\frac{3}{2}}} \\ &\quad \times \frac{(e^{2\eta_{i3}+\eta_{i1}} - e^{\eta_{i3}+\eta_{i1}} - e^{2\eta_{i3}} - 2e^{\eta_{i3}} - 1)}{2[(e^{\eta_{i1}} - 1)^2 e^{2\eta_{i3}} + 2(e^{\eta_{i1}} + 1) e^{\eta_{i3}} + 1]^{\frac{3}{2}}} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \pi_{i0}}{\partial^2 \eta_{i3}} &= \frac{e^{-\eta_{i3}-\eta_{i1}}}{2} - \frac{e^{-\eta_{i3}-\eta_{i1}}}{2\sqrt{(e^{\eta_{i1}} - 1)^2 e^{2\eta_{i3}} + 2(e^{\eta_{i1}} + 1) e^{\eta_{i3}} + 1}} \\ &\quad - \frac{e^{-\eta_{i1}} (e^{\eta_{i3}+\eta_{i1}} + e^{\eta_{i3}} + 1) (e^{\eta_{i3}+2\eta_{i1}} - 2e^{\eta_{i3}+\eta_{i1}} + e^{\eta_{i3}} + e^{\eta_{i1}} + 1)}{2[(e^{\eta_{i1}} - 1)^2 e^{2\eta_{i3}} + 2(e^{\eta_{i1}} + 1) e^{\eta_{i3}} + 1]^{\frac{3}{2}}}. \end{aligned}$$



# Capitolo 4

## Studi di simulazione ed applicazione dei modelli ad un caso reale

In questo capitolo si valuteranno, attraverso alcuni studi di simulazione, le proprietà distributive, per il modello di Richardson *et al.* (2017), dello stimatore di massima verosimiglianza,  $\hat{\gamma} = (\hat{\gamma}_0, \dots, \hat{\gamma}_{p-2})$ , dello stimatore con distorsione in media ridotta,  $\hat{\gamma}^* = (\hat{\gamma}_0^*, \dots, \hat{\gamma}_{p-2}^*)$ , dello stimatore con distorsione in mediana ridotta,  $\tilde{\gamma} = (\tilde{\gamma}_0, \dots, \tilde{\gamma}_{p-2})$  e delle relative versioni con parametro di disturbo alternativo (2.12). Queste ultime verranno indicate con  $\hat{\gamma}^A$ ,  $\hat{\gamma}^{A*}$  e  $\tilde{\gamma}^A$ . In ciascuna simulazione sono stati fissati a priori i parametri  $\gamma$  e  $\beta$ , la numerosità campionaria  $n$  e il vettore relativo al fattore di esposizione binario  $t_i$ . È stata inoltre generata casualmente la matrice di disegno  $X$ , la quale è stata mantenuta fissa durante la simulazione. Con questa configurazione, sono state generati 10000 campioni casuali della variabile risposta  $Y$  e, per ciascuno di questi, sono stati calcolati i sei stimatori. Le proprietà di ciascun stimatore sono state valutate attraverso cinque indici: la distorsione relativa percentuale stimata (Relative bias), lo *standard error* della stimatore (SE), ossia il valore medio, nel corso delle 10000 simulazioni, della radice quadrata degli elementi diagonali dell'inversa della matrice di informazione attesa stimata, la versione empirica della radice dell'errore quadratico medio (RMSE), la percentuale empirica di sottostima (PU%) e la copertura empi-

rica degli intervalli alla Wald con livello di confidenza nominale 95% (Wald).

Infine, nel paragrafo 4.4, è analizzato il comportamento dei diversi stimatori su un insieme di dati reali.

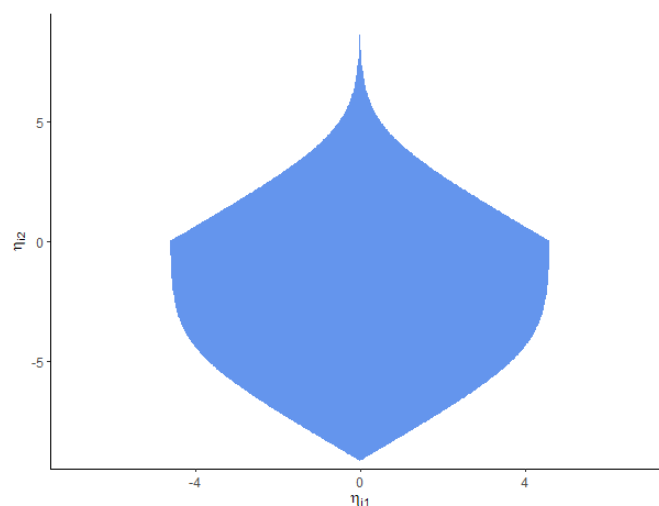
## 4.1 Scelta di $\gamma$ e $\beta$

Un passaggio sul quale è importante soffermarsi ulteriormente è quello della scelta dei vettori  $\gamma$  e  $\beta$  per ciascuna simulazione. Risulta, infatti, fondamentale assicurarsi che questi non portino a campioni in cui i valori di  $\pi_{i0}$  e  $\pi_{i1}$  sono troppo vicini a 0 o 1. Per questo si è deciso di utilizzare un approccio simile a quello proposto da Sur e Candès (2018), calcolando i valori di  $\eta_{i1}$  e  $\eta_{i2}$  per i quali  $(\pi_{i0}, \pi_{i1}) \in [0.01, 0.99]$ . Dato che non risulta immediato identificare in modo analitico questi vincoli, essi sono stati ottenuti numericamente ed è possibile osservarli graficamente nella Figura 4.1. Per ciascuno studio di simulazione sono stati quindi utilizzati dei predittori lineari che soddisfano la condizione

$$(\eta_{i1}, \eta_{2i}) \in A$$

con  $i = 1, \dots, n$  e  $A = \{(\eta_{i1}, \eta_{2i}) : (\pi_{i0}, \pi_{i1}) \in [0.01, 0.99]\}$ .

**Figura 4.1:** Rappresentazione in  $\mathbb{R}^2$  dell'insieme  $A$



## 4.2 Simulazioni con uguale numero di trattati e non trattati

Come primo passo consideriamo due casi bilanciati, con  $n = 60$ , in cui le osservazioni si dividono equamente tra trattati, ossia  $t_i = 1$ , e non trattati. Inoltre, sono presenti due ulteriori esplicative  $x_{i1}$  e  $x_{i2}$  generate in modo casuale da due normali standard tra loro indipendenti. Per la prima simulazione, come vettori dei coefficienti, sono stati utilizzati  $\gamma = (-0.5, -0.1, 0.6)$  e  $\beta = (-1.4, 0.8, 1.1)$ . Da questa scelta dei parametri si ottiene, infatti, che

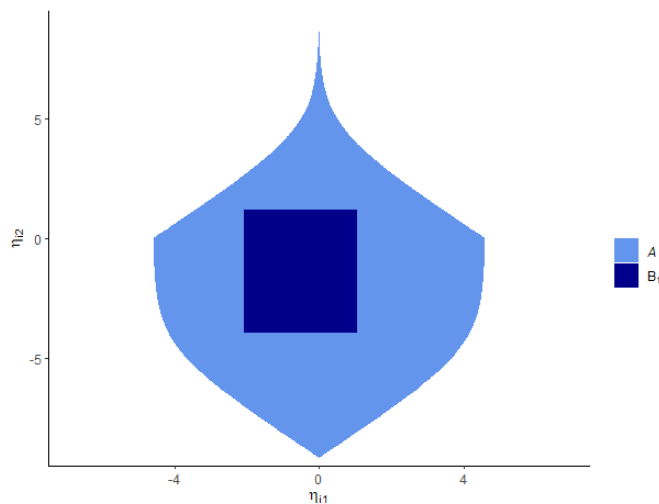
$$P_X \{(\eta_{i1}, \eta_{i2}) \in B_1\} = 0.99,$$

con  $B_1 = \{(\eta_{i1}, \eta_{i2}) : \eta_{i1} \in (-2.07, 1.06) \text{ e } \eta_{i2} \in (-3.98, 1.17)\}$ . Una volta generata la matrice del disegno è stato effettuato un controllo sui valori assunti dalle variabili  $x_{i1}$  e  $x_{i2}$  in modo da avere la garanzia che predittori lineari ad esse associati appartengano all'insieme  $A$ . Dato che, come si può osservare nella Figura 4.2, l'insieme  $B_1$  è compreso in  $A$ , questa condizione è stata ottenuta in modo computazionalmente efficiente, senza dover simulare molti valori di  $x_{i1}$  e  $x_{i2}$ . Infine, la matrice del disegno è stata mantenuta fissa nel corso delle 10000 simulazioni.

Analizzando i risultati riportati nella Tabella 4.1, è possibile notare come  $\tilde{\gamma}^A$  e  $\tilde{\gamma}$  abbiano una percentuale di sottostima più vicina, in generale, al 50% rispetto agli altri stimatori. Confrontati con gli stimatori di massima verosimiglianza, anche distorsione relativa, SE e RMSE risultano inferiori. Per quanto riguarda gli stimatori  $\hat{\gamma}^{A*}$  e  $\hat{\gamma}^*$ , essi presentano valori di distorsione relativa, SE e RMSE sensibilmente inferiori, rispetto agli altri metodi proposti, oltre che una discreta centratura attorno al 50%. La copertura degli intervalli alla Wald non risulta invece particolarmente diversa tra i sei stimatori.

Si consideri ora un secondo caso in cui  $\gamma = (0.7, -0.4, 0.3)$  e  $\beta = (-0.9, -1.1, 0.9)$ . Per questa combinazione dei parametri si ha che

$$P_X \{(\eta_{i1}, \eta_{i2}) \in B_2\} = 0.99,$$

**Figura 4.2:** Rappresentazione in  $\mathbb{R}^2$  dell'insieme  $B_1$ 

con  $B_2 = \{(\eta_{i1}, \eta_{2i}) : \eta_{i1} \in (-0.59, 1.99) \text{ e } \eta_{i2} \in (-3.84, 2.03)\}$ . Una volta generata la matrice del disegno è stato effettuato un controllo sui valori assunti dalle variabili  $x_{i1}$  e  $x_{i2}$  in modo da avere la garanzia che predittori lineari ad esse associati appartengano all'insieme  $A$ . Dato che, come si può osservare nella Figura 4.3, l'insieme  $B_2$  è compreso in  $A$ , questa condizione è stata ottenuta in modo computazionalmente efficiente, senza dover simulare molti valori di  $x_{i1}$  e  $x_{i2}$ . Anche in questo frangente la matrice del disegno è stata mantenuta fissa nel corso delle 10000 simulazioni.

Osservando i risultati riportati nella Tabella 4.2, si nota come le affermazioni fatte per la prima simulazione siano valide anche per la seconda. Si può quindi concludere che, nei casi considerati, i metodi di riduzione della distorsione in media e in mediana sono efficaci nel migliorare le proprietà di centratura dello stimatore di massima verosimiglianza.

Infine, è possibile effettuare un confronto empirico riguardante la stabilità computazione dei metodi proposti, valutando in quante occasioni gli algoritmi di stima non hanno raggiunto convergenza nel corso delle 10000 simulazioni. In particolare sono stati scartati tutti i casi in cui l'algoritmo ha raggiunto il numero massimo di iterazioni, oppure almeno uno dei coefficienti stimati è risultato, in termini di valore assoluto, maggiore di 6. Questo

**Tabella 4.1:** Risultati della prima simulazione per i sei diversi stimatori  $\hat{\gamma}$ ,  $\hat{\gamma}^*$ ,  $\tilde{\gamma}$ ,  $\hat{\gamma}^A$ ,  $\hat{\gamma}^{A*}$  e  $\tilde{\gamma}^A$ . Gli standard error di simulazione per SE, Relative Bias, RMSE e Wald sono riportati tra parentesi. Gli standard error di simulazione per PU% risultano pari a 0.5.

Stimatore		PU%	Relative Bias	SE	RMSE	Wald
$\hat{\gamma}$	$\hat{\gamma}_0$	52.7	-15.22 (1.16)	0.514 (0.002)	0.584 (0.006)	96.62 (0.18)
	$\hat{\gamma}_1$	51.7	-41.47 (6.00)	0.516 (0.001)	0.601 (0.006)	94.62 (0.23)
	$\hat{\gamma}_2$	46.0	16.17 (0.87)	0.446 (0.001)	0.529 (0.006)	95.51 (0.21)
$\hat{\gamma}^*$	$\hat{\gamma}_0^*$	47.7	-0.70 (0.96)	0.468 (0.001)	0.478 (0.005)	97.37 (0.16)
	$\hat{\gamma}_1^*$	49.5	-4.44 (5.00)	0.482 (0.001)	0.500 (0.005)	96.72 (0.18)
	$\hat{\gamma}_2^*$	53.2	0.96 (0.70)	0.414 (0.001)	0.423 (0.005)	96.80 (0.18)
$\tilde{\gamma}$	$\tilde{\gamma}_0$	50.6	-7.60 (1.05)	0.481 (0.001)	0.529 (0.007)	96.92 (0.17)
	$\tilde{\gamma}_1$	50.2	-15.13 (5.40)	0.493 (0.001)	0.541 (0.007)	96.06 (0.19)
	$\tilde{\gamma}_2$	50.1	6.48 (0.76)	0.424 (0.001)	0.459 (0.006)	96.45 (0.18)
$\hat{\gamma}^A$	$\hat{\gamma}_0^A$	53.0	-15.88 (1.18)	0.517 (0.002)	0.596 (0.007)	96.83 (0.18)
	$\hat{\gamma}_1^A$	51.4	-40.03 (6.27)	0.540 (0.001)	0.628 (0.007)	95.36 (0.21)
	$\hat{\gamma}_2^A$	45.5	19.18 (0.92)	0.467 (0.001)	0.565 (0.008)	96.11 (0.19)
$\hat{\gamma}^{A*}$	$\hat{\gamma}_0^{A*}$	47.8	-1.16 (0.96)	0.474 (0.001)	0.482 (0.006)	97.65 (0.15)
	$\hat{\gamma}_1^{A*}$	48.2	2.61 (5.15)	0.506 (0.001)	0.515 (0.005)	97.32 (0.16)
	$\hat{\gamma}_2^{A*}$	53.0	2.20 (0.73)	0.434 (0.001)	0.439 (0.005)	97.34 (0.16)
$\tilde{\gamma}^A$	$\tilde{\gamma}_0^A$	50.1	-8.32 (1.05)	0.491 (0.001)	0.527 (0.006)	97.38 (0.16)
	$\tilde{\gamma}_1^A$	49.2	-9.61 (5.56)	0.519 (0.001)	0.556 (0.006)	96.77 (0.18)
	$\tilde{\gamma}_2^A$	49.4	8.77 (0.80)	0.447 (0.001)	0.480 (0.006)	97.10 (0.17)

valore è stato scelto in quanto il rischio relativo associato a tali stime risulta maggiore di 403 se la stima ha segno positivo e minore di 0.0025 per le stime negative. Tali valori risultano quindi chiaramente estremi e non compatibili con la struttura delle due simulazioni. Nella Tabella 4.3 si può notare come, nella parametrizzazione proposta da Ricardson *et al.* (2017), lo stimatore con correzione in media sia meno robusto rispetto alle due alternative. Questa situazione non si verifica nella parametrizzazione alternativa (2.12), la quale è in generale più robusta in quanto presenta un numero di mancate convergenze sensibilmente inferiore.

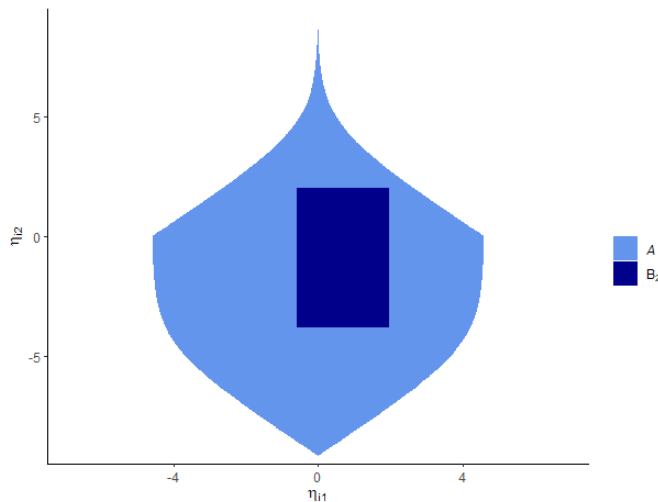
**Tabella 4.2:** Risultati della seconda simulazione per i sei diversi stimatori  $\hat{\gamma}$ ,  $\hat{\gamma}^*$ ,  $\tilde{\gamma}$ ,  $\hat{\gamma}^A$ ,  $\hat{\gamma}^{A*}$  e  $\tilde{\gamma}^A$ . Gli standard error di simulazione per SE, Relative Bias, RMSE e Wald sono riportati tra parentesi. Gli standard error di simulazione per PU% risultano pari a 0.5.

Stimatore		PU%	Relative Bias	SE	RMSE	Wald
$\hat{\gamma}$	$\hat{\gamma}_0$	44.4	16.82(0.78)	0.476 (0.002)	0.557 (0.007)	96.61 (0.18)
	$\hat{\gamma}_1$	56.4	-35.24 (1.54)	0.519 (0.002)	0.634 (0.007)	95.22 (0.21)
	$\hat{\gamma}_2$	47.9	20.87 (1.78)	0.462 (0.001)	0.538 (0.005)	92.84 (0.26)
$\hat{\gamma}^*$	$\hat{\gamma}_0^*$	50.9	2.76 (0.63)	0.430 (0.001)	0.438 (0.005)	97.08 (0.17)
	$\hat{\gamma}_1^*$	48.1	-5.85 (1.25)	0.466 (0.001)	0.502 (0.005)	96.16 (0.19)
	$\hat{\gamma}_2^*$	53.5	-0.41 (1.48)	0.416 (0.001)	0.445 (0.004)	94.54 (0.23)
$\tilde{\gamma}$	$\tilde{\gamma}_0$	48.2	8.27 (0.69)	0.445 (0.001)	0.488 (0.007)	97.07 (0.17)
	$\tilde{\gamma}_1$	51.4	-14.71 (1.36)	0.482 (0.001)	0.546 (0.006)	95.93 (0.20)
	$\tilde{\gamma}_2$	50.2	9.00 (1.58)	0.432 (0.001)	0.474 (0.005)	94.41 (0.23)
$\hat{\gamma}^A$	$\hat{\gamma}_0^A$	44.7	16.36 (0.79)	0.479 (0.002)	0.565 (0.007)	96.42 (0.19)
	$\hat{\gamma}_1^A$	55.9	-36.89 (1.63)	0.537 (0.002)	0.669 (0.008)	95.99 (0.20)
	$\hat{\gamma}_2^A$	47.0	26.93 (1.89)	0.481 (0.002)	0.573 (0.006)	93.64 (0.24)
$\hat{\gamma}^{A*}$	$\hat{\gamma}_0^{A*}$	51.6	1.68 (0.63)	0.435 (0.001)	0.445 (0.006)	97.05 (0.17)
	$\hat{\gamma}_1^{A*}$	47.8	-5.49 (1.29)	0.482 (0.001)	0.516 (0.006)	97.01 (0.17)
	$\hat{\gamma}_2^{A*}$	53.0	3.11 (1.55)	0.433 (0.001)	0.465 (0.005)	95.56 (0.17)
$\tilde{\gamma}^A$	$\tilde{\gamma}_0^A$	48.9	7.07 (0.70)	0.451 (0.001)	0.491 (0.007)	97.01 (0.17)
	$\tilde{\gamma}_1^A$	51.2	-15.57 (1.41)	0.501 (0.001)	0.569 (0.007)	96.89 (0.17)
	$\tilde{\gamma}_2^A$	49.5	15.07 (1.66)	0.452 (0.001)	0.501 (0.005)	95.37 (0.21)

**Tabella 4.3:** Numero di mancate convergenze su 10000 replicazioni per le due simulazioni

	$\hat{\gamma}$	$\hat{\gamma}^*$	$\tilde{\gamma}$	$\hat{\gamma}^A$	$\hat{\gamma}^{A*}$	$\tilde{\gamma}^A$
Simulazione 1	36	136	28	15	6	6
Simulazione 2	69	226	79	25	17	17



**Figura 4.3:** Rappresentazione in  $\mathbb{R}^2$  dell'insieme  $B_2$ 

### 4.3 Simulazioni con numero di trattati e non trattati sbilanciato

Spesso, in ambito applicativo, il numero di osservazioni per le quali il fattore di esposizione binario  $t_i$  risulta pari a 1 non coincide con quello delle osservazioni con  $t_i$  uguale a 0. Per questa ragione in seguito vengono valutate, attraverso due studi di simulazione, le proprietà dei vari stimatori nel caso in cui il numero di soggetti non trattati differisca da quello dei trattati. Per favorire il confronto con il caso bilanciato, le configurazioni utilizzate per i parametri  $\gamma$  e  $\beta$  sono quelle del paragrafo precedente. Per la prima delle due simulazioni sono stati quindi impiegati i vettori  $\gamma = (-0.5, -0.1, 0.6)$  e  $\beta = (-1.4, 0.8, 1.1)$  mentre per la seconda i coefficienti risultano pari a  $\gamma = (0.7, -0.4, 0.3)$  e  $\beta = (-0.9, -1.1, 0.9)$ . Come numerosità campionaria si è utilizzato, anche in questo frangente,  $n = 60$  di cui solo  $1/3$  con fattore di esposizione binario pari ad 1. I valori delle variabili esplicative  $x_{i1}$  e  $x_{i2}$  sono gli stessi delle simulazioni del paragrafo precedente.

Nelle Tabelle 4.5 e 4.6 è possibile osservare come, anche nel caso sbilanciato, valgano le considerazioni fatte nel paragrafo precedente. In particolare

lo stimatore con distorsione in media ridotta presenta, in generale, valori più contenuti, in termini di distorsione in media, di SE e di RMSE, rispetto ai metodi alternativi in entrambe le parametrizzazioni. Al contrario, lo stimatore con distorsione in mediana ridotta si dimostra preferibile in termini di percentuale di sottostima, con valori prossimi al 50%. Lo stimatore di massima verosimiglianza presenta invece una distorsione in media maggiore e una percentuale di sottostima più distante dal 50%, rispetto agli altri metodi considerati, in entrambe le parametrizzazioni. La copertura effettiva degli intervalli alla Wald si discosta sensibilmente da quella nominale per tutti e sei gli stimatori considerati.

Infine, per quanto riguarda la stabilità computazionale dei metodi proposti, in Tabella 4.4 si può notare come, anche nel caso sbilanciato, gli algoritmi di stima raggiungano più spesso la convergenza nella parametrizzazione (2.12).

**Tabella 4.4:** Numero di mancate convergenze su 10000 replicazioni per le due simulazioni con sbilanciamento tra il numero di osservazioni con  $t_i = 0$  e  $t_i = 1$ .

	$\hat{\gamma}$	$\hat{\gamma}^*$	$\tilde{\gamma}$	$\hat{\gamma}^A$	$\hat{\gamma}^{A*}$	$\tilde{\gamma}^A$
Simulazione 1 caso sbilanciato	120	178	66	112	49	34
Simulazione 2 caso sbilanciato	124	218	71	23	40	18

**Tabella 4.5:** Risultati della prima simulazione, con sbilanciamento tra trattati e non trattati, per i sei diversi stimatori  $\hat{\gamma}$ ,  $\hat{\gamma}^*$ ,  $\tilde{\gamma}$ ,  $\hat{\gamma}^A$ ,  $\hat{\gamma}^{A*}$  e  $\tilde{\gamma}^A$ . Gli standard error di simulazione per SE, Relative Bias, RMSE e Wald sono riportati tra parentesi. Gli standard error di simulazione per PU% risultano pari a 0.5.

Stimatore		PU%	Relative Bias	SE	RMSE	Wald
$\hat{\gamma}$	$\hat{\gamma}_0$	56.1	-40.01 (1.43)	0.610 (0.003)	0.741 (0.010)	97.01 (0.17)
	$\hat{\gamma}_1$	50.9	-15.12 (7.27)	0.592 (0.002)	0.728 (0.008)	94.87 (0.22)
	$\hat{\gamma}_2$	43.7	23.88 (1.05)	0.518 (0.002)	0.648 (0.008)	96.01 (0.20)
$\hat{\gamma}^*$	$\hat{\gamma}_0^*$	45.5	-3.89 (1.12)	0.530 (0.002)	0.559 (0.008)	96.25 (0.19)
	$\hat{\gamma}_1^*$	50.0	-7.77 (5.83)	0.538 (0.001)	0.583 (0.007)	97.16 (0.17)
	$\hat{\gamma}_2^*$	53.3	1.33 (0.82)	0.463 (0.001)	0.492 (0.006)	96.85 (0.17)
$\tilde{\gamma}$	$\tilde{\gamma}_0$	50.4	-20.62 (1.29)	0.561 (0.002)	0.652 (0.010)	96.55 (0.18)
	$\tilde{\gamma}_1$	50.5	-6.18 (6.44)	0.559 (0.002)	0.644 (0.008)	96.55 (0.18)
	$\tilde{\gamma}_2$	49.6	8.44 (0.92)	0.484 (0.002)	0.554 (0.008)	96.62 (0.18)
$\hat{\gamma}^A$	$\hat{\gamma}_0^A$	56.1	-39.83 (1.43)	0.611 (0.003)	0.742 (0.010)	96.95 (0.17)
	$\hat{\gamma}_1^A$	50.1	-4.44 (7.35)	0.608 (0.002)	0.735 (0.008)	95.59 (0.21)
	$\hat{\gamma}_2^A$	43.6	25.28 (1.08)	0.537 (0.002)	0.667 (0.008)	96.11 (0.19)
$\hat{\gamma}^{A*}$	$\hat{\gamma}_0^{A*}$	45.6	-3.46 (1.11)	0.534 (0.002)	0.553 (0.007)	96.59 (0.18)
	$\hat{\gamma}_1^{A*}$	49.4	2.17 (5.84)	0.556 (0.001)	0.584 (0.006)	98.00 (0.14)
	$\hat{\gamma}_2^{A*}$	52.8	2.36 (0.84)	0.485 (0.001)	0.502 (0.006)	96.97 (0.17)
$\tilde{\gamma}^A$	$\tilde{\gamma}_0^A$	50.6	-20.87 (1.29)	0.570 (0.003)	0.655 (0.010)	96.81 (0.18)
	$\tilde{\gamma}_1^A$	49.6	2.82 (6.48)	0.580 (0.002)	0.648 (0.007)	97.34 (0.16)
	$\tilde{\gamma}_2^A$	49.2	10.59 (0.94)	0.507 (0.002)	0.569 (0.008)	96.74 (0.18)

**Tabella 4.6:** Risultati della seconda simulazione, con sbilanciamento tra trattati e non trattati, per i sei diversi stimatori  $\hat{\gamma}$ ,  $\hat{\gamma}^*$ ,  $\tilde{\gamma}$ ,  $\hat{\gamma}^A$ ,  $\hat{\gamma}^{A*}$  e  $\tilde{\gamma}^A$ . Gli standard error di simulazione per SE, Relative Bias, RMSE e Wald sono riportati tra parentesi. Gli standard error di simulazione per PU% risultano pari a 0.5.

Stimatore		PU%	Relative Bias	SE	RMSE	Wald
$\hat{\gamma}$	$\hat{\gamma}_0$	49.3	3.25 (0.71)	0.446 (0.001)	0.495 (0.005)	95.88 (0.20)
	$\hat{\gamma}_1$	56.7	-38.56 (1.54)	0.517 (0.002)	0.635 (0.007)	94.77 (0.22)
	$\hat{\gamma}_2$	46.3	29.63 (1.81)	0.457 (0.001)	0.550 (0.006)	94.48 (0.23)
$\hat{\gamma}^*$	$\hat{\gamma}_0^*$	49.9	1.30 (0.59)	0.414 (0.001)	0.415 (0.004)	97.14 (0.17)
	$\hat{\gamma}_1^*$	47.2	-4.77 (1.23)	0.471 (0.001)	0.494 (0.005)	96.35 (0.19)
	$\hat{\gamma}_2^*$	53.7	-0.18 (1.44)	0.417 (0.001)	0.432 (0.004)	96.41 (0.19)
$\tilde{\gamma}$	$\tilde{\gamma}_0$	49.4	2.45 (0.63)	0.423 (0.001)	0.445 (0.005)	96.68 (0.18)
	$\tilde{\gamma}_1$	50.5	-15.69 (1.33)	0.485 (0.001)	0.537 (0.005)	95.85 (0.20)
	$\tilde{\gamma}_2$	50.7	10.79 (1.57)	0.430 (0.001)	0.472 (0.005)	95.92 (0.20)
$\hat{\gamma}^A$	$\hat{\gamma}_0^A$	49.7	2.42 (0.72)	0.448 (0.001)	0.502 (0.005)	95.73 (0.20)
	$\hat{\gamma}_1^A$	55.9	-39.42 (1.62)	0.533 (0.002)	0.666 (0.008)	95.39 (0.21)
	$\hat{\gamma}_2^A$	46.2	33.22 (1.89)	0.473 (0.001)	0.575 (0.007)	94.86 (0.22)
$\hat{\gamma}^{A*}$	$\hat{\gamma}_0^{A*}$	50.9	0.02 (0.60)	0.420 (0.001)	0.417 (0.004)	97.41 (0.16)
	$\hat{\gamma}_1^{A*}$	46.7	-4.03 (1.26)	0.486 (0.001)	0.505 (0.005)	97.24 (0.16)
	$\hat{\gamma}_2^{A*}$	53.6	1.82 (1.48)	0.433 (0.001)	0.444 (0.005)	96.96 (0.17)
$\tilde{\gamma}^A$	$\tilde{\gamma}_0^A$	50.2	1.10 (0.64)	0.431 (0.001)	0.448 (0.005)	96.78 (0.18)
	$\tilde{\gamma}_1^A$	50.2	-15.80 (1.39)	0.503 (0.001)	0.558 (0.006)	96.65 (0.18)
	$\tilde{\gamma}_2^A$	50.2	14.37 (1.63)	0.449 (0.001)	0.490 (0.005)	96.53 (0.18)

## 4.4 Caso studio

L'obiettivo di questo paragrafo è valutare su un insieme di dati reali, reperibili in Agresti (2019, Esercizio 4.16), il comportamento dei vari metodi per l'inferenza sul logaritmo del rischio relativo esposti nella tesi.

In particolare, i dati utilizzati si riferiscono ad uno studio atto a valutare l'insorgenza di faringodinia (mal di gola) in 33 soggetti sottoposti ad una operazione chirurgica con anestesia totale. La variabile di interesse  $y$  è quindi di tipo dicotomico (0 = assenza di faringodinia al termine dell'operazione, 1 = presenza di faringodinia al termine dell'operazione). Le variabili esplicative disponibili sono  $x_1$  = 'durata in minuti dell'operazione' e  $t$  = 'tipo di strumento utilizzato per la respirazione assistita' (0 = maschera laringea, 1 = tubo endotracheale). L'obiettivo finale dell'analisi è quello di valutare se l'utilizzo del tubo endotracheale per la respirazione assistita possa ridurre il rischio di insorgenza di faringodinia nei pazienti in seguito all'operazione. Dato che, come è possibile notare dalla Figura 4.4, la durata totale dell'operazione sembra influenzare il fenomeno di interesse, anch'essa è stata utilizzata nel corso dell'analisi.

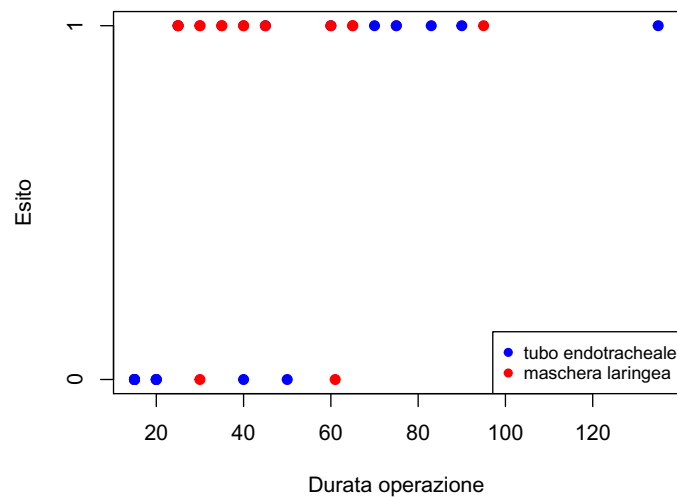
In questo caso, quindi, la struttura ipotizzata per il rischio relativo risulta essere

$$\log(\text{RR}_i) = \gamma_0 + \gamma_1 x_{i1},$$

per  $i = 1, \dots, 33$ .

Come è possibile notare nella Tabella 4.7 tutti i metodi considerati tendono ad evidenziare come l'utilizzo del tubo endotracheale porti ad una diminuzione del rischio di faringodinia al termine dell'operazione. Infatti, la stima puntuale del coefficiente  $\gamma_0$  è sempre negativa e significativamente diversa da zero al 95%. La stima puntuale del coefficiente  $\gamma_1$  è invece sempre positiva anche se non risulta significativa al livello 0.05 in alcuni dei metodi considerati. Nonostante i modelli portino a conclusioni almeno qualitativamente simili, è possibile osservare alcune differenze importanti. In primo luogo, le stime ottenute attraverso la regressione logistica  $\hat{\gamma}^L = (\hat{\gamma}_0^L, \hat{\gamma}_1^L)$  non sem-

**Figura 4.4:** Effetto della durata del trattamento e del tipo di strumento utilizzato per la respirazione assistita sul rischio di insorgenza di faringodinia in seguito all'operazione



brano fornire una approssimazione adeguata del rischio relativo, con valori stimati molto distanti da quelli ottenuti negli altri modelli e *standard error* nettamente maggiori. Questo fenomeno ha anche un'importante conseguenza sugli intervalli alla Wald, i quali risultano più ampi e quindi meno informativi. Per quanto riguarda il modello proposto da Richardson *et al.* (2017) e il modello con parametrizzazione alternativa (2.12), è possibile osservare come, utilizzando gli stimatori con distorsione in media ridotta, si ottenga un sensibile restringimento dei valori stimati e dei relativi *standard error* verso lo zero. Gli stimatori con distorsione in mediana ridotta non sembrano invece avere, in questo caso, un comportamento marcatamente diverso dagli stimatori di massima verosimiglianza. Infine, è importante sottolineare come, nella Tabella 4.7, non sia stata inclusa la regressione log-binomiale. Ciò è dovuto al fatto che non è stato possibile ottenere una stima per i coefficienti del modello nonostante l'obiettivo semplicità del caso analizzato.

Tabella 4.7

Stimatore	Stima	Standard error	Intervallo alla Wald 95%		
			Limite inferiore	Limite superiore	
$\hat{\gamma}^L$	$\hat{\gamma}_0^L$	-6.125	2.783	-11.579	-0.671
	$\hat{\gamma}_1^L$	0.098	0.061	-0.022	0.217
$\hat{\gamma}$	$\hat{\gamma}_0$	-2.047	0.900	-3.811	-0.283
	$\hat{\gamma}_1$	0.029	0.014	0.002	0.056
$\hat{\gamma}^*$	$\hat{\gamma}_0^*$	-1.824	0.839	-3.470	-0.179
	$\hat{\gamma}_1^*$	0.024	0.013	-0.001	0.049
$\tilde{\gamma}$	$\tilde{\gamma}_0$	-1.993	0.860	-3.679	-0.307
	$\tilde{\gamma}_1$	0.027	0.013	0.001	0.053
$\hat{\gamma}^A$	$\hat{\gamma}_0^A$	-2.002	0.905	-3.776	-0.228
	$\hat{\gamma}_1^A$	0.028	0.014	0.001	0.056
$\hat{\gamma}^{A*}$	$\hat{\gamma}_0^{A*}$	-1.828	0.860	-3.513	-0.144
	$\hat{\gamma}_1^{A*}$	0.024	0.013	-0.001	0.050
$\tilde{\gamma}^A$	$\tilde{\gamma}_0^A$	-1.999	0.877	-3.718	-0.279
	$\tilde{\gamma}_1^A$	0.027	0.013	0.001	0.054





# Conclusioni

L'obiettivo di questa tesi è stato quello di implementare i metodi di riduzione della distorsione dello stimatore di massima verosimiglianza per il modello di Richardson *et al.* (2017). Per tale modello, inoltre, è stata proposta una differente specificazione del parametro di disturbo. Quest'ultima è sicuramente più ostica per quanto riguarda l'aspetto interpretativo, ma allo stesso tempo risulta definita, senza bisogno di prolungamenti per continuità, in ogni combinazione dei predittori lineari ad essa associati. In aggiunta, l'aspetto interpretativo non sembra avere particolare importanza, in quanto l'obiettivo finale dell'inferenza statistica è, nei casi considerati, il logaritmo del rischio relativo il quale è parametrizzato in modo equivalente nei due modelli.

Dagli studi di simulazione effettuati nel Capitolo 4 risulta evidente come i metodi di riduzione della distorsione in media e mediana siano efficaci nel migliorare le proprietà dello stimatore di massima verosimiglianza in entrambe le parametrizzazioni. In particolare, lo stimatore con distorsione in mediana ridotta presenta una notevole centratura attorno al vero valore del parametro e una distorsione in media nettamente inferiore a quella dello stimatore di massima verosimiglianza. Il metodo di Firth (1993) è, invece, efficace nel ridurre la distorsione in media dello stimatore di massima verosimiglianza. Questo si traduce anche in un generale miglioramento della centratura dello stimatore.

Oltre a ciò, la parametrizzazione alternativa non sembra influenzare significativamente le stime associate al logaritmo del rischio relativo. Tale fenomeno, congiuntamente alla maggior robustezza degli algoritmi di stima

osservata nel Capitolo 4, rende il modello con parametrizzazione alternativa particolarmente attraente in ambito applicativo.

Si conclude, infine, osservando come i modelli trattati in questo lavoro siano applicabili in contesti relativamente ristretti. Possibili sviluppi futuri potrebbero quindi focalizzarsi sull'implementazione di metodi non parametrici per la modellazione della componente di disturbo. La metodologia utilizzata da Richardson *et al.* (2017), inoltre, potrebbe essere estesa al caso con fattore di esposizione continuo.

# Appendice A

## Dettagli tecnici

### A.1 Legame tra $\pi_{i0}$ , $\pi_{i1}$ e i parametri $\eta_{i1}$ e $\eta_{i2}$ .

Affinché, ad ogni  $(\eta_{i1}, \eta_{i2}) \in \mathbb{R}^2$  corrispondano valori di  $(\pi_{i0}, \pi_{i1}) \in (0, 1)^2$  è necessario che siano verificate le seguenti disequaglianze

$$0 < \pi_{i0} < 1 \quad \text{se } \eta_{i1} < 0, \quad (\text{A.1})$$

$$0 < \pi_{i0} < e^{-\eta_{i1}} \quad \text{se } \eta_{i1} > 0. \quad (\text{A.2})$$

Considerando in un primo momento la (A.1) si ottiene che

$$0 \leq \frac{-(e^{\eta_{i1}} + 1) e^{\eta_{i2}} + \sqrt{e^{2\eta_{i2}} (e^{\eta_{i1}} + 1)^2 + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}})}}{2e^{\eta_{i1}} (1 - e^{\eta_{i2}})} \leq 1. \quad (\text{A.3})$$

Se  $\eta_{i2} > 0$ , la (A.3) è equivalente a

$$2e^{\eta_{i1}} (1 - e^{\eta_{i2}}) \leq -e^{\eta_{i2}} (1 + e^{\eta_{i1}}) + \sqrt{e^{2\eta_{i2}} (1 + e^{\eta_{i1}})^2 + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}})} \leq 0.$$

La disequaglianza a destra è sempre verificata in quanto

$$e^{2\eta_{i2}} (1 + e^{\eta_{i1}})^2 \geq e^{2\eta_{i2}} (1 + e^{\eta_{i1}})^2 + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}}),$$

ossia

$$0 \geq 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}}),$$

è sempre vera se  $\eta_{i2} > 0$ .

La disuguaglianza di sinistra è anch'essa sempre verificata in quanto

$$-e^{\eta_{i2}} (1 + e^{\eta_{i1}}) + \sqrt{e^{2\eta_{i2}} (1 + e^{\eta_{i1}})^2 + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}})} \geq 2e^{\eta_{i1}} (1 - e^{\eta_{i2}}),$$

ossia

$$\sqrt{e^{2\eta_{i2}} (1 + e^{\eta_{i1}})^2 + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}})} \geq 2e^{\eta_{i1}} (1 - e^{\eta_{i2}}) + e^{\eta_{i2}} (1 + e^{\eta_{i1}}),$$

può essere spezzata in due casi

A) Se  $2e^{\eta_{i1}} (1 - e^{\eta_{i2}}) + e^{\eta_{i2}} (1 + e^{\eta_{i1}}) \leq 0$ .

In questo caso la disuguaglianza è sempre vera per via del codominio della radice quadrata.

B) Se  $2e^{\eta_{i1}} (1 - e^{\eta_{i2}}) + e^{\eta_{i2}} (1 + e^{\eta_{i1}}) \geq 0$ .

Anche in questo caso la disuguaglianza è verificata infatti:

$$\begin{aligned} e^{2\eta_{i2}} (1 + e^{\eta_{i1}})^2 + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}}) &\geq 4e^{2\eta_{i1}} (1 - e^{\eta_{i2}})^2 + \\ &4e^{\eta_{i1} + \eta_{i2}} (1 + e^{\eta_{i1}}) \times \\ &(1 - e^{\eta_{i2}}) + e^{2\eta_{i2}} (1 + e^{\eta_{i1}})^2, \end{aligned}$$

che si traduce nella condizione

$$4e^{2\eta_{i1}} (e^{\eta_{i2}} - 1) \geq 0, \tag{A.4}$$

la quale è sempre verificata per  $e^{\eta_{i2}} \geq 1$  ossia per  $\eta_{i2} > 0$  come da ipotesi.

Se invece  $\eta_{i2} < 0$  il sistema di disequazioni da soddisfare affinché  $\pi_{i0} \in (0, 1)$  assume la forma

$$0 \leq -(e^{\eta_{i1}} + 1) e^{\eta_{i2}} + \sqrt{e^{2\eta_{i2}} (1 + e^{\eta_{i1}})^2 + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}})} \leq 2e^{\eta_{i1}} (1 - e^{\eta_{i2}}),$$

ed anche in questo caso il sistema è sempre verificato, in quanto

A)

$$\sqrt{e^{2\eta_{i2}} (1 + e^{\eta_{i1}})^2 + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}})} \geq (e^{\eta_{i1}} + 1) e^{\eta_{i2}},$$

ossia

$$4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}}) \geq 0,$$

è sempre vera se  $\eta_{i2} < 0$ .

B) In modo analogo

$$2e^{\eta_{i1}} (1 - e^{\eta_{i2}}) + (1 + e^{\eta_{i1}}) e^{\eta_{i2}} \geq \sqrt{e^{2\eta_{i2}} (1 + e^{\eta_{i1}})^2 + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}})},$$

ossia

$$4e^{2\eta_{i1}} (1 - e^{\eta_{i2}}) \geq 0,$$

è sempre verificata per  $\eta_{i2} < 0$ .

Per la (A.2), il sistema di equazioni assume invece la forma

$$0 \leq \frac{-(e^{\eta_{i1}} + 1) e^{\eta_{i2}} + \sqrt{e^{2\eta_{i2}} (e^{\eta_{i1}} + 1)^2 + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}})}}{2e^{\eta_{i1}} (1 - e^{\eta_{i2}})} \leq e^{-\eta_{i1}}. \quad (\text{A.5})$$

In questo caso si mostra, con argomenti del tutto analoghi a quelli usati nella (A.3), che la diseuguaglianza di sinistra della (A.5) risulta sempre verificata. Per quanto riguarda quella di destra si ha invece

$$-(e^{\eta_{i1}} + 1) e^{\eta_{i2}} + \sqrt{(e^{\eta_{i1}} + 1)^2 e^{2\eta_{i2}} + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}})} \leq 2(1 - e^{\eta_{i2}}),$$

ossia

$$\sqrt{(e^{\eta_{i1}} + 1)^2 e^{2\eta_{i2}} + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}})} \leq 2 + e^{\eta_{i2}} (e^{\eta_{i1}} - 1), \quad (\text{A.6})$$

dato che  $2 + e^{\eta_{i2}} (e^{\eta_{i1}} - 1) \geq 0$  se  $\eta_{i1} > 0$  è possibile elevare al quadrato entrambi i membri della (A.6) mantenendo inalterato il verso della diseuguaglianza

$$(e^{\eta_{i1}} + 1)^2 e^{2\eta_{i2}} + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}}) \leq 4 + 2e^{\eta_{i2}} (e^{\eta_{i1}} - 1) + e^{2\eta_{i2}} (e^{\eta_{i1}} - 1)^2,$$

dopo opportune semplificazioni si ottiene la condizione sufficiente affinché la (A.2) sia soddisfatta

$$2e^{\eta_{i2}} (2 - e^{\eta_{i2} + \eta_{i1}}) - 4 \leq 0,$$

ossia

$$\eta_{i1} \geq -\log 2.$$

Questa è sempre verificata in quanto, per ipotesi,  $\eta_{i1} > 0$ . Si ha quindi che, quando  $\eta_{i2} \neq 0$ ,  $\eta_{i2}$  e  $\eta_{i1}$  vengono effettivamente mappati in coppie di probabilità  $(\pi_{i0}, \pi_{i1})$  comprese tra 0 e 1.

Si consideri infine il caso in cui  $\eta_{i2} \rightarrow 0$ . In questo frangente  $\pi_{i0}$  assume la forma indeterminata  $\frac{0}{0}$ . Applicando la regola di De l'Hôpital si ottiene che

$$\lim_{\eta_{i2} \rightarrow 0} \frac{-(e^{\eta_{i1}} + 1) e^{\eta_{i2}} + \sqrt{e^{2\eta_{i2}} (e^{\eta_{i1}} + 1)^2 + 4e^{\eta_{i1} + \eta_{i2}} (1 - e^{\eta_{i2}})}}{2e^{\eta_{i1}} (1 - e^{\eta_{i2}})} = \frac{1}{1 + e^{\eta_{i1}}}, \quad (\text{A.7})$$

ed è quindi possibile prolungare  $\pi_{i0}$  per continuità. Di per sé tale prolungamento per continuità non porta, dal punto di vista computazionale, a particolari problematiche se la stima di massima verosimiglianza viene ottenuta direttamente dalla massimizzazione della verosimiglianza. E' invece più problematica in metodi dove è necessario calcolare in modo esplicito le derivate prime e seconde della funzione di verosimiglianza.

## A.2 Legame tra $\pi_{i0}$ , $\pi_{i1}$ e i parametri $\eta_{i1}$ e $\eta_{i3}$ .

Affinché ad ogni coppia di parametri  $(\eta_{i1}, \eta_{i3}) \in \mathbb{R}^2$  corrispondano valori di  $\pi_{i0}$  e  $\pi_{i1}$  entrambi compresi tra 0 e 1 è necessario che siano soddisfatte le seguenti disuguaglianze

$$0 < \pi_{i0}(\eta_{i1}, \eta_{i3}) < 1 \quad \text{se } \eta_{i1} < 0, \quad (\text{A.8})$$

$$0 < \pi_{i0}(\eta_{i1}, \eta_{i3}) < e^{-\eta_{i1}} \quad \text{se } \eta_{i1} > 0. \quad (\text{A.9})$$

In particolare, si consideri innanzitutto il caso in cui  $\eta_{i1} < 0$ . L'insieme di disuguaglianze (A.8) risulta pari a

$$0 < \frac{[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})] - \sqrt{[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2 - 4e^{\eta_{i1}+2\eta_{i3}}}}{2e^{\eta_{i1}+\eta_{i3}}} < 1,$$

e può essere riscritto nella forma

$$0 < [1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})] - \sqrt{[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2 - 4e^{\eta_{i1}+2\eta_{i3}}} < 2e^{\eta_{i1}+\eta_{i3}}.$$

La disuguaglianza di sinistra verificata in quanto

$$[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2 - 4e^{\eta_{i1}+2\eta_{i3}} < [1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2,$$

è sempre vera dato che  $-4e^{\eta_{i1}+2\eta_{i3}} < 0$  per qualsiasi valore di  $(\eta_{i1}, \eta_{i3})$ . La disuguaglianza di destra è anch'essa sempre verificata. Infatti

$$[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})] - \sqrt{[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2 - 4e^{\eta_{i1}+2\eta_{i3}}} < 2e^{\eta_{i1}+\eta_{i3}},$$

può essere riscritta come

$$\sqrt{[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2 - 4e^{\eta_{i1}+2\eta_{i3}}} > 1 + e^{\eta_{i3}}(1 - e^{\eta_{i1}}), \quad (\text{A.10})$$

dato che per ipotesi  $\eta_{i1} < 0$  entrambe le componenti della (A.10) sono positive, esse possono essere elevate al quadrato senza modificare la disuguaglianza stessa. Si ha quindi che

$$[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2 - 4e^{\eta_{i1}+2\eta_{i3}} > [1 + e^{\eta_{i3}}(1 - e^{\eta_{i1}})]^2. \quad (\text{A.11})$$

Dopo alcune semplificazioni algebriche la (A.11) risulta pari a

$$4e^{\eta_{i1}+\eta_{i3}} > 0,$$

ed è chiaramente sempre soddisfatta.

Nel caso in cui  $\eta_{i1} \geq 0$  il sistema di equazioni da essere verificato assume la forma

$$0 < \frac{[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})] - \sqrt{[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2 - 4e^{\eta_{i1}+2\eta_{i3}}}}{2e^{\eta_{i1}+\eta_{i3}}} < e^{-\eta_{i1}},$$

ossia

$$0 < [1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})] - \sqrt{[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2 - 4e^{\eta_{i1}+2\eta_{i3}}} < 2e^{\eta_{i3}}.$$

Con argomenti del tutto analoghi a quelli utilizzati per il caso  $\eta_{i1} < 0$  si dimostra che la disuguaglianza di sinistra risulta sempre verificata.

La disuguaglianza di destra assume invece la forma

$$[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})] - \sqrt{[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2 - 4e^{\eta_{i1}+2\eta_{i3}}} < 2e^{\eta_{i3}},$$

ossia

$$\sqrt{[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2 - 4e^{\eta_{i1}+2\eta_{i3}}} > [1 - e^{\eta_{i3}}(1 - e^{\eta_{i1}})]. \quad (\text{A.12})$$

Dato che per ipotesi  $\eta_{i1} > 0$ , entrambi gli elementi della (A.12) sono positivi e possono quindi essere elevati al quadrato senza modificare i valori per cui è verificata la disuguaglianza. Si ha quindi che

$$[1 + e^{\eta_{i3}}(1 + e^{\eta_{i1}})]^2 - 4e^{\eta_{i1}+2\eta_{i3}} > [1 - e^{\eta_{i3}}(1 - e^{\eta_{i1}})]^2,$$

dopo opportune semplificazioni si ottiene la disuguaglianza

$$4e^{\eta_{i3}} > 0,$$

la quale risulta sempre verificata.

In conclusione si ha quindi che ad ogni coppia di valori  $(\eta_{i1}, \eta_{i3}) \in \mathbb{R}^2$  sono associate delle probabilità  $(\pi_{i0}, \pi_{i1}) \in (0, 1)^2$ .



# Appendice B

## Codice R utilizzato

### B.1 Funzione per la stima dei modelli sviluppati nella tesi

---

```
# x matrix of explanatory variables
# y response variable
# t binary exposition factor
# param: nuisance parameter, possible values are "standard" or "
  alternative"
# method: estimator, possible values are "Mle","Firth" and "Median"
rrmod <- function(x,y,t, param = "standard", method= "Mle", ep=1e-8,maxit
  =50, start=NULL)
{
  if( sum(param ==c("standard","alternative"))==0)
  {
    return(errorCondition("Wrong parameter specification. Allowed parameters
are 'standard' and 'alternative'"))
  }
  if( sum(method ==c("Mle","Firth","Median"))==0)
  {
    return(errorCondition("Unknown method specification. Allowed methods are
```

```
'Mle', 'Firth' and 'Median'"))
}
# re-ordering the data
y <- c(y[t==0], y[t==1])
x <- as.matrix(rbind( as.matrix(x[t==0,]), as.matrix(x[t==1,])))

# getting indicies
n <- nrow(x)
nex <- n-sum(t)
p <- ncol(x)
if(is.null(start))
{
start<- rep(0,2*p)
}

# defining functions for probabilities
# function for Richardson's nuisance parameter
prob.standard <- function(eta1,eta2)
{
p0 <- rep(-1,length(eta1))
num <- (- (exp(eta1)+1)*exp(eta2)+ sqrt(exp(2*eta2)*(exp(eta1)+1)^2+4*exp(
eta1+eta2)*(1-exp(eta2))))
den <- (2*exp(eta1)*(1-exp(eta2)))
p0 <- num/den
p0[abs(eta2)<1e-8] <- 1/(1+exp(eta1[abs(eta2)<1e-8]))
p1 <- p0*exp(eta1)
p0[p0<1e-8]<- 1e-8
p0[p0>(1-1e-8)]<- 1-1e-8
p1[p1<1e-8]<- 1e-8
p1[p1>(1-1e-8)]<- 1-1e-8
return(rbind(p0,p1))
}
```

```

# function for the alternative nuisance parameter
prob.alternative<- function(eta1,eta2)
{
p0 <- rep(0,length(eta1))
num <- (1+exp(eta2)*(1+exp(eta1))-sqrt((1+exp(eta2)*(1+exp(eta1)))^2-4*exp
      (eta1+2*eta2)))
num[num<1e-8]<-1e-8 # control too small numerators
den<- 2*exp(eta1+eta2)
den[den<1e-8]<-1e-8
p0 <- exp( log(num)-log(den))
p1 <- p0*exp(eta1)
p0[p0<1e-8]<- 1e-8
p0[p0>(1-1e-8)]<- 1-1e-8
p1[p1<1e-8]<- 1e-8
p1[p1>(1-1e-8)]<- 1-1e-8
return(rbind(p0,p1))
}
# derivatives of prob.standard
# derivative of pi0 with respect to eta1
p0.dp0.eta1.std <- function(eta1, eta2)
{
dp0_eta1 <- -exp(eta2-eta1)/(2*(exp(eta2)-1))+
exp(eta2)/((exp(eta2)-1)*sqrt(4*exp(eta2+eta1)+(exp(eta1)-1)^2*exp(2*eta2)
))+
(exp(-eta1)*(1-exp(eta1))*exp(2*eta2))/(2*(exp(eta2)-1)*sqrt(4*exp(eta2+
eta1)+(exp(eta1)-1)^2*exp(2*eta2)))
# extension for continuity
dp0_eta1[abs(eta2)<ep] <- -exp(eta1[abs(eta2)<ep])/(exp(eta1[abs(eta2)<ep]
)+1)^2
return(dp0_eta1)
}
# derivative of pi0 with respect to eta2
# written in a different way compared to thesis

```

```

p0.dp0.eta2.std <- function(eta1,eta2)
{
dp0_eta2 <- -((exp(eta1)+1)*exp(eta2))/(2*exp(eta1)*(exp(eta2)-1)^2)+
exp(eta2)/((exp(eta2)-1)^2*sqrt(4*exp(eta2+eta1)+(exp(eta1)-1)^2*exp(2*
eta2)))+
(exp(-eta1)*(exp(2*eta1)+1)*exp(2*eta2))/(2*(exp(eta2)-1)^2*sqrt(4*exp(
eta2+eta1)+(exp(eta1)-1)^2*exp(2*eta2)))
# extension for continuity
dp0_eta2[abs(eta2)<ep] <- exp(eta1[abs(eta2)<ep])/(exp(eta1[abs(eta2)<ep])
+1)^3
return(dp0_eta2)
}
# Second derivative of pi0 with respect to eta1
p0.d2p0.eta1.std <- function(eta1,eta2)
{
d2p0_eta1 <- exp(eta2-eta1)/(2*(exp(eta2)-1))+
(exp(2*eta2)*(exp(eta2+eta1)-exp(eta2)+2)*(exp(eta2+eta1)-exp(eta2)-2*exp(
eta1)))/
(2*(exp(eta2)-1)*(4*exp(eta2+eta1)+(exp(eta1)-1)^2*exp(2*eta2))^(3/2)) +
(-(1-exp(eta1))*exp(2*eta2-eta1)-exp(2*eta2))/(2*(exp(eta2)-1)*sqrt(4*exp(
eta2+eta1)+(exp(eta1)-1)^2*exp(2*eta2)))
# extension for continuity
d2p0_eta1[abs(eta2)<ep] <- ((exp(eta1[abs(eta2)<ep])-1)*exp(eta1[abs(eta2)<
ep]))/(exp(eta1[abs(eta2)<ep])+1)^3
return(d2p0_eta1)
}
# Second derivative of pi0 with respect to eta2
p0.d2p0.eta2.std <- function(eta1,eta2)
{
d2p0_eta2 <- ((exp(eta1)+1)*exp(2*eta2-eta1))/(exp(eta2)-1)^3-((exp(eta1)
+1)*exp(eta2-eta1))/(2*(exp(eta2)-1)^2)+
(-exp(2*eta2)-exp(eta2))/((exp(eta2)-1)^3*sqrt(4*exp(eta2+eta1)+(exp(eta1)
-1)^2*exp(2*eta2)))+

```

```

-((exp(2*eta1)+1)*exp(2*eta2-eta1))/((exp(eta2)-1)^3*sqrt(4*exp(eta2+eta1)
+exp(eta1)-1)^2*exp(2*eta2)))+
-((exp(2*eta1)+1)*(4*exp(eta2+eta1)+2*(exp(eta1)-1)^2*exp(2*eta2))*exp(2*
eta2-eta1))/
(4*(exp(eta2)-1)^2*(4*exp(eta2+eta1)+(exp(eta1)-1)^2*exp(2*eta2))^(3/2))+
-(exp(eta2)*(4*exp(eta2+eta1)+2*(exp(eta1)-1)^2*exp(2*eta2)))/
(2*(exp(eta2)-1)^2*(4*exp(eta2+eta1)+(exp(eta1)-1)^2*exp(2*eta2))^(3/2))
# extension for continuity
d2p0_eta2[abs(eta2)<ep] <- -((exp(eta1[abs(eta2)<ep])-1)^2*exp(eta1[abs(
eta2)<ep]))/(exp(eta1[abs(eta2)<ep])+1)^5
return(d2p0_eta2)
}
# Second derivative of p0 respect to eta1 and eta2
p0.d2p0.eta1.eta2.std <- function(eta1,eta2)
{
d2p0_eta12 <- exp(2*eta2-eta1)/(2*(exp(eta2)-1)^2)-exp(eta2-eta1)/(2*(exp(
eta2)-1))+
((exp(eta1)-1)*exp(3*eta2-eta1)-2*exp(eta2))/
(2*(exp(eta2)-1)^2*sqrt(4*exp(eta2+eta1)+(exp(eta1)-1)^2*exp(2*eta2)))+
(1-exp(eta1))*exp(2*eta2-eta1)/
((exp(eta2)-1)*sqrt(4*exp(eta2+eta1)+(exp(eta1)-1)^2*exp(2*eta2)))+
-(exp(eta2)*(4*exp(eta2+eta1)+2*(exp(eta1)-1)^2*exp(2*eta2)))/
(2*(exp(eta2)-1)*(4*exp(eta2+eta1)+(exp(eta1)-1)^2*exp(2*eta2))^(3/2))+
-((1-exp(eta1))*(4*exp(eta2+eta1)+2*(exp(eta1)-1)^2*exp(2*eta2))*exp(2*
eta2-eta1))/
(4*(exp(eta2)-1)*(4*exp(eta2+eta1)+(exp(eta1)-1)^2*exp(2*eta2))^(3/2))
# extension for continuity
d2p0_eta12[abs(eta2)<ep] <- -(exp(eta1[abs(eta2)<ep])*(2*exp(eta1[abs(eta2)
]<ep])-1))/(exp(eta1[abs(eta2)<ep])+1)^4
return(d2p0_eta12)
}
# Derivative of prob.alternative
p0.dp0.eta1.alt <- function(eta1, eta2)

```

```

{
dp0_eta1 <- -(exp(-eta2-eta1)*(exp(eta2)+1))/2-(exp(-eta2-eta1)*
((exp(eta1)-1)*exp(2*eta2)+(-exp(eta1)-2)*exp(eta2)-1))/
(2*sqrt((exp(eta1)-1)^2*exp(2*eta2)+(2*exp(eta1)+2)*exp(eta2)+1))
return(dp0_eta1)
}
# derivative with respect to eta2
p0.dp0.eta2.alt <- function(eta1,eta2)
{
dp0_eta2 <- -exp(-eta2-eta1)/2+
(exp(-eta2-eta1)*(exp(eta1)+1)*exp(eta2)+1))/(2*sqrt((exp(eta1)-1)^2*
exp(2*eta2)+(2*exp(eta1)+2)*exp(eta2)+1))
return(dp0_eta2)
}
# Second derivative with respect to eta2
p0.d2p0.eta1.alt <- function(eta1,eta2)
{
d2p0_eta1<- (exp(-eta2-eta1)*(exp(eta2)+1))/2+
-(exp(-eta2-eta1)*(exp(eta2)+1)^2)/(2*sqrt((exp(eta1)-1)^2*
exp(2*eta2)+(2*exp(eta1)+2)*exp(eta2)+1))+((exp(eta2+eta1)-exp(eta2)+1)*
(exp(2*eta2+eta1)-exp(eta2+eta1)-exp(2*eta2)-2*exp(eta2)-1))/
(2*((exp(eta1)-1)^2*exp(2*eta2)+(2*exp(eta1)+2)*exp(eta2)+1)^(3/2))
return(d2p0_eta1)
}

# Still problems when eta2 is equal to zero
p0.d2p0.eta2.alt <- function(eta1,eta2)
{
d2p0_eta2<- exp(-eta2-eta1)/2+
-exp(-eta2-eta1)/(2*sqrt((exp(eta1)-1)^2*exp(2*eta2)+(2*exp(eta1)+2)*exp(
eta2)+1))+
-(exp(-eta1)*(exp(eta2+eta1)+exp(eta2)+1)*(exp(eta2+2*eta1)-
2*exp(eta2+eta1)+exp(eta2)+exp(eta1)+1)))/

```

```

(2*((exp(eta1)-1)^2*exp(2*eta2)+(2*exp(eta1)+2)*exp(eta2)+1)^(3/2))
return(d2p0_eta2)
}

# Derivative with respect to eta1 and eta2
p0.d2p0.eta1.eta2.alt <- function(eta1,eta2)
{
d2p0_eta12<- exp(-eta2-eta1)/2-(exp(-eta2-eta1)*(exp(2*eta2+eta1)-
exp(2*eta2)+1))/(2*sqrt((exp(eta1)-1)^2*exp(2*eta2)+(2*exp(eta1)+2)*exp(
eta2)+1))+
(exp(-eta1)*(exp(eta2+2*eta1)-2*exp(eta2+eta1)+exp(eta2)+exp(eta1)+1)*
(exp(2*eta2+eta1)-exp(eta2+eta1)-exp(2*eta2)-2*exp(eta2)-1))/
(2*((exp(eta1)-1)^2*exp(2*eta2)+(2*exp(eta1)+2)*exp(eta2)+1)^(3/2))
return(d2p0_eta12)
}

if(param == "standard")
{
prob<-prob.standard
p0.dp0.eta1<- p0.dp0.eta1.std
p0.dp0.eta2 <- p0.dp0.eta2.std
p0.d2p0.eta1<- p0.d2p0.eta1.std
p0.d2p0.eta2<- p0.d2p0.eta2.std
p0.d2p0.eta1.eta2<- p0.d2p0.eta1.eta2.std
}

else
{
prob<-prob.alternative
p0.dp0.eta1<- p0.dp0.eta1.alt
p0.dp0.eta2 <- p0.dp0.eta2.alt
p0.d2p0.eta1<- p0.d2p0.eta1.alt
p0.d2p0.eta2<- p0.d2p0.eta2.alt
}

```

```

p0.d2p0.eta1.eta2<- p0.d2p0.eta1.eta2.alt

}

# Key quantities
key_quantities <- function(pars)
{
  gammas <- pars[1:p]
  betas <- pars[-c(1:p)]
  eta1 <- drop(x%*%gammas)
  eta2 <- drop(x%*%betas)
  pi0 <- prob(eta1[1:nex],eta2[1:nex])[1,]
  pi1 <- prob(eta1[(nex+1):n],eta2[(nex+1):n])[2,]
  dp0_eta1 <- p0.dp0.eta1(eta1[1:nex],eta2[1:nex])
  dp0_eta2 <- p0.dp0.eta2(eta1[1:nex],eta2[1:nex])
  dp1_eta1 <- p0.dp0.eta1(eta1[(nex+1):n],eta2[(nex+1):n])*exp(eta1[(nex+1):
    n])+pi1
  dp1_eta2 <- p0.dp0.eta2(eta1[(nex+1):n],eta2[(nex+1):n])*exp(eta1[(nex+1):
    n])
  d2p0_eta1 <- p0.d2p0.eta1(eta1[1:nex],eta2[1:nex])
  d2p0_eta2 <- p0.d2p0.eta2(eta1[1:nex],eta2[1:nex])
  d2p0_eta12 <- p0.d2p0.eta1.eta2(eta1[1:nex],eta2[1:nex])
  d2p1_eta1 <- exp(eta1[(nex+1):n])*(p0.d2p0.eta1(eta1[(nex+1):n],eta2[(nex
    +1):n])+
  2* p0.dp0.eta1(eta1[(nex+1):n],eta2[(nex+1):n]))+pi1
  d2p1_eta2 <- exp(eta1[(nex+1):n])*p0.d2p0.eta2(eta1[(nex+1):n],eta2[(nex
    +1):n])
  d2p1_eta12 <-exp(eta1[(nex+1):n])*(p0.d2p0.eta1.eta2(eta1[(nex+1):n],eta2
    [(nex+1):n]))+dp1_eta2
  out <- list(betas=betas, gammas=gammas, eta1=eta1,eta2=eta2,
  pi0=pi0,pi1=pi1,dp0_eta1=dp0_eta1,dp0_eta2=dp0_eta2,
  dp1_eta1=dp1_eta1,dp1_eta2=dp1_eta2,d2p0_eta1=d2p0_eta1,
  d2p0_eta2=d2p0_eta2,d2p0_eta12=d2p0_eta12,d2p1_eta1=d2p1_eta1,

```



```

d2p1_eta2=d2p1_eta2,d2p1_eta12=d2p1_eta12)
return(out)
}
# Fisher expected information
information <- function(pars,fit = NULL)
{
  if(is.null(fit))
  {
    fit <- key_quantities(pars)
  }
  info<- with(fit,
  { info <- matrix(NA, ncol = 2*p, nrow = 2*p)
  # defining the blocks of the expected information
  i.gamma.gamma <- t(x[1:nex,])%*%(1/(pi0*(1-pi0))*(dp0_eta1)^2*x[1:nex,])+
  t(x[(nex+1):n,])%*%(1/(pi1*(1-pi1))*(dp1_eta1)^2*x[(nex+1):n,])
  i.beta.beta <- t(x[1:nex,])%*%(1/(pi0*(1-pi0))*(dp0_eta2)^2*x[1:nex,])+
  t(x[(nex+1):n,])%*%(1/(pi1*(1-pi1))*(dp1_eta2)^2*x[(nex+1):n,])
  i.gamma.beta <- t(x[1:nex,])%*%(1/(pi0*(1-pi0))*(dp0_eta1)*(dp0_eta2)*x[1:
    nex,])+
  t(x[(nex+1):n,])%*%(1/(pi1*(1-pi1))*(dp1_eta1)*(dp1_eta2)*x[(nex+1):n,])
  info <- as.matrix(rbind( cbind(i.gamma.gamma,i.gamma.beta), cbind(t(i.
    gamma.beta),i.beta.beta)))
  info
  })
  return(info)
}
# Score-function
score <- function(pars,fit=NULL)
{
  if(is.null(fit))
  {
    fit <- key_quantities(pars)

```

```

}

score_tot <-with( fit,
# score for each observation
# need to put as.matrix for 1d case
{
score_gamma_ob <- rbind( (y[1:nex]-pi0)/(pi0*(1-pi0))*dp0_eta1*as.matrix(x
  [1:nex,]),
(y[(nex+1):n]-pi1)/(pi1*(1-pi1))*dp1_eta1*as.matrix(x[(nex+1):n,]))
# Sum of individual scores for gamma
score_gamma <- apply(score_gamma_ob,2,sum)
score_beta_ob <- rbind( (y[1:nex]-pi0)/(pi0*(1-pi0))*dp0_eta2*as.matrix(x
  [1:nex,]),
(y[(nex+1):n]-pi1)/(pi1*(1-pi1))*dp1_eta2*as.matrix(x[(nex+1):n,]))
score_beta <- apply(score_beta_ob,2,sum)
score_tot <- c(score_gamma,score_beta)
score_tot
}
)
score_tot
}

# Pgamma matrix
Pgamma <- function(pars, fit = NULL)
{
if(is.null(fit))
{
fit <- key_quantities(pars)
}
Pg <- with(fit,{
# parts costant in the matrix
Pg <- array(data = NA, dim= c(2*p,2*p,p))
q1_p0 <- (1-2*pi0)/(pi0*(1-pi0))^2
q1_p1 <-(1-2*pi1)/(pi1*(1-pi1))^2

```

```

for(k in 1:p)
{
for(s in 1:p)
{
for(r in 1:p)
{
Pg[s,r,k] <- sum(q1_p0*(dp0_eta1)^3*x[1:nex,k]*x[1:nex,s]*x[1:nex,r])+
sum(q1_p1*(dp1_eta1)^3*x[(nex+1):n,k]*x[(nex+1):n,s]*
x[(nex+1):n,r])
Pg[s,r+p,k] <- Pg[r+p,s,k] <- sum(q1_p0*(dp0_eta1)^2*(dp0_eta2)*
x[1:nex,k]*x[1:nex,s]*x[1:nex,r])+
sum(q1_p1*(dp1_eta1)^2*(dp1_eta2)*
x[(nex+1):n,k]*x[(nex+1):n,s]*x[(nex+1):n,r])
Pg[s+p,r+p,k] <- sum(q1_p0*(dp0_eta1)*(dp0_eta2)^2*
x[1:nex,k]*x[1:nex,s]*x[1:nex,r])+
sum(q1_p1*(dp1_eta1)*(dp1_eta2)^2*
x[(nex+1):n,k]*x[(nex+1):n,s]*x[(nex+1):n,r])
}
}
}
Pg
})
return(Pg)
}

# Pbeta matrix

Pbeta <- function(pars, fit = NULL)
{
if(is.null(fit))
{
fit <- key_quantities(pars)

```

```

}
Pb <- with(fit,{
Pb <- array(data = NA, dim= c(2*p,2*p,p))
# parts costant in the matrix
q1_p0 <- (1-2*pi0)/(pi0*(1-pi0))^2
q1_p1 <- (1-2*pi1)/(pi1*(1-pi1))^2
for(k in 1:p)
{
for(s in 1:p)
{
for(r in 1:p)
{
Pb[s,r,k] <- sum(q1_p0*(dp0_eta1)^2*(dp0_eta2)*x[1:nex,k]*
x[1:nex,s]*x[1:nex,r])+sum(q1_p1*(dp1_eta1)^2*
(dp1_eta2)*x[(nex+1):n,k]*x[(nex+1):n,s]*x[(nex+1):n,r])
Pb[s,r+p,k] <- Pb[r+p,s,k] <- sum(q1_p0*(dp0_eta1)*(dp0_eta2)^2*
x[1:nex,k]*x[1:nex,s]*x[1:nex,r])+
sum(q1_p1*(dp1_eta1)*(dp1_eta2)^2*
x[(nex+1):n,k]*x[(nex+1):n,s]*x[(nex+1):n,r])
Pb[s+p,r+p,k] <- sum(q1_p0*(dp0_eta2)^3*x[1:nex,k]*x[1:nex,s]*x[1:nex,r])+
sum(q1_p1*(dp1_eta2)^3*x[(nex+1):n,k]*x[(nex+1):n,s]*
x[(nex+1):n,r])
}
}
}
Pb
})
return(Pb)
}
# Qgamma matrix
Qgamma <- function(pars, fit = NULL)
{
if(is.null(fit))

```

```

{
fit <- key_quantities(pars)
}
Qg <- with(fit,{
Qg <- array(data = NA, dim= c(2*p,2*p,p))
# parts costant in the matrix
q1_p0 <- (1-2*pi0)/(pi0*(1-pi0))^2
q1_p1 <- (1-2*pi1)/(pi1*(1-pi1))^2
q2_p0 <- 1/(pi0*(1-pi0))
q2_p1 <- 1/(pi1*(1-pi1))
for(k in 1:p)
{
for(s in 1:p)
{
for(r in 1:p)
{
Qg[s,r,k] <- -sum((q1_p0*(dp0_eta1)^3- q2_p0*(d2p0_eta1)*(dp0_eta1))*
x[1:nex,k]*x[1:nex,s]*x[1:nex,r])+
-sum((q1_p1*(dp1_eta1)^3-q2_p1*(d2p1_eta1)*(dp1_eta1))*
x[(nex+1):n,k]*x[(nex+1):n,s]*x[(nex+1):n,r])
Qg[s,r+p,k] <- Qg[r+p,s,k] <- -sum((q1_p0*(dp0_eta1)^2*(dp0_eta2)-
q2_p0*(d2p0_eta12)*(dp0_eta1) )*
x[1:nex,k]*x[1:nex,s]*x[1:nex,r])+
-sum((q1_p1*(dp1_eta1)^2*(dp1_eta2)-
q2_p1*(d2p1_eta12)*(dp1_eta1))*
x[(nex+1):n,k]*x[(nex+1):n,s]*x[(nex+1):n,r])
Qg[s+p,r+p,k] <- -sum((q1_p0*(dp0_eta1)*(dp0_eta2)^2-q2_p0*(d2p0_eta2)*(
dp0_eta1))*
x[1:nex,k]*x[1:nex,s]*x[1:nex,r])-sum((q1_p1*(dp1_eta1)*
(dp1_eta2)^2-q2_p1*(d2p1_eta2)*(dp1_eta1) )*
x[(nex+1):n,k]*x[(nex+1):n,s]*x[(nex+1):n,r])
}
}
}
}
}

```

```

}
Qg
})
return(Qg)
}

# Qbeta matrix
Qbeta <- function(pars, fit = NULL)
{
  if(is.null(fit))
  {
    fit <- key_quantities(pars)
  }
  Qb <- with(fit,{
  Qb <- array(data = NA, dim= c(2*p,2*p,p))
  q1_p0 <- (1-2*pi0)/(pi0*(1-pi0))^2 # parts constant in the matrix
  q1_p1 <- (1-2*pi1)/(pi1*(1-pi1))^2
  q2_p0 <- 1/(pi0*(1-pi0))
  q2_p1 <- 1/(pi1*(1-pi1))
  for(k in 1:p)
  {
    for(s in 1:p)
    {
      for(r in 1:p)
      {
        Qb[s,r,k] <- -sum((q1_p0*(dp0_eta1)^2*(dp0_eta2) - q2_p0*(d2p0_eta1)*
          (dp0_eta2))*x[1:nex,k]*x[1:nex,s]*x[1:nex,r])+
          -sum((q1_p1*(dp1_eta1)^2*(dp1_eta2) - q2_p1*(d2p1_eta1)*
          (dp1_eta2))*x[(nex+1):n,k]*x[(nex+1):n,s]*x[(nex+1):n,r])
        Qb[s,r+p,k] <- Qb[r+p,s,k] <- -sum((q1_p0*(dp0_eta1)*(dp0_eta2)^2-
          q2_p0*(d2p0_eta12)*(dp0_eta2) )*
          x[1:nex,k]*x[1:nex,s]*x[1:nex,r]) - sum((q1_p1*(dp1_eta1)*
          (dp1_eta2)^2 - q2_p1*(d2p1_eta12)*(dp1_eta2) )*

```

```

x[(nex+1):n,k]*x[(nex+1):n,s]*x[(nex+1):n,r]
Qb[s+p,r+p,k] <- -sum((q1_p0*(dp0_eta2)^3-q2_p0*(d2p0_eta2)*(dp0_eta2))*
x[1:nex,k]*x[1:nex,s]*x[1:nex,r])+
-sum((q1_p1*(dp1_eta2)^3-q2_p1*(d2p1_eta2)*(dp1_eta2) ) *
x[(nex+1):n,k]*x[(nex+1):n,s]*x[(nex+1):n,r]
}
}
}
Qb
})
return(Qb)
}
# Matrix F1
F1 <- function(pars, fit = NULL)
{
if(is.null(fit))
{
fit <- key_quantities(pars)
}
f1 <- rep(NA,2*p)
p_gamma <- Pgamma(fit=fit)
p_beta <- Pbeta(fit=fit)
q_gamma <- Qgamma(fit=fit)
q_beta <- Qbeta(fit=fit)
info_inv <- solve(information(fit = fit))
for(k in 1:p)
{
f1[k] <- sum(diag(info_inv**%(p_gamma[, ,k]+q_gamma[, ,k])))
f1[k+p] <- sum(diag(info_inv**%(p_beta[, ,k]+q_beta[, ,k])))
}
return(f1)
}
# h function

```

```

h.function <- function(pars, fit = NULL)
{
  if(is.null(fit))
  {
    fit <- key_quantities(pars)
  }
  h <- with(fit,{
  h <- array(data = NA, dim= c(2*p,2*p,2*p))
  info.inv <- solve(information(fit=fit))
  for( k in 1:(2*p) )
  {
    h[,,k] <- info.inv[,k]%*%t(info.inv[,k])/info.inv[k,k]
  }
  h
})
  return(h)
}

# Creating F2 a function that calculate all the 2*(p-1) F2 vectors
F2.tilde <- function(pars, fit = NULL)
{
  if(is.null(fit))
  {
    fit <- key_quantities(pars)
  }
  f2 <- matrix(NA,ncol = 2*p, nrow = 2*p )
  f2.tilde <- rep(NA, 2*p)
  p_gamma <- Pgamma(fit=fit)
  p_beta <- Pbeta(fit=fit)
  q_gamma <- Qgamma(fit=fit)
  q_beta <- Qbeta(fit=fit)
  h<- h.function(fit=fit)
  inv.info <- solve(information(fit=fit))
  for(k in 1:p)

```



```
{
for(j in 1:(2*p))
{
f2[j,k] <- sum(diag(h[, , j]**%(p_gamma[, , k]/3+q_gamma[, , k]/2)))
f2[j,k+p] <- sum(diag(h[, , j]**%(p_beta[, , k]/3+q_beta[, , k]/2)))
}
}
for(k in 1:(2*p))
{
f2.tilde[k] <- inv.info[k]**%f2[k,]
}
return(f2.tilde)
}
#Estimation
max_step_factor<-10
epsilon <- 1e-5
slowit <- 1
par <- start ## initial value
# Controll for mle, mean bias reduction and median
# bias reduction
controlA <- 0
controlB <- 0
# Mean bias reduction
if(method=="Firth")
{
controlA <- 1
}
# Median bias reduction
if(method=="Median")
{
controlA <- 1
controlB <- 1
}
}
```

```

adjusted_grad <- score(par)+1/2*F1(par)*controlA-
information(par)%*%F2.tilde(par)*controlB
step_par <- solve(information(par))%*% adjusted_grad
for (iter in seq.int(maxit))
{
step_factor <- 0
testhalf <- TRUE
while (testhalf & step_factor < max_step_factor)
{
step_par_previous <- step_par
par <- par + slowit*2^(-step_factor)*step_par
adjusted_grad <- score(par)+1/2*F1(par)*controlA-information(par)%*%F2.
  tilde(par)*controlB
step_par <- solve(information(par)) %*% adjusted_grad
if (step_factor == 0 & iter == 1)
{
testhalf <- TRUE
}
else
{
testhalf <- sum(abs(step_par),na.rm=TRUE) > sum(abs(step_par_previous),na.
  rm = TRUE)
}
step_factor <- step_factor + 1
}

if ( (sum(abs(step_par),na.rm = TRUE) < epsilon) | (any(abs(par)>15)>0) )
  {
break
}
}

if(iter >= maxit| (any(abs(par)>15)>0))

```

```
{
convergence <- 0
warning("optimization failed to converge")
}
else
{
convergence <- 1
}
Point.est <- par
var.est <- solve(information(Point.est))
se <- sqrt(diag(var.est))
return(list(Point.est=Point.est, se=se,convergence=convergence))
}
```

---

## B.2 Funzione per gli studi di simulazione

---

```
source("~/Funzione_stima.R")
# Data generation function
genData <- function(nsim,true_gammas,true_betas,x,t)
{
prob.calc<- function(eta1,eta2)
{
p0 <- rep(-1,length(eta1))
num <- (- (exp(eta1)+1)*exp(eta2)+ sqrt(exp(2*eta2)*(exp(eta1)+1)^2+4*exp(
eta1+eta2)*(1-exp(eta2))))
den <- (2*exp(eta1)*(1-exp(eta2)))
p0 <- num/den
p0[abs(eta2)<1e-8] <- 1/(1+exp(eta1[abs(eta2)<1e-8]))
p1 <- p0*exp(eta1)
p0[p0<1e-8]<- 1e-8
p0[p0>(1-1e-8)]<- 1-1e-8
p1[p1<1e-8]<- 1e-8
```

```

p1[p1>(1-1e-8)]<- 1-1e-8
return(rbind(p0,p1))
}
n <- length(x[,1])
sim <- matrix(NA,n, ncol = nsim)
eta1 <- drop(x%%true_gammas)
eta2 <- drop(x%%true_betas)
pr <- rep(NA,n)
pr[t==0]<- prob.calc(eta1[t==0],eta2[t==0])[1,]
pr[t==1]<- prob.calc(eta1[t==1],eta2[t==1])[2,]
# simulate nsim samples of y
for(j in 1:nsim)
{
sim[,j] <- rbinom(n=n,1,prob =pr )
}
sim
}
# Simulation function
simRichardson<- function(nsim,n,true_gammas,true_betas,setSeed=1234,prop.
  esp=0.5)
{
require(brm) # loading Richardson package for a comparision with mle
set.seed(setSeed)
p <- length(true_gammas)
t <- c(rep(0,n-round(n*prop.esp)),rep(1,round(n*prop.esp)))
x <- rep(1,n)
for(k in 2:p){ x <- cbind(x,rnorm(n))}
x <- as.matrix(x)
# error messages
messagemle<- messagebr<-messagebrm<-messagembr <- messageAltM<-messageAltF
  <- messageAltMBR <- list()
# Generating the data with fixed model matrix

```

```
responses=genData(nsim,true_gammas,true_betas,x,t) # funzione che genera
  le risopste
# Defining other quantities of interest
mle <- semle<-brm <- sebrm<- meanBR <- semeanBR <- medianbr<- semedian<-
  altm<- sealtm <-
altf<- sealtf <- altmbr <- sealtmbr <- matrix(NA,nsim,2*p) # righe e
  variabili p
outmle <- outmeanBr <- outaltm <- outaltf <-outmbr <- outaltmbr <- rep(NA
  , nsim)
# defining a loop in order to get the estimates in each sample
for(j in 1:nsim)
{
y <- responses[,j]
# mle Richardson
fit1=try(rrmod(x,y,t),T)
if(is.character(fit1)) {
messagemle[[j]] <- fit1
}
if(!is.character(fit1)){
mle[j,]= fit1$Point.est
semle[j,]<- fit1$se
outmle[j]<- fit1$convergence
}
# Firth Richardson
fit2 <- try(rrmod(x,y,t, method = "Firth")
,T)
if(is.character(fit2)) {
messagebr[[j]] <- fit2
}
if(!is.character(fit2)){
meanBR[j,]=fit2$Point.est
semeanBR[j,]=fit2$se
outmeanBr[j]<- fit2$convergence
```

```
}  
# Richardson package  
fit3 <- try(brm(y,t,va=x,param = "RR")  
,T)  
if(is.character(fit3)) {  
messagebrm[[j]] <- fit3  
}  
if(!is.character(fit3)){  
brm[j,]=fit3$point.est  
sebrm[j,]=fit3$se.est  
}  
# mle alternative parametrization  
fit4 <- try(rrmod(x,y,t, param = "alternative")  
,T)  
if(is.character(fit4)) {  
messageAltM[[j]] <- fit4  
}  
if(!is.character(fit4)){  
altm[j,]=fit4$Point.est  
sealtm[j,]=fit4$se  
outaltm[j]<- fit4$convergence  
}  
# Firth alternative parametrization  
fit5 <- try(rrmod(x,y,t, param = "alternative", method = "Firth")  
,T)  
if(is.character(fit5)) {  
messageAltF[[j]] <- fit5  
}  
if(!is.character(fit5)){  
altf[j,]=fit5$Point.est  
sealtf[j,]=fit5$se  
outaltf[j]<- fit5$convergence  
}
```

```
# Median Richardson parametrization
fit6 <- try(rrmod(x,y,t,method = "Median")
,T)
if(is.character(fit6)) {
  messagebr[[j]] <- fit6
}
if(!is.character(fit6)){
  medianbr[j,]=fit6$Point.est
  semedian[j,]=fit6$se
  outmbr[j]<- fit6$convergence
}
# Median alternative parametrization
fit7 <- try(rrmod(x,y,t,param = "alternative", method = "Median")
,T)
if(is.character(fit7)) {
  messageAltMBR[[j]] <- fit7
}
if(!is.character(fit7)){
  altmbr[j,]=fit7$Point.est
  sealtmbr[j,]=fit7$se
  outaltmbr[j]<- fit7$convergence
}
cat("Simulation",j,"of",nsim,"\n")
}
list(nsim=nsim,n=n,x=x,t=t,responses=responses,setSeed=setSeed,
mle=mle,semle=semle,messagele=messagele,
meanBR=meanBR,semeanBR=semeanBR,messagebr=messagebr,
medianbr=medianbr,semedian=semedian,altmbr=altmbr,sealtmbr=sealtmbr,
brm=brm,sebrm=sebrm, messagebrm=messagebrm,altm=altm,sealtm=sealtm,
messageAltM=messageAltM,altf=altf,sealtf=sealtf,messageAltF=messageAltF,
messagebr=messagebr,messageAltMBR=messageAltMBR,
true_gammas=true_gammas, true_betas=true_betas,
outmle=outmle,outmeanBr=outmeanBr,outaltf=outaltf,outaltm=outaltm,
```

```

outmbr=outmbr,outaltmbr=outaltmbr)
}
# Function for Wald coverage
Wald.coverage <- function(sim,level=0.975)
{
p <- length(sim$true_gammas)
s <- 6
true_gammas.matrix <- matrix(rep(sim$true_gammas,sim$nsim),ncol=length(sim
  $x[1,]),byrow = T)
mle.nex <- apply(sim$mle,1,function(x) all(abs(x)<s))
mle.cov <- apply(abs((sim$mle[mle.nex,1:p]-true_gammas.matrix[mle.nex,])/
  sim$semle[mle.nex,1:p])<qnorm(level),2,mean, na.rm = T)
firth.nex <- apply(sim$meanBR,1,function(x) all(abs(x)<s))
firth.cov <- apply(abs((sim$meanBR[firth.nex,1:p]-true_gammas.matrix[firth
  .nex,])/sim$semeanBR[firth.nex,1:p])<qnorm(level),2,mean,na.rm=T)
brm.nex <- apply(sim$brm,1,function(x) all(abs(x)<s))
brm.cov <- apply(abs((sim$brm[brm.nex,1:p]-true_gammas.matrix[brm.nex,])/
  sim$sebrm[brm.nex,1:p])<qnorm(level),2,mean, na.rm = T)
med.nex <- apply(sim$medianbr,1,function(x) all(abs(x)<s))
median.cov <- apply(abs((sim$medianbr[med.nex,1:p]-true_gammas.matrix[med.
  nex,])/sim$semedian[med.nex,1:p])<qnorm(level),2,mean,na.rm=T)
alt.fir.nex <- apply(sim$altf,1,function(x) all(abs(x)<s))
alt.fir.cov <- apply(abs((sim$altf[alt.fir.nex,1:p]-true_gammas.matrix[alt
  .fir.nex,])/sim$sealtf[alt.fir.nex,1:p])<qnorm(level),2,mean, na.rm =
  T)
alt.nex <- apply(sim$altm,1,function(x) all(abs(x)<s))
alt.cov <- apply(abs((sim$altm[alt.nex,1:p]-true_gammas.matrix[alt.nex,])/
  sim$sealtm[alt.nex,1:p])<qnorm(level),2,mean, na.rm = T)
alt.med.nex <- apply(sim$altmbr,1,function(x) all(abs(x)<s))
alt.med.cov <- apply(abs((sim$altmbr[alt.med.nex,1:p]-true_gammas.matrix[
  alt.med.nex,])/sim$sealtmbr[alt.med.nex,1:p])<qnorm(level),2,mean, na.
  rm = T)

```



```

rbind(mle.cov, firth.cov, median.cov, brm.cov, alt.cov, alt.fir.cov, alt.med.
      cov)
}
# Function for relative bias
Relative.bias <- function(sim)
{
true_gammas.matrix <- matrix(rep(sim$true_gammas, sim$nsim), ncol=length(sim
  $x[1,]), byrow = T)
p <- length(sim$true_gammas)
s <- 6
mle.nex <- apply(sim$mle, 1, function(x) all(abs(x) < s))
mle <- 100*(apply(sim$mle[mle.nex&sim$outmle, 1:p] - true_gammas.matrix[mle.
  nex&sim$outmle, ], 2, mean, na.rm = T)/abs(sim$true_gammas))
brm.nex <- apply(sim$brm, 1, function(x) all(abs(x) < s))
brm <- 100*(apply(sim$brm[brm.nex, 1:p] - true_gammas.matrix[brm.nex, ], 2, mean
  , na.rm = T)/abs(sim$true_gammas))
firth.nex <- apply(sim$meanBR, 1, function(x) all(abs(x) < s))
firth <- 100*(apply(sim$meanBR[firth.nex&sim$outmeanBr, 1:p] - true_gammas.
  matrix[firth.nex&sim$outmeanBr, ], 2, mean, na.rm=T)/abs(sim$true_gammas))
med.nex <- apply(sim$medianbr, 1, function(x) all(abs(x) < s))
medianBr <- 100*(apply(sim$medianbr[med.nex, 1:p] - true_gammas.matrix[med.
  nex, ], 2, mean, na.rm=T)/abs(sim$true_gammas))
alt.nex <- apply(sim$altm, 1, function(x) all(abs(x) < s))
alt <- 100*(apply(sim$altm[alt.nex&sim$outaltm, 1:p] - true_gammas.matrix[alt
  .nex&sim$outaltm, ], 2, mean, na.rm=T)/abs(sim$true_gammas))
alt.fir.nex <- apply(sim$altf, 1, function(x) all(abs(x) < s))
alt.firth <- 100*(apply(sim$altf[alt.fir.nex&sim$outaltf, 1:p] - true_gammas.
  matrix[alt.fir.nex&sim$outaltf, ], 2, mean, na.rm=T)/abs(sim$true_gammas))
alt.med.nex <- apply(sim$altmbr, 1, function(x) all(abs(x) < s))
alt.medianBr <- 100*(apply(sim$altmbr[alt.med.nex, 1:p] - true_gammas.matrix[
  alt.med.nex, ], 2, mean, na.rm=T)/abs(sim$true_gammas))
out <- rbind(mle, brm, firth, medianBr, alt, alt.firth, alt.medianBr)
out

```

```

}
# Relative bias std-err
Relative.bias.std <- function(sim)
{
p <- length(sim$true_gammas)
s <- 6
true_gammas.matrix <- matrix(rep(c(sim$true_gammas,sim$true_betas ),sim$
  nsim),ncol=length(sim$true_gammas)+length(sim$true_betas),byrow = T)
mle.nex <- apply(sim$mle,1,function(x) all(abs(x)<s))
mle <- (apply((100*(sim$mle- true_gammas.matrix )/abs(true_gammas.matrix))
  [mle.nex&sim$outmle,1:p],2,var, na.rm =T))
brm.nex <- apply(sim$brm,1,function(x) all(abs(x)<s))
brm <- (apply((100*(sim$brm- true_gammas.matrix)/ abs(true_gammas.matrix))
  [brm.nex,1:p],2,var,na.rm = T))
firth.nex <- apply(sim$meanBR,1,function(x) all(abs(x)<s))
firth<- (apply((100*(sim$meanBR- true_gammas.matrix)/ abs(true_gammas.
  matrix))[firth.nex&sim$outmeanBr,1:p],2,var,na.rm=T))
med.nex <- apply(sim$medianbr,1,function(x) all(abs(x)<s))
medianBr <- (apply((100*(sim$medianbr- true_gammas.matrix)/ abs(true_
  gammas.matrix))[med.nex,1:p],2,var,na.rm=T))
alt.nex <- apply(sim$altm,1,function(x) all(abs(x)<s))
alt <- (apply((100*(sim$altm- true_gammas.matrix)/ abs(true_gammas.matrix)
  )[alt.nex&sim$outaltm,1:p],2,var,na.rm=T))
alt.fir.nex <- apply(sim$altf,1,function(x) all(abs(x)<s))
alt.firth <- (apply((100*(sim$altf- true_gammas.matrix)/ abs(true_gammas.
  matrix))[alt.fir.nex&sim$outaltf,1:p],2,var,na.rm=T))
alt.med.nex <- apply(sim$altmbr,1,function(x) all(abs(x)<s))
alt.medianBr <- (apply((100*(sim$altmbr - true_gammas.matrix)/ abs(true_
  gammas.matrix))[alt.med.nex,1:p],2,var,na.rm=T))
out <- sqrt(rbind(mle,brm,firth,medianBr,alt,alt.firth,alt.medianBr)/sim$
  nsim)
out
}

```

```

# Empirical RMSE Function
RMSE <- function(sim)
{
p <- length(sim$true_gammas)
s <-6
true_gammas.matrix <- matrix(rep(c(sim$true_gammas,sim$true_betas ),sim$
  nsim),ncol=length(sim$true_gammas)+length(sim$true_betas),byrow = T)
mle.nex <- apply(sim$mle,1,function(x) all(abs(x)<s))
mle <- (apply((sim$mle-true_gammas.matrix)[mle.nex&sim$outmle,1:p]^2,2,
  mean, na.rm =T))
mle.var <- (apply((sim$mle-true_gammas.matrix)[mle.nex&sim$outmle,1:p
  ]^2,2,var, na.rm =T))
brm.nex <- apply(sim$brm,1,function(x) all(abs(x)<s))
brm <- (apply((sim$brm-true_gammas.matrix)[brm.nex,1:p]^2,2,mean,na.rm = T
  ))
brm.var <- (apply((sim$brm-true_gammas.matrix)[brm.nex,1:p]^2,2,var,na.rm
  = T))
firth.nex <- apply(sim$meanBR,1,function(x) all(abs(x)<s))
firth<-(apply((sim$meanBR-true_gammas.matrix)[firth.nex&sim$outmeanBr,1:p
  ]^2,2,mean,na.rm=T))
firth.var<-(apply((sim$meanBR-true_gammas.matrix)[firth.nex&sim$outmeanBr
  ,1:p]^2,2,var,na.rm=T))
med.nex <- apply(sim$medianbr,1,function(x) all(abs(x)<s))
medianBr <- (apply((sim$medianbr-true_gammas.matrix)[med.nex,1:p]^2,2,mean,
  na.rm=T))
medianBr.var <- (apply((sim$medianbr-true_gammas.matrix)[med.nex,1:p]^2,2,
  var,na.rm=T))
alt.nex <- apply(sim$altm,1,function(x) all(abs(x)<s))
alt <- (apply((sim$altm-true_gammas.matrix)[alt.nex&sim$outaltm,1:p]^2,2,
  mean,na.rm=T))
alt.var <- (apply((sim$altm-true_gammas.matrix)[alt.nex&sim$outaltm,1:p
  ]^2,2,var,na.rm=T))
alt.fir.nex <- apply(sim$altf,1,function(x) all(abs(x)<s))

```

```

alt.firth <- (apply((sim$altf-true_gammas.matrix)[alt.fir.nex&sim$outaltf
,1:p]^2,2,mean,na.rm=T))
alt.firth.var <- (apply((sim$altf-true_gammas.matrix)[alt.fir.nex&sim$
outaltf,1:p]^2,2,var,na.rm=T))
alt.med.nex <- apply(sim$altmbr,1,function(x) all(abs(x)<s))
alt.medianBr <- (apply((sim$altmbr-true_gammas.matrix)[alt.med.nex,1:p
]^2,2,mean,na.rm=T))
alt.medianBr.var <- (apply((sim$altmbr-true_gammas.matrix)[alt.med.nex,1:p
]^2,2,var,na.rm=T))
out <- rbind(mle,brm,firth,medianBr,alt,alt.firth,alt.medianBr)
out.std <- rbind(mle.var,brm.var,firth.var,medianBr.var,alt.var,alt.firth.
var,alt.medianBr.var)/sim$nsim*1/
(2*sqrt(rbind(mle,brm,firth,medianBr,alt,alt.firth,alt.medianBr)))^2
return( list( RMSE = sqrt(out), RMSE.std = sqrt(out.std))
}
# Percentage of under estimation
Perc.under <- function(sim)
{
s <- 6
p <- length(sim$true_gammas)
true_gammas.matrix <- matrix(rep(sim$true_gammas,sim$nsim),ncol=length(sim
$x[1,]),byrow = T)
mle.nex <- apply(sim$mle,1,function(x) all(abs(x)<s))&sim$outmle
brm.nex <- apply(sim$brm,1,function(x) all(abs(x)<s))
firth.nex <- apply(sim$meanBR,1,function(x) all(abs(x)<s))&sim$outmeanBr
med.nex <- apply(sim$medianbr,1,function(x) all(abs(x)<s))&sim$outmbr
alt.nex <- apply(sim$altm,1,function(x) all(abs(x)<s))&sim$outaltm
alt.fir.nex <- apply(sim$altf,1,function(x) all(abs(x)<s))&sim$outaltf
alt.med.nex <- apply(sim$altmbr,1,function(x) all(abs(x)<s))&sim$outaltmbr
id <- mle.nex&firth.nex&med.nex&alt.nex&alt.fir.nex&alt.med.nex
mle <- 100*(apply(sim$mle[id,1:p]<true_gammas.matrix[id,1:p],2,mean, na.rm
=T))

```

```

brm <- 100*(apply(sim$brm[id,1:p]<true_gammas.matrix[id,1:p],2,mean,na.rm=
  T))
firth<- 100*(apply(sim$meanBR[id,1:p]<true_gammas.matrix[id,1:p],2,mean,na
  .rm=T))
medianBr <- 100*(apply(sim$medianbr[id,1:p]<true_gammas.matrix[id,1:p],2,
  mean,na.rm=T))
alt <- 100*(apply(sim$altm[id,1:p]<true_gammas.matrix[id,1:p],2,mean,na.rm
  =T))
alt.firth <- 100*(apply(sim$altf[id,1:p]<true_gammas.matrix[id,1:p],2,mean
  , na.rm=T))
alt.medianBr <- 100*(apply(sim$altmbr[id,1:p]<true_gammas.matrix[id,1:p
  ],2,mean, na.rm=T))
out <- rbind(mle,brm,firth,medianBr,alt,alt.firth, alt.medianBr)
out
}
# number of finite samples in the simulation
number.finite.sample <- function(sim)
{
s <- 6
p <- length(sim$true_gammas)
true_gammas.matrix <- matrix(rep(sim$true_gammas,sim$nsim),ncol=length(sim
  $x[1,]),byrow = T)
mle.nex <- sum(apply(sim$mle,1,function(x) all(abs(x)<s)&(!any(is.na(x))))
  &sim$outmle,na.rm = T)
brm.nex <- sum(apply(sim$brm,1,function(x) all(abs(x)<s)&(!any(is.na(x))))
  ,na.rm = T)
firth.nex <- sum(apply(sim$meanBR,1,function(x) all(abs(x)<s)&(!any(is.na(
  x))))&sim$outmeanBr,na.rm = T)
med.nex <- sum(apply(sim$medianbr,1,function(x) all(abs(x)<s)&(!any(is.na(
  x))))&sim$outmbr,na.rm = T)
alt.nex <- sum(apply(sim$altm,1,function(x) all(abs(x)<s)&(!any(is.na(x)))
  )&sim$outaltm,na.rm = T)

```

```

alt.fir.nex <- sum(apply(sim$altf,1,function(x) all(abs(x)<s)&(!any(is.na(
  x))))&sim$outaltf,na.rm = T)
alt.med.nex <- sum(apply(sim$altmbr,1,function(x) all(abs(x)<s)&(!any(is.
  na(x))))&sim$outaltmbr,na.rm = T)
out <-c(mle.nex,brm.nex,firth.nex,
med.nex,alt.nex,alt.fir.nex,
alt.med.nex)
names(out)<- c("mle","brm","firth", "median", "alt","altF","altM")
out
}
# Estimator standard error
SE <- function(sim)
{
p <- length(sim$true_gammas)
s <-6
mle.nex <- apply(sim$mle,1,function(x) all(abs(x)<s))
mle <- apply(sim$semle[mle.nex,],2,mean,na.rm=T)
brm.nex <- apply(sim$brm,1,function(x) all(abs(x)<s))
firth.nex <- apply(sim$meanBR,1,function(x) all(abs(x)<s))
firth<- apply(sim$semeanBR[firth.nex,],2,mean,na.rm=T)
med.nex <- apply(sim$medianbr,1,function(x) all(abs(x)<s))
medianBr <- apply(sim$semedian[med.nex,],2,mean,na.rm=T)
alt.nex <- apply(sim$altm,1,function(x) all(abs(x)<s))
alt <- apply(sim$sealtm[alt.nex,],2,mean,na.rm=T)
alt.fir.nex <- apply(sim$altf,1,function(x) all(abs(x)<s))
alt.firth <- apply(sim$sealtf[alt.fir.nex,],2,mean,na.rm=T)
alt.med.nex <- apply(sim$medianAlt,1,function(x) all(abs(x)<s))
alt.medianBr <- apply(sim$semedianAlt[alt.med.nex,],2,mean,na.rm=T)
out <- rbind(mle,firth,medianBr,alt,alt.firth,alt.medianBr)
out
}
# Stderr of estimator standard error
SE.se <- function(sim)

```

```

{
p <- length(sim$true_gammas)
s <- 6
mle.nex <- apply(sim$mle,1,function(x) all(abs(x)<s))
mle <- apply(sim$semle[mle.nex,],2,var,na.rm=T)/sum(mle.nex, na.rm = T)
firth.nex <- apply(sim$meanBR,1,function(x) all(abs(x)<s))
firth <- apply(sim$semeanBR[firth.nex,],2,var,na.rm=T)/sum(firth.nex, na.rm
= T)
med.nex <- apply(sim$medianbr,1,function(x) all(abs(x)<s))
medianBr <- apply(sim$semedian[med.nex,],2,var,na.rm=T)/sum(med.nex, na.rm
= T)
alt.nex <- apply(sim$altm,1,function(x) all(abs(x)<s))
alt <- apply(sim$sealtm[alt.nex,],2,var,na.rm=T)/sum(alt.nex, na.rm = T)
alt.fir.nex <- apply(sim$altf,1,function(x) all(abs(x)<s))
alt.firth <- apply(sim$sealtf[alt.fir.nex,],2,var,na.rm=T)/sum(alt.fir.nex
, na.rm = T)
alt.med.nex <- apply(sim$medianAlt,1,function(x) all(abs(x)<s))
alt.medianBr <- apply(sim$semedianAlt[alt.med.nex,],2,var,na.rm=T)/sum(alt
.med.nex, na.rm = T)
out <- sqrt(rbind(mle,firth,medianBr,alt,alt.firth,alt.medianBr))
out
}
# Simulations
# simulation 1 Tesi
true_g1 <- c(-0.5,-0.1,0.6)
true_b1 <- c(-1.4, 0.8,1.1)
sim1 <- simRichardson(nsim = 10000,n=60,true_g1,true_b1, setSeed = 1)
save(sim1,file = "Simulazione1")
load("Simulazione1")
Perc.under(sim1)
Relative.bias(sim1)
Relative.bias.std(sim1)
Wald.coverage(sim1)

```

```
RMSE(sim1)
sim1$med
# simulation 2 Tesi
true_g2 <- c(0.7, -0.4, 0.3)
true_b2 <- c(-0.9, -1.1, 0.9)
sim2 <- simRichardson(nsim = 10000, n=60, true_g2, true_b2, setSeed = 1)
save(sim2, file = "Simulazione2")
load("Simulazione2")
Perc.over(sim2)
Relative.bias(sim2)
Relative.bias.std(sim2)
Wald.coverage(sim2)
RMSE(sim2)
# simulation 3 Tesi
true_g3 <- c(-0.5, -0.1, 0.6)
true_b3 <- c(-1.4, 0.8, 1.1)
sim3 <- simRichardson(nsim = 10^4, n=60, true_g3, true_b3, prop.esp = 1/3,
  setSeed = 1)
save(sim3, file = "Simulazione3")
load("Simulazione3")
Perc.over(sim3)
Relative.bias(sim3)
Relative.bias.std(sim3)
Wald.coverage(sim3)
RMSE(sim3)
# simualtion 4 Tesi
true_g4 <- c(0.7, -0.4, 0.3)
true_b4 <- c(-0.9, -1.1, 0.9)
sim4 <- simRichardson(nsim = 10000, n=60, true_g4, true_b4, setSeed = 1, prop
  .esp = 1/3)
save(sim4, file = "Simulazione4")
load("Simulazione4")
Perc.over(sim4)
```



---

`Relative.bias(sim4)`

`Relative.bias.std(sim4)`

`Wald.coverage(sim4)`

`RMSE(sim4)`

---



# Bibliografia

Agresti, A (2019). *An Introduction to Categorical Data Analysis*. 3rd ed. John Wiley & Sons, New Jersey.

Barros, A.J. & Hirakata, V.N (2003). Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology* **3**, (21).

Carter, R. E., Lipsitz, S. R. & Tilley, B. C. (2005). Quasi-likelihood estimation for relative risk regression models. *Biostatistics* **6**, 39-44.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27-38.

Kenne Pagui, E. C., Salvan, A. & Sartori, N. (2017). Median bias reduction of maximum likelihood estimates. *Biometrika* **104**, 923-38.

Kenne Pagui, E. C., Salvan, A. & Sartori, N. (2020). Efficient implementation of median bias reduction. URL: <https://arxiv.org/pdf/2004.08630.pdf>

Kosmidis, I. (2014). Bias in parametric estimation: Reduction and useful side-effects. *WIREs Computational Statistics* **6**, 185-96.

Kosmidis, I. & Firth, D. (2010) A generic algorithm for reducing bias in parametric estimation. *Electronic Journal of Statistics* **4**, 1097-1112.

Lumley, T., Kronmal, R. & Ma, S. (2006). Relative risk regression in medical research: models, contrasts, estimators, and algorithms. *UW Biostatistics Working Paper Series*.

McCullagh, P. (1987). *Tensor Methods in Statistics*, Chapman and Hall, London.

McNutt, L.A., Wu, C., Xue, X. & Hafner, J.P (2003). Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American Journal of Epidemiology* **157**, 940-43.

Pace, L. & Salvan, A. (2001). *Introduzione alla Statistica II*, Cedam, Padova.

- 
- Pace, L. & Salvan, A. (1996). *Teoria della Statistica* Cedam, Padova.
- Richardson, T. S., Robins, J. M. & Wang, L. (2017). On modeling and estimation for the relative risk and risk difference. *Journal of the American Statistical Association* **112**, 1121-30.
- Sur P. & Candès E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences Jul 2019*, **116** 14516-25.
- Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika* **43**, 353-60.
- Zhang, J. & Yu K.F. (1998) What's relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *Journal of American Medical Association* **280**,1690–1701.
- Zou G. A. (2004) Modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* **159**, 702-6.