

Università degli studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Magistrale in  
Scienze statistiche



**METODI EMPIRICI DI RIDUZIONE DELLA DISTORSIONE IN  
MODELLI LINEARI GENERALIZZATI E LORO ESTENSIONI**

Relatore Prof. Nicola Sartori  
Dipartimento di Scienze Statistiche

Laureando Nicolò Noventa  
Matricola N 1202670

Anno Accademico 2019/2020

# Indice

<b>Introduzione</b>	<b>4</b>
<b>1 Inferenza di verosimiglianza</b>	<b>6</b>
1.1 Modelli statistici: definizioni . . . . .	6
1.2 Stimatori e proprietà . . . . .	7
1.3 Funzione di verosimiglianza . . . . .	8
1.4 Quantità di verosimiglianza e proprietà . . . . .	9
1.4.1 Funzione <i>score</i> . . . . .	10
1.4.2 Informazione osservata e informazione attesa . . . . .	10
1.5 Risultati distributivi . . . . .	11
<b>2 Modelli lineari generalizzati e sovradisersione</b>	<b>12</b>
2.1 Famiglie di dispersione esponenziale . . . . .	12
2.2 Modelli lineari generalizzati . . . . .	13
2.3 Distribuzione binomiale e modello di regressione logistica . . . . .	14
2.4 Distribuzione e modello di regressione Poisson . . . . .	16
2.5 Sovradispersione . . . . .	17
2.6 Distribuzione e modello di regressione Gamma . . . . .	18
2.7 Distribuzione e modello di regressione binomiale negativo . . . . .	20
<b>3 Metodi empirici di riduzione della distorsione</b>	<b>25</b>
3.1 Riduzione della distorsione . . . . .	25
3.2 Metodo empirico di riduzione della distorsione . . . . .	27
3.2.1 Specificazione del modello . . . . .	28
3.2.2 Assunzioni . . . . .	29
3.3 Modifica della funzione di stima per la riduzione della distorsione	31

3.3.1	Distorsione asintotica . . . . .	31
3.3.2	Famiglia di modificazioni delle funzioni di stima per la riduzione della distorsione . . . . .	32
3.4	Modificazione empirica per la correzione della distorsione . . . . .	34
3.5	Distribuzione asintotica dello stimatore RBM e inferenza . . . . .	35
3.6	Penalizzazione empirica della funzione obiettivo per la riduzione della distorsione . . . . .	37
3.7	Riduzione della distorsione e selezione del modello basata sulla divergenza di Kullback-Leibler . . . . .	38
3.8	Implementazione . . . . .	39
3.9	Riduzione della distorsione in modelli lineari generalizzati . . . . .	40
<b>4</b>	<b>Applicazione di metodi di riduzione della distorsione in modelli lineari generalizzati e loro estensioni</b> . . . . .	<b>42</b>
4.1	I dati Ants . . . . .	42
4.2	I dati Crabs . . . . .	45
4.3	Applicazione dei metodi di riduzione della distorsione . . . . .	47
4.4	Studi di simulazione . . . . .	50
4.5	Studio dei tempi di esecuzione dei metodi . . . . .	63
	<b>Conclusioni</b> . . . . .	<b>69</b>
<b>A</b>	<b>Quantità per la penalizzazione empirica</b> . . . . .	<b>75</b>
A.1	Quantità necessarie per le penalizzazioni in modelli lineari generalizzati . . . . .	75
A.2	Espressione di $j^{(G)}(\cdot)$ , $\Psi(\cdot)$ , $\tilde{\Psi}_r(\cdot)$ e $u_r(\cdot)$ per alcuni modelli di regressione . . . . .	76
A.2.1	Modello di regressione logistica con funzione legame logistica . . . . .	76
A.2.2	Modello di regressione Poisson con funzione legame logaritmica . . . . .	77
A.2.3	Modello di regressione Gamma con funzione legame logaritmica . . . . .	78
A.2.4	Modello di regressione Binomiale negativo con funzione legame logaritmica . . . . .	79



# Introduzione

Nei problemi di stima parametrica regolare, lo stimatore di massima verosimiglianza è trovato come soluzione dell'equazione di stima basata sulla funzione *score*. Esso gode di importanti proprietà tra cui la non distorsione asintotica. Tuttavia per campioni finiti non è in generale garantita la non distorsione. Viene quindi illustrato un metodo generale, proposto in Kosmidis e Lunardon (2020), per la riduzione della distorsione asintotica per gli stimatori che vengono ottenuti dalla risoluzione di equazioni di stima non distorte o tramite la massimizzazione di opportune funzioni obiettivo. Il metodo consiste in una modificazione empirica dell'equazione di stima che dipende solo dalle prime due derivate dei contributi delle equazioni di stima.

Il nuovo stimatore ha maggiore applicabilità rispetto ai precedenti metodi per la riduzione della distorsione presenti in letteratura poiché può essere applicato a modelli in cui la verosimiglianza è parzialmente specificata o troppo onerosa da calcolare. Questo è in contrasto con i consueti metodi per la riduzione della distorsione che richiedono ricampionamento o il calcolo di valori attesi di prodotti di derivate parziali della funzione di stima. Se lo stimatore è di tipo  $M$ , ossia se è ottenuto massimizzando una funzione obiettivo, allora, lo stimatore  $M$  ridotto per la distorsione può essere ottenuto tramite la massimizzazione di un'opportuna funzione obiettivo penalizzata. Questa penalizzazione è strettamente connessa ai criteri di informazione basati sulla divergenza di Kullback-Leibler, stabilendo un legame tra riduzione della distorsione nella stima e selezione del modello.

L'obiettivo della tesi è illustrare e confrontare il metodo empirico proposto in Kosmidis e Lunardon (2020) con i classici metodi di riduzione della distorsione presenti in letteratura, con particolare attenzione ai modelli di

regressione gamma e binomiali negativi. Nel primo capitolo sono presentate alcune nozioni teoriche riguardanti la specificazione dei modelli statistici e la stima nell'ambito dell'inferenza di verosimiglianza. Nel secondo capitolo si affronta inizialmente il problema della regressione nell'ambito dei modelli lineari generalizzati. Successivamente viene illustrata la specificazione dei modelli di regressione gamma e binomiale negativo per gestire la sovradi dispersione. Nel terzo capitolo viene discusso il problema della riduzione della distorsione. In un primo momento vengono confrontati i metodi di riduzione della distorsione noti in letteratura, in seguito vengono definite le assunzioni e le quantità necessarie all'ottenimento dello stimatore ridotto per la distorsione empirica. Infine, nel quarto capitolo sono presentati i risultati di alcuni studi di simulazione, condotti in modelli di regressione gamma e binomiale negativo, volti alla verifica empirica delle proprietà possedute dagli stimatori ridotti per la distorsione e alla valutazione dell'efficienza computazionale delle procedure di stima. Gli studi di simulazione e le applicazioni sono effettuate con R (R Core Team, [2019](#)).

# Capitolo 1

## Inferenza di verosimiglianza

L'inferenza statistica si basa sull'assunzione che i dati osservati  $y$  siano una realizzazione di un vettore casuale  $Y$ , la cui distribuzione di probabilità è ignota. L'obiettivo dell'inferenza è ricostruire la distribuzione di  $Y$  sulla base dei dati. In questo capitolo vengono presentati, in linea generale, i concetti teorici sottostanti all'inferenza basata sulla verosimiglianza che saranno utilizzati nel seguito del lavoro.

### 1.1 Modelli statistici: definizioni

Sia  $y$  un vettore di realizzazioni, un modello statistico  $\mathcal{F}$  è un insieme di funzioni di probabilità o, nel caso continuo, di funzioni di densità di probabilità  $p_Y(y; \theta)$ , dove  $\theta$  è un parametro reale, possibilmente vettoriale. In letteratura si specifica con

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

dove  $\theta$  rappresenta una qualche quantità caratterizzante l'esperimento casuale e assume valori in  $\Theta$ , detto spazio parametrico. Nei modelli statistici parametrici  $\Theta \subseteq \mathbb{R}^d$ , per un qualche  $d \in \mathbb{N}^+$ . Il modello  $\mathcal{F}$  è detto correttamente specificato se  $p^0(\cdot) \in \mathcal{F}$ , dove  $p^0(y)$  è la vera e ignota densità di  $Y$ , variabile aleatoria di cui i dati  $y$  si assumono realizzazione. In tal caso, il valore  $\theta_0$  tale che  $p_Y(y; \theta_0) = p^0(y)$  si dice vero valore del parametro. Il pa-

parametro  $\theta$  è **identificabile** se esiste una corrispondenza biunivoca tra  $\Theta$  e le distribuzioni in  $\mathcal{F}$ .

## 1.2 Stimatori e proprietà

Sia  $\mathcal{F}$  un modello statistico parametrico per i dati  $y$ , con funzione del modello  $p_Y(y; \theta)$ , con  $\theta \in \Theta \subseteq \mathbb{R}^p$  e  $\mathcal{Y}$  spazio campionario. L'obiettivo della stima puntuale è individuare il valore di  $\theta$  che meglio di altri giustifica i dati osservati  $y$ . Si utilizza un'applicazione  $\hat{\theta} : \mathcal{Y} \rightarrow \Theta$  definita dallo spazio campionario allo spazio parametrico, che corrisponde ad ogni elemento di  $y \in \mathcal{Y}$  un valore in  $\theta \in \Theta$ . Il valore  $\hat{\theta} = \hat{\theta}(y)$  è detto stima di  $\theta$ . Il confronto tra stimatori alternativi può essere basata sullo studio della loro distribuzione campionaria e su alcune loro proprietà desiderabili.

- Uno stimatore  $\hat{\theta}$  si dice **non distorto** per  $\theta$  se  $E_\theta(\hat{\theta}) = \theta$  per ogni  $\theta$ . La distorsione di uno stimatore può essere definita come  $b(\theta) = E_\theta(\hat{\theta}) - \theta$ . Uno stimatore non distorto ha la propria distribuzione centrata sul parametro  $\theta$  che si vuole stimare.
- L'**errore quadratico medio** di uno stimatore  $\hat{\theta}$  è la quantità  $MSE_\theta(\hat{\theta}) = E_\theta(\hat{\theta} - \theta)^2$ , che può essere scomposta in  $MSE_\theta(\hat{\theta}) = Var_\theta(\hat{\theta}) + b(\theta)^2$ . Se  $\hat{\theta}_1$  e  $\hat{\theta}_2$  sono due stimatori del parametro  $\theta$ , lo stimatore  $\hat{\theta}_1$  è più efficiente dello stimatore  $\hat{\theta}_2$  se  $MSE_\theta(\hat{\theta}_1) < MSE_\theta(\hat{\theta}_2)$  per ogni  $\theta$ .
- Uno stimatore non distorto  $\hat{\theta}$  è **efficiente** tra gli stimatori non distorti per  $\theta$  in un modello  $\mathcal{F}$  se si ha che  $Var_\theta(\hat{\theta}) < Var_\theta(\bar{\theta})$  per ogni stimatore non distorto  $\bar{\theta}$ .

Alcune proprietà asintotiche di uno stimatore sono valide quando  $n \rightarrow \infty$ . Assumendo che  $y = (y_1, \dots, y_n)$  sia un campione casuale semplice di numerosità  $n$ , è auspicabile che il comportamento degli stimatori migliorino con l'aumentare della numerosità campionaria.

- Uno stimatore  $\hat{\theta}$  si dice **asintoticamente non distorto** per  $\theta$  se all'aumentare di  $n$  la distorsione tende a 0:

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



- Uno stimatore  $\hat{\theta}$  si dice **consistente in media quadratica** per  $\theta$  se all'aumentare di  $n$  l'errore quadratico medio tende a 0:

$$\lim_{n \rightarrow \infty} MSE_{\theta}(\hat{\theta}) = \lim_{n \rightarrow \infty} E_{\theta}(\hat{\theta} - \theta)^2 = 0, \quad \theta \in \Theta.$$

- Uno stimatore  $\hat{\theta}$  si dice **consistente** per  $\theta$  se, per ogni  $\varepsilon > 0$  fissato risulta,

$$\lim_{n \rightarrow \infty} \Pr_{\theta}(|\hat{\theta} - \theta| < \varepsilon) = 1, \quad \theta \in \Theta.$$

Uno stimatore consistente in media quadratica, è anche consistente.

- Una successione di variabili casuali  $(Y_n)_{n \in \mathbb{N}}$  con funzione di ripartizione  $P_n$  si dice convergere in distribuzione alla variabile casuale  $Y$  con funzione di ripartizione  $P$ , cioè  $Y_n \xrightarrow{d} Y$  se  $\lim_{n \rightarrow \infty} P_n(y) = P(y)$  esiste per ogni  $y \in \mathbb{R}$  in cui  $P$  è continua. Uno stimatore  $\hat{\theta}$  di  $\theta$  si dice **asintoticamente normale** se al divergere di  $n$  lo stimatore standardizzato converge in distribuzione alla normale standard  $N(0, 1)$ :

$$\frac{\hat{\theta} - E_{\theta}(\hat{\theta})}{\sqrt{\text{Var}_{\theta}(\hat{\theta})}} \xrightarrow{d} Z \sim N(0, 1), \quad \theta \in \Theta.$$

## 1.3 Funzione di verosimiglianza

Sia  $\mathcal{F}$  un modello statistico parametrico correttamente specificato per i dati  $y$  e con funzione del modello  $p_Y(y; \theta)$ , con  $\theta \in \Theta \subseteq \mathbb{R}^p$ . Si consideri  $p_Y(y; \theta)$  come funzione solamente di  $\theta$ , con  $y$  fissato al valore osservato. La funzione  $L : \Theta \rightarrow \mathbb{R}^+$  definita da

$$L(\theta) = c(y)p_Y(y; \theta) \propto p_Y(y; \theta) \tag{1.1}$$

è detta funzione di verosimiglianza di  $y$ , per  $\theta \in \Theta$ . Sia nel caso continuo che nel caso discreto, la funzione di verosimiglianza segnala che, alla luce delle osservazioni, i valori più credibili del parametro sono quelli per cui  $L(\theta)$  è prossima al suo valore massimo.

Spesso, per le procedure di inferenza, risulta utile operare con la **funzione di log-verosimiglianza**

$$l(\theta) = \log L(\theta),$$

con  $l(\theta) = -\infty$  se  $L(\theta) = 0$ .

Nel caso di campionamento casuale semplice con numerosità  $n$  o di osservazioni indipendenti, aventi come distribuzione marginale  $p_{Y_i}(y_i; \theta)$ , la (1.1) si può scrivere come

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

e di conseguenza la log-verosimiglianza diventa

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

La funzione di verosimiglianza gode delle seguenti proprietà:

- $L(\theta)$  è **invariante rispetto a trasformazioni biunivoche di  $y$** . Per ciò, se invece di osservare  $y$  si è osservato  $t = t(y)$ , con  $t(\cdot)$  una trasformazione biunivoca con inversa  $y = y(t)$ , allora  $L(\theta; y) = L(\theta; y(t))$ .
- $L(\theta)$  è **invariante rispetto a riparametrizzazioni di  $\mathcal{F}$** . Data la riparametrizzazione definita dalla trasformazione biunivoca  $\omega(\theta)$  con inversa  $\theta(\omega)$ , le procedure inferenziali basate su  $\omega \in \Omega$  risultano equivalenti a quelle basate sulla parametrizzazione originale  $\theta \in \Theta$ , cioè:

$$L^\Omega(\omega) = L^\Theta(\theta(\omega)),$$

dove  $L^\Omega(\omega)$  e  $L^\Theta(\theta(\omega))$  rappresentano rispettivamente la verosimiglianza nella parametrizzazione  $\omega$  e  $\theta$  rispettivamente.

## 1.4 Quantità di verosimiglianza e proprietà

Si definisce, **stima di massima verosimiglianza** di  $\theta$ , un valore  $\hat{\theta}$  tale per cui  $L(\hat{\theta}) \geq L(\theta)$ ,  $\forall \theta \in \Theta$ . Per determinare la stima di massima verosimiglianza di  $\theta$  risulta equivalente massimizzare la funzione  $l(\theta)$ . Se  $\hat{\theta}(Y)$  è unico, allora la variabile casuale  $\hat{\theta} = \hat{\theta}(Y)$  è detta stimatore di massima verosimiglianza. Tale stimatore gode della proprietà di **equivarianza rispetto a riparametrizzazioni**. Ovvero, se  $\omega$  costituisce una riparametrizzazione di  $\mathcal{F}$ , si avrà che  $\hat{\omega} = \omega(\hat{\theta})$  e  $\hat{\theta} = \theta(\hat{\omega})$ .

### 1.4.1 Funzione *score*

Si definisce **funzione *score***, il vettore delle derivate prime della funzione di log-verosimiglianza

$$l_{\theta}(\theta) = (l_1, \dots, l_p)^T = \left( \frac{\partial l(\theta)}{\partial \theta_1}, \dots, \frac{\partial l(\theta)}{\partial \theta_p} \right)^T. \quad (1.2)$$

Sotto condizioni di regolarità (si veda Azzalini, 2001, Paragrafo 3.3.3), la stima di massima verosimiglianza è individuata come unica soluzione dell'equazione

$$l_{\theta}(\theta) = 0,$$

chiamata equazione di verosimiglianza. Se l'equazione di verosimiglianza non contempla soluzioni analitiche, algoritmi iterativi come Newton-Raphson (Pace e Salvan, 1996, Paragrafo 4.2) possono essere utilizzati per risolvere numericamente l'equazione.

### 1.4.2 Informazione osservata e informazione attesa

Si definisce **matrice di informazione osservata** la matrice  $p \times p$  delle derivate seconde di  $l(\theta)$  cambiate di segno,

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

Infine, si dice **informazione attesa** la matrice  $p \times p$  del valore atteso delle derivate seconde di  $l(\theta)$  cambiate di segno

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

dove  $E_{\theta}(\cdot)$  indica il valore atteso rispetto alla distribuzione di  $p_Y(y; \theta)$ . L'informazione attesa viene spesso utilizzata per esprimere la varianza della distribuzione asintotica di  $\hat{\theta}$ . L'informazione attesa, che dipende dal parametro ma non dal campione, è una misura dell'informazione portata da un generico campione per il dato esperimento, mentre l'informazione osservata, che dipende solo dal campione, misura l'informazione portata dal campione osservato.

## 1.5 Risultati distributivi

Sotto le stesse condizioni di regolarità in Azzalini (2001, Paragrafo 3.3.3), valgono alcuni risultati distributivi asintotici del prim'ordine per lo stimatore  $\hat{\theta}$  e per altre importanti quantità. In particolare, assumendo che  $n \rightarrow \infty$  e che  $\theta$  sia il vero valore del parametro, vale l'approssimazione normale per la distribuzione dello stimatore di massima verosimiglianza

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

Nella distribuzione limite,  $i(\theta)$  può essere sostituita dalla stima  $i(\hat{\theta})$  oppure  $j(\hat{\theta})$ .

Quando la funzione *score* (1.2) è valutata al vero valore del parametro  $\theta$  si ottiene

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

A partire da questi risultati asintotici, si possono definire le seguenti statistiche asintoticamente equivalenti

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

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

$$W_u(\theta) = l_\theta(\theta)^\top i(\theta)^{-1} l_\theta(\theta),$$

definite rispettivamente **log-rapporto di verosimiglianza**, **test di Wald** e **test score** e aventi distribuzione asintotica  $\chi_p^2$ , dove  $p$  indica la dimensione del parametro  $\theta$ .

Nel caso in cui  $\theta$  sia un parametro scalare ( $p = 1$ ), le versioni unilaterali di tali statistiche sono rispettivamente:

$$r(\theta) = \text{sign}(\hat{\theta} - \theta) \sqrt{W(\theta)},$$

che prende il nome di **radice con segno del rapporto di verosimiglianza**,

$$r_e(\theta) = (\hat{\theta} - \theta) \sqrt{j(\hat{\theta})},$$

$$r_u(\theta) = l_\theta(\theta) \sqrt{i(\theta)^{-1}},$$

che si distribuiscono asintoticamente come una  $N(0, 1)$ , per  $\theta \in \Theta$ .

## Capitolo 2

# Modelli lineari generalizzati e sovra-dispersione

In molte applicazioni statistiche è di interesse studiare la relazione tra una variabile risposta e altre variabili definendo un modello di regressione. Quando i valori della risposta  $y_1, \dots, y_n$  rappresentano il risultato di conteggi in cui totale non è prefissato, un modello statistico naturale si basa sulla **distribuzione di Poisson**. La distribuzione di Poisson assume l'identità tra media e varianza della risposta, tuttavia i dati possono presentare **sovra-dispersione**, soprattutto per piccoli campioni, ovvero la varianza empirica è superiore a quella teorica. La **distribuzione binomiale negativa** è una naturale estensione della distribuzione di Poisson e la sua specificazione contempla la presenza di un parametro di dispersione. In questo capitolo sono presentati i dettagli teorici per la specificazione dei modelli di regressione basati sulle distribuzioni binomiale, Poisson, gamma e binomiale negativa.

### 2.1 Famiglie di dispersione esponenziale

La funzione di probabilità, o di probabilità, di una variabile aleatoria  $Y$  appartiene alla **famiglia di dispersione esponenziale** univariata se è esprimibile nella forma

$$p(y; \theta, \phi) = \exp\left\{\frac{\theta y - b(\theta)}{a(\phi)} + c(y, \phi)\right\}, \quad (2.1)$$

con  $y \in S \subseteq \mathbb{R}$ , detto supporto di  $Y$ , e  $a(\phi) > 0$ . Il parametro  $\theta \in \Theta \subseteq \mathbb{R}$  è detto parametro naturale, mentre  $\phi$  è detto parametro di dispersione. Se si specifica l'espressione delle funzioni  $a(\cdot)$  e  $b(\cdot)$ , si ottiene una specifica distribuzione di dispersione esponenziale. Indicando con  $\theta(\mu)$  la funzione inversa che esprime  $\theta$  in funzione di  $\mu$ , sono valide le seguenti espressioni

$$E(Y) = \mu(\theta) = b'(\theta),$$

$$Var(Y) = a(\phi)b''(\theta)|_{\theta=\theta(\mu)} = a(\phi)v(\mu),$$

dove  $b'(\cdot)$  e  $b''(\cdot)$  indicano rispettivamente la derivata prima e seconda di  $b(\theta)$  e  $v(\mu)$  è detta funzione di varianza. Le espressioni appena enunciate definiscono una riparametrizzazione  $(\mu, \phi)$  di una famiglia di dispersione esponenziale. Per indicare la distribuzione di  $Y$  si può utilizzare la seguente notazione compatta

$$Y \sim DE_1(\mu, a(\phi)v(\mu)). \quad (2.2)$$

## 2.2 Modelli lineari generalizzati

I **modelli lineari generalizzati** per risposte univariate (Nelder e Wedderburn, 1972) furono introdotti come estensione del modello di regressione lineare normale per trattare risposte con distribuzione diversa dalla normale.

Sia  $y = (y_1, \dots, y_n)$  l'osservazione della variabile risposta relativa a  $n$  unità statistiche. Si assume che  $y_i$  sia realizzazione di una variabile aleatoria  $Y_i$ ,  $i$ -esima componente della variabile risposta  $Y = (Y_1, \dots, Y_n)$ . Per ogni unità si ha a disposizione il valore di  $p$  variabili esplicative  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ ,  $i = 1, \dots, n$ . Ogni modello lineare generalizzato è definito da tre caratteristiche:

- **Componente aleatoria:** le variabili  $Y_1, \dots, Y_n$  sono variabili aleatorie indipendenti con media  $E(Y_i) = \mu_i$  e varianza  $Var(Y_i) = a_i(\phi)v(\mu_i)$  con distribuzione appartenente alla famiglia di dispersione esponenziale univariata, usando la notazione in (2.2)

$$Y_i \sim DE_1(\mu_i, a_i(\phi)v(\mu_i)), \quad i = 1, \dots, n$$

- **Predittore lineare:** per un vettore  $\beta = (\beta_1, \dots, \beta_p)^T$  di coefficienti di regressione e una matrice del modello  $X$ ,  $n \times p$ , il predittore lineare è  $\eta = X\beta$  con componenti  $\eta_i = x_i\beta$ , dove  $x_i$  è l' $i$ -esima riga di  $X$ .
- **Funzione di legame:** è una funzione nota e invertibile  $g(\cdot)$  che mette in relazione la media  $\mu_i$  al predittore  $\eta_i$  tramite l'equazione  $g(\mu_i) = x_i\beta = \eta_i$ . La funzione di legame canonica è quella che pone il predittore lineare uguale al parametro naturale della distribuzione di dispersione esponenziale di  $Y_i$ .

## 2.3 Distribuzione binomiale e modello di regressione logistica

In teoria della probabilità la **distribuzione binomiale** è una distribuzione di probabilità discreta che descrive il numero di successi in un processo di Bernoulli, ovvero la variabile aleatoria  $Y = X_1 + X_2 + \dots + X_m$  che somma  $m$  variabili aleatorie indipendenti di uguale distribuzione di Bernoulli  $\mathcal{B}(\pi)$ . La distribuzione binomiale  $\mathcal{B}(m, \pi)$  è caratterizzata da due parametri:

- $m$  il numero di prove effettuate;
- $\pi$  la probabilità di successo della singola prova di Bernoulli  $X_i$  (con  $0 \leq \pi \leq 1$ )

Sia  $Y \sim \mathcal{B}(m, \pi)$ , la sua funzione di probabilità è

$$p_Y(y; \pi) = \binom{m}{y} \pi^y (1 - \pi)^{m-y} \quad y = 0, 1, \dots, m. \quad (2.3)$$

La media e la varianza della variabile casuale sono

$$E_\pi(Y) = m\pi, \quad V_\pi(Y) = m\pi(1 - \pi).$$

Sia  $S = \frac{y}{m}$  una variabile casuale sul supporto  $\{0, 1/m, 2/m, \dots, 1\}$ , la sua funzione di probabilità è

$$p(s; m, \pi) = \binom{m}{ms} \exp \left\{ ms \log \left( \frac{\pi}{1 - \pi} \right) + m \log (1 - \pi) \right\},$$

che è esprimibile come la (2.1) nella forma

$$p(s; m, \pi) = \exp \left\{ \frac{s\theta - \log [1 + \exp(\theta)]}{1/m} + \log \binom{m}{ms} \right\},$$

dove  $\theta = \log \frac{\pi}{1-\pi}$ ,  $b(\theta) = \log(1 + e^\theta)$ ,  $a(\phi) = 1/m$  e  $c(s; \phi) = \log \binom{m}{ms}$ .

Per dati di conteggio con numerosità complessiva prefissata, un possibile modello statistico è il **modello di regressione logistica**. Il modello appartiene alla classe dei modelli lineari generalizzati (si veda Paragrafo 2.2) e assume che  $y_1, \dots, y_n$  siano realizzazioni di variabili casuali indipendenti  $Y_1, \dots, Y_n$  tali che  $Y_i \sim DE_1(\pi_i, \frac{\pi_i(1-\pi_i)}{m_i})$ . La probabilità di successo di  $Y_i$  è determinata da un insieme di  $p$  variabili  $x_j$ ,  $j \in 1, \dots, p$ . Nel caso si utilizzi la funzione di legame logistica, l'espressione che lega le quantità è

$$\log \frac{\pi_i}{1 - \pi_i} = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} = x_i \beta = \eta_i.$$

I parametri  $\beta_1, \beta_2, \dots, \beta_p$  devono essere stimati attraverso i dati. Sotto ipotesi di indipendenza e utilizzando la funzione di legame canonica, la funzione di verosimiglianza è

$$l(\beta) = \sum_{i=1}^n m_i (y_i x_i \beta - \log(1 + \exp(x_i \beta))). \quad (2.4)$$

La stima di massima verosimiglianza  $(\hat{\beta}_1, \dots, \hat{\beta}_p)$  è ottenuta come soluzione dell'equazioni di verosimiglianza

$$\frac{\partial l(\beta)}{\partial \beta_r} = \sum_{i=1}^n m_i x_{ir} \left( y_i - \frac{\exp(x_i \beta)}{1 + \exp(x_i \beta)} \right) = 0, \quad r \in \{1, \dots, p\},$$

o alternativamente massimizzando la (2.4) rispetto ai parametri  $(\beta_1, \dots, \beta_p)$ .

Le derivate parziali fino al terzo ordine sono

$$\begin{aligned} \frac{\partial^2 l(\beta)}{\partial \beta_r \partial \beta_s} &= \sum_{i=1}^n -\frac{m_i x_{ir} x_{is} \exp(x_i \beta)}{(1 + \exp(x_i \beta))^2}, \quad r, s \in \{1, \dots, p\}, \\ \frac{\partial^3 l(\beta)}{\partial \beta_r \partial \beta_s \partial \beta_t} &= \sum_{i=1}^n \frac{m_i x_{ir} x_{is} x_{it} \exp(x_i \beta) (\exp(x_i \beta) - 1)}{(1 + \exp(x_i \beta))^3}, \quad r, s, t \in \{1, \dots, p\}. \end{aligned}$$



L'espressione dell'informazione attesa è

$$i(\beta) = \begin{bmatrix} \sum_{i=1}^n \frac{m_i x_{i1} x_{i1} e^{x_i \beta}}{(1+e^{x_i \beta})^2} & \cdots & \sum_{i=1}^n \frac{m_i x_{ip} x_{i1} e^{x_i \beta}}{(1+e^{x_i \beta})^2} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n \frac{m_i x_{ip} x_{ip} e^{x_i \beta}}{(1+e^{x_i \beta})^2} & \cdots & \sum_{i=1}^n \frac{m_i x_{ip} x_{ip} e^{x_i \beta}}{(1+e^{x_i \beta})^2} \end{bmatrix}.$$

## 2.4 Distribuzione e modello di regressione Poisson

La **distribuzione di Poisson** è una distribuzione di probabilità discreta che esprime le probabilità per il numero di incidenze o di arrivi che occorrono successivamente ed indipendentemente in un dato intervallo di tempo. Sia  $Y \sim P(\mu)$ , la sua funzione di probabilità è

$$p_Y(y; \mu) = \frac{e^{-\mu} \mu^y}{y!} \quad y = 0, 1, \dots \quad (2.5)$$

La media e la varianza della variabile casuale sono

$$E_\mu(Y) = \mu, \quad V_\mu(Y) = \mu.$$

La funzione di probabilità di  $Y$  appartiene ad una famiglia di dispersione esponenziale infatti la (2.5) può essere scritta come

$$p_Y(y; \mu) = \exp\left\{y \log(\mu) - \mu - \log(y!)\right\}.$$

Utilizzando la parametrizzazione canonica  $\theta = \log(\mu)$ , la funzione di densità può essere riscritta come la (2.1)

$$p_Y(y; \theta) = \exp\left\{y\theta - e^\theta - \log(y!)\right\}$$

con  $b(\theta) = e^\theta$ ,  $a(\phi) = 1$ ,  $c(y; \phi) = -\log(y!)$  e  $v(\mu) = \mu$ .

Per dati di conteggio con numerosità complessiva non prefissata, un modello statistico elementare è il **modello di regressione di Poisson**. Il modello appartiene alla classe dei modelli lineari generalizzati (si veda Paragrafo 2.2) e assume che  $y_1, \dots, y_n$  siano realizzazioni di variabili casuali indipendenti  $Y_1, \dots, Y_n$ , tali che  $Y_i \sim DE_1(\mu_i, \mu_i)$  per  $i \in \{1, \dots, n\}$ . La media di  $Y_i$

è determinata da un insieme di  $p$  variabili  $x_j$ ,  $j \in 1, \dots, p$ . L'espressione che lega le quantità è

$$g(E(Y_i)) = g(\mu_i) = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} = x_i \beta = \eta_i,$$

con funzione di legame canonica logaritmica.

Sotto ipotesi di campionamento casuale semplice e utilizzando la funzione di legame canonica, la funzione di verosimiglianza è

$$l(\beta) = \sum_{i=1}^n y_i x_i \beta - n e^{x_i \beta}. \quad (2.6)$$

La stima di massima verosimiglianza  $(\hat{\beta}_1, \dots, \hat{\beta}_p)$  è ottenuta come soluzione dell'equazioni di verosimiglianza

$$\frac{\partial l(\beta)}{\partial \beta_r} = \sum_{i=1}^n x_{ir} (y_i - e^{x_i \beta}) = 0, \quad r \in \{1, \dots, p\},$$

o alternativamente massimizzando la (2.6) rispetto ai parametri  $(\beta_1, \dots, \beta_p)$ .

Le derivate parziali fino al terzo ordine sono

$$\begin{aligned} \frac{\partial^2 l(\beta)}{\partial \beta_r \partial \beta_s} &= \sum_{i=1}^n -x_{ir} x_{is} e^{x_i \beta}, \quad r, s \in \{1, \dots, p\}, \\ \frac{\partial^3 l(\beta)}{\partial \beta_r \partial \beta_s \partial \beta_t} &= \sum_{i=1}^n -x_{ir} x_{is} x_{it} e^{x_i \beta}, \quad r, s, t \in \{1, \dots, p\}. \end{aligned}$$

L'espressione dell'informazione osservata è

$$i(\beta) = \begin{bmatrix} \sum_{i=1}^n -x_{i1} x_{i1} e^{x_i \beta} & \dots & \sum_{i=1}^n -x_{i1} x_{ip} e^{x_i \beta} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n -x_{ip} x_{i1} e^{x_i \beta} & \dots & \sum_{i=1}^n -x_{ip} x_{ip} e^{x_i \beta} \end{bmatrix}.$$

## 2.5 Sovradispersione

Nell'utilizzare distribuzioni ad un parametro, come per la Poisson o la binomiale, la varianza della risposta è determinata dalla media. Tuttavia in

molte applicazioni, la relazione media-varianza fallisce poichè i dati presentano una variabilità campionaria superiore a quella prevista dal modello. In tal caso vi è **sovradisersione** nei dati, si vedano Cox (1983) e Hinde e Demétrio (1998). La sovradisersione ha ripercussioni sugli errori standard, che possono essere sottostimati, falsando l'interpretazione dei risultati. Conseguentemente, numerosi modelli e associati metodi di stima sono stati proposti per gestire la sovradisersione. La sovradisersione in modelli di Poisson è discussa, per esempio, in Breslow (1984) e Lawless (1987).

In un modello minimale di campionamento casuale semplice Poisson per i dati  $(y_1, \dots, y_n)$ , in assenza di variabili concomitanti, un test per saggiare l'adeguatezza del modello di Poisson contro alternative di sovradisersione, può essere condotto utilizzando statistica di Pearson equivalente asintoticamente alla devianza nulla nel modello Poisson

$$\chi_{oss}^2 = \sum_{i=1}^n \frac{(y_i - \bar{y}_n)^2}{\bar{y}_n},$$

con distribuzione nulla approssimata  $\chi_{n-1}^2$ . L'approssimazione è ritenuta soddisfacente se  $\bar{y}_n$  è maggiore di 3. Si noti che la statistica  $\chi_{oss}^2/(n-1)$  è il rapporto tra la varianza campionaria corretta e la media campionaria ed è nota come indice di dispersione di Fisher.

## 2.6 Distribuzione e modello di regressione Gamma

La **distribuzione Gamma** è una distribuzione di probabilità continua, definita come la somma di variabili aleatorie indipendenti e identicamente distribuite con distribuzione esponenziale. A seconda delle esigenze, la distribuzione Gamma può essere parametrizzata in modi diversi, ad esempio, una comune è quella che contempla la presenza dei parametri  $(\alpha, \beta)$ , che indicano rispettivamente i parametri di forma e scala. Se  $Y \sim Ga(\alpha, \beta)$ , la sua funzione di densità è

$$p_Y(y; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}, \quad y \in \mathbb{R}^+,$$

dove  $\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt$  è la funzione Gamma di Eulero.

Nell'ambito dei modelli di regressione risulta più comodo utilizzare la parametrizzazione  $(\alpha, \mu)$ , in questo caso i parametri sono quelli di forma e di media. Se  $Y \sim Ga(\alpha, \mu)$ , la sua funzione di densità è

$$p_Y(y; \alpha, \mu) = \frac{\left(\frac{\alpha}{\mu}\right)^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\frac{\alpha y}{\mu}}, \quad y \in \mathbb{R}^+,$$

che è esprimibile come la (2.1) nella forma

$$p_Y(y; \alpha, \mu) = \exp \left\{ \alpha \left( -\frac{1}{\mu} y - \log \mu \right) + (\alpha - 1) \log y + \alpha \log \alpha - \log \Gamma(\alpha) \right\},$$

con parametro naturale  $\theta = -1/\mu = -\beta/\alpha$ ,  $\phi = 1/\alpha$ ,  $a_i(\phi) = \phi$ ,  $b(\theta) = -\log(-\theta)$  e  $c(y_i, \phi) = (1/\phi - 1) \log y_i - (1/\phi) \log \phi - \log \Gamma(1/\phi)$ . Utilizzando quest'ultima parametrizzazione, la media e la varianza della variabile casuale sono

$$E_{\alpha\mu}(Y) = \mu, \quad V_{\alpha\mu}(Y) = \frac{\mu^2}{\alpha}.$$

Per dati di natura continua, un possibile modello statistico è il **modello di regressione Gamma**. Il modello appartiene alla classe dei modelli lineari generalizzati (si veda Paragrafo 2.2) e assume che  $y_1, \dots, y_n$  siano realizzazioni di variabili casuali indipendenti  $Y_1, \dots, Y_n$  tali che  $Y_i \sim DE_1(\mu_i, \frac{\mu_i^2}{\alpha})$ , per  $i \in \{1, \dots, n\}$ . La media di  $Y_i$  è determinata da un insieme di  $p$  variabili  $x_j$ ,  $j \in 1, \dots, p$ . Nel caso si utilizzi la funzione di legame logaritmica, l'espressione che lega le quantità è

$$\log(E(Y_i)) = \log(\mu_i) = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} = x_i \beta = \eta_i.$$

Sotto ipotesi di campionamento casuale semplice e utilizzando la funzione di legame canonica, la funzione di verosimiglianza è

$$l(\beta, \alpha) = n\alpha \log(\alpha) - n \log \Gamma(\alpha) + \sum_{i=1}^n \left( (\alpha - 1) \log(y_i) - \alpha \log(e^{\beta' x_i}) - \frac{\alpha y_i}{e^{\beta' x_i}} \right). \quad (2.7)$$

La stima di massima verosimiglianza  $(\hat{\beta}_1, \dots, \hat{\beta}_p, \hat{\alpha})$  è ottenuta come soluzione dell'equazioni di verosimiglianza

$$\frac{\partial l(\beta, \alpha)}{\partial \beta_r} = \alpha \sum_{i=1}^n x_{ir} \left( \frac{y_i}{e^{\beta' x_i}} - 1 \right) = 0, \quad r \in \{1, \dots, p\},$$

$$\frac{\partial l(\beta, \alpha)}{\partial \alpha} = n(\log(\alpha) + 1) - n \frac{\partial \log \Gamma(\alpha)}{\partial \alpha} + \sum_{i=1}^n \left( \log \left( \frac{y_i}{e^{\beta' x_i}} \right) - \frac{y_i}{e^{\beta' x_i}} \right) = 0,$$

o alternativamente massimizzando la (2.7) rispetto ai parametri  $(\beta_1, \dots, \beta_p, \alpha)$ .

Le derivate parziali fino al terzo ordine sono

$$\begin{aligned}\frac{\partial^2 l(\beta, \alpha)}{\partial \beta_r \partial \beta_s} &= -\alpha \sum_{i=1}^n \frac{x_{ir} x_{is} y_i}{e^{\beta' x_i}}, \quad r, s \in \{1, \dots, p\}, \\ \frac{\partial^2 l(\beta, \alpha)}{\partial \beta_r \partial \alpha} &= -\sum_{i=1}^n x_{ir} \left(1 - \frac{y_i}{e^{\beta' x_i}}\right), \quad r \in \{1, \dots, p\}, \\ \frac{\partial^2 l(\beta, \alpha)}{\partial \alpha^2} &= \frac{n}{\alpha} - n \frac{\partial^2 \log \Gamma(\alpha)}{\partial \alpha^2}, \\ \frac{\partial^3 l(\beta, \alpha)}{\partial \beta_r \partial \beta_s \partial \beta_t} &= \alpha \sum_{i=1}^n \frac{x_{ir} x_{is} x_{it} y_i}{e^{\beta' x_i}}, \quad r, s, t \in \{1, \dots, p\}, \\ \frac{\partial^3 l(\beta, \alpha)}{\partial \beta_r \partial \beta_s \partial \alpha} &= -\sum_{i=1}^n \frac{x_{ir} x_{is} y_i}{e^{\beta' x_i}}, \quad r, s \in \{1, \dots, p\}, \\ \frac{\partial^3 l(\beta, \alpha)}{\partial \beta_r \partial \alpha^2} &= 0, \quad r \in \{1, \dots, p\}, \\ \frac{\partial^3 l(\beta, \alpha)}{\partial \alpha^3} &= -\frac{n}{\alpha^2} - n \frac{\partial^3 \log \Gamma(\alpha)}{\partial \alpha^3}.\end{aligned}$$

L'espressione dell'informazione attesa è

$$i(\beta, \alpha) = \begin{bmatrix} \alpha \sum_{i=1}^n x_{i1} x_{i1} & \dots & \alpha \sum_{i=1}^n x_{i1} x_{ip} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \alpha \sum_{i=1}^n x_{ip} x_{i1} & \dots & \alpha \sum_{i=1}^n x_{ip} x_{ip} & 0 \\ 0 & \dots & 0 & -\frac{n}{\alpha} + n \frac{\partial^2 \log \Gamma(\alpha)}{\partial \alpha^2} \end{bmatrix}.$$

## 2.7 Distribuzione e modello di regressione binomiale negativo

La **distribuzione binomiale negativa** è una distribuzione di probabilità discreta che modella il numero di fallimenti in una sequenza di prove indipendenti prima di un determinato (non casuale) numero di successi. Formalmente, data una sequenza di prove Bernoulli indipendenti e identicamente distribuite con probabilità di successo  $\pi$  e fallimento  $1 - \pi$ , si osserva la sequenza finché un predefinito numero  $\kappa$  di successi viene raggiunto. La variabile casuale  $Y$  che conta il numero di fallimenti ha distribuzione binomiale

negativa  $Y \sim \text{NB}(\pi, \kappa)$  con funzione di probabilità

$$\Pr(Y = y; \pi, \kappa) = \binom{y + \kappa - 1}{y} \pi^\kappa (1 - \pi)^y, \quad y = 0, 1, \dots \quad (2.8)$$

La media e la varianza della variabile casuale sono

$$E_{\pi\kappa}(Y) = \frac{\kappa(1 - \pi)}{\pi}, \quad V_{\pi\kappa}(Y) = \frac{\kappa(1 - \pi)}{\pi^2}.$$

Un'altra parametrizzazione nota in letteratura considera  $X$  come il numero delle prove che occorrono prima del  $\kappa$ -esimo successo. In questo caso  $X = Y + \kappa$  e la funzione di probabilità è

$$\Pr(X = x; \pi, \kappa) = \binom{x - 1}{\kappa - 1} \pi^\kappa (1 - \pi)^{x - \kappa}, \quad x = \kappa, \kappa + 1, \dots \quad (2.9)$$

. La media e la varianza della variabile casuale sono

$$E_{\pi\kappa}(X) = \frac{\kappa}{\pi}, \quad V_{\pi\kappa}(X) = \frac{\kappa(1 - \pi)}{\pi^2}.$$

La distribuzione binomiale negativa può anche essere vista come una generalizzazione della distribuzione di Poisson attraverso una mistura con una distribuzione gamma. Se  $\lambda \sim \text{Gamma}(\frac{1}{\alpha}, \frac{1}{\alpha\mu})$  e  $Y|\lambda \sim \text{Poisson}(\lambda)$ , allora la funzione di probabilità marginale di  $Y$  è

$$\begin{aligned} P(Y = y; \mu, \alpha) &= \int_0^\infty P(Y = y|\lambda)P(\lambda = \lambda)d\lambda \\ &= \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!\Gamma(\frac{1}{\alpha})} \left(\frac{1}{\alpha\mu}\right)^{\frac{1}{\alpha}} e^{-\frac{\lambda}{\alpha\mu}} \lambda^{\frac{1}{\alpha}-1} d\lambda \\ &= \frac{\left(\frac{1}{\alpha\mu}\right)^{\frac{1}{\alpha}}}{y!\Gamma(\frac{1}{\alpha})} \int_0^\infty \lambda^{\frac{1}{\alpha}+y-1} e^{-\lambda(\frac{1}{\alpha\mu}+1)} d\lambda \\ &= \frac{\Gamma(\frac{1}{\alpha} + y)}{\Gamma(y + 1)\Gamma(\frac{1}{\alpha})} \frac{\left(\frac{1}{\alpha\mu}\right)^{\frac{1}{\alpha}}}{\left(\frac{1}{\alpha\mu} + 1\right)^{\frac{1}{\alpha}+y}} \\ &= \frac{\Gamma(y + \alpha^{-1})}{\Gamma(y + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu}\right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1} + \mu}\right)^y. \end{aligned}$$

Si noti che l'ultima espressione è un caso particolare della (2.8), ponendo  $\pi = \frac{1}{1+\alpha\mu}$  e  $\kappa = \frac{1}{\alpha}$ . Nell'ambito dei modelli di regressione si preferisce utilizzare la parametrizzazione  $(\mu, \alpha)$  (si veda Lawless, 1987 e Cameron e Trivedi, 2013 Paragrafo 3.3) che indica rispettivamente i parametri di media e di precisione. La distribuzione assume una relazione quadratica tra media e varianza

$$E_{\mu\alpha}(Y) = \mu, \quad V_{\mu\alpha}(Y) = \mu + \alpha\mu^2.$$

Inoltre ponendo  $\alpha = 0$  si ricade nella distribuzione di Poisson. La distribuzione binomiale negativa non appartiene alla famiglia di dispersione esponenziale a causa della presenza di  $\alpha$ , come ovvia conseguenza il modello di regressione basato sulla distribuzione binomiale negativa non appartiene ai modelli lineari generalizzati.

Per dati di conteggio con numerosità complessiva non prefissata, un modello statistico che tiene in considerazione la presenza di sovradisersione è il **modello di regressione binomiale negativo**. Il modello assume che  $y_1, \dots, y_n$  siano realizzazioni di variabili casuali indipendenti  $Y_1, \dots, Y_n$  tali che  $Y_i \sim \text{NB}(\mu_i, \alpha)$  per  $i \in \{1, \dots, n\}$ . Tipicamente il parametro di precisione  $\alpha$  (o di dispersione  $\kappa$ ) viene assunto comune tra le variabili. La media di  $Y_i$  è determinata da un insieme di  $p$  variabili  $x_j$ ,  $j \in 1, \dots, p$ . L'espressione che lega le quantità è tipicamente

$$E_{\mu_i\alpha}(Y_i) = \mu_i = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}) = \exp((x_i)\beta) = \exp(\eta_i).$$

I parametri  $\beta_1, \beta_2, \dots, \beta_p, \alpha$  sono parametri che devono essere stimati dai dati e spesso  $x_1 = 1_n$ , in questo caso  $\beta_1$  è chiamato intercetta. Usando questa notazione, il modello di regressione binomiale per l'osservazione  $i$  è

$$\Pr(Y = y_i; \mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1}) \Gamma(y_i + 1)} \left( \frac{1}{1 + \alpha\mu_i} \right)^{\alpha^{-1}} \left( \frac{\alpha\mu_i}{1 + \alpha\mu_i} \right)^{y_i}. \quad (2.10)$$

Sotto ipotesi di campionamento casuale semplice e utilizzando la funzione di legame logaritmo, la funzione di verosimiglianza è

$$l(\beta, \alpha) = l(\beta, \alpha) = \sum_{i=1}^n \left\{ \sum_{j=0}^{y_i-1} \log(1 + j\alpha) - (y_i + \alpha^{-1}) \log(1 + \alpha\mu_i) + y_i \log(\mu_i) \right\}. \quad (2.11)$$

La stima di massima verosimiglianza  $(\hat{\beta}_1, \dots, \hat{\beta}_p, \hat{\alpha})$  è ottenuta come soluzione dell'equazioni di verosimiglianza

$$\frac{\partial l(\beta, \alpha)}{\partial \beta_r} = \sum_{i=1}^n \frac{x_{ir}(y_i - \mu_i)}{1 + \alpha\mu_i} = 0, \quad r \in \{1, \dots, p\},$$

$$\frac{\partial l(\beta, \alpha)}{\partial \alpha} = \sum_{i=1}^n \left\{ \alpha^{-2} \left( \log(1 + \alpha\mu_i) - \sum_{j=0}^{y_i-1} \frac{1}{j + \alpha^{-1}} \right) + \frac{y_i - \mu_i}{\alpha(1 + \alpha\mu_i)} \right\} = 0,$$

o alternativamente massimizzando la (2.11) rispetto ai parametri  $(\beta_1, \dots, \beta_p, \alpha)$ .

Le derivate parziali fino al terzo ordine sono

$$\frac{\partial^2 l(\beta, \alpha)}{\partial \beta_r \partial \beta_s} = - \sum_{i=1}^n \frac{\mu_i(1 + \alpha y_i) x_{ir} x_{is}}{(1 + \alpha\mu_i)^2}, \quad r, s \in \{1, \dots, p\},$$

$$\frac{\partial^2 l(\beta, \alpha)}{\partial \beta_r \partial \alpha} = - \sum_{i=1}^n \frac{\mu_i(y_i - \mu_i) x_{ir}}{(1 + \alpha\mu_i)^2}, \quad r \in \{1, \dots, p\},$$

$$\frac{\partial^2 l(\beta, \alpha)}{\partial \alpha^2} = - \sum_{i=1}^n \left\{ \sum_{j=0}^{y_i-1} \left( \frac{j}{1 + \alpha j} \right)^2 + 2\alpha^{-3} \log(1 + \alpha\mu_i) - \frac{2\alpha^{-2}\mu_i}{1 + \alpha\mu_i} - \frac{(y_i + \alpha^{-1})\mu_i^2}{(1 + \alpha\mu_i)^2} \right\},$$

$$\frac{\partial^3 l(\beta, \alpha)}{\partial \beta_r \partial \beta_s \partial \beta_t} = \sum_{i=1}^n \frac{(\alpha y_i + 1)\mu_i(\alpha\mu_i - 1) x_{ir} x_{is} x_{it}}{(\alpha\mu_i + 1)^3}, \quad r, s, t \in \{1, \dots, p\},$$

$$\frac{\partial^3 l(\beta, \alpha)}{\partial \beta_r \partial \beta_s \partial \alpha} = \sum_{i=1}^n \frac{\mu_i(\alpha y_i \mu_i + 2\mu_i - y_i) x_{ir} x_{is}}{(\alpha\mu_i + 1)^3}, \quad r, s \in \{1, \dots, p\},$$

$$\frac{\partial^3 l(\beta, \alpha)}{\partial \beta_r \partial \alpha^2} = \sum_{i=1}^n \frac{2\mu_i^2(y_i - \mu_i) x_{ir}}{(\alpha\mu_i + 1)^3}, \quad r \in \{1, \dots, p\},$$

$$\frac{\partial^3 l(\beta, \alpha)}{\partial \alpha^3} = \sum_{i=1}^n \left\{ \sum_{j=0}^{y_i-1} \frac{2j^3}{(\alpha j + 1)^3} + \frac{6 \log(\alpha\mu_i + 1)}{\alpha^4} - \frac{4\mu_i}{\alpha^3(\alpha\mu_i + 1)} - \frac{3\mu_i^2}{\alpha^2(\alpha\mu_i + 1)^2} - \frac{2\left(\frac{1}{\alpha} + y_i\right)\mu_i^3}{(\alpha\mu_i + 1)^3} \right\}.$$

L'espressione dell'informazione attesa è

$$i(\beta, \alpha) = \begin{bmatrix} \sum_{i=1}^n \frac{\mu_i x_{i1} x_{i1}}{(1 + \alpha\mu_i)} & \cdots & \sum_{i=1}^n \frac{\mu_i x_{i1} x_{ip}}{(1 + \alpha\mu_i)} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \sum_{i=1}^n \frac{\mu_i x_{ip} x_{i1}}{(1 + \alpha\mu_i)} & \cdots & \sum_{i=1}^n \frac{\mu_i x_{ip} x_{ip}}{(1 + \alpha\mu_i)} & 0 \\ 0 & \cdots & 0 & i_{\alpha\alpha}(\beta, \alpha) \end{bmatrix},$$



dove

$$i_{\alpha\alpha}(\beta, \alpha) = \sum_{i=1}^n \left\{ E \left( \sum_{j=0}^{y_i-1} \left( \frac{j}{1+\alpha j} \right)^2 \right) + 2\alpha^{-3} \log(1 + \alpha\mu_i) - \frac{2\alpha^{-2}\mu_i}{1 + \alpha\mu_i} - \frac{(\mu_i + \alpha^{-1})\mu_i^2}{(1 + \alpha\mu_i)^2} \right\}.$$

La quantità  $E \left( \sum_{j=0}^{y_i-1} \left( \frac{j}{1+\alpha j} \right)^2 \right)$  viene calcolata per via numerica, sfruttando la definizione di valore atteso

$$E \left( \sum_{j=0}^{y_i-1} \left( \frac{j}{1+\alpha j} \right)^2 \right) = \sum_{y_i=0}^{+\infty} \left( p_{Y_i}(y_i; \mu_i, \alpha) \sum_{j=0}^{y_i-1} \left( \frac{j}{1+\alpha j} \right)^2 \right).$$

Per ottenere un'approssimazione viene scelto un intervallo  $[0, y_{max}]$  tale che  $p_{Y_i}(y_i; \mu_i, \alpha) \sum_{j=0}^{y_i-1} \left( \frac{j}{1+\alpha j} \right)^2 \approx 0$ ,  $\forall y_i > y_{max}$ . Per ogni valore appartenente all'intervallo vengono calcolate le grandezze  $p_{Y_i}(y_i; \mu_i, \alpha) \sum_{j=0}^{y_i-1} \left( \frac{j}{1+\alpha j} \right)^2$ , che successivamente vengono sommate al fine di ottenere l'approssimazione finale.

# Capitolo 3

## Metodi empirici di riduzione della distorsione

### 3.1 Riduzione della distorsione

La **riduzione della distorsione** è una tematica che coinvolge la ricerca dai primi giorni della letteratura statistica. Come notato in Kosmidis (2014), la maggior parte dei metodi partono da uno stimatore  $\hat{\theta}$  e producono, direttamente o indirettamente, uno stimatore  $\tilde{\theta}$  di un parametro ignoto  $\theta$  tramite la risoluzione di un'equazione del tipo

$$\hat{\theta} - \tilde{\theta} = B_G(\bar{\theta}), \quad (3.1)$$

rispetto a  $\bar{\theta}$ . Tipicamente  $G$  è l'ignota distribuzione congiunta che ha generato i dati,  $B_G(\theta) = E_G(\hat{\theta} - \theta)$  è la funzione distorsione, e  $\bar{\theta}$  è il valore per il quale si assume che  $\hat{\theta}$  converga in probabilità con l'aumentare dell'informazione in possesso per l'inferenza su  $\theta$ . Le problematiche nella risoluzione dell'equazione (3.1) sono molteplici poichè da una parte  $G$  è ignota, oppure i valori attesi rispetto a  $G$  non sono disponibili in forma chiusa, dall'altra parte anche  $\bar{\theta}$  è ignoto.

I metodi di riduzione della distorsione come la **modificazione della funzione score** in Firth (1993), l'**inferenza indiretta** in Gourieroux *et al.* (1993), **bootstrap iterato** in Kuk (1995) e Guerrier *et al.* (2019) e la **riduzione della distorsione in mediana** in Kenne Pagui *et al.* (2017), assumo-

no che il modello sia correttamente e interamente specificato, nel senso che  $G$  risulta dal modello assunto per un specifico valore dei parametri. Questa assunzione è necessaria affinché si possa calcolare il valore atteso dei prodotti delle derivate parziali della funzione di log-verosimiglianza, o per generare dal modello nel caso di ricampionamento. Altri metodi di riduzione della distorsione, come **correzione della distorsione asintotica** Efron (1975), **bootstrap** (Efron e Tibshirani, 1986; Martin, 1992) e **jackknife** (Quenouille, 1956; Efron, 1982) possono essere applicati anche a modelli parzialmente specificati. In questo senso sono strumenti che hanno reso più efficaci la stima nei modelli in cui alcune assunzioni sono difficili da giustificare, oppure perchè la funzione di verosimiglianza è impraticabile o difficilmente calcolabile; si veda Wedderburn (1974) per la **quasiverosimiglianza**, Liang e Zeger (1986) per le **equazioni di stima generalizzate**, Lindsay (1988) e Varin e Vidoni (2005) per i metodi della **verosimiglianza composita**.

Un'altro criterio per classificare i metodi di riduzione della distorsione viene dato in Kosmidis (2014). Se il metodo per la riduzione della distorsione viene ottenuto sottraendo  $B_G(\bar{\theta})$  a  $\hat{\theta}$  allora il metodo opera in maniera esplicita, mentre se la (3.1) viene risolta sostituendo  $B_G(\bar{\theta})$  con  $\hat{B}_G(\tilde{\theta})$ , per uno stimatore  $\hat{B}_G$  della funzione di stima, allora lo stimatore opera in maniera implicita.

I metodi di riduzione della distorsione possono essere classificati anche in accordo al modo in cui viene ottenuta l'approssimazione della funzione di distorsione in (3.1), analiticamente o tramite ricampionamento. L'implementazione classica della correzione asintotica della distorsione e modificazione della funzione score approssimano  $B_G(\theta)$  con una funzione  $\tilde{b}(\theta)$  tale che  $B_G(\theta) = \tilde{b}(\theta) + O(n^{-3/2})$ , dove  $n$  è una misura dell'informazione in possesso per l'inferenza su  $\theta$ , tipicamente la numerosità campionaria. Dall'altra parte, jackknife, bootstrap, bootstrap iterato e inferenza diretta, generalmente approssimano la distorsione simulando campioni dal modello assunto o da uno stimatore di  $G$ , come la funzione di ripartizione empirica. Tipicamente i metodi basati sul ricampionamento sono computazionalmente più intensivi dei metodi analitici e richiedono attenzione riguardo all'approssimazione della soluzione in (3.1) poichè lo stimatore basato sulle simulazioni di  $B_G(\theta)$  non è sempre differenziabile rispetto a  $\theta$ .

Le versioni classiche della correzione della distorsione asintotica e della modificazione della funzione score in Efron (1975) e Firth (1993), rispettivamente, richiedono il calcolo di valori attesi, rispetto alla distribuzione del modello, di prodotti di derivate parziali della funzione di log-verosimiglianza. Questi valori attesi possono essere impossibili da calcolare o onerosi dal punto di vista computazionale per modelli in cui la log-verosimiglianza è impraticabile, perciò possono essere applicati a modelli relativamente semplici.

Infine, ad eccezione dell'approccio di modificazione dello score in Firth (1993), tutti i metodi di riduzione della distorsione sopra descritti richiedono la stima originale di  $\hat{\theta}$  e non possono operare senza di essa. Per questa ragione eriditano direttamente i problemi di instabilità che  $\hat{\theta}$  potrebbe avere. Per esempio, nella regressione multinomiale logistica, c'è sempre una probabilità non nulla di separazione dei dati (Albert e Anderson, 1984) che risultano in stime di massima verosimiglianza infinite. Quindi, correzione della distorsione asintotica, bootstrap, bootstrap iterato e jackknife non possono essere applicati.

## 3.2 Metodo empirico di riduzione della distorsione

In Kosmidis e Lunardon (2020) viene presentato il metodo **RBM** (*reduced bias M-estimation*) per la **riduzione della distorsione asintotica** di stimatori di tipo  $M$ , che derivano dalla massimizzazione di un'appropriata funzione obiettivo o dalla risoluzione di un'equazione di stima non distorta. Come l'approccio della modificazione dello score in Firth (1993), il metodo consiste in una modificazione additiva alla funzione di stima non distorta che è limitata in probabilità e risulta in uno stimatore con distorsione di ordine asintotico più piccolo dello stimatore di tipo  $M$  iniziale. La differenza chiave è che la modificazione empirica consiste solo delle prime due derivate dei contributi della funzione di stima, non richiedendo intensità di calcolo con valori attesi o il calcolo dello stimatore  $M$  su tutti i campioni simulati. Specificatamente alla classificazione vista nella sezione precedente:

- il metodo può essere applicato a modelli parzialmente specificati;

- usa un'approssimazione analitica della funzione di stima che dipende solo dalle prime due derivate dei contributi della funzione di stima;
- non dipende dallo stimatore originale;
- non richiede onerosità computazionale e il calcolo di valori attesi.

Nel metodo della modificazione della funzione score in Firth (1993), la risoluzione dell'equazione di stima modificata non corrisponde sempre alla massimizzazione di una funzione obiettivo. Si veda Kosmidis e Firth (2009), per le condizioni necessarie e sufficienti per l'esistenza di una verosimiglianza penalizzata per la riduzione della distorsione nei modelli lineari generalizzati. Nel metodo di riduzione della distorsione empirica in Kosmidis e Lunardon (2020) c'è sempre questa corrispondenza. Inoltre, viene mostrato che la funzione obiettivo penalizzata per la riduzione della distorsione è connessa al criterio di informazione per la selezione del modello basata sulla divergenza di Kullback-Leibler. Le funzioni dei parametri e dei dati che sono usate per la riduzione della distorsione e della selezione del modello differiscono solo di una costante nota. Questa peculiarità stabilisce per la prima volta un forte legame tra riduzione della distorsione nella stima e selezione del modello. Inoltre lo stimatore *RBM* ha la stessa distribuzione asintotica, e quindi anche la stessa efficienza, dello stimatore *M* iniziale.

### 3.2.1 Specificazione del modello

Si suppone di osservare i valori  $y_1, \dots, y_k$  di una sequenza di vettori casuali  $Y_1, \dots, Y_k$  con  $y_i = (y_{i1}, \dots, y_{ic_i})^\top \in \mathcal{Y} \subset \mathbb{R}^c$ , possibilmente con vettore di covariate  $x_1, \dots, x_k$ , con  $x_i = (x_{i1}, \dots, x_{iq_i})^\top \in \mathcal{X} \subset \mathbb{R}^{q_i}$ . Un tipico scopo della modellazione statistica è di stimare un sottovettore di un ignoto vettore dei parametri  $\theta \in \Theta \subset \mathbb{R}^p$  usando i dati  $y_1, \dots, y_k$  e  $x_1, \dots, x_k$ . Questo è comunemente ottenuto attraverso un vettore di  $p$  funzioni di stima  $\sum_{i=1}^k \psi^i(\theta) = \left( \sum_{i=1}^k \psi_1^i(\theta), \dots, \sum_{i=1}^k \psi_p^i(\theta) \right)^\top$ , dove  $\psi^i(\theta) = \psi(\theta, Y_i, x_i)$  e  $\psi_r^i(\theta) = \psi_r(\theta, Y_i, x_i)$ , ( $r = 1, \dots, p$ ). In particolare,  $\theta$  è stimato dallo stimatore di tipo  $\hat{\theta}$ , che risulta dalla soluzione del sistema di equazioni

$$\sum_{i=1}^k \psi^i(\theta) = 0_p, \quad (3.2)$$

rispetto a  $\theta$ . Esempi di metodi di stima che cadono in questa formulazione sono la quasi-verosimiglianza (Wedderburn, 1974) e le funzioni di stima generalizzate (Liang e Zeger, 1986). Un modo per derivare le equazioni di stima è attraverso una tipica assunzione sulla distribuzione di  $Y_i$ , si assume che appartenga ad una famiglia di distribuzione per cui la sua funzione di ripartizione è  $F_i(y_i|x_i, \theta)$ . Lo stimatore  $\hat{\theta}$  può quindi essere trovato come il valore che massimizza la funzione obiettivo

$$l(\theta) = \sum_{i=1}^k \log f_i(y_i|x_i, \theta) \quad (3.3)$$

rispetto a  $\theta$ , dove  $f_i(y_i|x_i, \theta)$  è una funzione di densità. Se si utilizza la funzione obiettivo (3.3), allora le funzioni di stima in (3.2) sono  $\psi^i(\theta) = \nabla \log f_i(y_i|x_i, \theta)$ , assumendo che il gradiente esista in  $\Theta$ . Esempi di metodi che utilizzano una funzione obiettivo come in (3.3) sono la massima verosimiglianza e massima verosimiglianza composta (si veda Varin *et al.*, 2011).

### 3.2.2 Assunzioni

Le assunzioni necessarie allo sviluppo del metodo della riduzione della distorsione empirica sono date in Kosmidis e Lunardon (2020) e riportate di seguito.

- A1 Consistenza:

$$\hat{\theta} \xrightarrow{p} \bar{\theta},$$

dove  $\bar{\theta}$  è tale che  $E_G(\psi^i) = 0_p$  per ogni  $i \in \{1, \dots, k\}$ , con  $\psi^i = \psi^i(\bar{\theta})$  e  $E_G(\cdot)$  indica il valore atteso rispetto all'ignota distribuzione congiunta  $G$ . Inoltre, si assume che  $\hat{\theta} - \bar{\theta} = O_p(n^{-1/2})$ .

- A2 Differenziabilità: le derivate di  $\psi_r^i(\theta)$  esistono fino al quinto ordine. In particolare,

$$l_{R_a}(\theta) = \sum_{i=1}^k \frac{\partial^{a-1} \psi_{r_1}^i(\theta)}{\partial \theta^{r_2} \dots \partial \theta^{r_a}},$$

esiste per ogni insieme  $R_a = \{r_1, \dots, r_a\}$ , con  $r_j \in \{1, \dots, p\}$  e  $a \in \{1, \dots, 5\}$ , con la convenzione che  $l_r(\theta) = \sum_{i=1}^k \psi_r^i(\theta)$ . Le componenti di  $\theta$  sono individuate dall'apice.

- A3 Ordine di grandezza asintotico delle derivate centrate della funzione di stima:

$$H_{R_a} = l_{R_a} - \mu_{R_a} = O_p(n^{1/2}),$$

dove  $\mu_{R_a} = E_G(l_{R_a})$ ,  $l_{R_a} = l_{R_a}(\bar{\theta})$  e  $\mu_{R_a} = \mu_{R_a}(\bar{\theta})$  esiste per ogni  $a \in \{1, \dots, 5\}$ .

- A4 Ordine di grandezza asintotico dei cumulanti della funzione di stima e delle sue derivate:

$$\nu_{R_{a_1}, S_{a_2}, \dots, T_{a_b}} = \begin{cases} O(n^{(b-1)/2}), & \text{se } b \text{ è dispari} \\ O(n^{b/2}), & \text{altrimenti} \end{cases},$$

dove  $\nu_{R_{a_1}, S_{a_2}, \dots, T_{a_b}} = E_G(H_{R_{a_1}} H_{S_{a_2}} \cdots H_{T_{a_b}})$  sono i cumulanti congiunti delle funzioni di stima e le loro derivate, con  $R_{a_1}, S_{a_2}, \dots, T_{a_b}$  sottoinsiemi degli interi  $a_1, a_2, \dots, a_b > 0$ .

- A5 La matrice con elementi  $\mu_{rs}$  ( $r, s = 1, \dots, p$ ) è invertibile.

L'assunzione A1 riguarda la non distorsione della funzione di stima e la consistenza dello stimatore di tipo  $M$ . La consistenza può essere dimostrata sotto deboli condizioni su  $G$  e sulla non distorsione asintotica della funzione di stima, si vedano van der Van Der Vaart (2000) e Huber e Ronchetti (2009), per la dimostrazione sulla consistenza degli stimatori di tipo  $M$ . Sotto questa assunzione,  $\hat{\theta} \xrightarrow{p} \bar{\theta}$  con un ordine di  $O(n^{-1/2})$ .

L'assunzione A2 riguarda la differenziabilità della funzione di stima rispetto agli elementi allo spazio parametrico. Le derivate vengono utilizzate in uno sviluppo stocastico di Taylor per l'ottenimento dell'aggiustamento empirico per la riduzione della distorsione della funzione di stima. L'assunzione è valida anche per funzioni di stima generiche, come quella che viene prodotta con l'approccio di stima basato sulla quasi-verosimiglianza, o con la stima basata su equazioni di stima generalizzate, o stima di massima verosimiglianza composita. L'assunzione di differenziabilità non è valida per modelli per i quali la funzione di stima o alcune delle sue componenti non sono differenziabili per qualche punto dello spazio parametrico.

Le assunzioni A3 e A4 riguardano l'esistenza dei valori attesi, rispetto alla distribuzione  $G$ , dei prodotti delle funzioni di stima e le loro derivate.

Infine, l'assunzione A5 riguarda l'invertibilità del valore atteso dello jacobiano della funzione di stima, l'assunzione è tipicamente utilizzata per i processi di stima basati sulla massima verosimiglianza (si veda per esempio Boos e Stefanski (2013)).

### 3.3 Modifica della funzione di stima per la riduzione della distorsione

#### 3.3.1 Distorsione asintotica

Sotto le assunzioni del Paragrafo 3.2.2 si può mostrare (si veda Kosmidis e Lunardon, 2020) che la distorsione di  $\hat{\theta}$  è

$$E_G(\hat{\theta} - \bar{\theta}) = O(n^{-1}). \quad (3.4)$$

Il risultato garantisce che la distorsione da  $\bar{\theta}$  converge a 0 al crescere dell'informazione in possesso per l'inferenza su  $\theta$ . Per campioni finiti in generale non è garantita la non distorsione. Usando le stesse assunzioni è possibile sviluppare l'equazione (3.4) in forma più esplicita in modo che  $E_G(\hat{\theta} - \bar{\theta}) = b(\bar{\theta}) + O(n^{-3/2})$ , dove  $b(\bar{\theta})$  dipende dai momenti congiunti della funzione di stima e dalle loro derivate sotto  $G$ . In pratica lo stimatore  $\hat{\theta}$  viene rimpiazzato con uno stimatore  $\hat{\theta} - b(\hat{\theta})$ , sulla base del fatto che sotto l'ipotesi A1  $b(\hat{\theta}) \xrightarrow{p} b(\bar{\theta})$ , con la speranza che il nuovo stimatore si comporti meglio in termini di distorsione.

Quando  $\hat{\theta}$  viene ottenuto tramite massima verosimiglianza e il modello è correttamente specificato, lo stimatore  $\hat{\theta} - b(\hat{\theta})$  ha migliori proprietà asintotiche. Per un modello correttamente specificato si intende un modello nel quale la vera e ignota distribuzione congiunta  $G(z_1, \dots, z_m)$  si suppone che appartenga ad una particolare famiglia di distribuzioni completamente specificata da  $F_i(y_i, x_i, \theta)$  quando si costruisce la (3.2). In questi casi è possibile procedere a stimare  $b(\theta)$  e successivamente calcolare  $\hat{\theta} - b(\hat{\theta})$  con i dati, perchè i valori attesi sono rispetto all'assunta distribuzione del modello. Questa è la base di alcuni famosi approcci per la riduzione della distorsione, come in Firth (1993), Kosmidis e Firth (2009) e Kenne Pagui *et al.* (2017).



In contesti più generali dove il modello può essere solo parzialmente specificato, stime di  $b(\theta)$  usando  $F_i(y_i, x_i, \theta)$  non solo non riducono la distorsione, ma possono anche inflazionarla; si veda Lunardon e Scharfstein (2017) dove si studia l'impatto dell'uso di funzioni di distorsione errate nella stima dei log-rapporti delle quote in modelli per dati panel. Inoltre, anche se il ricercatore può assumere che  $F_i(y_i, x_i, \theta)$  è correttamente specificata, ci sono contesti in cui i consueti metodi di riduzione della distorsione non possono essere utilizzati poiché  $F_i(y_i, x_i, \theta)$  potrebbe essere impossibile da calcolare o impraticabile. In questi casi, la stima e l'inferenza può essere basata su pseudo verosimiglianze o funzioni di stima.

### 3.3.2 Famiglia di modificazioni delle funzioni di stima per la riduzione della distorsione

Supponiamo che le assunzioni del Paragrafo 3.2.2 siano valide e consideriamo lo stimatore  $\tilde{\theta}$  che si ottiene come soluzione dell'equazione di stima modificata

$$\sum_{i=1}^k \psi^i(\theta) + A(\theta) = 0_p, \quad (3.5)$$

dove sia  $A(\theta) = A(\theta, Y)$  e le sue derivate rispetto a  $\theta$  sono  $O_p(1)$  al crescere di  $n$ . Usando le assunzioni A1-A3 e la notazione indiciale, con gli indici che prendono valori nell'insieme  $\{1, \dots, p\}$  e assumendo che quando un indice è ripetuto sia sottointesa la sommatoria, un'espansione simile a McCullagh (2018, Paragrafo 7.3) e Pace e Salvan (1997, Paragrafo 9.4) può essere utilizzata per mostrare che  $0_p = \sum_{i=1}^k \psi^i(\tilde{\theta}) + A(\tilde{\theta})$  intorno a  $\tilde{\theta}$  risulta in un'espansione stocastica di Taylor per  $\tilde{\theta} - \bar{\theta}$  (si veda Kosmidis e Lunardon,

2020) nella forma

$$\begin{aligned}
 \tilde{\theta}^r - \bar{\theta}^r = & H^r + H^a H_a^r + \frac{1}{2} H^a H^b \mu_{ab}^r + A^r + \\
 & + H^a H_a^b H_b^r + \frac{1}{2} H^a H^b H_c^r \mu_{ab}^c + \frac{1}{2} H^a H^b H_b^c \mu_{ac}^r + \\
 & + \frac{1}{2} H^a H^b H_a^c \mu_{cb}^r + \frac{1}{4} H^a H^b H^c \mu_{bc}^d \mu_{ad}^r + \frac{1}{4} H^a H^b H^c \mu_{ab}^d \mu_{dc}^r + \\
 & + \frac{1}{2} H^a H^b H_{ab}^r + \frac{1}{6} H^a H^b H^c \mu_{abc}^r + \\
 & + A^a H_a^r + \frac{1}{2} A^a H^b \mu_{ba}^r + \frac{1}{2} A^a H^b \mu_{ab}^r + A_a^r H^a + O_p(n^{-2}),
 \end{aligned} \tag{3.6}$$

dove  $H_{r_1 \dots r_a}^r = -\mu^{rs} H_{sr_1 \dots r_a}$ ,  $\mu_{r_1 \dots r_a}^r = -\mu^{rs} \mu_{sr_1 \dots r_a}$ , e  $A_{r_1 \dots r_a}^r = -\mu^{rs} A_{sr_1 \dots r_a}$ , con  $\mu^{rs}$  che denota la matrice con inversa  $\mu_{rs}$  (assunzione A5) e  $A_{r_1 \dots r_a} = \partial^{a-1} A_{r_1}(\theta) / \partial \theta^{r_2} \dots \partial \theta^{r_a}$ .

Prendendo il valore atteso sotto  $G$  in entrambi i membri dell'equazione (3.6) ed utilizzando l'assunzione A4, la distorsione di  $\tilde{\theta}$  è

$$E_G(\tilde{\theta}^r - \bar{\theta}^r) = -\mu^{ra} E_G(A_a) + \frac{1}{2} \mu^{ra} \mu^{bc} (\nu_{ab,c} - \mu^{de} \nu_{c,e} \mu_{abd}) + O(n^{-3/2}),$$

dove si sottointende che tutti i termini del membro di destra sono calcolati in  $\bar{\theta}$ . L'espressione della distorsione sopra implica che una possibile scelta di  $A$  in (3.5) può essere tale che

$$E_G(A_r) = \frac{1}{2} \mu^{ab} (2\nu_{ra,b} - \mu^{cd} \nu_{b,d} \mu_{rac}) + O(n^{-1/2}). \tag{3.7}$$

Questa produrrà uno stimatore con distorsione  $E_G(\tilde{\theta}^r - \bar{\theta}^r)$  più piccola, asintoticamente, della distorsione di  $\hat{\theta}$  in (3.4). Perciò, l'espressione (3.7) definisce una famiglia di modificazioni per la funzione di stima per la riduzione della distorsione.

Un candidato per  $A$  può essere un vettore che ha come  $r$ -esima componente il primo termine del membro di destra di (3.7). Se assumiamo che il modello  $F_i(y_i, x_i, \theta)$  specifica interamente  $G$  e il metodo di stima è tramite massima verosimiglianza, allora sono valide anche le identità di Bartlett  $\mu_{cd} + \nu_{c,d} = 0$  (si veda Pace e Salvani, 1996, Paragrafo 5.4.3), allora la modificazione (3.7) diventa  $\mu^{ab} (2\nu_{ra,b} + \mu_{rab}) / 2$ . L'ultima espressione coincide con la modificazione dell'equazione di stima proposta in Firth (1993).

Nei contesti dove l'ignota distribuzione  $G$  è parzialmente specificata attraverso  $F_i(y_i, x_i, \theta)$ , il valore atteso del membro di destra di (3.7) non può essere calcolato. Nel paragrafo seguente viene quindi proposta una famiglia di modificazioni che può sempre essere implementata e produce una riduzione della distorsione degli stimatori di tipo  $M$ .

### 3.4 Modificazione empirica per la correzione della distorsione

Si considerino le quantità  $l_{r|s}(\theta) = l_r(\theta)l_s(\theta)$  e  $l_{rs|t}(\theta) = l_{rs}(\theta)l_t(\theta)$  tali che  $E_G(l_{r|s}) = \nu_{r,s}$  e  $E_G(l_{rs|t}) = \nu_{rs,t}$ , con  $H_{r|s} = l_{r|s} - \nu_{r,s} = O_p(n^{1/2})$  e  $H_{rs|t} = l_{rs|t} - \nu_{rs,t} = O_p(n^{1/2})$ . Allora usando le assunzioni del Paragrafo 3.2.2 è possibile dimostrare (si veda Kosmidis e Lunardon, 2020) che l'equazione (3.7) è soddisfatta da una modificazione tale che

$$A_r(\theta) = \frac{1}{2}l^{ab}(\theta) \{2l_{ra|b}(\theta) - l^{cd}(\theta)l_{b|d}(\theta)l_{rac}(\theta)\}, \quad (3.8)$$

dove  $l^{st}(\theta)$  è l'inversa della matrice  $l_{st}(\theta)$ . La forma matriciale dell'espressione (3.8) ha come  $r$ -esimo elemento del vettore di modificazione per la correzione della distorsione empirica

$$A_r(\theta) = -\text{tr} \{j(\theta)^{-1}d_r(\theta)\} - \frac{1}{2} \text{tr} \left[ j(\theta)^{-1}e(\theta) \{j(\theta)^{-1}\}^\top u_r(\theta) \right], \quad (3.9)$$

dove  $u_r(\theta) = \sum_{i=1}^k \nabla \nabla^\top \psi_r^i(\theta)$  e  $j(\theta)$  è l'informazione osservata, che ha come  $s$ -esima riga  $-\sum_{i=1}^k \nabla \psi_s^i(\theta)$ , è supposta invertibile ma non necessariamente simmetrica. Le matrici  $p \times p$   $e(\theta)$  e  $d_r(\theta)$  corrispondono alle quantità  $l_{r|s}(\theta)$  e  $l_{rs|t}(\theta)$  e devono essere definite per poter utilizzare l'aggiustamento in (3.8).

Sotto ipotesi di indipendenza delle variabili  $Y_1, \dots, Y_k$  sotto  $G$ , allora

$$\nu_{r,s} = \sum_{i=1}^k E_G(\psi_r^i \psi_s^i) + \sum_{i \neq j} E_G(\psi_r^i) E_G(\psi_s^j). \quad (3.10)$$

L'assunzione A1 della non distorsione della funzione di stima implica che il secondo termine del secondo membro della (3.10) è nullo. Inoltre,  $l_{r|s} = \sum_{i=1}^k \psi_r^i \psi_s^i$  è tale che  $E_G(l_{r|s}) = \nu_{r,s}$  e sotto l'assunzione A3,  $H_{r|s} = l_{r|s} - \nu_{r,s} = O_p(n^{1/2})$ .

In forma matriciale  $e(\theta)$  può essere espresso come una matrice  $p \times p$  data dal prodotto matriciale  $\Psi(\theta)^\top \Psi(\theta)$ , dove  $\Psi(\theta)$  è una matrice  $n \times p$  con  $(i, s)$ -esimo elemento  $\partial\psi_r^i/\partial\theta^s$ . In altre parole, l'elemento  $(s, t)$  di  $e(\theta)$  è

$$[e(\theta)]_{st} = \sum_{i=1}^k \psi_s^i(\theta) \psi_t^i(\theta). \quad (3.11)$$

Con un'argomentazione simile si può mostrare che  $d_r(\theta)$  può essere espresso come una matrice  $p \times p$  data dal prodotto matriciale  $\tilde{\Psi}_r(\theta)^\top \Psi(\theta)$ , dove  $\tilde{\Psi}_r(\theta)$  è una matrice  $n \times p$  con  $(i, s)$ -esimo elemento  $\partial\psi_r^i/\partial\theta^s$ . In altre parole, l'elemento  $(s, t)$  di  $d_r(\theta)$  è

$$[d_r(\theta)]_{st} = \sum_{i=1}^k \left\{ \frac{\partial}{\partial\theta^s} \psi_r^i(\theta) \right\} \psi_t^i(\theta). \quad (3.12)$$

L'espressioni appena enunciate possono essere utilizzate direttamente nel caso si disponga di realizzazioni indipendenti di vettori casuali, che possono condividere componenti. Esempi di queste applicazioni sono l'equazioni di stima generalizzate in Liang e Zeger (1986) per la stima dei parametri di regressione per risposte correlate, e l'approccio basato sulla verosimiglianza composita in Varin e Czado (2010) per la stima in modelli probit multivariati con effetti casuali che tengono in considerazione la struttura longitudinale dei dati in presenza di risposta binaria o categoriale ordinale. Le espressioni di  $j^{(G)}(\theta)$ ,  $\Psi(\theta)$ ,  $\tilde{\Psi}_r(\theta)$  e  $u_r(\theta)$  per modelli di regressione logistica, modelli di regressione Poisson, modelli di regressione gamma e modelli di regressione binomiale negativo sono dati nell'Appendice A.2.

### 3.5 Distribuzione asintotica dello stimatore RBM e inferenza

Sotto le assunzioni 3.2.2, utilizzando un approccio simile a Stefanski e Boos (2002) nello sviluppo in serie stocastica di Taylor (3.6) con  $A(\theta) = 0_p$ , risulta che lo stimatore di tipo  $M$   $\hat{\theta}$  è tale che

$$Q(\bar{\theta})^{1/2}(\hat{\theta} - \bar{\theta}) \xrightarrow{d} N_p(0_p, I_p)$$

al crescere di  $n$ , dove  $Q(\theta) = V(\theta)^{-1}$  con  $V(\theta) = B(\theta)^{-1}M(\theta)\{B(\theta)^{-1}\}^\top$ . Nell'ultima espressione,  $M(\theta) = \mathbb{E}_G \left[ \sum_{i=1}^k \sum_j^k \psi^i(\theta) \{\psi^j(\theta)\}^\top \right]$  e  $B(\theta)$  è una matrice  $p \times p$  con  $r$ -esima riga  $-\sum_{i=1}^k \mathbb{E}_G \{\nabla \psi_r^i(\theta)\}$ .

Dallo sviluppo in serie stocastica di Taylor (3.6) si nota che  $\tilde{\theta} - \bar{\theta}$  e  $\hat{\theta} - \bar{\theta}$  hanno esattamente lo stesso termine  $O_p(n^{-1/2})$  nella loro espansione perchè  $A(\theta) = O_p(1)$ . Perciò, lo stimatore *RBM*  $\tilde{\theta}$  è tale che

$$Q(\bar{\theta})^{1/2}(\tilde{\theta} - \bar{\theta}) \xrightarrow{d} N_p(0_p, I_p). \quad (3.13)$$

Un'implicazione della (3.13) è che  $\hat{V}(\theta) = j(\theta)^{-1}e(\theta)\{j(\theta)^{-1}\}^\top$  calcolato in  $\theta = \tilde{\theta}$  è uno stimatore consistente della matrice di varianza covarianza di  $\tilde{\theta}$ . Da qui è possibile notare l'analogia con Stefanski e Boos (2002) nel quale viene usato uno stimatore  $\hat{V}(\hat{\theta})$  per la stima della matrice di varianza covarianza di  $\hat{\theta}$ . Inoltre l'espressione per  $\hat{V}(\theta)$  appare inalterata nel secondo termine del membro di destra dell'espressione (3.9) per la modificazione empirica per la riduzione della distorsione. Come conseguenza, il valore di  $V(\theta)$  che stima gli standard error dei parametri, è generalmente disponibile agli ultimi step del processo iterativo per risolvere le equazioni di stima modificate. In aggiunta, se il modello è correttamente specificato e  $l(\theta)$  in (3.2) è la log-verosimiglianza, allora la seconda identità di Bartlett implica che  $M(\theta) = B(\theta)$ , cioè  $Q(\theta)$  è l'informazione attesa. Come risultato, lo stimatore *RBM* è asintoticamente efficiente quanto lo stimatore di massima verosimiglianza e lo stimatore ridotto per la distorsione in media di Firth (1993).

Un'ulteriore implicazione di (3.13) riguarda le procedure inferenziali basate sulle approssimazioni asintotiche. Per i test di ipotesi e per le regioni di confidenza per modelli parametrici è possibile costruire quantità pivotali alla Wald o basate sullo score generalizzato nella forma

$$\begin{aligned} W_{(e)}(\theta) &= (\tilde{\theta} - \theta)^\top \{\hat{V}(\tilde{\theta})\}^{-1}(\tilde{\theta} - \theta), \\ W_{(s)}(\theta) &= \left\{ \sum_{i=1}^k \psi^i(\theta) + A(\theta) \right\}^\top \{e(\tilde{\theta})\}^{-1} \left\{ \sum_{i=1}^k \psi^i(\theta) + A(\theta) \right\}, \end{aligned} \quad (3.14)$$

che hanno, asintoticamente, distribuzione  $\chi_p^2$ . Queste quantità sono le dirette estensioni delle classiche quantità pivotali alla Wald e basate sullo score

generalizzato che sono utilizzate quando si utilizzano stimatori di tipo  $M$  (si veda Boos, 1992, per una discussione riguardo il test basato sullo score generalizzato quando si utilizzano stimatori di tipo  $M$ ).

### 3.6 Penalizzazione empirica della funzione obiettivo per la riduzione della distorsione

Quando la stima avviene tramite la massimizzazione di (3.3) la matrice  $j(\theta)$  è simmetrica. Allora tramite un'ulteriore ipotesi riguardo al comportamento di  $l_{r|s}$  nella differenziazione, è possibile dimostrare che la modificazione empirica in (3.8) corrisponde sempre ad una penalità additiva nella funzione obiettivo usata nel processo di stima. In particolare,  $l_{r|s}$  ammette la seguente regola di differenziazione a catena

$$\frac{\partial}{\partial \theta^a} l_{r|s} = l_{ar|s} + l_{r|as}, \tag{3.15}$$

quindi la stima ottenuta tramite la modificazione empirica per la riduzione della distorsione è formalmente equivalente alla massimizzazione di una funzione obiettivo penalizzata della forma

$$l(\theta) - \frac{1}{2} \text{tr} \{ j(\theta)^{-1} e(\theta) \}, \tag{3.16}$$

supponendo che il massimo esista. Questo risultato garantisce semplicità nell'implementazione della riduzione della distorsione per una ampia classe di modelli e metodi di stima rispetto ai metodi basati sulla modificazione della funzione di stima *score* (si veda Firth, 1993), che funziona nei casi in cui  $l(\theta)$  è una funzione di verosimiglianza di un modello correttamente specificato e non sempre garantisce l'esistenza di una funzione di verosimiglianza penalizzata per la riduzione della distorsione. La matrice  $e(\theta)$  che è stata ottenuta sotto le assunzioni di indipendenza del Paragrafo 3.4 soddisfa la condizione (3.15).

### 3.7 Riduzione della distorsione e selezione del modello basata sulla divergenza di Kullback–Leibler

Sia  $l(\theta)$  la funzione di verosimiglianza basata sul supposto modello parametrico. Takeuchi (1976) ha dimostrato che

$$-2l(\hat{\theta}) + 2 \operatorname{tr} \left\{ j(\hat{\theta})^{-1} e(\hat{\theta}) \right\} \quad (3.17)$$

è uno stimatore della divergenza di Kullback-Leibler attesa della vera e ignota distribuzione  $G$ , supposto un modello  $F$ , dove  $\hat{\theta}$  è lo stimatore di massima verosimiglianza. L' espressione (3.17) è nota come criterio di informazione di Takeuchi (TIC), che al contrario del criterio di informazione di Akaike (Akaike, 1974), è più robusto rispetto a violazioni della corretta specificazione del modello.

La selezione del modello da un insieme di modelli parametrici avviene calcolando  $\hat{\theta}$  per ogni modello e selezionando il modello con minor valore di TIC (3.17), o equivalentemente, il modello con maggior valore di

$$l(\hat{\theta}) - \operatorname{tr} \left\{ j(\hat{\theta})^{-1} e(\hat{\theta}) \right\}. \quad (3.18)$$

Un confronto diretto delle espressioni (3.18) e (3.16) rivela la stretta connessione tra riduzione della distorsione dello stimatore di massima verosimiglianza e selezione del modello. Sia la riduzione della distorsione, sia la selezione del modello basata sul TIC fanno riferimento alla stessa penalizzazione basata sulla traccia di  $\{j(\theta)^{-1}e(\theta)\}$ , ma differiscono per intensità della penalizzazione; la riduzione della distorsione è ottenuta usando metà di questa penalizzazione, mentre una valida selezione del modello richiede una penalizzazione più forte usando una volta l'intera penalità.

Come discusso nel Paragrafo 3.5, lo stimatore *RBM*  $\tilde{\theta}$  ha la stessa distribuzione asintotica di  $\hat{\theta}$ . Allora la derivazione del TIC funziona anche quando si considera lo stimatore *RBM* al posto dello stimatore di massima verosimiglianza. Come risultato, TIC e AIC calcolati in  $\tilde{\theta}$  sono asintoticamente equivalenti alla loro classica versione calcolata nella stima di massima verosimiglianza. Lo stesso è valido per lo stimatore ridotto per la distorsione di Firth (1993). In altre parole, la selezione del modello basata su TIC può

essere ottenuta selezionando il modello con maggior valore di

$$l(\tilde{\theta}) - \text{tr} \left\{ j(\tilde{\theta})^{-1} e(\tilde{\theta}) \right\}, \quad (3.19)$$

mentre per la selezione del modello basata su AIC seleziona il modello con maggior valore di  $l(\tilde{\theta}) - p$ . Le quantità in (3.19) sono disponibili una volta che (3.16) è stata massimizzata per ottenere lo stimatore *RBM*. Per attuare selezione del modello è necessario modificare da 1/2 a 1, il coefficiente associato alla traccia  $\{j(\theta)^{-1}e(\theta)\}$  dopo la massimizzazione.

Varin e Vidoni (2005) hanno sviluppato un procedura di selezione del modello dove  $l(\theta)$  è una verosimiglianza composita. Il criterio di informazione per verosimiglianze composite (CLIC) derivato ha la stessa forma funzionale del TIC in (3.17). Quindi la connessione tra selezione del modello e riduzione della distorsione esiste anche quando  $l(\theta)$  è il logaritmo di una verosimiglianza composita.

## 3.8 Implementazione

La soluzione dell'equazioni di stima modificate per la correzione della distorsione è tipicamente non disponibile in forma chiusa e procedure iterative sono utilizzate per approssimare la soluzione. Una generale procedura iterativa che può essere utilizzata deriva da una modificazione dell'algoritmo Newton-Raphson secondo il quale è previsto che l' $u$ -esima iterazione aggiorna la stima corrente  $\theta^{(u)}$  nel nuovo valore  $\theta^{(u+1)}$  tramite

$$\theta^{(u+1)} := \theta^{(u)} + a_u \left\{ j(\theta^{(u)}) \right\}^{-1} \left\{ \sum_{i=1}^k \psi^i(\theta^{(u)}) + A(\theta^{(u)}) \right\}, \quad (3.20)$$

dove  $a_u$  è una sequenza deterministica di costanti positive che possono essere utilizzate per implementare veri schemi per il controllo dell'intensità della modifica, come *step-halving*. Dall'espressione (3.20) ponendo  $a_u = 1$  e  $A(\theta) = 0 \forall \theta$  si torna al classico Newton-Raphson. Lo stimatore di massima verosimiglianza, che è trovato come soluzione dell'equazione di stima score  $\sum_{i=1}^k \psi^i(\theta) = 0_p$  è l'ovvio candidato come valore di partenza della procedura quasi Newton-Raphson. Un criterio di arresto ragionevole può essere



$|\theta^{(u+1)} - \theta^{(u)}|/a_u < \varepsilon$  o  $\left\| \sum_{i=1}^k \psi^i(\theta^{(u)}) + A(\theta^{(u)}) \right\|_1 < \varepsilon$ , per qualche  $\varepsilon > 0$ , dove  $\|\cdot\|_1$  è la norma L1.

Tipicamente, le procedure quasi Newton-Raphson ha velocità di convergenza del primo ordine per la soluzione dell'equazione di stima modificata, in confronto al secondo ordine della procedura piena ma sono generalmente più robuste. Il vantaggio dell'uso della procedura quasi Newton-Raphson riguarda il fatto che le quantità richieste per l'implementazione in (3.20) sono disponibili non appena un'implementazione della modifica empirica per la riduzione della distorsione viene eseguita. La modificazione empirica per la riduzione della distorsione dipende solo dalle derivate delle funzioni di stima, questo fatto permette implementazioni generali nelle quali  $\partial\psi_r^i(\theta)/\partial\theta_s$  e  $\partial^2\psi_r^i(\theta)/\partial\theta_s\partial\theta_t$  ( $r, s, t = 1, \dots, p$ ) possono essere calcolati tramite tecniche di differenziazione automatica. Queste derivate possono essere combinate tra loro per calcolare  $u_r(\theta)$ ,  $j(\theta)$ ,  $e(\theta)$  e  $d_r(\theta)$ . In aggiunta, applicazioni di controllo numerico per l'inversione delle matrici possono essere utilizzate nell'implementazione di (3.9).

### 3.9 Riduzione della distorsione in modelli lineari generalizzati

Si consideri una sequenza di  $n$  vettori  $(y_1, x_1^\top)^\top, \dots, (y_n, x_n^\top)^\top$ , dove  $y_i \in \mathcal{Y} \subset \mathbb{R}^p$ . Si suppone che  $y_1, \dots, y_n$  siano realizzazioni di variabili casuali  $Y_1, \dots, Y_n$ , che condizionatamente ai vettori  $x_1, \dots, x_n$  sono assunti indipendenti e distribuiti in accordo ad uno specifico modello lineare generalizzato. L' $i$ -esimo contributo alla log-verosimiglianza è

$$\log f_i(y_i | x_i, \beta, \phi) = \frac{m_i}{\phi} \{y_i \theta_i - \kappa(\theta_i) - c_1(y_i)\} - \frac{1}{2} a \left( -\frac{m_i}{\phi} \right), \quad (3.21)$$

per funzioni sufficientemente regolari  $\kappa(\cdot)$ ,  $c_1(\cdot)$ , e  $a(\cdot)$ , dove la media e varianza di  $Y_i$  è associata a  $\beta$  e  $x_i$  come  $\mu_i = d\kappa(\theta_i)/d\theta_i = h(\eta_i)$  e  $\text{var}(Y_i | x_i) = \phi v(\mu_i)/m_i$ , rispettivamente, con  $\eta_i = x_i^\top \beta$ , e una funzione sufficientemente liscia  $h(\cdot)$  ( $i = 1, \dots, n$ ). Il parametro  $\phi$  è il noto o ignoto parametro di dispersione,  $m_1, \dots, m_n$  sono i pesi osservati, e  $v(\mu_i) = d^2\kappa(\theta_i)/d\theta_i^2$  è la funzione

varianza. Modelli lineari generalizzati che ricadono in questa applicazione sono il modello di regressione logistica, i modelli di Poisson log-lineari, i modelli di regressione gamma e il modello di regressione lineare normale introdotti nel Capitolo 2.

La riduzione della distorsione dello stimatore di massima verosimiglianza  $\hat{\beta}$  e  $\hat{\phi}$  per modelli lineari generalizzati interamente specificati è stata intensamente studiata. Gli studi di riferimento sono: Cordeiro e McCullagh (1991), che hanno derivato uno stimatore in forma chiusa per la distorsione di  $\hat{\beta}$  e  $\hat{\phi}$  e che viene sottratto alle stime; Kosmidis e Firth (2009), che hanno mostrato che lo stimatore dei minimi quadrati iterati ripesati risulta dalla soluzione dell'equazione score modificata per la riduzione della distorsione in Firth (1993), e derivano una condizione necessaria e sufficiente per l'esistenza di una verosimiglianza penalizzata per la riduzione della distorsione in modelli lineari generalizzati; e Kosmidis *et al.* (2020), che hanno derivato la connessione tra riduzione della distorsione in media e in mediana nei modelli lineari generalizzati e hanno proposto un'implementazione unificata attraverso l'utilizzo di una procedura quasi *Fisher scoring*.

In contrasto ai metodi di riduzione della distorsione in modelli lineari generalizzati proposti in Kosmidis e Firth (2009), la modificazione empirica per la riduzione della distorsione eseguita alla funzione score corrisponde sempre a una verosimiglianza penalizzata funzione di  $\beta$  e  $\phi$ . In accordo con il Paragrafo 3.6, se  $\phi$  è ignoto, gli unici ingredienti richiesti nella penalizzazione sono l'informazione osservata di  $\beta$  e  $\phi$ ,  $j^{(G)}(\beta, \phi)$ , e la somma dei prodotti incrociati del gradiente di (3.21) tra le osservazioni,  $e^{(G)}(\beta, \phi)$ . La verosimiglianza penalizzata è allora

$$\sum_{i=1}^n \log f_i(y_i | x_i, \beta, \phi) - \frac{1}{2} \text{tr} \left[ \{j^{(G)}(\beta, \phi)\}^{-1} e^{(G)}(\beta, \phi) \right].$$

La forma chiusa per  $j^{(G)}(\beta, \phi)$  e  $e^{(G)}(\beta, \phi)$  sono date nell'Appendice A.1. Se  $\phi$  è fissato, come per le distribuzioni binomiale e Poisson, allora la penalità coinvolge solo il blocco  $(\beta, \beta)$  di  $j^{(G)}(\beta, \phi)$  e  $e^{(G)}(\beta, \phi)$ .

## Capitolo 4

# Applicazione di metodi di riduzione della distorsione in modelli lineari generalizzati e loro estensioni

Lo scopo di questo capitolo è di valutare l'applicazione dei metodi di riduzione della distorsione discussi nel capitolo precedente. Inizialmente sono condotte alcune brevi procedure di inferenza di verosimiglianza basate sui modelli di regressione gamma e binomiale negativo. Nel seguito i metodi di riduzione della distorsione sono confrontati in alcuni studi di simulazione, condotti in modelli di regressione gamma e binomiale negativo, volti alla verifica empirica delle proprietà possedute dagli stimatori e alla valutazione dell'efficienza computazionale dei processi di stima. Gli studi di simulazione e le applicazioni sono effettuate con R (R Core Team, 2019). Vengono riportate nell'Appendice B le funzioni utilizzate per la stima e la simulazione.

### 4.1 I dati *Ants*

I dati **Ants** (Mackisack, 1994) si riferiscono ad un esperimento riguardo le preferenze alimentari di una particolare specie di formiche onnivore, la *Iridomyrmex purpureus*. Per riprodurre ciò che accade in un parco australiano in presenza di escursionisti, sono stati preparati diversi tipi di esche costituite

da combinazioni di pane, condimento e burro. Per ogni tipo di esca si sono effettuati 2 esperimenti, per un totale di  $n = 48$  esperimenti. La Tabella 4.1 contiene le prime righe del dataset.

La variabile di interesse è il numero di formiche che raggiungono l'esca in cinque minuti **Ants**. Le variabili concomitanti sono:

- il tipo di pane: **Bread**, variabile categoriale con 4 modalità (Rye, Wholemeal, Multi-grain, White);
- il tipo di condimento: **Filling**, variabile categoriale con tre modalità (Vegemite, PeanutButter, Ham);
- l'utilizzo del burro: **Butter**, variabile dicotomica (yes, no).

**Tabella 4.1:** Struttura del dataset **Ants**

Ants	Bread	Filling	Butter
22	Rye	Vegemite	no
18	Rye	Vegemite	yes
27	Rye	PeanutButter	no
43	Rye	PeanutButter	yes
68	Rye	Ham	no
44	Rye	Ham	yes
⋮	⋮	⋮	⋮

Volendo confrontare i modelli di regressione gamma e binomiale negativo, sono stati stimati alcuni modelli e sono stati selezionati i due con l'AIC più piccolo. I modelli in questione hanno come variabile risposta **Ants** e come covariate **Filling** e **Butter**.

$$Y_i \sim \text{Ga}(\mu_i, \alpha), \quad Y_i \perp Y_j, \quad \forall i \neq j;$$

$$\log(E(Y_i)) = \log(\mu_i) = \eta_i;$$

$$\eta_i = x_i \beta = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}, \quad i = 1, \dots, n,$$

$$\begin{aligned}
Y_i &\sim \text{Bineg}(\mu_i, \alpha), \quad Y_i \perp Y_j, \quad \forall i \neq j; \\
\log(E(Y_i)) &= \log(\mu_i) = \eta_i; \\
\eta_i &= x_i \beta = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}, \quad i = 1, \dots, n.
\end{aligned}$$

dove  $Y$ ,  $x_2$ ,  $x_3$  e  $x_4$  sono rispettivamente la variabile **Ants** e le variabili indicatrici relative alle modalità di **Filling** e **Butter**. La Tabella 4.2 contiene le statistiche di sintesi dell'adattamento dei modelli. Viene riportato nel Codice 4.1 i comandi per adattare i modelli.

**Tabella 4.2: Ants.** Sintesi dell'adattamento dei modelli gamma e binomiale negativo  $Y \sim \text{Filling-Peanut} + \text{Filling-Ham} + \text{Butter}$ . I valori in parentesi sono gli standard error stimati.

Modello gamma		Modello binomiale negativo	
$\hat{\beta}_1$	3.342 (0.094)	$\hat{\beta}_1$	3.337 (0.1)
$\hat{\beta}_2$	0.217 (0.116)	$\hat{\beta}_2$	0.221 (0.121)
$\hat{\beta}_3$	0.604 (0.116)	$\hat{\beta}_3$	0.608 (0.119)
$\hat{\beta}_4$	0.238 (0.094)	$\hat{\beta}_4$	0.244 (0.097)
$\hat{\alpha}$	6.398 (1.273)	$\hat{\alpha}$	11.303 (3.231)
Devianza residua	7.697	Devianza residua	54.851
Gradi di libertà	44	Gradi di libertà	44
AIC	410.53	AIC	402.34

**Codice 4.1: Ants.** Stima dei modelli gamma e binomiale negativo  $Y \sim \text{Filling-Peanut} + \text{Filling-Ham} + \text{Butter}$

```

library(MASS)
source("generazione.R")

ants=read.table("ants.dat",header=T)
ants$Bread=as.factor(ants$Bread)
ants$Butter=as.factor(ants$Butter)
ants$Filling=as.factor(ants$Filling)
y=ants$Ant_count

```

```

glm_gamma=glm(y~ants$Filling+ants$Butter+ants$Bread, family=Gamma(link='log
  '))
glm_step=stepAIC(glm_gamma,direction = 'both')
glm_nb=glm.nb(y~ants$Filling+ants$Butter)

```

## 4.2 I dati Crabs

Si considerino i dati **Crabs**, presentati in Agresti (2015). I dati sono relativi ad uno studio sul comportamento riproduttivo dei limuli. La riproduzione dei limuli avviene in primavera, quando le maree sono alte. La femmina emerge dall'acqua con un maschio attaccato ad una spina posteriore e depone le uova in buche profonde scavate nella sabbia. Le uova vengono fecondate esternamente e, in questa fase, può accadere che anche altri maschi, detti satelliti, si raccolgano attorno alla coppia contribuendo alla fecondazione delle uova.

La variabile di interesse è il numero di satelliti della femmina  $y$ , misurata su  $n = 173$  unità statistiche. La struttura del dataset, limitata alle prime osservazioni è presentata nella Tabella 4.3.

$y$	Weight	Width	Color	Spine
8	3.05	28.3	2	3
0	1.55	22.5	3	3
9	2.30	26.0	1	1
0	2.10	24.8	3	3
4	2.60	26.0	3	3
0	2.10	23.8	2	3
⋮	⋮	⋮	⋮	⋮

**Tabella 4.3:** Struttura del dataset **Crabs**

Le variabili concomitanti sono:

- il colore della femmina: **color**, variabile categoriale con 4 modalità (1, chiaro; 2, medio; 3, abbastanza scuro; 4, scuro);
- la condizione delle spine posteriori: **spine** fattore con 3 livelli (1, entrambe in buono stato; 2, una in buono stato e una danneggiata; 3, entrambe danneggiate);
- la larghezza del carapace: **width** variabile quantitativa continua (in cm);
- il peso: **weight**, variabile quantitativa continua (in kg).

Volendo adattare un modello di regressione binomiale negativo, sono stati stimati alcuni modelli ed è stato selezionati quello con l'AIC più piccolo.

La specificazione del modello selezionato è la seguente

$$\begin{aligned}
 Y_i &\sim \text{Bineg}(\mu_i, \alpha), \quad Y_i \perp Y_j, \quad \forall i \neq j; \\
 \log(E(Y_i)) &= \log(\mu_i) = \eta_i; \\
 \eta_i &= \beta_1 + \beta_2 x_{i2}, \quad i = 1, \dots, n.
 \end{aligned}$$

dove  $Y$  è la variabile che indica il numero di satelliti e  $x_2$  è la variabile che indica il peso del granchio. Nel Codice 4.2 sono riportati i comandi per adattare il modello e nella Tabella 4.4 le statistiche di sintesi del modello selezionato

**Tabella 4.4: Crabs.** Statistiche di sintesi del modello binomiale negativo  $Y \sim \text{Weight}$ . I valori in parentesi sono gli standard error stimati.

Modello binomiale negativo	
$\hat{\beta}_1$	-0.8646 (0.4048)
$\hat{\beta}_2$	0.7603 (0.1578)
$\hat{\alpha}$	0.931 (0.168 )
Devianza residua	196.16
Gradi di libertà	171
AIC	754.64

---

**Codice 4.2: Crabs.** Stima del modello binomiale negativo  $Y \sim \text{Weight}$ .

---

```
library(MASS)
source("generazione.R")
source("EBRN.R")
source('brnb11.R')

crabs=read.table("crabs.dat",header=T)
crabs$color=as.factor(crabs$color)
colnames(crabs)=c("id","y","x1","x2","x3","x4")

glm_nb_full=glm.nb(y~., data=crabs)
glm_nb_int=glm.nb(y~1, data=crabs)
glm_nb_step=stepAIC(glm_nb_full,formula(glm_nb_int))

b=coef(glm_nb_step)
a=glm_nb_step$theta
```

---

### 4.3 Applicazione dei metodi di riduzione della distorsione

In questo Paragrafo vengono mostrate delle applicazioni di metodi di riduzione della distorsione per i dati presentati nei Paragrafi 4.1 e 4.2. In particolare, gli stimatori utilizzati per il confronto sono:

- lo stimatore di massima verosimiglianza **ML**;
- lo stimatore ottenuto dalla risoluzione della funzione score modificata per la riduzione della distorsione empirica **EBR**, proposto in Kosmidis e Lunardon (2020).



- lo stimatore ottenuto dalla risoluzione della funzione score modificata in media per la distorsione **ASmean**, proposto in Firth (1993) e Kosmidis e Firth (2009);
- lo stimatore ottenuto dalla risoluzione della funzione score modificata in mediana per la distorsione **ASmedian**, proposto Kenne Pagui *et al.* (2017);
- lo stimatore ottenuto attraverso la diretta sottrazione della distorsione stimata allo stimatore di massima verosimiglianza **correction**, proposto in Cordeiro e McCullagh (1991).

Nell'Appendice B sono forniti i due sorgenti **EBRGA.R** e **EBRNB.R** per ottenere le stime ridotte per la distorsione empirica in modelli di regressione gamma e binomiale negativo. I sorgenti per ottenere le altre stime ridotte per la distorsione in modelli di regressione gamma e binomiali negativi sono forniti rispettivamente in Kosmidis (2020) e Kenne Pagui (2020).

Per quanto riguarda il confronto tra le stime dei coefficienti  $\beta$  dei modelli di regressione, ci si attende che quella ottenuta con la massima verosimiglianza sia già essenzialmente non distorta, quindi l'effetto della modificazione potrebbe non essere evidente. Il parametro di dispersione (o di precisione) è spesso distorto, quindi ci si aspetta che l'effetto dei metodi sia più intensa nella sua stima.

Si riportano nelle Tabelle 4.5, 4.6 e 4.7 i risultati ottenuti e nel Codice 4.3 i comandi per ottenere le stime.

**Tabella 4.5:** **Ants**, stime ridotte per la distorsione nel modello gamma  $Y \sim \text{Filling-Peanut} + \text{Filling-Ham} + \text{Butter}$

		$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\alpha$
n= 48	ML	3.342	0.218	0.604	0.238	6.398
	ASmean	3.349	0.218	0.604	0.238	6.396
	ASmedian	3.348	0.218	0.604	0.238	6.397
	correction	3.349	0.218	0.6073	0.238	6.397
	EBR	3.347	0.216	0.599	0.239	5.172

**Tabella 4.6: Ants**, stime ridotte per la distorsione nel modello binomiale negativo  
 $Y \sim \text{Filling-Peanut} + \text{Filling-Ham} + \text{Butter}$

		$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\alpha$
n= 48	ML	3.3368	0.2213	0.6078	0.2436	0.0885
	ASmean	3.3430	0.2207	0.6069	0.2428	0.1031
	ASmedian	3.3413	0.2208	0.6071	0.2429	0.1031
	correction	3.3418	0.2211	0.6073	0.2434	0.0978
	EBR	3.3428	0.2189	0.6025	0.2441	0.0911

**Tabella 4.7: Crabs**, stime ridotte per la distorsione nel modello binomiale negativo  
 $y \sim \text{Weight}$

		$\beta_1$	$\beta_2$	$\alpha$
n= 173	ML	-0.8647	0.7603	1.0740
	ASmean	-0.8767	0.7686	1.1050
	ASmedian	-0.8767	0.7664	1.1048
	correction	-0.8696	0.7656	1.0885
	EBR	-0.8330	0.7494	1.0792

**Codice 4.3:** Stime ridotte per la distorsione in **Ants** e **Crabs**

```

library(MASS)
source("generazione.R")
source("EBRNB.R")
source('brnb11.R')
library(rootSolve)

crabs=read.table("crabs.dat",header=T)
crabs$color=as.factor(crabs$color)
y=crabs$y
colnames(crabs)=c("id","y","x1","x2","x3","x4")
glm_nb_full=glm.nb(y~., data=crabs)
glm_nb_int=glm.nb(y~1, data=crabs)
glm_nb_step=stepAIC(glm_nb_full,formula(glm_nb_int))
X=model.matrix(glm_nb_step)

```

```
s=single_estimation_all_method(data=list(X=X,y=y))
ants=read.table("ants.dat",header=T)
ants$Bread=as.factor(ants$Bread)
ants$Butter=as.factor(ants$Butter)
ants$Filling=as.factor(ants$Filling)
y=ants$Ant_count
glm_nb=glm.nb(y~ants$Filling+ants$Butter)
X=model.matrix(glm_nb)
s=single_estimation_all_method(list(X=X,y=y))

source("EBRGA.R")
glm_gamma=glm(y~ants$Filling+ants$Butter+ants$Bread, family=Gamma(link='log
'))
glm_step=stepAIC(glm_gamma,direction = 'both')
X=model.matrix(glm_step)
s=single_estimation_all_method(list(X=X,y=y))
```

## 4.4 Studi di simulazione

Partendo dai datasets **Ants** e **Crabs** si sono condotti degli studi di simulazione volti alla verifica empirica delle proprietà possedute dagli stimatori proposti nel Paragrafo 4.3. In particolare, per i dati Ants sono stati condotti due studi di simulazione basati sui modelli di regressione gamma e binomiale negativo, per i dati Crabs è stato condotto un solo studio di simulazione basato sul modello binomiale negativo.

Inizialmente sono stati adattati i modelli di regressione binomiale negativo e gamma selezionati nei Paragrafi 4.1 e 4.2. Per ogni modello sono state prodotte 5000 replicazioni Monte Carlo, generando dalla distribuzione corrispondente, con parametri fissati al valore assunto dalla stima di massima verosimiglianza. Per ogni campione sono state calcolate le stime con i vari metodi. Successivamente gli stimatori sono stati messi a confronto in termini di distorsione, probabilità di sottostima, errore quadratico medio e livello di copertura empirica degli intervalli alla Wald. Ci si attende che gli stimatori

proposti per la riduzione della distorsione siano migliori dello stimatore di massima verosimiglianza in termini di distorsione e probabilità di sottostima. Si riportano nelle Tabelle 4.8, 4.9 e 4.10 i risultati ottenuti e nei Codici 4.4, 4.5 e 4.6, i comandi per compiere i tre studi di simulazione.

**Tabella 4.8: Ants.** Sintesi dello studio di simulazione del modello gamma  $Y \sim \text{Filling-Peanut} + \text{Filling-Ham} + \text{Butter}$ . Vero valore del parametro  $(\beta_1, \beta_2, \beta_3, \beta_4, \alpha) = (3.3421, 0.2178, 0.6043, 0.2382, 6.3983)$ .

			$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\alpha$
n= 48	BIAS	ML	-0.0067	0.0005	0.0011	0.0008	0.8742
		ASmean	-0.0002	0.0005	0.0011	0.0008	0.8723
		ASmedian	-0.0018	0.0005	0.0011	0.0008	0.8731
		correction	-0.0007	0.0005	0.0011	0.0008	0.8726
		EBR	-0.0016	0.0004	0.0010	0.0007	0.1541
PU	ML	0.5208	0.4996	0.4956	0.4984	0.3162	
	ASmean	0.4962	0.4996	0.4956	0.4984	0.3164	
	ASmedian	0.5012	0.4996	0.4956	0.4984	0.3164	
	correction	0.4978	0.4996	0.4956	0.4984	0.3164	
	EBR	0.5022	0.4992	0.4962	0.4974	0.5104	
RMSE	ML	0.0129	0.0201	0.0200	0.0133	3.3604	
	ASmean	0.0129	0.0201	0.0200	0.0133	3.3570	
	ASmedian	0.0129	0.0201	0.0200	0.0133	3.3584	
	correction	0.0129	0.0201	0.0200	0.0133	3.3575	
	EBR	0.0130	0.0201	0.0201	0.0133	2.1120	
CI	ML	0.9340	0.9254	0.9270	0.9314	0.9448	
	ASmean	0.9372	0.9254	0.9272	0.9316	0.9446	
	ASmedian	0.9362	0.9254	0.9270	0.9316	0.9448	
	correction	0.9370	0.9254	0.9270	0.9316	0.9448	
	EBR	0.9490	0.9398	0.9424	0.9442	0.9324	

**Codice 4.4: Ants.** Studio di simulazione del modello gamma  $Y \sim \text{Filling-Peanut} + \text{Filling-Ham} + \text{Butter}$

```
library(MASS)
source("generazione.R")
```

**Tabella 4.9: Ants.** Sintesi dello studio di simulazione del modello binomiale negativo  $Y \sim \text{Filling-Peanut} + \text{Filling-Ham} + \text{Butter}$ . Vero valore del parametro  $(\beta_1, \beta_2, \beta_3, \beta_4, \alpha) = (3.337, 0.221, 0.608, 0.244, 11.30251)$ .

			$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\alpha$
n= 48	BIAS	ML	-0.0030	-0.0041	0.0004	0.0001	-0.0091
		ASmean	0.0020	-0.0043	0.0000	0.0000	0.0020
		ASmedian	0.0004	-0.0042	0.0002	0.0000	0.0020
		correction	0.0016	-0.0043	0.0000	0.0000	-0.0005
		EBR	0.0011	-0.0043	0.0000	-0.0001	-0.0012
	PU	ML	0.5164	0.5158	0.5034	0.4984	0.6850
		ASmean	0.4972	0.5162	0.5042	0.4992	0.4998
		ASmedian	0.5038	0.5156	0.5038	0.4992	0.5004
		correction	0.4992	0.5158	0.5046	0.4988	0.5396
		EBR	0.5022	0.5180	0.5028	0.4994	0.5514
	RMSE	ML	0.01007	0.01456	0.01420	0.00939	0.00058
		ASmean	0.01005	0.01454	0.01418	0.00938	0.00059
		ASmedian	0.01006	0.01455	0.01420	0.00939	0.00059
		correction	0.01005	0.01454	0.01418	0.00938	0.00057
		EBR	0.01005	0.01454	0.01419	0.00939	0.00056
CI	ML	0.9350	0.9360	0.9344	0.9354	0.8616	
	ASmean	0.9482	0.9482	0.9458	0.9486	0.8874	
	ASmedian	0.9484	0.9482	0.9458	0.9486	0.8874	
	correction	0.9460	0.9448	0.9438	0.9460	0.8828	
	EBR	0.9454	0.9452	0.9430	0.9448	0.9176	

```
source("EBRGA.R")
library(rootSolve)

ants=read.table("ants.dat",header=T)
ants$Bread=as.factor(ants$Bread)
ants$Butter=as.factor(ants$Butter)
ants$Filling=as.factor(ants$Filling)
y=ants$Ant_count
glm_gamma=glm(y~ants$Filling+ants$Butter+ants$Bread, family=Gamma(link='log
```

**Tabella 4.10: Crabs.** Sintesi dello studio di simulazione del modello binomiale negativo  $y \sim \text{Weight}$ . Vero valore del parametro  $(\beta_1, \beta_2, \alpha) = (-0.8647, 0.7603, 1.0740)$ .

			$\beta_1$	$\beta_2$	$\alpha$
n= 173	BIAS	ML	-0.0007	-0.0033	-0.0143
		ASmean	-0.0059	0.0019	0.0081
		ASmedian	-0.0071	0.0004	0.0080
		correction	-0.0053	0.0016	-0.0011
		EBR	0.0067	-0.0035	-0.0030
	PU	ML	0.5048	0.5026	0.5488
		ASmean	0.5094	0.4864	0.4948
		ASmedian	0.5108	0.4912	0.4950
		correction	0.5094	0.4866	0.5150
		EBR	0.4966	0.5022	0.5186
	RMSE	ML	0.15548	0.02357	0.02422
		ASmean	0.15565	0.02359	0.02497
		ASmedian	0.15595	0.02361	0.02496
		correction	0.15562	0.02358	0.02441
		EBR	0.15542	0.02357	0.02436
CI	ML	0.9416	0.9386	0.9296	
	ASmean	0.9432	0.9414	0.9334	
	ASmedian	0.9434	0.9418	0.9336	
	correction	0.9428	0.9404	0.9310	
	EBR	0.9414	0.9384	0.9370	

```

' ))
glm_step=stepAIC(glm_gamma,direction = 'both')
b=coef(glm_step)
a=gamma.shape(glm_step)$alpha
X=model.matrix(glm_step)
truth=c(b,a)

datasets=sim.data.gamma(truth,data=list(X=X,y=y),n=48,Nsim=5000,seed=800)
simulazioni=multi_estimation_all_method(datasets)

```

---

```
e1=elaboration(na.omit(simulazioni),truth)
```

---

**Codice 4.5: Ants.** Studio di simulazione del modello binomiale negativo  $Y \sim$   
Filling-Peanut + Filling-Ham + Butter

---

```
library(MASS)
source("generazione.R")
source("EBRNB.R")
source('brnb11.R')
ants=read.table("ants.dat",header=T)
ants$Bread=as.factor(ants$Bread)
ants$Butter=as.factor(ants$Butter)
ants$Filling=as.factor(ants$Filling)
y=ants$Ant_count
glm_nb=glm.nb(y~ants$Filling+ants$Butter)
b=coef(glm_nb)
a=glm_nb$theta
X=model.matrix(glm_nb)
truth=c(b,1/a)
datasets=sim.data(truth,data=list(X=X,y=y),n=48,Nsim=5000,seed=800)
single_estimation_all_method(list(X=datasets[[4]]$X,y=datasets[[3]]$y))
simulazioni=multi_estimation_all_method(datasets)
e1=elaboration(simulazioni,truth)
```

---

**Codice 4.6: Crabs.** Studio di simulazione del modello binomiale negativo  $y \sim$   
Weight

---

```
library(MASS)
source("generazione.R")
source("EBRNB.R")
source('brnb11.R')
crabs=read.table("crabs.dat",header=T)
crabs$color=as.factor(crabs$color)
colnames(crabs)=c("id","y","x1","x2","x3","x4")
y=crabs$y
glm_nb_full=glm.nb(y~., data=crabs)
glm_nb_int=glm.nb(y~1, data=crabs)
glm_nb_step=stepAIC(glm_nb_full,formula(glm_nb_int))
```

```

b=coef(glm_nb_step)
a=glm_nb_step$theta
X=model.matrix(glm_nb_step)
datasets=sim.data(truth,data=list(X=X,y=y),n=173,Nsim=5000,seed=800)
simulazioni=multi_estimation_all_method(datasets)
e1=elaboration(simulazioni,truth)

```

Successivamente è stato condotto uno studio di simulazione su dei dati simulati, per verificare le proprietà possedute dagli stimatori in modelli di regressione con molti parametri. Inizialmente è stata generata una variabile risposta  $y$  tale che  $y_i \sim \text{Bineg}(\alpha = 1, \pi = 0.4)$ ,  $i \in \{1, \dots, 1000\}$  e delle variabili esplicative  $x_1, \dots, x_{100}$  da una distribuzione di Poisson tali che  $x_{ij} \sim P(\lambda = j)$ ,  $i \in \{1, \dots, 1000\}$ . Utilizzando come variabile dipendente  $y$  e come variabili indipendenti  $x_1, \dots, x_{100}$ , e includendo un'intercetta, viene stimato un modello di regressione binomiale negativo ottenendo con la massima verosimiglianza le stime  $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p, \hat{\alpha})$ . I valori dei coefficienti stimati sono utilizzati per generare 500 campioni  $z_j$ ,  $j \in \{1, \dots, 500\}$ , tali che  $z_{ij} \sim \text{Bineg}(\exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_{100} x_{i100}), \hat{\alpha})$   $i \in \{1, \dots, 1000\}$ . Per ogni campione sono state calcolate le stime e successivamente gli stimatori sono stati messi a confronto in termini di distorsione, probabilità di sottostima, errore quadratico medio e livello di copertura empirica degli intervalli alla Wald. Si riportano nella Tabella 4.11 i risultati ottenuti e nel Codice 4.7 i comandi per compiere lo studio di simulazione.

Dai risultati emerge che gli stimatori per la riduzione della distorsione hanno caratteristiche preferibili rispetto allo stimatore di massima verosimiglianza. La stima di massima verosimiglianza dei coefficienti  $\beta$  è già essenzialmente non distorta e l'effetto dei metodi non è intenso (si vedano le Figure 4.1, 4.2). Le stime del parametro di dispersione  $\alpha$  con gli stimatori ridotti per la distorsione si sono mostrate migliori rispetto a quella basata sulla massima verosimiglianza, poichè hanno mostrato migliori proprietà in termini di distorsione, probabilità di sottostima, errore quadratico medio e copertura empirica degli intervalli alla Wald. In particolare il metodo analitico *ASmedian* si è mostrato più efficace in termini di probabilità di sottostima, mentre il metodo empirico *EBR* ha mostrato buoni comportamenti in termini di copertura empirica degli intervalli alla Wald (si vedano Figure 4.3 e 4.4). Gli



Tabella 4.11: Risultati dello studio di simulazione con  $n=1000$  e  $p=100$ 

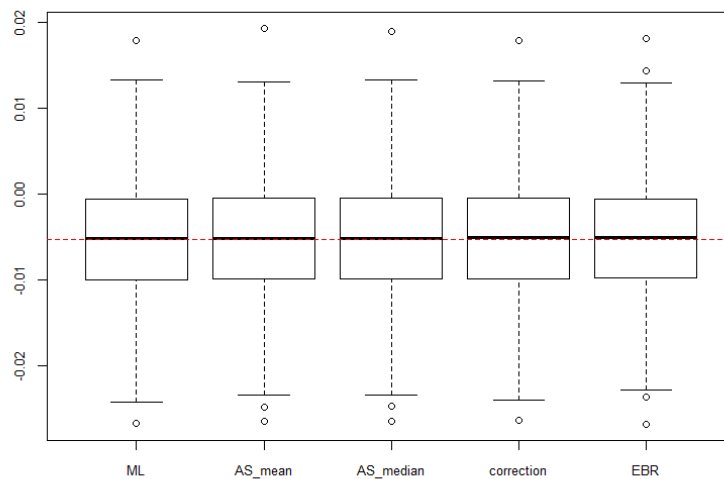
		$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$	$\beta_9$	$\beta_{10}$	$\beta_{11}$	$\beta_{12}$
BIAS	ML	-0.0119	0	-3e-04	3e-04	3e-04	0	9e-04	3e-04	2e-04	0	0	4e-04	-2e-04
	ASmean	-0.0089	0	-2e-04	1e-04	4e-04	2e-04	9e-04	4e-04	3e-04	1e-04	0	5e-04	-2e-04
	ASmedian	-0.0059	0	-2e-04	1e-04	4e-04	1e-04	9e-04	4e-04	3e-04	1e-04	1e-04	5e-04	-2e-04
	correction	-0.0034	1e-04	-2e-04	3e-04	3e-04	1e-04	9e-04	3e-04	3e-04	0	1e-04	5e-04	-2e-04
	EBR	0.0154	1e-04	-2e-04	6e-04	1e-04	0	9e-04	2e-04	2e-04	-2e-04	1e-04	4e-04	-2e-04
PU	ML	0.554	0.498	0.506	0.49	0.51	0.496	0.452	0.498	0.468	0.508	0.49	0.46	0.51
	ASmean	0.538	0.488	0.5	0.502	0.506	0.488	0.448	0.486	0.474	0.492	0.496	0.45	0.498
	ASmedian	0.522	0.49	0.502	0.502	0.508	0.492	0.446	0.488	0.472	0.502	0.494	0.454	0.506
	correction	0.51	0.488	0.498	0.488	0.51	0.496	0.452	0.496	0.464	0.506	0.49	0.46	0.508
	EBR	0.43	0.49	0.514	0.468	0.516	0.498	0.45	0.504	0.474	0.518	0.494	0.456	0.508
RMSE	ML	0.0083	0	1e-04	1e-04	0	0	0	1e-04	1e-04	1e-04	1e-04	1e-04	0
	ASmean	0.0084	0	1e-04	1e-04	0	0	0	1e-04	1e-04	1e-04	1e-04	1e-04	0
	ASmedian	0.0083	0	1e-04	1e-04	0	0	0	1e-04	1e-04	1e-04	1e-04	1e-04	0
	correction	0.0081	0	0	1e-04	0	0	0	1e-04	1e-04	1e-04	1e-04	1e-04	0
	EBR	0.0083	0	0	1e-04	0	0	0	0	0	1e-04	1e-04	1e-04	0
CI	ML	0.934	0.902	0.934	0.932	0.94	0.966	0.926	0.928	0.922	0.926	0.912	0.916	0.942
	ASmean	0.95	0.932	0.952	0.944	0.948	0.972	0.94	0.95	0.944	0.942	0.924	0.936	0.966
	ASmedian	0.95	0.932	0.952	0.946	0.948	0.972	0.942	0.95	0.946	0.944	0.924	0.936	0.968
	correction	0.946	0.92	0.944	0.938	0.948	0.97	0.936	0.948	0.94	0.936	0.924	0.93	0.964
	EBR	0.944	0.922	0.95	0.948	0.95	0.974	0.94	0.948	0.95	0.938	0.916	0.938	0.96
		$\beta_{13}$	$\beta_{14}$	$\beta_{15}$	$\beta_{16}$	$\beta_{17}$	$\beta_{18}$	$\beta_{19}$	$\beta_{20}$	$\beta_{21}$	$\beta_{22}$	$\beta_{23}$	$\beta_{24}$	$\beta_{25}$
BIAS	ML	3e-04	3e-04	-3e-04	1e-04	-9e-04	2e-04	4e-04	4e-04	1e-04	-5e-04	-2e-04	3e-04	-4e-04
	ASmean	2e-04	3e-04	-3e-04	2e-04	-9e-04	2e-04	4e-04	1e-04	1e-04	-4e-04	-1e-04	1e-04	-5e-04
	ASmedian	2e-04	3e-04	-3e-04	2e-04	-9e-04	2e-04	3e-04	1e-04	1e-04	-4e-04	-1e-04	1e-04	-5e-04
	correction	3e-04	3e-04	-3e-04	2e-04	-9e-04	3e-04	4e-04	3e-04	1e-04	-5e-04	-1e-04	2e-04	-5e-04
	EBR	4e-04	3e-04	-1e-04	2e-04	-9e-04	4e-04	4e-04	5e-04	0	-5e-04	-2e-04	3e-04	-5e-04
PU	ML	0.466	0.494	0.528	0.474	0.532	0.476	0.48	0.476	0.488	0.536	0.492	0.492	0.546
	ASmean	0.474	0.498	0.528	0.474	0.532	0.486	0.49	0.486	0.488	0.526	0.49	0.506	0.552
	ASmedian	0.472	0.5	0.528	0.472	0.534	0.482	0.492	0.488	0.486	0.536	0.494	0.502	0.552
	correction	0.47	0.496	0.528	0.468	0.536	0.476	0.486	0.478	0.49	0.536	0.486	0.496	0.548
	EBR	0.458	0.488	0.528	0.48	0.526	0.476	0.486	0.474	0.488	0.538	0.498	0.478	0.558
RMSE	ML	0	1e-04	0	1e-04	0	0	1e-04	1e-04	0	1e-04	0	1e-04	0
	ASmean	1e-04	1e-04	0	1e-04	0	0	1e-04	1e-04	0	1e-04	0	1e-04	0
	ASmedian	1e-04	1e-04	0	1e-04	0	0	1e-04	1e-04	0	1e-04	0	1e-04	0
	correction	0	1e-04	0	1e-04	0	0	1e-04	1e-04	0	0	0	1e-04	0
	EBR	0	1e-04	0	1e-04	0	0	1e-04	1e-04	0	0	0	0	0
CI	ML	0.936	0.908	0.928	0.922	0.93	0.932	0.908	0.898	0.934	0.922	0.94	0.94	0.926
	ASmean	0.95	0.926	0.938	0.936	0.948	0.954	0.928	0.92	0.944	0.94	0.96	0.952	0.948
	ASmedian	0.95	0.928	0.938	0.936	0.95	0.954	0.93	0.92	0.944	0.94	0.96	0.952	0.948
	correction	0.948	0.92	0.938	0.934	0.94	0.952	0.922	0.918	0.942	0.938	0.95	0.948	0.936
	EBR	0.946	0.908	0.944	0.938	0.944	0.942	0.926	0.918	0.94	0.94	0.96	0.95	0.944

		$\beta_{26}$	$\beta_{27}$	$\beta_{28}$	$\beta_{29}$	$\beta_{30}$	$\beta_{31}$	$\beta_{32}$	$\beta_{33}$	$\beta_{34}$	$\beta_{35}$	$\beta_{36}$	$\beta_{37}$	$\beta_{38}$
BIAS	ML	-2e-04	1e-04	-1e-04	-1e-04	-3e-04	-6e-04	-7e-04	-1e-04	5e-04	-8e-04	0	1e-04	1e-04
	ASmean	-1e-04	0	0	-1e-04	-3e-04	-6e-04	-7e-04	0	5e-04	-7e-04	-1e-04	1e-04	2e-04
	ASmedian	-1e-04	0	0	-1e-04	-3e-04	-6e-04	-7e-04	0	5e-04	-7e-04	-1e-04	1e-04	2e-04
	correction	-2e-04	1e-04	0	-1e-04	-3e-04	-6e-04	-6e-04	-1e-04	5e-04	-7e-04	-1e-04	2e-04	3e-04
	EBR	-3e-04	2e-04	0	-1e-04	-2e-04	-5e-04	-5e-04	-2e-04	5e-04	-7e-04	-1e-04	2e-04	2e-04
PU	ML	0.526	0.478	0.51	0.502	0.518	0.53	0.54	0.518	0.48	0.544	0.496	0.5	0.508
	ASmean	0.512	0.48	0.506	0.502	0.53	0.536	0.538	0.512	0.482	0.53	0.496	0.5	0.502
	ASmedian	0.512	0.48	0.506	0.5	0.526	0.532	0.538	0.516	0.48	0.534	0.504	0.498	0.498
	correction	0.52	0.476	0.506	0.506	0.516	0.532	0.538	0.518	0.484	0.534	0.502	0.496	0.498
	EBR	0.528	0.474	0.502	0.5	0.512	0.514	0.538	0.522	0.476	0.534	0.49	0.49	0.498
RMSE	ML	0	0	1e-04	0	0	0	1e-04	1e-04	0	0	0	0	0
	ASmean	0	0	1e-04	1e-04	0	0	1e-04	1e-04	0	0	1e-04	0	1e-04
	ASmedian	0	0	1e-04	0	0	0	1e-04	1e-04	0	0	1e-04	0	0
	correction	0	0	1e-04	0	0	0	0	1e-04	0	0	0	0	0
	EBR	0	0	1e-04	0	0	0	0	1e-04	0	0	0	0	0
CI	ML	0.94	0.94	0.924	0.938	0.926	0.924	0.93	0.92	0.932	0.92	0.924	0.938	0.938
	ASmean	0.954	0.958	0.936	0.946	0.946	0.934	0.936	0.934	0.948	0.942	0.944	0.954	0.954
	ASmedian	0.954	0.958	0.936	0.95	0.948	0.938	0.936	0.936	0.95	0.946	0.946	0.954	0.954
	correction	0.95	0.95	0.932	0.944	0.944	0.934	0.936	0.934	0.942	0.934	0.936	0.948	0.948
	EBR	0.956	0.952	0.93	0.942	0.946	0.936	0.932	0.932	0.948	0.94	0.94	0.956	0.954
		$\beta_{39}$	$\beta_{40}$	$\beta_{41}$	$\beta_{42}$	$\beta_{43}$	$\beta_{44}$	$\beta_{45}$	$\beta_{46}$	$\beta_{47}$	$\beta_{48}$	$\beta_{49}$	$\beta_{50}$	$\beta_{51}$
BIAS	ML	-2e-04	2e-04	3e-04	3e-04	4e-04	3e-04	3e-04	-4e-04	-5e-04	-3e-04	-3e-04	-1e-04	0
	ASmean	-2e-04	2e-04	2e-04	4e-04	4e-04	4e-04	2e-04	-4e-04	-6e-04	0	-2e-04	-1e-04	-1e-04
	ASmedian	-3e-04	2e-04	2e-04	4e-04	4e-04	3e-04	2e-04	-5e-04	-6e-04	-1e-04	-2e-04	-2e-04	-1e-04
	correction	-3e-04	1e-04	2e-04	3e-04	4e-04	2e-04	3e-04	-3e-04	-5e-04	-2e-04	-2e-04	-1e-04	-2e-04
	EBR	-3e-04	1e-04	3e-04	2e-04	4e-04	1e-04	4e-04	-2e-04	-4e-04	-4e-04	-3e-04	-2e-04	-1e-04
PU	ML	0.518	0.486	0.466	0.478	0.494	0.472	0.486	0.534	0.548	0.514	0.508	0.508	0.506
	ASmean	0.514	0.486	0.468	0.474	0.49	0.474	0.484	0.53	0.554	0.51	0.496	0.51	0.504
	ASmedian	0.52	0.484	0.47	0.478	0.496	0.472	0.482	0.53	0.56	0.512	0.502	0.514	0.512
	correction	0.52	0.49	0.466	0.478	0.494	0.472	0.486	0.53	0.548	0.508	0.506	0.512	0.514
	EBR	0.514	0.478	0.462	0.492	0.484	0.48	0.468	0.506	0.542	0.518	0.5	0.504	0.516
RMSE	ML	1e-04	0	0	1e-04	1e-04	0	0	1e-04	0	0	0	1e-04	0
	ASmean	1e-04	0	0	1e-04	1e-04	1e-04	1e-04	1e-04	1e-04	0	1e-04	1e-04	0
	ASmedian	1e-04	0	0	1e-04	1e-04	1e-04	1e-04	1e-04	0	0	0	1e-04	0
	correction	1e-04	0	0	1e-04	1e-04	0	0	1e-04	0	0	0	1e-04	0
	EBR	1e-04	0	0	1e-04	1e-04	0	0	1e-04	0	0	0	0	0
CI	ML	0.924	0.932	0.928	0.93	0.934	0.932	0.934	0.912	0.922	0.936	0.928	0.916	0.922
	ASmean	0.946	0.95	0.938	0.95	0.946	0.946	0.944	0.944	0.93	0.946	0.948	0.944	0.932
	ASmedian	0.946	0.95	0.938	0.952	0.946	0.946	0.946	0.944	0.93	0.95	0.948	0.948	0.934
	correction	0.934	0.944	0.934	0.944	0.938	0.942	0.936	0.932	0.928	0.946	0.938	0.932	0.93
	EBR	0.94	0.946	0.932	0.952	0.942	0.946	0.938	0.932	0.934	0.944	0.942	0.942	0.934

		$\beta_{52}$	$\beta_{53}$	$\beta_{54}$	$\beta_{55}$	$\beta_{56}$	$\beta_{57}$	$\beta_{58}$	$\beta_{59}$	$\beta_{60}$	$\beta_{61}$	$\beta_{62}$	$\beta_{63}$	$\beta_{64}$
BIAS	ML	-1e-04	-2e-04	-4e-04	-4e-04	2e-04	1e-04	5e-04	-8e-04	-2e-04	1e-04	-5e-04	4e-04	-4e-04
	ASmean	-2e-04	-2e-04	-3e-04	-4e-04	1e-04	1e-04	4e-04	-7e-04	-1e-04	2e-04	-6e-04	5e-04	-4e-04
	ASmedian	-2e-04	-3e-04	-3e-04	-3e-04	1e-04	1e-04	4e-04	-7e-04	-1e-04	2e-04	-6e-04	5e-04	-4e-04
	correction	-1e-04	-3e-04	-3e-04	-3e-04	1e-04	1e-04	4e-04	-8e-04	0	1e-04	-5e-04	5e-04	-4e-04
	EBR	-1e-04	-4e-04	-4e-04	-2e-04	1e-04	0	5e-04	-9e-04	0	0	-4e-04	5e-04	-4e-04
PU	ML	0.516	0.526	0.514	0.536	0.516	0.472	0.456	0.53	0.486	0.498	0.516	0.476	0.516
	ASmean	0.53	0.522	0.516	0.532	0.52	0.48	0.464	0.524	0.482	0.502	0.528	0.47	0.516
	ASmedian	0.538	0.524	0.514	0.536	0.518	0.48	0.466	0.532	0.482	0.502	0.526	0.47	0.516
	correction	0.514	0.528	0.506	0.53	0.516	0.478	0.456	0.532	0.478	0.498	0.52	0.47	0.516
	EBR	0.518	0.538	0.516	0.528	0.514	0.476	0.458	0.534	0.474	0.502	0.522	0.472	0.522
RMSE	ML	1e-04	1e-04	1e-04	0	0	1e-04	0	1e-04	0	0	0	0	1e-04
	ASmean	1e-04	1e-04	1e-04	0	0	1e-04	0	1e-04	0	0	0	0	1e-04
	ASmedian	1e-04	1e-04	1e-04	0	0	1e-04	0	1e-04	0	0	0	0	1e-04
	correction	1e-04	1e-04	1e-04	0	0	1e-04	0	1e-04	0	0	0	0	0
	EBR	1e-04	1e-04	1e-04	0	0	1e-04	0	1e-04	0	0	0	0	0
CI	ML	0.932	0.912	0.916	0.942	0.93	0.922	0.94	0.93	0.946	0.944	0.926	0.91	0.93
	ASmean	0.942	0.922	0.934	0.964	0.94	0.938	0.948	0.946	0.96	0.958	0.934	0.928	0.952
	ASmedian	0.946	0.926	0.934	0.966	0.942	0.938	0.948	0.946	0.962	0.962	0.936	0.928	0.952
	correction	0.942	0.916	0.928	0.958	0.938	0.934	0.946	0.938	0.956	0.95	0.932	0.92	0.942
	EBR	0.942	0.936	0.932	0.962	0.938	0.936	0.946	0.948	0.966	0.956	0.928	0.922	0.956
		$\beta_{65}$	$\beta_{66}$	$\beta_{67}$	$\beta_{68}$	$\beta_{69}$	$\beta_{70}$	$\beta_{71}$	$\beta_{72}$	$\beta_{73}$	$\beta_{74}$	$\beta_{75}$	$\beta_{76}$	$\beta_{77}$
BIAS	ML	-1e-04	-4e-04	-1e-04	5e-04	0	-2e-04	-1e-04	3e-04	-2e-04	1e-04	-4e-04	-6e-04	0
	ASmean	-1e-04	-3e-04	-4e-04	6e-04	1e-04	-1e-04	1e-04	1e-04	0	4e-04	-3e-04	-4e-04	-2e-04
	ASmedian	-1e-04	-3e-04	-3e-04	5e-04	0	-2e-04	1e-04	1e-04	0	3e-04	-3e-04	-4e-04	-2e-04
	correction	-1e-04	-3e-04	-3e-04	5e-04	-1e-04	-1e-04	-1e-04	1e-04	-1e-04	3e-04	-4e-04	-4e-04	-2e-04
	EBR	-1e-04	-3e-04	0	5e-04	-3e-04	-2e-04	-3e-04	3e-04	-2e-04	2e-04	-5e-04	-4e-04	-1e-04
PU	ML	0.494	0.522	0.488	0.5	0.496	0.528	0.486	0.5	0.488	0.486	0.5	0.522	0.504
	ASmean	0.498	0.524	0.504	0.494	0.49	0.522	0.486	0.528	0.482	0.458	0.49	0.512	0.506
	ASmedian	0.498	0.522	0.504	0.496	0.5	0.526	0.486	0.518	0.482	0.466	0.494	0.512	0.51
	correction	0.492	0.52	0.502	0.5	0.498	0.528	0.494	0.506	0.48	0.476	0.5	0.512	0.512
	EBR	0.488	0.516	0.496	0.498	0.516	0.54	0.52	0.51	0.486	0.478	0.498	0.522	0.502
RMSE	ML	1e-04	0	1e-04	0	1e-04	0	0	0	0	0	0	1e-04	0
	ASmean	1e-04	0	1e-04	0	1e-04	0	0	0	1e-04	0	0	1e-04	0
	ASmedian	1e-04	0	1e-04	0	1e-04	0	0	0	0	0	0	1e-04	0
	correction	1e-04	0	1e-04	0	1e-04	0	0	0	0	0	0	1e-04	0
	EBR	1e-04	0	0	0	1e-04	0	0	0	0	0	0	0	0
CI	ML	0.922	0.938	0.91	0.936	0.92	0.946	0.926	0.922	0.932	0.938	0.93	0.912	0.918
	ASmean	0.934	0.958	0.934	0.946	0.932	0.958	0.946	0.944	0.948	0.954	0.94	0.93	0.942
	ASmedian	0.938	0.958	0.938	0.946	0.932	0.962	0.946	0.948	0.95	0.954	0.94	0.93	0.942
	correction	0.93	0.952	0.93	0.946	0.928	0.956	0.94	0.944	0.942	0.944	0.94	0.924	0.932
	EBR	0.936	0.958	0.932	0.946	0.926	0.962	0.944	0.942	0.942	0.944	0.944	0.934	0.938

		$\beta_{78}$	$\beta_{79}$	$\beta_{80}$	$\beta_{81}$	$\beta_{82}$	$\beta_{83}$	$\beta_{84}$	$\beta_{85}$	$\beta_{86}$	$\beta_{87}$	$\beta_{88}$	$\beta_{89}$
BIAS	ML	4e-04	1e-04	-1e-04	-1e-04	0	-4e-04	1e-04	0	5e-04	-3e-04	1e-04	1e-04
	ASmean	4e-04	2e-04	-1e-04	-2e-04	1e-04	-3e-04	0	2e-04	6e-04	-4e-04	1e-04	3e-04
	ASmedian	4e-04	2e-04	-1e-04	-2e-04	0	-3e-04	0	2e-04	5e-04	-4e-04	1e-04	2e-04
	correction	3e-04	3e-04	0	-1e-04	0	-3e-04	1e-04	0	5e-04	-3e-04	1e-04	2e-04
	EBR	3e-04	4e-04	1e-04	-1e-04	0	-5e-04	2e-04	-3e-04	4e-04	-2e-04	2e-04	2e-04
PU	ML	0.486	0.494	0.498	0.518	0.496	0.512	0.498	0.514	0.478	0.55	0.494	0.45
	ASmean	0.49	0.488	0.494	0.51	0.492	0.512	0.508	0.5	0.47	0.552	0.492	0.438
	ASmedian	0.49	0.486	0.494	0.512	0.498	0.508	0.508	0.506	0.47	0.55	0.496	0.436
	correction	0.486	0.49	0.5	0.518	0.498	0.51	0.498	0.516	0.478	0.55	0.49	0.444
	EBR	0.494	0.482	0.492	0.508	0.498	0.512	0.49	0.518	0.476	0.532	0.504	0.448
RMSE	ML	0	1e-04	0	0	1e-04	1e-04	0	1e-04	1e-04	0	0	0
	ASmean	0	1e-04	1e-04	0	1e-04	1e-04	0	1e-04	1e-04	0	0	0
	ASmedian	0	1e-04	1e-04	0	1e-04	1e-04	0	1e-04	1e-04	0	0	0
	correction	0	1e-04	0	0	1e-04	1e-04	0	1e-04	1e-04	0	0	0
	EBR	0	1e-04	0	0	1e-04	0	0	0	1e-04	0	0	0
CI	ML	0.94	0.932	0.926	0.922	0.91	0.926	0.922	0.922	0.9	0.962	0.924	0.934
	ASmean	0.948	0.94	0.94	0.952	0.938	0.938	0.948	0.938	0.922	0.97	0.942	0.946
	ASmedian	0.948	0.94	0.942	0.952	0.938	0.938	0.95	0.938	0.922	0.97	0.942	0.948
	correction	0.948	0.938	0.936	0.942	0.936	0.936	0.936	0.934	0.914	0.968	0.938	0.942
	EBR	0.946	0.934	0.942	0.946	0.93	0.936	0.948	0.938	0.924	0.974	0.936	0.94
		$\beta_{90}$	$\beta_{91}$	$\beta_{92}$	$\beta_{93}$	$\beta_{94}$	$\beta_{95}$	$\beta_{96}$	$\beta_{97}$	$\beta_{98}$	$\beta_{99}$	$\beta_{100}$	$\alpha$
BIAS	ML	-1e-04	2e-04	5e-04	1e-04	3e-04	-2e-04	2e-04	-2e-04	-2e-04	5e-04	-3e-04	-0.2625
	ASmean	1e-04	3e-04	7e-04	0	3e-04	-1e-04	3e-04	-1e-04	-3e-04	5e-04	-2e-04	0.0535
	ASmedian	1e-04	3e-04	6e-04	0	3e-04	-1e-04	3e-04	-1e-04	-4e-04	5e-04	-2e-04	0.0305
	correction	1e-04	3e-04	4e-04	0	3e-04	-1e-04	3e-04	-1e-04	-4e-04	6e-04	-2e-04	-0.0319
	EBR	-1e-04	2e-04	1e-04	0	2e-04	-1e-04	4e-04	-1e-04	-5e-04	6e-04	-2e-04	-0.0741
PU	ML	0.496	0.502	0.482	0.488	0.488	0.488	0.496	0.496	0.5	0.46	0.538	0.996
	ASmean	0.492	0.504	0.472	0.498	0.484	0.48	0.492	0.49	0.508	0.466	0.536	0.338
	ASmedian	0.494	0.5	0.474	0.498	0.488	0.482	0.49	0.494	0.508	0.462	0.532	0.412
	correction	0.488	0.5	0.482	0.492	0.488	0.478	0.486	0.494	0.512	0.45	0.536	0.634
	EBR	0.5	0.506	0.504	0.5	0.496	0.478	0.484	0.478	0.516	0.456	0.532	0.768
RMSE	ML	1e-04	1e-04	0	0	1e-04	1e-04	0	0	1e-04	0	0	0.078
	ASmean	1e-04	1e-04	0	1e-04	1e-04	1e-04	0	1e-04	1e-04	0	0	0.0172
	ASmedian	1e-04	1e-04	0	0	1e-04	1e-04	0	1e-04	1e-04	0	0	0.0146
	correction	1e-04	0	0	0	1e-04	1e-04	0	0	0	0	0	0.0135
	EBR	1e-04	0	0	0	1e-04	1e-04	0	0	0	0	0	0.0169
CI	ML	0.92	0.926	0.938	0.94	0.91	0.904	0.918	0.936	0.924	0.928	0.928	0.188
	ASmean	0.928	0.944	0.948	0.954	0.928	0.934	0.93	0.952	0.936	0.936	0.944	0.274
	ASmedian	0.928	0.944	0.95	0.954	0.928	0.938	0.932	0.952	0.936	0.938	0.944	0.278
	correction	0.924	0.934	0.946	0.95	0.918	0.93	0.928	0.95	0.93	0.932	0.94	0.246
	EBR	0.922	0.94	0.958	0.944	0.922	0.926	0.924	0.952	0.934	0.94	0.946	0.834

stimatori proposti sembrano quindi essere preferibili rispetto allo stimatore di massima verosimiglianza, poichè il loro effetto desiderato si manifesta spesso.

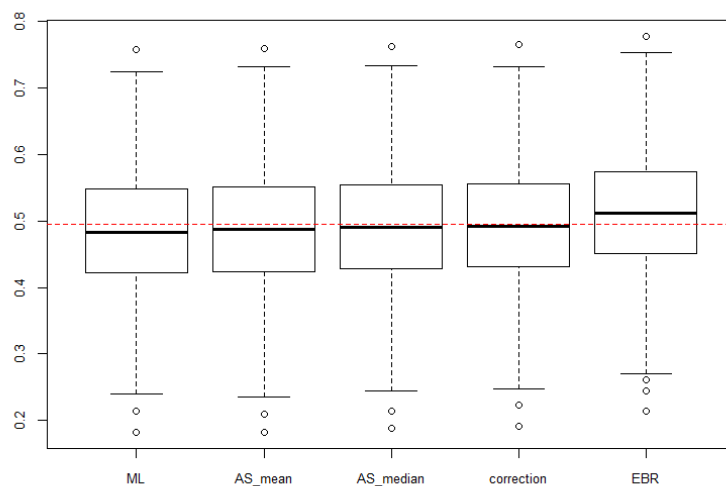


**Figura 4.1:** Boxplot delle stime  $\beta_0$  per ogni metodo. In rosso il vero valore del parametro.

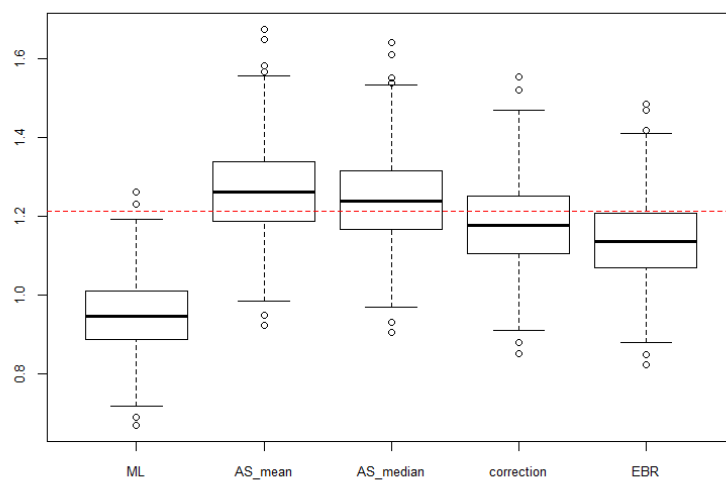
**Codice 4.7:** Studio di simulazione con  $n=1000$   $p=100$

```
rm(list=ls())  
library(MASS)  
source("generazione.R")  
source("EBRNB.R")  
source('brnb11.R')
```

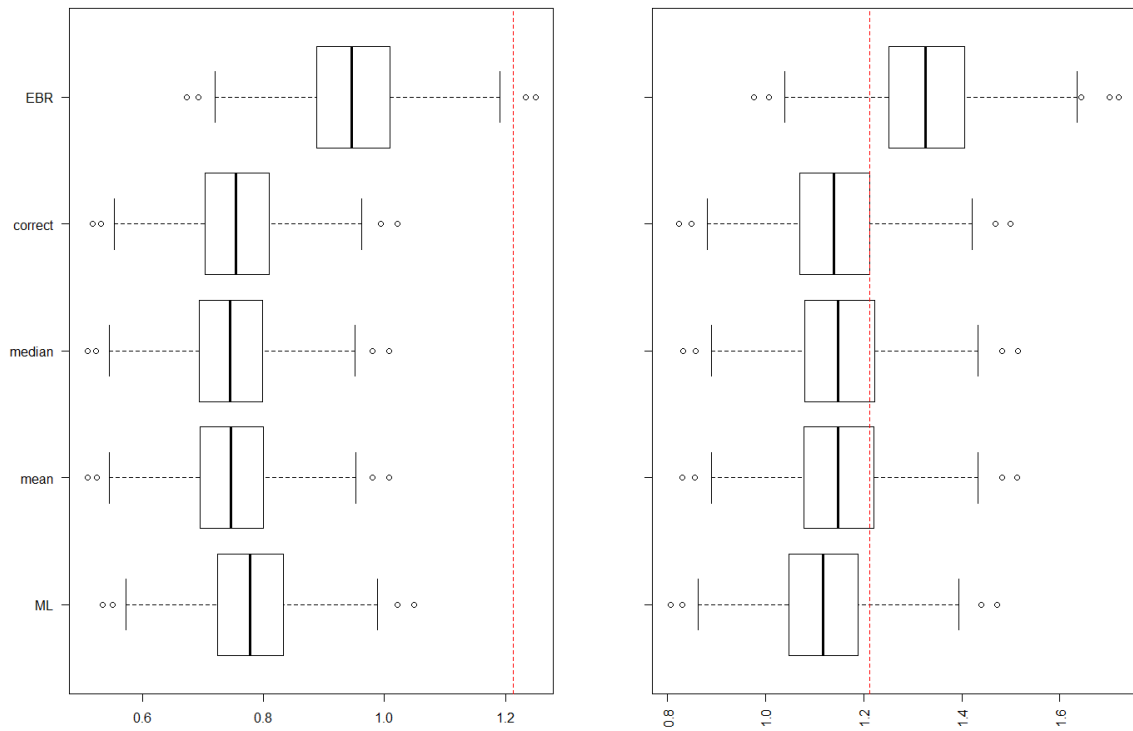
```
library(MASS)  
times=NULL  
p=100  
n=1000
```



**Figura 4.2:** Boxplot delle stime  $\beta_1$  per ogni metodo. In rosso il vero valore del parametro.



**Figura 4.3:** Boxplot delle stime  $\alpha$  per ogni metodo. In rosso il vero valore del parametro.



**Figura 4.4:** Boxplot degli estremi inferiori e superiori degli intervalli alla Wald delle stime  $\alpha$

```

y=runif(n,n,size=1,prob = 0.4)
XG=matrix(rpois(n*p,lambda = 1:p),n)
colnames(XG)=c(paste(rep('Var'),1:p))
XG2=cbind(1,XG)
glm_nb=glm.nb(y~XG[,1:p])

b=coef(glm_nb)
a=glm_nb$theta
truth=c(b,a)
X=model.matrix(glm_nb)

datasets=sim.data(truth,data=list(X=X,y=y),n=1000,Nsim=500,seed=800)

```

```

simulazioni=multi_estimation_all_method(datasets)
e=elaboration(simulazioni,truth)

boxplot(simulazioni[seq(102,51000,102),1:5],names = c('ML', 'AS_mean', 'AS_
median', 'correction', 'EBR'))
abline(h=a,col='red',lty=2,cex=1.5,horizontal=TRUE)

boxplot(simulazioni[seq(1,51000,102),1:5],names = c('ML', 'AS_mean', 'AS_
median', 'correction', 'EBR'))
abline(h=b[1],col='red',lty=2,horizontal=TRUE)

boxplot(simulazioni[seq(2,51000,102),1:5],names = c('ML', 'AS_mean', 'AS_
median', 'correction', 'EBR'))
abline(h=b[2],col='red',lty=2,horizontal=TRUE)

par(mfrow=c(1,2))

boxplot(simulazioni[seq(102,51000,102),6:10],names = c('ML', 'mean', 'median
', 'correction', 'EBR'),las=2,horizontal=TRUE)
abline(v=a,col='red',lty=2)

boxplot(simulazioni[seq(102,51000,102),11:15],names = NA,las=2,horizontal=
TRUE)
abline(v=a,col='red',lty=2)

```

## 4.5 Studio dei tempi di esecuzione dei metodi

In questa sezione sono presentati i risultati di uno studio volto al confronto dell'efficienza computazionale dei processi di stima per l'ottenimento degli stimatori ridotti per la distorsione utilizzati nel Paragrafo 4.3. Il confronto è basato sullo studio dei tempi medi di esecuzione degli algoritmi proposti, al crescere del numero di parametri in modelli di regressione binomiale negativi.

Per ogni dimensione del parametro  $p \in \{1, 2, 5, 10, 15, 20, 30, 40, 50, 70, 80\}$ , viene inizialmente generata casualmente una variabile risposta  $z$  tale che  $z_i \sim \text{Bineg}(\alpha = 0.5, \pi = 0.4)$ ,  $i \in \{1, \dots, 1000\}$  e delle variabili esplicati-



ve  $x_1, \dots, x_p$  da una distribuzione di Poisson tali che  $x_{ij} \sim P(\lambda = j)$ ,  $i \in \{1, \dots, 1000\}$ . Utilizzando come variabile dipendente  $z$  e come variabili indipendenti  $x_1, \dots, x_p$ , aggiungendo anche un'intercetta, viene stimato un modello di regressione binomiale negativo con il metodo della massima verosiglianza e sono state ottenute le stime  $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p, \hat{\alpha})$ . I valori dei coefficienti stimati sono utilizzati per generare un altro campione  $y$  da una distribuzione binomiale negativa, tale che  $y_i \sim \text{Bineg}(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}, \hat{\alpha})$ ,  $i \in \{1, \dots, 1000\}$ . Il nuovo campione viene regredito con le variabili esplicative  $x_0, x_1, \dots, x_p$  con i metodi di stima, con valore di partenza dell'algoritmo  $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p, \hat{\alpha})$ . Il processo viene replicato 50 volte e vengono memorizzati i tempi di medi di esecuzione.

In questo caso i metodi *ML* e *EBR* si basano sulla stessa implementazione proposta nell'Appendice B, perciò i loro tempi di esecuzione sono confrontabili. I metodi *ASmean*, *ASmedian* e *correction* sono stati implementati in maniera diversa, perciò i loro tempi di esecuzione non sono stati confrontati con quelli degli altri due metodi. Nelle Figure 4.5 e 4.6 vengono riportati i tempi medi di esecuzione all'aumentare del numero di parametri per i due gruppi di metodi e nel Codice 4.8 i comandi per replicare lo studio.

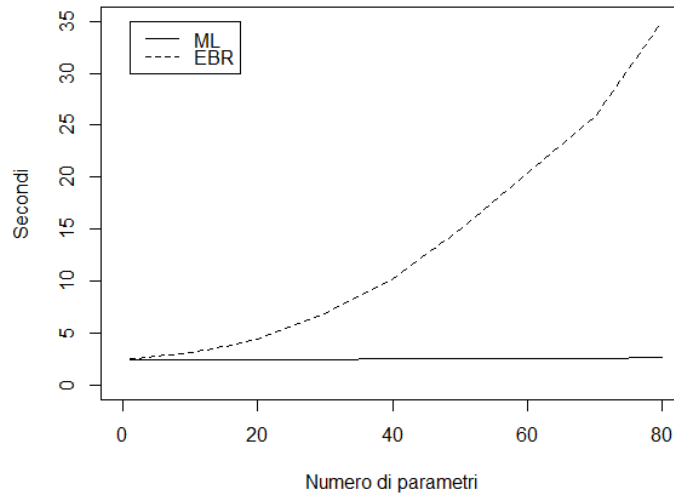
Dai risultati emerge l'inefficienza del metodo di riduzione della distorsione che mostra tempi di esecuzione esplosivi al crescere del numero di parametri. Il comportamento è inatteso, poichè ci si aspettava che il metodo fosse prestazionale poichè basato solo su quantità empiriche facilmente reperibili. Questo può essere dovuto al costo computazionale per l'ottenimento delle quantità necessarie alla modificazione empirica in (3.9), che risulta più grande di quello della semplice funzione *score*, o alla forma della modificazione, che diventa sempre più irregolare all'aumentare del numero di parametri, con conseguenti problemi di risoluzione del sistema di equazioni. Per quanto riguarda il confronto degli altri metodi di riduzione della distorsione, dai risultati emerge che gli algoritmi per l'ottenimento degli stimatori *correction* e *ASmean* richiedono meno tempo di esecuzione rispetto a quello per *ASmedian*.

---

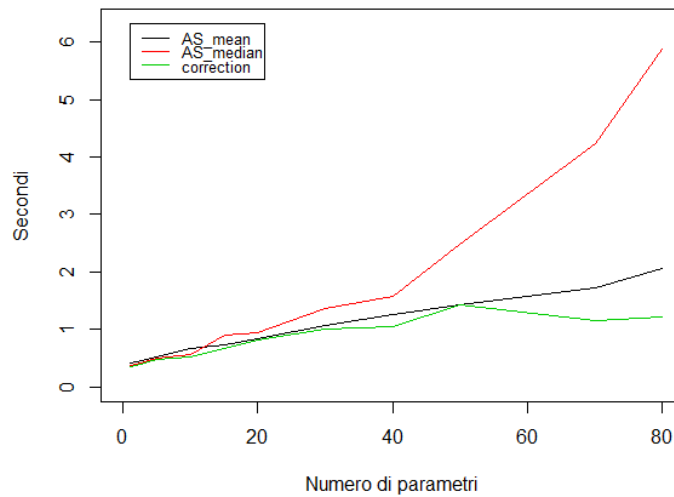
**Codice 4.8:** Studio dell'efficienza dei metodi

---

```
library(MASS)
source("generazione.R")
source("EBRNB.R")
```



**Figura 4.5:** Tempi medi di esecuzione dei metodi *ML*, *EBR* all'aumentare del numero di parametri del modello



**Figura 4.6:** Tempi medi di esecuzione dei metodi *ASmean*, *ASmedian* e *correction* all'aumentare del numero di parametri del modello

```
source('brnb11.R')

all_times=NULL
q=400
n=1000
a=0.5
p=0.4
y_start=rnbinom(n=n,size=1/a,prob = p)
XG=matrix(rpois(n*q,lambda = 1:q),n)
colnames(XG)=c(paste(rep('Var'),1:q))

for(k in 1:50){
times=NULL
for (p in c(1,2,5,10,15,20,30,40,50,70,80)){
  print(p)
  t=NULL
  glm_nb=glm.nb(y_start~XG[,1:p])
  XG2=cbind(1,XG[,1:p])
  beta=glm_nb$coefficients
  alpha=1/glm_nb$theta
  truth=c(beta,alpha)
  y_sim=simulate_nb(truth,XG2)
  colnames(XG2)=c('Intercept',paste(rep('Var'),1:p))

  start.time <- Sys.time()
  proc=try(process(truth,data=list(y=y_sim,X=XG2),type = 'ML'),silent =
    TRUE)
  if (failed <- inherits(proc, "try-error")) {
    t=c(t,NA)
  }
  else{
    end.time <- Sys.time()
    t=c(t,end.time - start.time)
  }

  start.time <- Sys.time()
}
```

```
proc=try(process(truth,data=list(y=y_sim,X=XG2),type = 'EBR'),silent =
  TRUE)
if (failed <- inherits(proc, "try-error")) {
  t=c(t,NA)
}
else{
  end.time <- Sys.time()
  t=c(t,end.time - start.time)
}

start.time <- Sys.time()
proc=try(process(truth,data=list(y=y_sim,X=XG2),type = 'AS_mean'),silent
  = TRUE)
if (failed <- inherits(proc, "try-error")) {
  t=c(t,NA)
}
else{
  end.time <- Sys.time()
  t=c(t,end.time - start.time)
}

start.time <- Sys.time()
proc=try(process(truth,data=list(y=y_sim,X=XG2),type = 'AS_median'),
  silent = TRUE)
if (failed <- inherits(proc, "try-error")) {
  t=c(t,NA)
}
else{
  end.time <- Sys.time()
  t=c(t,end.time - start.time)
}

start.time <- Sys.time()
proc=try(process(truth,data=list(y=y_sim,X=XG2),type = 'correction'),
  silent = TRUE)
if (failed <- inherits(proc, "try-error")) {
```

```
t=c(t,NA)
}
else{
  end.time <- Sys.time()
  t=c(t,end.time - start.time)
}

times=cbind(times,t)
}
all_times=rbind(all_times,times)
}

time_e=NULL
for (j in 1:5){
  time_e=rbind(time_e,colMeans(all_times[seq(j,NROW(all_times),5),]))
}

plot(c(1,2,5,10,15,20,30,40,50,70,80),time_e[1,],type='l',xlab='Numero di
  parametri',ylab='Secondi',xlim=c(0,80),ylim=c(0,max(time_e)))
points(c(1,2,5,10,15,20,30,40,50,70,80),time_e[2,],type='l',add=T,lty=2)
legend(1, max(time_e[1:2,]), legend=c("ML", "EBR"),
  col=c(1), lty=1:2, cex=1)

plot(c(1,2,5,10,15,20,30,40,50,70,80),time_e[3,],type='l',xlab='Numero di
  parametri',ylab='Secondi',xlim=c(0,80),ylim=c(0,max(time_e[3:5,])))
for (j in 4:5){
  points(c(1,2,5,10,15,20,30,40,50,70,80),times[j,],type='l',col=j-2,add=T)
}
legend(1, max(time_e[3:5,]), legend=c("AS_mean", "AS_median","correction")
  ,
  col=c(1:3), lty=1, cex=0.8)
```

# Conclusioni

In questa tesi si è trattato il problema della riduzione della distorsione nel contesto dei modelli di regressione. Lo studio è stato focalizzato sui modelli di regressione gamma e binomiale negativo, in particolare nella stima dei loro parametri di dispersione.

L'obiettivo principale del lavoro è stato confrontare e valutare il metodo empirico proposto in Kosmidis e Lunardon (2020) con lo stimatore di massima verosimiglianza e i classici metodi di riduzione della distorsione presenti in letteratura. Il confronto tra gli stimatori è basato su degli studi di simulazione condotti in tre modelli di regressione binomiale negativi e un modello di regressione gamma. Alla luce delle simulazioni effettuate emerge che lo stimatore ridotto per la distorsione empirica ha caratteristiche preferibili rispetto allo stimatore di massima verosimiglianza, specialmente per il parametro di dispersione, poichè oltre ad aver ridotto la distorsione, ha mostrato buone proprietà in termini di probabilità di sottostima e di copertura degli intervalli alla Wald.

Per quanto riguarda il confronto con gli altri metodi di riduzione della distorsione, dai risultati emerge che le stime di massima verosimiglianza dei coefficienti  $\beta$  di regressione sono poco distorte e l'effetto dei metodi di riduzione della distorsione è simile e poco intenso. Al contrario, le stime di massima verosimiglianza dei parametri di dispersione  $\alpha$  sono distorte e gli stimatori ridotti per la distorsione si sono mostrati più opportuni rispetto a quella basata sulla massima verosimiglianza, poichè hanno mostrato migliori proprietà in termini di distorsione, probabilità di sottostima, errore quadratico medio e copertura empirica degli intervalli. In particolare il metodo analitico di riduzione della distorsione in mediana si è mostrato più efficace in

termini di probabilità di sottostima, mentre il metodo di riduzione della distorsione empirica ha mostrato buoni comportamenti in termini di copertura empirica degli intervalli alla Wald. Date le potenzialità emerse dagli studi di simulazione effettuati, si può dunque ritenere che il metodo proposto in Kosmidis e Lunardon (2020) costituisca una valida alternativa ai suoi principali competitori, poiché è risultato un valido strumento per ridurre la distorsione dello stimatore di massima verosimiglianza. Come possibile sviluppo futuro, sarebbe utile costruire una procedura di selezione automatica delle variabili basata sul TIC per i modelli di regressione.

Dallo studio dell'efficienza computazionale dei metodi di riduzione della distorsione è emersa l'inefficienza del metodo di riduzione della distorsione empirica che mostra tempi medi di esecuzione esplosivi al crescere del numero di parametri. Il comportamento è inatteso, poiché ci si aspettava che il metodo fosse prestazionale poiché basato solo su quantità empiriche facilmente reperibili. Questo può essere dovuto al costo computazionale per l'ottenimento delle quantità necessarie alla modificazione empirica, che risulta più grande di quello della semplice funzione *score*, o alla forma della modificazione, che diventa sempre più irregolare all'aumentare del numero di parametri, con conseguenti problemi di risoluzione del sistema di equazioni. Se avessi avuto più tempo, avrei potuto studiare il motivo dell'inefficienza computazionale del metodo empirico di riduzione della distorsione, per comprendere se la parte predominante del tempo di esecuzione è dovuta al reperimento delle quantità necessarie alla modificazione o alla risoluzione del sistema di equazioni di stima modificate o ad una implementazione meno efficiente rispetto agli altri metodi.

Alla luce di queste considerazioni, si rendono necessarie ulteriori valutazioni delle proprietà possedute dello stimatore ridotto per la distorsione empirica, specialmente in presenza di molti parametri. Sarebbe interessante studiarne le caratteristiche in modelli di regressione che contemplano la presenza di più parametri di dispersione, come ad esempio quelli in ambito genomico.

# Bibliografia

- Agresti, Alan (2015). *Foundations of linear and generalized linear models*. John Wiley & Sons.
- Akaike, Hirotugu (1974). «A new look at the statistical model identification». In: *IEEE transactions on automatic control* 19.6, 716–723.
- Albert, Adelin e John A Anderson (1984). «On the existence of maximum likelihood estimates in logistic regression models». In: *Biometrika* 71.1, 1–10.
- Azzalini, Adelchi (2001). *Inferenza statistica: Una presentazione basata sul concetto di verosimiglianza*. Springer Science & Business Media.
- Boos, Dennis D (1992). «On generalized score tests». In: *The American Statistician* 46.4, 327–333.
- Boos, Dennis D e Stefanski (2013). *Essential statistical inference*. Springer.
- Breslow, Norman E (1984). «Extra-Poisson variation in log-linear models». In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 33.1, 38–44.
- Cameron, A Colin e Pravin K Trivedi (2013). *Regression analysis of count data*. Vol. 53. Cambridge university press.
- Cordeiro, Gauss M e Peter McCullagh (1991). «Bias correction in generalized linear models». In: *Journal of the Royal Statistical Society: Series B (Methodological)* 53.3, 629–643.
- Cox, David R (1983). «Some remarks on overdispersion». In: *Biometrika* 70.1, 269–274.
- Efron, Bradley (1975). «Defining the curvature of a statistical problem (with applications to second order efficiency)». In: *The Annals of Statistics* 3.6, 1189–1242.



- Efron, Bradley (1982). *The jackknife, the bootstrap and other resampling plans*. SIAM.
- Efron, Bradley e Robert Tibshirani (1986). «Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy». In: *Statistical science*, 54–75.
- Firth, David (1993). «Bias reduction of maximum likelihood estimates». In: *Biometrika*, 27–38.
- Gourieroux, Christian, Alain Monfort e Eric Renault (1993). «Indirect inference». In: *Journal of applied econometrics* 8.S1, S85–S118.
- Guerrier, Stéphane, Elise Dupuis-Lozeron, Yanyuan Ma e Maria-Pia Victoria-Feser (2019). «Simulation-based bias correction methods for complex models». In: *Journal of the American Statistical Association* 114.525, 146–157.
- Hinde, John e Clarice GB Demétrio (1998). «Overdispersion: models and estimation». In: *Computational Statistics and Data Analysis* 27.2, 151–170.
- Huber, Peter J e Elvezio M Ronchetti (2009). «Robust statistics. 2009». In: *Hoboken: John Wiley & Sons* 2.
- Kenne Pagui, Euloge Clovis (2020). *brnb.R*. <https://github.com/eulogepagui/brglm2>.
- Kenne Pagui, Euloge Clovis, Alessandra Salvan e Nicola Sartori (2017). «Median bias reduction of maximum likelihood estimates». In: *Biometrika* 104.4, 923–938.
- Kosmidis, Ioannis (2014). «Bias in parametric estimation: reduction and useful side-effects». In: *Wiley Interdisciplinary Reviews: Computational Statistics* 6.3, 185–196.
- (2020). *brglm2, Bias Reduction in Generalized Linear Models*. R package version 0.6.2.
- Kosmidis, Ioannis e David Firth (2009). «Bias reduction in exponential family nonlinear models». In: *Biometrika* 96.4, 793–804.
- Kosmidis, Ioannis, Euloge Clovis Kenne Pagui e Nicola Sartori (2020). «Mean and median bias reduction in generalized linear models». In: *Statistics and Computing* 30.1, 43–59.

- Kosmidis, Ioannis e Nicola Lunardon (2020). «Empirical bias-reducing adjustments to estimating functions». In: *arXiv preprint arXiv:2001.03786*.
- Kuk, Anthony YC (1995). «Asymptotically unbiased estimation in generalized linear models with random effects». In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57.2, 395–407.
- Lawless, Jerald F (1987). «Negative binomial and mixed Poisson regression». In: *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 209–225.
- Liang, Kung-Yee e Scott L Zeger (1986). «Longitudinal data analysis using generalized linear models». In: *Biometrika* 73.1, 13–22.
- Lindsay, Bruce G (1988). «Composite likelihood methods». In: *Contemporary mathematics* 80.1, 221–239.
- Lunardon, N e D Scharfstein (2017). «Comment on ‘small sample GEE estimation of regression parameters for longitudinal data’». In: *Statistics in medicine* 36.22, 3596–3600.
- Mackisack, Margaret (1994). «What is the use of experiments conducted by statistics students». In: *Journal of Statistics Education* 2.1, 1–15.
- Martin, Michael A (1992). «On the double bootstrap». In: *Computing science and statistics*. Springer, 73–78.
- McCullagh, Peter (2018). *Tensor methods in statistics*. Courier Dover Publications.
- Nelder, John Ashworth e Robert WM Wedderburn (1972). «Generalized linear models». In: *Journal of the Royal Statistical Society: Series A (General)* 135.3, 370–384.
- Pace, Luigi e Alessandra Salvan (1996). *Introduzione alla statistica: Inferenza, verosimiglianza, modelli.-2001.-xvi, 422 p.* Cedam.
- (1997). «Principles of statistical inference from a Neo-Fisherian perspective». In: 4, xviii, 535.
- Quenouille, Maurice H (1956). «Notes on bias in estimation». In: *Biometrika* 43.3/4, 353–360.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.

- Stefanski, Leonard A e Dennis D Boos (2002). «The calculus of M-estimation». In: *The American Statistician* 56.1, 29–38.
- Takeuchi, Kei (1976). «The distribution of information statistics and the criterion of goodness of fit of models». In: *Mathematical Science* 153, 12–18.
- Van Der Vaart, Aad W (2000). «Asymptotic statistics». In: 3.
- Varin, Cristiano e Claudia Czado (2010). «A mixed autoregressive probit model for ordinal longitudinal data». In: *Biostatistics* 11.1, 127–138.
- Varin, Cristiano, Nancy Reid e David Firth (2011). «An overview of composite likelihood methods». In: *Statistica Sinica*, 5–42.
- Varin, Cristiano e Paolo Vidoni (2005). «A note on composite likelihood inference and model selection». In: *Biometrika* 92.3, 519–528.
- Wedderburn, Robert WM (1974). «Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method». In: *Biometrika* 61.3, 439–447.

# Appendice A

## Quantità per la penalizzazione empirica

### A.1 Quantità necessarie per le penalizzazioni in modelli lineari generalizzati

$$j^{(G)} = \begin{bmatrix} j_{\beta\beta} & j_{\beta\phi} \\ j_{\beta\phi}^\top & j_{\phi\phi}^{(G)} \end{bmatrix} \quad \text{e} \quad e^{(G)} = \begin{bmatrix} e_{\beta\beta} & e_{\beta\phi}^{(G)} \\ \{e_{\beta\phi}^{(G)}\}^T & e_{\phi\phi}^{(G)} \end{bmatrix},$$

dove

$$\begin{aligned} j_{\beta\beta} &= \frac{1}{\phi} X^\top Q X, & j_{\phi\phi}^{(G)} &= \frac{1}{\phi^3} \mathbf{1}_n^\top (R - A') \mathbf{1}_n + \frac{1}{2\phi^4} \mathbf{1}_n^\top A'' \mathbf{1}_n, & j_{\beta\phi} &= \frac{1}{\phi^2} X^\top \tilde{W} \mathbf{1}_n \\ e_{\beta\beta} &= \frac{1}{\phi^2} X^\top \tilde{W}^2 X, & e_{\phi\phi}^{(G)} &= \frac{1}{4\phi^4} \mathbf{1}_n^\top (R - A')^2 \mathbf{1}_n, & e_{\beta\phi}^{(G)} &= \frac{1}{2\phi^3} X^\top \tilde{W} (R - A') \mathbf{1}_n. \end{aligned}$$

Nell'espressione sopra,  $\mathbf{1}_n$  è un vettore di uni  $n$ -dimensionale. Le matrici  $n \times n$  diagonali  $Q$ ,  $\tilde{W}$ ,  $R$ ,  $A'$ ,  $A''$  hanno come  $i$ -esimo elemento  $q_i = b_i d_i - b'_i (y_i - \mu_i)$ ,  $\tilde{g}_i = b_i (y_i - \mu_i)$ ,  $r_i = -2m_i (y_i \theta_i - b(\theta_i) - c_1(y_i))$  (devianza residua),  $a'_i = m_i a'(-m_1/\phi)$ ,  $a''_i = m_i^2 a''(-m_1/\phi)$ , rispettivamente, dove  $b_i = m_i d_i / v_i$ ,  $b'_i = m_i (d'_i / v_i - d_i^2 v'_i / v_i^2)$ ,  $d'_i = d^2 \mu_i / d\eta_i^2$ ,  $v'_i = dv_i / d\mu_i$  ( $i = 1, \dots, n$ ), e  $a'(u) = da(u)/du$ ,  $a''(u) = d^2 a(u)/du^2$

## A.2 Espressione di $j^{(G)}(\cdot)$ , $\Psi(\cdot)$ , $\tilde{\Psi}_r(\cdot)$ e $u_r(\cdot)$ per alcuni modelli di regressione

### A.2.1 Modello di regressione logistica con funzione legame logistica

$$\begin{aligned}
 j^{(G)}(\beta) &= \begin{bmatrix} \sum_{i=1}^n \frac{m_i x_{i1} x_{i1} e^{\beta' x_i}}{(1+e^{\beta' x_i})^2} & \cdots & \sum_{i=1}^n \frac{m_i x_{ip} x_{i1} e^{\beta' x_i}}{(1+e^{\beta' x_i})^2} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n \frac{m_i x_{ip} x_{i1} e^{\beta' x_i}}{(1+e^{\beta' x_i})^2} & \cdots & \sum_{i=1}^n \frac{m_i x_{ip} x_{ip} e^{\beta' x_i}}{(1+e^{\beta' x_i})^2} \end{bmatrix}, \\
 \Psi(\beta) &= \begin{bmatrix} m_1 x_{11} \left( y_1 - \frac{e^{\beta' x_1}}{1+e^{\beta' x_1}} \right) & \cdots & m_1 x_{1p} \left( y_1 - \frac{e^{\beta' x_1}}{1+e^{\beta' x_1}} \right) \\ \vdots & \ddots & \vdots \\ m_n x_{n1} \left( y_n - \frac{e^{\beta' x_n}}{1+e^{\beta' x_n}} \right) & \cdots & m_n x_{np} \left( y_n - \frac{e^{\beta' x_n}}{1+e^{\beta' x_n}} \right) \end{bmatrix}, \\
 \tilde{\Psi}_r(\beta) &= \begin{bmatrix} -\frac{m_1 x_{1r} x_{i1} e^{\beta' x_1}}{(1+e^{\beta' x_1})^2} & \cdots & -\frac{m_1 x_{1r} x_{ip} e^{\beta' x_1}}{(1+e^{\beta' x_1})^2} \\ \vdots & \ddots & \vdots \\ -\frac{m_n x_{nr} x_{n1} e^{\beta' x_n}}{(1+e^{\beta' x_n})^2} & \cdots & -\frac{m_n x_{nr} x_{np} e^{\beta' x_n}}{(1+e^{\beta' x_n})^2} \end{bmatrix} r \in \{1, \dots, p\}, \\
 u_r(\beta) &= \begin{bmatrix} \sum_{i=1}^n \frac{m_i x_{ir} x_{i1} x_{i1} e^{\beta' x_i} (e^{\beta' x_i} - 1)}{(1+e^{\beta' x_i})^3} & \cdots & \sum_{i=1}^n \frac{m_i x_{ir} x_{i1} x_{ip} e^{\beta' x_i} (e^{\beta' x_i} - 1)}{(1+e^{\beta' x_i})^3} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n \frac{m_i x_{ir} x_{ip} x_{i1} e^{\beta' x_i} (e^{\beta' x_i} - 1)}{(1+e^{\beta' x_i})^3} & \cdots & \sum_{i=1}^n \frac{m_i x_{ir} x_{ip} x_{ip} e^{\beta' x_i} (e^{\beta' x_i} - 1)}{(1+e^{\beta' x_i})^3} \end{bmatrix} r \in \{1, \dots, p\}.
 \end{aligned}$$

### A.2.2 Modello di regressione Poisson con funzione legame logaritmica

$$\begin{aligned}
 j^{(G)}(\beta) &= \begin{bmatrix} \sum_{i=1}^n -x_{i1}x_{i1}e^{\beta'x_i} & \dots & \sum_{i=1}^n -x_{i1}x_{ip}e^{\beta'x_i} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n -x_{ip}x_{i1}e^{\beta'x_i} & \dots & \sum_{i=1}^n -x_{ip}x_{ip}e^{\beta'x_i} \end{bmatrix}, \\
 \Psi(\beta) &= \begin{bmatrix} x_{11}(y_1 - e^{\beta'x_1}) & \dots & x_{1p}(y_1 - e^{\beta'x_1}) \\ \vdots & \ddots & \vdots \\ x_{n1}(y_n - e^{\beta'x_n}) & \dots & x_{np}(y_n - e^{\beta'x_n}) \end{bmatrix}, \\
 \tilde{\Psi}_r(\beta) &= \begin{bmatrix} -x_{1r}x_{11}e^{\beta'x_1} & \dots & -x_{1r}x_{1p}e^{\beta'x_1} \\ \vdots & \ddots & \vdots \\ -x_{nr}x_{n1}e^{\beta'x_n} & \dots & -x_{nr}x_{np}e^{\beta'x_n} \end{bmatrix} r \in \{1, \dots, p\}, \\
 u_r(\beta) &= \begin{bmatrix} \sum_{i=1}^n -x_{ir}x_{i1}x_{i1}e^{\beta'x_i} & \dots & \sum_{i=1}^n -x_{ir}x_{i1}x_{ip}e^{\beta'x_i} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n -x_{ir}x_{ip}x_{i1}e^{\beta'x_i} & \dots & \sum_{i=1}^n -x_{ir}x_{ip}x_{ip}e^{\beta'x_i} \end{bmatrix} r \in \{1, \dots, p\}.
 \end{aligned}$$

### A.2.3 Modello di regressione Gamma con funzione legame logaritmica

$$\begin{aligned}
 j^{(G)}(\beta, \alpha) &= \begin{bmatrix} \alpha \sum_{i=1}^n \frac{x_{i1}x_{i1}y_i}{e^{\beta'x_i}} & \cdots & \alpha \sum_{i=1}^n \frac{x_{i1}x_{ip}y_i}{e^{\beta'x_i}} & \sum_{i=1}^n x_{1r} \left(1 - \frac{y_i}{e^{\beta'x_i}}\right) \\ \vdots & \ddots & \vdots & \vdots \\ \alpha \sum_{i=1}^n \frac{x_{ip}x_{i1}y_i}{e^{\beta'x_i}} & \cdots & \alpha \sum_{i=1}^n \frac{x_{ip}x_{ip}y_i}{e^{\beta'x_i}} & \sum_{i=1}^n x_{1p} \left(1 - \frac{y_i}{e^{\beta'x_i}}\right) \\ \sum_{i=1}^n x_{11} \left(1 - \frac{y_i}{e^{\beta'x_i}}\right) & \cdots & \sum_{i=1}^n x_{1p} \left(1 - \frac{y_i}{e^{\beta'x_i}}\right) & -\frac{n}{\alpha} + n \frac{\partial^2 \log \Gamma(\alpha)}{\partial \alpha^2} \end{bmatrix}, \\
 \Psi(\beta, \alpha) &= \begin{bmatrix} \alpha x_{11} \left(\frac{y_1}{e^{\beta'x_1}} - 1\right) & \cdots & \alpha x_{1p} \left(\frac{y_1}{e^{\beta'x_1}} - 1\right) & \log(\alpha) + 1 - \frac{\partial \log \Gamma(\alpha)}{\partial \alpha} + \left(\log\left(\frac{y_1}{e^{\beta'x_1}}\right) - \frac{y_1}{e^{\beta'x_1}}\right) \\ \vdots & \ddots & \vdots & \vdots \\ \alpha x_{n1} \left(\frac{y_n}{e^{\beta'x_n}} - 1\right) & \cdots & \alpha x_{np} \left(\frac{y_n}{e^{\beta'x_n}} - 1\right) & \log(\alpha) + 1 - \frac{\partial \log \Gamma(\alpha)}{\partial \alpha} + \left(\log\left(\frac{y_n}{e^{\beta'x_n}}\right) - \frac{y_n}{e^{\beta'x_n}}\right) \end{bmatrix}, \\
 \tilde{\Psi}_r(\beta, \alpha) &= \begin{bmatrix} -\alpha \frac{x_{1r}x_{11}y_1}{e^{\beta'x_1}} & \cdots & -\alpha \frac{x_{1r}x_{1p}y_1}{e^{\beta'x_1}} & -x_{1r} \left(1 - \frac{y_1}{e^{\beta'x_1}}\right) \\ \vdots & \ddots & \vdots & \vdots \\ -\alpha \frac{x_{nr}x_{n1}y_n}{e^{\beta'x_n}} & \cdots & -\alpha \frac{x_{nr}x_{np}y_n}{e^{\beta'x_n}} & -x_{nr} \left(1 - \frac{y_n}{e^{\beta'x_n}}\right) \end{bmatrix} \quad r \in \{1, \dots, p\}, \\
 u_r(\beta, \alpha) &= \begin{bmatrix} \alpha \sum_{i=1}^n \frac{x_{ir}x_{i1}x_{i1}y_i}{e^{\beta'x_i}} & \cdots & \alpha \sum_{i=1}^n \frac{x_{ir}x_{i1}x_{ip}y_i}{e^{\beta'x_i}} & -\sum_{i=1}^n \frac{x_{ir}x_{i1}y_i}{e^{\beta'x_i}} \\ \vdots & \ddots & \vdots & \vdots \\ \alpha \sum_{i=1}^n \frac{x_{ir}x_{ip}x_{i1}y_i}{e^{\beta'x_i}} & \cdots & \alpha \sum_{i=1}^n \frac{x_{ir}x_{ip}x_{ip}y_i}{e^{\beta'x_i}} & -\sum_{i=1}^n \frac{x_{ir}x_{ip}y_i}{e^{\beta'x_i}} \\ -\sum_{i=1}^n \frac{x_{ir}x_{i1}y_i}{e^{\beta'x_i}} & \cdots & -\sum_{i=1}^n \frac{x_{ir}x_{ip}y_i}{e^{\beta'x_i}} & 0 \end{bmatrix} \quad r \in \{1, \dots, p\}, \\
 \tilde{\Psi}_{p+1}(\beta, \alpha) &= \begin{bmatrix} -x_{11} \left(1 - \frac{y_1}{e^{\beta'x_1}}\right) & \cdots & -x_{1p} \left(1 - \frac{y_1}{e^{\beta'x_1}}\right) & \frac{1}{\alpha} - \frac{\partial^2 \log \Gamma(\alpha)}{\partial \alpha^2} \\ \vdots & \ddots & \vdots & \vdots \\ -x_{n1} \left(1 - \frac{y_n}{e^{\beta'x_n}}\right) & \cdots & -x_{np} \left(1 - \frac{y_n}{e^{\beta'x_n}}\right) & \frac{1}{\alpha} - \frac{\partial^2 \log \Gamma(\alpha)}{\partial \alpha^2} \end{bmatrix}, \\
 u_{p+1}(\beta) &= \begin{bmatrix} -\sum_{i=1}^n \frac{x_{i1}x_{i1}y_i}{e^{\beta'x_i}} & \cdots & -\sum_{i=1}^n \frac{x_{i1}x_{ip}y_i}{e^{\beta'x_i}} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -\sum_{i=1}^n \frac{x_{ip}x_{i1}y_i}{e^{\beta'x_i}} & \cdots & -\sum_{i=1}^n \frac{x_{ip}x_{ip}y_i}{e^{\beta'x_i}} & 0 \\ 0 & \cdots & 0 & -\frac{n}{\alpha^2} - n \frac{\partial^3 \log \Gamma(\alpha)}{\partial \alpha^3} \end{bmatrix}.
 \end{aligned}$$

### A.2.4 Modello di regressione Binomiale negativo con funzione legame logaritmica

$$\begin{aligned}
 j^{(G)}(\beta, \alpha) &= \begin{bmatrix} \sum_{i=1}^n \frac{\mu_i(1+\alpha y_i)x_{i1}x_{i1}}{(1+\alpha\mu_i)^2} & \cdots & \sum_{i=1}^n \frac{\mu_i(1+\alpha y_i)x_{i1}x_{ip}}{(1+\alpha\mu_i)^2} & \sum_{i=1}^n \frac{\mu_i(y_i-\mu_i)x_{i1}}{(1+\alpha\mu_i)^2} \\ \vdots & \ddots & \vdots & \vdots \\ \sum_{i=1}^n \frac{\mu_i(1+\alpha y_i)x_{ip}x_{i1}}{(1+\alpha\mu_i)^2} & \cdots & \sum_{i=1}^n \frac{\mu_i(1+\alpha y_i)x_{ip}x_{ip}}{(1+\alpha\mu_i)^2} & \sum_{i=1}^n \frac{\mu_i(y_i-\mu_i)x_{ip}}{(1+\alpha\mu_i)^2} \\ \sum_{i=1}^n \frac{\mu_i(y_i-\mu_i)x_{i1}}{(1+\alpha\mu_i)^2} & \cdots & \sum_{i=1}^n \frac{\mu_i(y_i-\mu_i)x_{ip}}{(1+\alpha\mu_i)^2} & j_{\alpha\alpha}^{(G)}(\beta, \alpha) \end{bmatrix}, \\
 \Psi(\beta, \alpha) &= \begin{bmatrix} \frac{x_{11}(y_1-\mu_1)}{1+\alpha\mu_1} & \cdots & \frac{x_{1p}(y_1-\mu_1)}{1+\alpha\mu_1} & \left\{ \alpha^{-2} \left( \log(1+\alpha\mu_1) - \sum_{j=0}^{y_1-1} \frac{1}{j+\alpha-1} \right) + \frac{y_1-\mu_1}{\alpha(1+\alpha\mu_1)} \right\} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{x_{n1}(y_n-\mu_n)}{1+\alpha\mu_n} & \cdots & \frac{x_{np}(y_n-\mu_n)}{1+\alpha\mu_n} & \left\{ \alpha^{-2} \left( \log(1+\alpha\mu_n) - \sum_{j=0}^{y_n-1} \frac{1}{j+\alpha-1} \right) + \frac{y_n-\mu_n}{\alpha(1+\alpha\mu_n)} \right\} \end{bmatrix}, \\
 \tilde{\Psi}_r(\beta, \alpha) &= \begin{bmatrix} -\frac{\mu_1(1+\alpha y_1)x_{1r}x_{i1}}{(1+\alpha\mu_1)^2} & \cdots & -\frac{\mu_1(1+\alpha y_1)x_{1r}x_{ip}}{(1+\alpha\mu_1)^2} & -\frac{\mu_1(y_1-\mu_1)x_{1r}}{(1+\alpha\mu_1)^2} \\ \vdots & \ddots & \vdots & \vdots \\ -\frac{\mu_n(1+\alpha y_n)x_{nr}x_{n1}}{(1+\alpha\mu_n)^2} & \cdots & -\frac{\mu_n(1+\alpha y_n)x_{nr}x_{np}}{(1+\alpha\mu_n)^2} & -\frac{\mu_n(y_n-\mu_n)x_{nr}}{(1+\alpha\mu_n)^2} \end{bmatrix} \quad r \in \{1, \dots, p\}, \\
 u_r(\beta, \alpha) &= \begin{bmatrix} \sum_{i=1}^n \frac{(\alpha y_i+1)\mu_i(\alpha\mu_i-1)x_{ir}x_{i1}x_{i1}}{(\alpha\mu_i+1)^3} & \cdots & \sum_{i=1}^n \frac{(\alpha y_i+1)\mu_i(\alpha\mu_i-1)x_{ir}x_{i1}x_{ip}}{(\alpha\mu_i+1)^3} & \sum_{i=1}^n \frac{\mu_i(\alpha y_i\mu_i+2\mu_i-y_i)x_{ir}x_{i1}}{(a\mu_i+1)^3} \\ \vdots & \ddots & \vdots & \vdots \\ \sum_{i=1}^n \frac{(\alpha y_i+1)\mu_i(\alpha\mu_i-1)x_{ir}x_{ip}x_{i1}}{(\alpha\mu_i+1)^3} & \cdots & \sum_{i=1}^n \frac{(\alpha y_i+1)\mu_i(\alpha\mu_i-1)x_{ir}x_{ip}x_{ip}}{(\alpha\mu_i+1)^3} & \sum_{i=1}^n \frac{\mu_i(\alpha y_i\mu_i+2\mu_i-y_i)x_{ir}x_{ip}}{(a\mu_i+1)^3} \\ \sum_{i=1}^n \frac{\mu_i(\alpha y_i\mu_i+2\mu_i-y_i)x_{ir}x_{i1}}{(a\mu_i+1)^3} & \cdots & \sum_{i=1}^n \frac{\mu_i(\alpha y_i\mu_i+2\mu_i-y_i)x_{ir}x_{ip}}{(a\mu_i+1)^3} & \sum_{i=1}^n \frac{2\mu_i^2(y_i-\mu_i)x_{ir}}{(\alpha\mu_i+1)^3} \end{bmatrix} \\
 r &\in \{1, \dots, p\}, \\
 \tilde{\Psi}_{p+1}(\beta, \alpha) &= \begin{bmatrix} -\frac{\mu_1(y_1-\mu_1)x_{11}}{(1+\alpha\mu_1)^2} & \cdots & -\frac{\mu_1(y_1-\mu_1)x_{1p}}{(1+\alpha\mu_1)^2} & -\sum_{j=0}^{y_1-1} \left( \frac{j}{1+\alpha j} \right)^2 - \frac{2\log(1+\alpha\mu_1)}{\alpha^3} + \frac{2\alpha^{-2}\mu_1}{1+\alpha\mu_1} + \frac{(y_1+\alpha^{-1})\mu_1^2}{(1+\alpha\mu_1)^2} \\ \vdots & \ddots & \vdots & \vdots \\ -\frac{\mu_n(y_n-\mu_n)x_{n1}}{(1+\alpha\mu_n)^2} & \cdots & -\frac{\mu_n(y_n-\mu_n)x_{np}}{(1+\alpha\mu_n)^2} & -\sum_{j=0}^{y_n-1} \left( \frac{j}{1+\alpha j} \right)^2 - \frac{2\log(1+\alpha\mu_n)}{\alpha^3} + \frac{2\alpha^{-2}\mu_n}{1+\alpha\mu_n} + \frac{(y_n+\alpha^{-1})\mu_n^2}{(1+\alpha\mu_n)^2} \end{bmatrix}, \\
 u_{p+1}(\beta) &= \begin{bmatrix} \sum_{i=1}^n \frac{\mu_i(\alpha y_i\mu_i+2\mu_i-y_i)x_{i1}x_{i1}}{(a\mu_i+1)^3} & \cdots & \sum_{i=1}^n \frac{\mu_i(\alpha y_i\mu_i+2\mu_i-y_i)x_{i1}x_{ip}}{(a\mu_i+1)^3} & \sum_{i=1}^n \frac{2\mu_i^2(y_i-\mu_i)x_{i1}}{(\alpha\mu_i+1)^3} \\ \vdots & \ddots & \vdots & \vdots \\ \sum_{i=1}^n \frac{\mu_i(\alpha y_i\mu_i+2\mu_i-y_i)x_{ip}x_{i1}}{(a\mu_i+1)^3} & \cdots & \sum_{i=1}^n \frac{\mu_i(\alpha y_i\mu_i+2\mu_i-y_i)x_{ip}x_{ip}}{(a\mu_i+1)^3} & \sum_{i=1}^n \frac{2\mu_i^2(y_i-\mu_i)x_{ip}}{(\alpha\mu_i+1)^3} \\ \sum_{i=1}^n \frac{2\mu_i^2(y_i-\mu_i)x_{i1}}{(\alpha\mu_i+1)^3} & \cdots & \sum_{i=1}^n \frac{2\mu_i^2(y_i-\mu_i)x_{ip}}{(\alpha\mu_i+1)^3} & u_{(p+1)\alpha\alpha}(\beta, \alpha) \end{bmatrix}, \\
 \text{dove } j_{\alpha\alpha}^{(G)}(\beta, \alpha) &= \sum_{i=1}^n \left\{ \sum_{j=0}^{y_i-1} \left( \frac{j}{1+\alpha j} \right)^2 + 2\alpha^{-3} \log(1+\alpha\mu_i) - \frac{2\alpha^{-2}\mu_i}{1+\alpha\mu_i} - \frac{(y_i+\alpha^{-1})\mu_i^2}{(1+\alpha\mu_i)^2} \right\}, \\
 u_{(p+1)\alpha\alpha}(\beta, \alpha) &= \sum_{i=1}^n \left\{ \sum_{j=0}^{y_i-1} \frac{2j^3}{(\alpha j+1)^3} + \frac{6\log(\alpha\mu_i+1)}{\alpha^4} - \frac{4\mu_i}{\alpha^3(\alpha\mu_i+1)} - \frac{3\mu_i^2}{\alpha^2(\alpha\mu_i+1)^2} - \frac{2\left(\frac{1}{\alpha}+y_i\right)\mu_i^3}{(\alpha\mu_i+1)^3} \right\}.
 \end{aligned}$$



# Appendice B

## Codice R utilizzato

### Codice B.1: File EBRNB.R

```
#logverosimiglianza negativa
nlogL=function(par,X,y){
  p = NCOL(X)
  n = NROW(X)
  beta = par[1:p]
  alpha = par[p+1]
  if(alpha<=0){
    return( -Inf)
  }
  eta = as.vector(X%*%beta)
  mu = exp(eta)
  return(-sum(dnbinom(x=y, size=1/alpha, mu=mu, log = TRUE)))
}

#score di verosimiglianza
score=function(par,X,y){
  p <- NCOL(X)
  beta <- par[1:p]
  alpha <- par[p+1]
  if(alpha<=0){
    return( -Inf)
  }
}
```

```

}
eta = as.vector(X%*%beta)
mu = exp(eta)
scorebeta=t(X)%*%((y-mu)/(1+alpha*mu))
scorealpha=sum(alpha^(-2)*log(1+alpha*mu)- (y+1/alpha)*mu/(alpha*mu+1)+
  sapply(y,function(x){if (x >= 1 ){
    z=0:(x-1)
    return(sum(z/(z*alpha+1)))}
  else {return(0)}}))

return(c(scorebeta,scorealpha))
}

#Informazione di Fisher con approssimazione del valore atteso
FisherInfo_ <- function(par,X,y_max=2500) {
  # par is beta,kappa
  # X is the design matrix
  p <- ncol(X)
  beta <- par[1:p]
  kappa <- par[p+1]
  eta <- as.vector(X%*%beta)
  mu <- exp(eta)
  i_bb <- t(X)%*%diag(mu/(1+kappa*mu))%*%X
  yval <- 2:y_max
  probs <- matrix(dnbinom(yval,mu=rep(mu,each=length(yval)),size=1/kappa),
    nrow=length(yval),byrow=FALSE)
  ff <- cumsum(((yval-1)^2/(1+kappa*(yval-1))^2)
  A <- colSums(probs*ff)
  i_kk <- sum(A+2/kappa^3*log(1+kappa*mu)-2*mu*kappa^-2/(1+kappa*mu)-mu^2*
    kappa^-1/(1+kappa*mu))
  out <- matrix(0,p+1,p+1)
  out[1:p,1:p] <- i_bb
  out[p+1,p+1] <- i_kk
  return(out)
}

```

```

#Informazione osservata
ObsInfo=function(par,X,y){
  p <- NCOL(X)
  beta <- par[1:p]
  alpha <- par[p+1]
  eta = as.vector(X%*%beta)
  mu = exp(eta)

  obsinfo=matrix(0,p+1,p+1)
  obsinfo[1:p,1:p]=t(X)%*%diag(mu*(1+alpha*y)/(1+alpha*mu)^2)%*%X
  obsinfo[p+1,1:p]=obsinfo[1:p,p+1]=t(X)%*%(mu*(y-mu)/(1+alpha*mu)^2)
  obsinfo[p+1,p+1]= sum(2*alpha^(-3)*log(1+alpha*mu)+
                        sapply(y,function(x){if (x >= 1){
                          z=0:(x-1)
                          return(sum((z/(1+alpha*z))^2))}
                        else{return(0)}}))
                        -2*alpha^(-2)*mu/(1+alpha*mu)
                        -(y+alpha^-1)*mu^2/(1+alpha*mu)^2)

  return(obsinfo)
}

e=function(par,X,y){
  p = NCOL(X)
  n = NROW(X)
  beta <- par[1:p]
  alpha <- par[p+1]
  eta = as.vector(X%*%beta)
  mu = exp(eta)
  psi=matrix(0,n,p+1)
  psi[1:n,1:p]=X*(y-mu)/(1+alpha*mu)

  q=sapply(1:n,function(i){
    if (y[i] >= 1 ){

```

```

        z=0:(y[i]-1)
        return(sum((z/(1+alpha*z))))
    }
    else {return(0)}
})

psi[1:n,p+1]=alpha^(-2)*log(1+alpha*mu) - (y+1/alpha)*mu/(alpha*mu+1)+q

return(t(psi)%*%psi)
}

dr=function(par,X,y,r){
  p = NCOL(X)
  n = NROW(X)
  beta <- par[1:p]
  alpha <- par[p+1]
  eta = as.vector(X%*%beta)
  mu = exp(eta)
  tildepsi=matrix(0,n,p+1)
  psi=matrix(0,n,p+1)
  psi[1:n,1:p]=X*(y-mu)/(1+alpha*mu)
  psi[1:n,p+1]=alpha^(-2)*log(1+alpha*mu) - (y+1/alpha)*mu/(alpha*mu+1)+
    sapply(y,function(x){if (x >= 1 ){
      z=0:(x-1)
      return(sum(z/(z*alpha+1)))}
      else {return(0)}})

  if (r<=p){

    tildepsi[1:n,1:p]=-mu*(1+alpha*y)*X[,r]*X/(1+alpha*mu)^2
    tildepsi[1:n,p+1]=-mu*(y-mu)*X[,r]/(1+alpha*mu)^2

  }
  else{

```

```

tildepsi[1:n,1:p]=-mu*(y-mu)*X/(1+alpha*mu)^2

q=sapply(y,function(x){if (x >= 1){
  z=0:(x-1)
  return(sum((z/(1+alpha*z))^2))}
else{return(0)}})

tildepsi[1:n,p+1]=sapply(1:n,function(i){
  if (y[i] >= 1 ){
    z=0:(y[i]-1)
    q=sum((z/(1+alpha*z))^2)
  }
  else {q=0}

  return(-q-2*alpha^(-3)*log(1+alpha*mu[i])
    +2*alpha^(-2)*mu[i]/(1+alpha*mu[i])
    +(y[i]+alpha^(-1))*mu[i]^2/(1+alpha*mu[i])^2)

})

}

return(crossprod(tildepsi,psi))
}

ur=function(par,X,y,r){
  p = NCOL(X)
  n = NROW(X)
  beta <- par[1:p]
  alpha <- par[p+1]
  eta = as.vector(X%*%beta)
  mu = exp(eta)
  ur_matrix=matrix(0,p+1,p+1)
  if (r <= p){

```

```

ur_matrix[1:p,1:p]=t(X)%%diag((alpha*y+1)*mu*(alpha*mu-1)*X[,r]/(
  alpha*mu+1)^3)%%X
ur_matrix[1:p,p+1]=ur_matrix[p+1,1:p]=t(X)%%(mu*(alpha*y*mu+2*mu-y)*X
  [,r]/(alpha*mu+1)^3)
ur_matrix[p+1,p+1]=sum(2*mu^2*(y-mu)*X[,r]/(alpha*mu+1)^3)
}
else{
ur_matrix[1:p,1:p]=t(X)%%diag(mu*(alpha*y*mu+2*mu-y)/(alpha*mu+1)^3)%
  %X
ur_matrix[1:p,p+1]=ur_matrix[p+1,1:p]=t(X)%%(2*mu^2*(y-mu)/(alpha*mu
  +1)^3)
ur_matrix[p+1,p+1]=sum(
  sapply(y,function(x){if (x >= 1){
    z=0:(x-1)
    return(sum((2*z^3/(1+alpha*z)^3)))}
    else{return(0)}))}
+6*log(alpha*mu+1)/alpha^4
-6*mu/(alpha^3*(alpha*mu+1))
-3*mu^2/(alpha^2*(alpha*mu+1)^2)
-2*(y+1/alpha)*mu^3/(alpha*mu+1)^3)

}

return(ur_matrix)
}

#Modificazione empirica per lo score di verosimiglianza
emp_mod_score <- function(par,X,y,inverse_ObsInfo){
  p=length(par)
  return(score(par,X,y)+sapply(1:p,function(x){-sum(diag(inverse_ObsInfo%*
    %dr(par,X,y,x))-0.5*sum(diag(inverse_ObsInfo%%e(par,X,y)%%t(
    inverse_ObsInfo)%%ur(par,X,y,x))))}))
}

```

```
}

#Funzione per stimare con uno dei metodi
process<- function(par,data,type="EBR",maxit=10000,tol=1e-04,alpha=0.05) {
  p=length(par)

  if (type=="EBR" || type == "ML"){

    inverse_obsinfo <- try(solve(ObsInfo(par,data$X,data$y)),silent = TRUE)
    if (failed <- inherits(inverse_obsinfo, "try-error")) {

      warning("failed to invert the information matrix: iteration stopped
              prematurely")
      break
    }

  }

  if (type == "ML"){
    adjScore = score(par,data$X,data$y)
  }
  else{
    adjScore = emp_mod_score(par,data$X,data$y,inverse_obsinfo)
  }
  step <- inverse_obsinfo%*%adjScore

  if(maxit <=0 )
    warning("meanBR and medianBR cannot be performed with maxit <= 0")
  iter <- 0
  if( maxit > 0 ) {
    for (iter in seq.int(maxit)) {
      stepPrev <- step
      stepFactor <- 0
      testhalf <- TRUE

      while (testhalf & stepFactor < 13) {
        par <- par + 10^(-stepFactor)*step
      }
    }
  }
}
```

```

inverse_obsinfo <- try(solve(ObsInfo(par,data$X,data$y)),silent =
  TRUE)
if (failed <- inherits(inverse_obsinfo, "try-error")) {
  warning("failed to invert the observed information matrix:
    iteration stopped prematurely")
  break
}

if (type == 'ML'){
  adjScore = score(par,data$X,data$y)
}
else{
  adjScore = emp_mod_score(par,data$X,data$y,inverse_obsinfo)
}
step <- inverse_obsinfo%*%adjScore
stepFactor <- stepFactor + 1
testhalf <- drop(crossprod(stepPrev) < crossprod(step))

}
if (failed | (all(abs(step) < tol))) {
  break
}
}
}
par <- drop(par)
converged <- maxit > iter
wald_sx=par-sqrt(diag(solve(FisherInfo_(par,data$X))))*qnorm(1-alpha/2)
wald_dx=par+sqrt(diag(solve(FisherInfo_(par,data$X))))*qnorm(1-alpha/2)
return(data.frame(par,wald_sx,wald_dx))
}
else{

data$X=data.frame(data$X)
data$X=data$X[!(apply(data$X,2,function(x){return(all(x==1))}))]
if (!is.vector(data$X)){

```



```
names=colnames(data$X)
data <- as.data.frame(cbind(data$y,data$X))
colnames(data)=c('y',names)
f=paste('y ~ ',paste(names,collapse=' + '))
}
else{
  data <- as.data.frame(cbind(data$y,data$X))
  colnames(data)=c('y',paste('x1'))
  f='y ~ x1'
}

if(type=='AS_mean'){
  fit <- brnb(f,link="log",transformation="identity",start = par[1:(p
-1)],
             data = data)
}

if(type=='AS_median'){
  fit <- brnb(f,link="log",transformation="identity",start = par[1:(p
-1)],
             type = "AS_median",data = data)
}

if(type=='correction'){
  fit <- brnb(f,link="log",transformation="identity",start = par[1:(p
-1)],
             type = "correction",data = data)
}

se=c(summary(fit)$coefficients[,2],summary(fit)$coef.dispersion[2])
return(as.data.frame(cbind(c(fit$coefficients,fit$dispersion),
                           sapply(1:p,function(j){par[j]-se[j]*qnorm(1-alpha/2)
                           })),
```

```
        sapply(1:p,function(j){par[j]+se[j]*qnorm(1-alpha/2)
        }))))
    }
}

#Funzione per stimare in un dataset e memorizzare gli estremi degli
  intervalli
single_estimation <- function(data,theta_init,type="ML",alpha=0.05){
  y=data$y
  X=as.matrix(data$X)
  n=nrow(X)
  p=ncol(X)+1
  proc <- try(process(theta_init,data,type=type),silent = TRUE)
  if (failed <- inherits(proc, "try-error")) {
    warning("Failed to estimate")
    d=data.frame(matrix(0,p,3))
    colnames(d)=c('par','wald_sx','wald_dx')
    return(d)
  }
  else{
    return(proc)
  }
}

#In input lista dei dataset
#In output la single estimation per ogni dataset
multi_estimation=function(dataall,theta_init,type="EBR"){
  require(plyr)
  out=ldply(.data=dataall,.fun=single_estimation,type=type,theta_init=
    theta_init)
  return(out)
}
```

```

#Per un dataset effettua le stime con tutti i metodi
single_estimation_all_method <- function(data,alpha=0.05){
  y=data$y
  X=as.matrix(data$X)
  n=nrow(X)
  p=ncol(X)+1
  d=data.frame(matrix(NA,p,15))
  colnames(d)=c('par_ML', 'par_AS_mean', 'par_AS_median', 'par_correction', '
    par_EBR',
                'waldsx_ML', 'waldsx_AS_mean', 'waldsx_AS_median', 'waldsx_
                correction', 'waldsx_EBR',
                'walddx_ML', 'walddx_AS_mean', 'walddx_AS_median', 'walddx_
                correction', 'walddx_EBR')
  glm_try <- try(glm.nb(data$y~-1+data$X),silent = TRUE)
  if (failed <- inherits(glm_try, "try-error")) {
    warning("Failed to estimate")
    return(d)
  }
  else{
    d[,1]=c(coef(glm_try),1/glm_try$theta)
    inverseinfo=try(solve(FisherInfo_(d[,1],data$X)),silent = TRUE)
    if (failed <- inherits(glm_try, "try-error")) {
      warning("Failed to solve")
      d[,6]=d[,11]=NA}
    else{
      d[,6]=d[,1]-sqrt(diag(inverseinfo))*qnorm(1-alpha/2)
      d[,11]=d[,1]+sqrt(diag(inverseinfo))*qnorm(1-alpha/2)
    }
    for (i in c('AS_mean', 'AS_median', 'correction', 'EBR')){
      proc=try(process(d[,1],data=data,type = i),silent = TRUE)
      if (failed <- inherits(proc, "try-error")) {
        if (i=='AS_mean'){
          d[,2]=d[,7]=d[,12]=NA
        }
        if (i=='AS_median'){

```

```
        d[,3]=d[,8]=d[,13]=NA
    }
    if (i=='correction'){
        d[,4]=d[,9]=d[,14]=NA
    }
    if (i=='EBR'){
        d[,5]=d[,10]=d[,15]=NA
    }
}
else{
    if (i=='AS_mean'){
        d[,2]=proc[,1]
        d[,7]=proc[,2]
        d[,12]=proc[,3]
    }
    if (i=='AS_median'){
        d[,3]=proc[,1]
        d[,8]=proc[,2]
        d[,13]=proc[,3]
    }
    if (i=='correction'){
        d[,4]=proc[,1]
        d[,9]=proc[,2]
        d[,14]=proc[,3]
    }
    if (i=='EBR'){
        d[,5]=proc[,1]
        d[,10]=proc[,2]
        d[,15]=proc[,3]
    }
}
}
return(d)
}
```

```
}

multi_estimation_all_method=function(dataall){
  require(plyr)
  out=ldply(.data=dataall,.fun=single_estimation_all_method)
  return(out)
}

#Elaborazione dei risultati
elaboration=function(output,true_theta){
  p=length(true_theta)
  nstime=NROW(output)
  pu=CI=eqm=bias=matrix(NA,5,p)
  stime=output[,1:5]
  walddx=output[,6:10]
  walddx=output[,11:15]

  for(j in 1:p){
    for(s in 1:5){
      stimeparametri=stime[seq(j,nstime,by = p),s]
      bias[s,j]=round(mean(stimeparametri)-true_theta[j],digits=4)
      pu[s,j]=mean(stimeparametri<=true_theta[j])

      CI[s,j]=1-mean((walddx[seq(j,nstime,by = p),s]>true_theta[j]))-mean(
        walddx[seq(j,nstime,by = p),s]<true_theta[j])
      eqm[s,j]=round(mean(stimeparametri-true_theta[j])^2,digits=4)+var(
        stimeparametri)
    }
  }
  return(rbind(bias,pu,eqm,CI))
}
```

```
#Input: par = vero valore del parametro
#       data = lista contenente la matrice del disegno X e la variabile
#           risposta y
#       n = dimensione dei campioni generati
#       Nsim = numero di campioni da simulare
#       seed = seme di generazione
simulate_nb <- function(pars, x) {
  n = nrow(x)
  p = ncol(x)
  beta = pars[-(p + 1)]
  eta = drop(x %*% beta)
  mu = exp(eta)
  alpha = pars[(p + 1)]
  rnbinom(n=n,size=1/alpha, mu=mu )
}

sim.data <- function(par,data,n,Nsim=10000,seed){
  require(MASS)
  beta=par[1:NCOL(data$X)]
  alpha=par[NCOL(data$X)+1]
  set.seed(seed)
  data.sim <- as.list(numeric(Nsim))

  for (i in 1:Nsim)
  {
    y <- rnbinom(n,size=1/alpha,mu=exp(data$X%*%beta))
    data.sim[[i]] <- list(id=i,y=y,X=data$X)
  }

  attr(data.sim,"beta")=beta
  attr(data.sim,"alpha")=alpha
  return(data.sim)
}
```

## Codice B.3: File EBRGA.R

```

#logverosimiglianza negativa
nlogL=function(par,X,y){
  p = NCOL(X)
  n = NROW(X)
  beta = par[1:p]
  alpha = par[p+1]
  eta = as.vector(X%*%beta)
  mu = exp(eta)
  return(-n*alpha*log(alpha)+alpha*sum(log(mu))+alpha*sum(y/mu)-(alpha-1)*
    sum(log(y))+n*lgamma(alpha))
}

#score di verosimiglianza
score=function(par,X,y){
  p = NCOL(X)
  n = NROW(X)
  beta = par[1:p]
  alpha = par[p+1]
  eta = as.vector(X%*%beta)
  mu = exp(eta)
  scoreb=-alpha*t(X)%*%(1-y/mu)
  scorea=n*(log(alpha)+1)-n*digamma(alpha)+sum(log(y/mu)-(y/mu))
  return(c(scoreb,scorea))
}

#Informazione di Fisher con approssimazione del valore atteso
FisherInfo_ <- function(par,X) {
  p = NCOL(X)
  n = NROW(X)
  beta = par[1:p]
  alpha = par[p+1]
  if(alpha<=0){
    return(-Inf)
  }
}

```

```

}
eta = as.vector(X**beta)
mu = exp(eta)
info=matrix(0,p+1,p+1)
info[1:p,1:p] <- alpha*t(X)**X
info[p+1,p+1] <- -n/alpha+n*trigamma(alpha)
return(info)
}

#Informazione osservata
ObsInfo=function(par,X,y){
  p <- NCOL(X)
  beta <- par[1:p]
  alpha <- par[p+1]
  eta = as.vector(X**beta)
  mu = exp(eta)
  n=NROW(X)
  obsinfo=matrix(0,p+1,p+1)
  obsinfo[1:p,1:p]=t(X)**diag(alpha*y/mu)**X
  obsinfo[p+1,1:p]=obsinfo[1:p,p+1]=t(X)**(1-y/mu)
  obsinfo[p+1,p+1]= -n/alpha+n*trigamma(alpha)

  return(obsinfo)
}

e=function(par,X,y){
  p = NCOL(X)
  n = NROW(X)
  beta <- par[1:p]
  alpha <- par[p+1]
  eta = as.vector(X**beta)
  mu = exp(eta)
  psi=matrix(0,n,p+1)
  psi[1:n,1:p]=X*(-alpha*(1-y/mu))
  psi[1:n,p+1]=log(alpha)+1-(y/mu)+log(y/mu)-digamma(alpha)

```



```
    return(t(psi)%*%psi)
  }

dr=function(par,X,y,r){
  p = NCOL(X)
  n = NROW(X)
  beta <- par[1:p]
  alpha <- par[p+1]
  eta = as.vector(X%*%beta)
  mu = exp(eta)
  tildepsi=matrix(0,n,p+1)
  psi=matrix(0,n,p+1)
  psi[1:n,1:p]=X*(-alpha*(1-y/mu))
  psi[1:n,p+1]=log(alpha)+1-(y/mu)+log(y/mu)-digamma(alpha)

  if (r<=p){

    tildepsi[1:n,1:p]=-alpha*y/mu*X[,r]*X
    tildepsi[1:n,p+1]=-(1-y/mu)*X[,r]
  }
  else{

    tildepsi[1:n,1:p]=-(1-y/mu)*X
    tildepsi[1:n,p+1]=1/alpha-trigamma(alpha)

  }
  return(crossprod(tildepsi,psi))
}

ur=function(par,X,y,r){
  p = NCOL(X)
  n = NROW(X)
  beta <- par[1:p]
  alpha <- par[p+1]
```

```

eta = as.vector(X%%beta)
mu = exp(eta)
ur_matrix=matrix(0,p+1,p+1)
if (r <= p){
  ur_matrix[1:p,1:p]=t(X)%%diag(alpha*X[,r]*y/mu)%%X
  ur_matrix[1:p,p+1]=ur_matrix[p+1,1:p]=t(X)%%(-y/mu*X[,r])
  ur_matrix[p+1,p+1]=0
}
else{
  ur_matrix[1:p,1:p]=t(X)%%diag(-y/mu)%%X
  ur_matrix[1:p,p+1]=ur_matrix[p+1,1:p]=0
  ur_matrix[p+1,p+1]=-n/alpha^2-n*psigamma(alpha,deriv=2)
}
return(ur_matrix)
}

#Modificazione empirica per lo score di verosimiglianza
emp_mod_score <- function(par,X,y){
  p = NCOL(X)
  solveinfo=solve(ObsInfo(par,X,y))
  return(score(par,X,y)+sapply(1:(p+1),function(x){-sum(diag(solveinfo%%
    dr(par,X,y,x)))-0.5*sum(diag(solveinfo%%e(par,X,y)%%t(solveinfo)%
    %ur(par,X,y,x))}))})
}

#Funzione per stimare con uno dei metodi
process<- function(par,data,type="EBR",alpha=0.05) {
  p=length(par)

  if (type == "ML"){

```

```

out <- try(multiroot(score,par,X=data$X,y=data$y),silent = TRUE)
if (failed <- inherits(out, "try-error")) {

  warning("Errore nell'ottimizzazione")
  break

}

}

if (type == 'EBR'){

  out <- try(multiroot(emp_mod_score,par,X=data$X,y=data$y),silent =
    TRUE)
  if (failed <- inherits(out, "try-error")) {

    warning("Errore nell'ottimizzazione")
    break

  }
}
par <- drop(out$root)
wald_sx=par-sqrt(diag(solve(FisherInfo_(par,data$X))))*qnorm(1-alpha/
  2)
wald_dx=par+sqrt(diag(solve(FisherInfo_(par,data$X))))*qnorm(1-alpha/
  2)
return(data.frame(par,wald_sx,wald_dx))
}

```

#Funzione per stimare in un dataset e memorizzare gli estremi degli intervalli

```

single_estimation <- function(data,theta_init,type="ML",alpha=0.05){
  y=data$y
  X=as.matrix(data$X)
  n=nrow(X)
  p=ncol(X)+1
  proc <- try(process(theta_init,data,type=type),silent = TRUE)

```

```
if (failed <- inherits(proc, "try-error")) {
  warning("Failed to estimate")
  d=data.frame(matrix(0,p,3))
  colnames(d)=c('par', 'wald_sx', 'wald_dx')
  return(d)
}
else{
  return(proc)
}
}

#In input lista dei dataset
#In output la single estimation per ogni dataset
multi_estimation=function(dataall,theta_init,type="EBR"){
  require(plyr)
  out=ldply(.data=dataall,.fun=single_estimation,type=type,theta_init=
    theta_init)
  return(out)
}

#Per un dataset effettua le stime con tutti i metodi
single_estimation_all_method <- function(data,alpha=0.05){
  y=data$y
  X=as.matrix(data$X)
  n=nrow(X)
  p=ncol(X)+1
  d=data.frame(matrix(NA,p,15))
  colnames(d)=c('par ML', 'par AS mean', 'par AS median', 'par_correction', '
    par_EBR',
```

```

      'waldsx_ML', 'waldsx_AS_mean', 'waldsx_AS_median', 'waldsx_
      correction', 'waldsx_EBR',
      'walddx_ML', 'walddx_AS_mean', 'walddx_AS_median', 'walddx_
      correction', 'walddx_EBR')
glm_try <- try(glm(data$y~-1+data$X, family = Gamma(link='log')), silent =
  TRUE)
if (failed <- inherits(glm_try, "try-error")) {
  warning("Failed to estimate")
  return(d)
}
else{

  d[,1]=c(coef(glm_try), gamma.shape(glm_try)$alpha)
  inverseinfo=try(solve(FisherInfo_(d[,1], data$X)), silent = TRUE)
  if (failed <- inherits(inverseinfo, "try-error")) {
    warning("Failed to solve")
    d[,6]=d[,11]=NA
    return(d)}
  else{
    d[,6]=d[,1]-sqrt(diag(inverseinfo))*qnorm(1-alpha/2)
    d[,11]=d[,1]+sqrt(diag(inverseinfo))*qnorm(1-alpha/2)
  }
}

proc=try(update(glm_try, method = "brglmFit", type = "AS_mean"), silent =
  TRUE)
if (failed <- inherits(proc, "try-error")) {
  d[,2]=d[,7]=d[,12]=NA

}

else{
  d[,2]=c(coef(proc), gamma.shape(proc)$alpha)
  inverseinfo=try(solve(FisherInfo_(d[,2], data$X)), silent = TRUE)
  if (failed <- inherits(inverseinfo, "try-error")) {
    warning("Failed to solve")

```

```
    d[,7]=d[,12]=NA
  }
  else{
    d[,7]=d[,2]-sqrt(diag(inverseinfo))*qnorm(1-alpha/2)
    d[,12]=d[,2]+sqrt(diag(inverseinfo))*qnorm(1-alpha/2)
  }
}

proc=try(update(glm_try, method = "brglmFit", type = "AS_median"),silent
          = TRUE)
if (failed <- inherits(proc, "try-error")) {
  d[,3]=d[,8]=d[,13]=NA
}

else{
  d[,3]=c(coef(proc),gamma.shape(proc)$alpha)
  inverseinfo=try(solve(FisherInfo(d[,3],data$X)),silent = TRUE)
  if (failed <- inherits(inverseinfo, "try-error")) {
    warning("Failed to solve")
    d[,8]=d[,13]=NA
  }
  else{
    d[,8]=d[,3]-sqrt(diag(inverseinfo))*qnorm(1-alpha/2)
    d[,13]=d[,3]+sqrt(diag(inverseinfo))*qnorm(1-alpha/2)
  }
}

proc=try(update(glm_try, method = "brglmFit", type = "correction"),
          silent = TRUE)
if (failed <- inherits(proc, "try-error")) {
  d[,4]=d[,9]=d[,14]=NA
}

else{
```

```

d[,4]=c(coef(proc),gamma.shape(proc)$alpha)
inverseinfo=try(solve(FisherInfo_(d[,4],data$X)),silent = TRUE)
if (failed <- inherits(inverseinfo, "try-error")) {
  warning("Failed to solve")
  d[,9]=d[,14]=NA
}
else{
  d[,9]=d[,4]-sqrt(diag(inverseinfo))*qnorm(1-alpha/2)
  d[,14]=d[,4]+sqrt(diag(inverseinfo))*qnorm(1-alpha/2)
}
}

proc=try(process(d[,1],data=data,type = 'EBR'),silent = TRUE)
if (failed <- inherits(proc, "try-error")) {
  d[,5]=d[,10]=d[,15]=NA
}

else{
  d[,5]=proc[,1]
  d[,10]=proc[,2]
  d[,15]=proc[,3]
}

return(d)
}

multi_estimation_all_method=function(dataall){
  require(plyr)
  out=ldply(.data=dataall,.fun=single_estimation_all_method)
  return(out)
}

#Elaborazione dei risultati
elaboration=function(output,true_theta){
  p=length(true_theta)

```

```
nstime=NROW(output)
pu=CI=eqm=bias=matrix(NA,5,p)
stime=output[,1:5]
waldsx=output[,6:10]
walddx=output[,11:15]

for(j in 1:p){
  for(s in 1:5){
    stimeparametri=stime[seq(j,nstime,by = p),s]
    bias[s,j]=round(mean(stimeparametri)-true_theta[j],digits=4)
    pu[s,j]=mean(stimeparametri<=true_theta[j])

    CI[s,j]=1-mean((waldsx[seq(j,nstime,by = p),s]>true_theta[j]))-mean(
      walddx[seq(j,nstime,by = p),s]<true_theta[j])
    eqm[s,j]=round(mean(stimeparametri-true_theta[j])^2,digits=4)+var(
      stimeparametri)

  }
}
return(rbind(bias,pu,eqm,CI))
}
```