

Università degli studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per le Tecnologie e le Scienze



RELAZIONE FINALE

ALCUNE ANOMALIE DEI TEST DI VEROSIMIGLIANZA

Relatore Prof. Alessandra Salvan
Dipartimento di Scienze Statistiche

Laureando Francesco Masiero
Matricola N 1103023

Anno Accademico 2016/2017

Indice

Introduzione	5
1 L'inferenza di verosimiglianza	7
1.1 Introduzione	7
1.2 Specificazione del modello	7
1.3 La funzione di verosimiglianza	8
1.3.1 Quantità di verosimiglianza	9
1.4 Risultati asintotici e test	10
1.5 Riparametrizzazioni	12
1.5.1 Invarianza dei test rispetto alla parametrizzazione . . .	13
1.6 Famiglie esponenziali e modelli lineari generalizzati (MLG) . .	14
1.6.1 Famiglie esponenziali di ordine uno	14
1.6.2 Modelli lineari generalizzati (MLG)	15
2 Anomalie del test di Wald	18
2.1 Comportamento anomalo nel modello di regressione logistica .	18
2.2 Il test di Wald in famiglie esponenziali di ordine uno	21
3 Il modello di Poisson	24
3.1 Il caso di osservazioni omogenee	24
3.2 Modelli di regressione di Poisson	30
3.2.1 Legame canonico	30
3.2.2 Legame identità	33
3.2.3 Legame radice quadrata	34

4	Le principali funzioni di legame per dati binari	37
4.1	Il caso di osservazioni omogenee	37
4.1.1	Parametrizzazione probit	37
4.1.2	Parametrizzazione log-log complementare	40
4.1.3	Parametrizzazione log-log	42
4.1.4	Confronto tra le parametrizzazioni	43
4.2	Anomalie nel modello di regressione probit	45
5	Conclusioni	47
	Bibliografia	49

Introduzione

Uno dei test statistici più utilizzati e comuni è il test di Wald, che per la sua semplicità e intuitività è spesso riportato di *default* negli *output* di *software* statistici. Nonostante la sua vasta diffusione, il test di Wald presenta alcune lacune, in primis la non invarianza rispetto a riparametrizzazioni. Un secondo aspetto problematico è la possibile non monotonicità all'allontanarsi della stima di massima verosimiglianza dal valore sotto l'ipotesi nulla. Tale fenomeno è stato evidenziato da Hauck e Donner (1977) per i test su un coefficiente di regressione nel modello di regressione logistica e ulteriormente esplorato da Væth (1985) nelle famiglie esponenziali di ordine uno. Lo scopo della relazione è fornire un approfondimento del 'fenomeno di Hauck-Donner' nell'ambito di modelli lineari generalizzati per dati binari e Poisson.

Il capitolo 1 richiama gli argomenti riguardanti la verosimiglianza, i test statistici e i modelli lineari generalizzati, che compaiono nel corso di tutta la trattazione, e per questo se ne presentano le nozioni principali e ne viene stabilita la notazione.

Il capitolo 2 analizza in dettaglio gli articoli di Hauck e Donner (1977) e Væth (1985). Il primo è relativo al comportamento anomalo del test di Wald nel modello di regressione logistica. Il secondo specifica le condizioni per cui il test di Wald si comporta in modo non coerente nella famiglia di dispersione esponenziale di ordine uno e nei modelli lineari generalizzati al variare, rispettivamente, della parametrizzazione e della funzione di legame.

Il capitolo 3 presenta in dettaglio il comportamento del test di Wald nel modello di Poisson, sia nel caso di osservazioni da popolazione omogenea, sia in ambito di regressione tramite un semplice esempio.

Il capitolo 4 amplia lo studio del 'fenomeno di Hauck-Donner' ai model-

li per dati binari, sia in ambito di osservazioni da popolazione omogenea che in ambito di regressione, alle parametrizzazioni e funzioni di legame più utilizzate.

Nel capitolo 5 si riassumono i risultati ottenuti.

Capitolo 1

L'inferenza di verosimiglianza

1.1 Introduzione

Il presente capitolo è sostanzialmente un richiamo ai concetti base dell'inferenza statistica basata sulla verosimiglianza e servirà a fissare le notazioni utilizzate nel seguito. La presentazione è basata su Pace e Salvani (2001, §1-6) e Pace e Salvani (1996, §1-3). Il sesto paragrafo è dedicato alle famiglie esponenziali e ai modelli lineari generalizzati (MLG), classe di modelli su cui è incentrata la discussione ed è basata su Pace e Salvani (1996, §5-6), Agresti (2015, §4-7), Pace e Salvani (2001, §8-10).

1.2 Specificazione del modello

L'inferenza statistica parte da una fondamentale idealizzazione, ossia che i dati osservati $y = (y_1, \dots, y_n)$ siano una determinazione di un vettore casuale $Y \sim P^0(y)$, con spazio campionario \mathcal{Y} , detto modello probabilistico generatore dei dati. Essendo $P^0(y)$ almeno in parte ignoto, l'inferenza statistica si pone come obiettivo la ricostruzione del modello probabilistico utilizzando l'informazione contenuta nei dati.

La prima operazione che consente di semplificare il problema è di restringere le forme possibili di $P^0(y)$, e quindi viene specificata una famiglia di distribuzioni \mathcal{F} , compatibili con il modello generatore dei dati. L'insieme di

distribuzioni appartenenti a \mathcal{F} costituisce il modello statistico. Nel caso in cui $P^0(y) \in \mathcal{F}$, si dice che il modello è correttamente specificato. Il modello \mathcal{F} può appartenere a una delle seguenti classi: modello non parametrico, modello semi-parametrico, modello parametrico. Nello sviluppo della tesi ci si occuperà solo di quest'ultima, la cui definizione formale è

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

per qualche $p \in \mathbb{N}^+$. Lo spazio Θ è detto spazio parametrico. Solitamente la corrispondenza tra Θ e \mathcal{F} è biunivoca. Questa proprietà è chiamata identificabilità.

1.3 La funzione di verosimiglianza

Assumendo valido il modello \mathcal{F} per i dati y con funzione del modello $p_Y(y; \theta)$, $\theta \in \Theta$, la funzione di verosimiglianza $L : \Theta \rightarrow \mathbb{R}^+$ definita da:

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

con $c(y)$ costante moltiplicativa non dipendente dai parametri. La verosimiglianza è una classe di funzioni equivalenti che differiscono solo per la costante moltiplicativa $c(y)$. La funzione di verosimiglianza riassume l'informazione dei dati y sul parametro θ . Nel caso in cui i dati siano realizzazioni di n variabili casuali indipendenti e identicamente distribuite (i.i.d.), la verosimiglianza assume la forma

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

con $p_{Y_i}(y_i; \theta)$ distribuzione marginale della variabile casuale Y_i ; in questo caso quindi la funzione di verosimiglianza è il prodotto delle verosimiglianze delle singole y_i .

Spesso si preferisce usare il logaritmo naturale della funzione di verosimiglianza in quanto permette una maggiore facilità di calcolo, pur mantenendo tutta l'informazione di $L(\theta)$. Si definisce funzione di log-verosimiglianza la funzione

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

Per convenzione si pone $l(\theta) = -\infty$ se $L(\theta) = 0$. Nel caso di osservazioni i.i.d. la log-verosimiglianza assume la forma

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

1.3.1 Quantità di verosimiglianza

Stima di massima verosimiglianza

Un valore $\hat{\theta}(y) = \hat{\theta} \in \Theta$ tale che $L(\hat{\theta}) \geq L(\theta)$ per ogni θ è detto stima di massima verosimiglianza di θ . In generale non è detto che $\hat{\theta}$ esista o sia unico. Se $\hat{\theta} = \hat{\theta}(y)$ esiste unico con probabilità uno, la variabile casuale $\hat{\theta}(Y)$ è detto stimatore di massima verosimiglianza.

La verosimiglianza si definisce regolare se:

- Θ è un sottoinsieme aperto di \mathbb{R}^p ;
- $l(\theta)$ è differenziabile almeno tre volte con derivate parziali continue in Θ ;
- il modello è identificabile;
- le funzioni di ripartizione/densità di probabilità hanno tutte lo stesso supporto non dipendente dal parametro.

Nel caso di verosimiglianza regolare, le informazioni sulla funzione di verosimiglianza sono contenute in $\hat{\theta}$ e nelle derivate parziali della funzione di verosimiglianza.

Funzione score

Dato il parametro $\theta = (\theta_1, \dots, \theta_p)$, il vettore delle derivate parziali

$$l_*(\theta) = \left(\frac{\partial l(\theta)}{\partial \theta_1}, \dots, \frac{\partial l(\theta)}{\partial \theta_p} \right) = \left[\frac{\partial l(\theta)}{\partial \theta_r} \right] = [l_r(\theta)]$$

è detto funzione *score*. Se il modello ha verosimiglianza regolare, la soluzione di $l_*(\theta) = 0$ spesso individua $\hat{\theta}$.

Informazione osservata e attesa

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

$$j(\theta) = -l_{**}(\theta) = -\left[\frac{\partial^2 l(\theta)}{\partial \theta_r \partial \theta_s}\right] = [j_{rs}(\theta)] \quad (1.1)$$

è detta matrice di informazione osservata, e altro non è che l'hessiano di $-l(\theta)$ e in quanto tale ne definisce la curvatura. Il valore atteso, sotto θ della matrice di informazione osservata è la matrice di informazione attesa

$$i(\theta) = E_\theta[j(\theta)] = [i_{rs}(\theta)]$$

Se si assume che si possa scambiare l'operazione di derivazione (rispetto a θ) e integrazione (rispetto a y), valgono anche le seguenti proprietà:

- sotto θ , ossia quando θ è il vero valore del parametro, il valore atteso della funzione score è nullo: $E_\theta[l_*(\theta)] = 0$ per ogni $\theta \in \Theta$;
- vale l'identità dell'informazione, ossia $i(\theta) = E_\theta[l_*(\theta)l_*(\theta)^\top]$ per ogni $\theta \in \Theta$. L'informazione attesa è quindi la matrice di covarianza del vettore score, e come tale è una matrice definita non negativa.

1.4 Risultati asintotici e test

Sotto tenui condizioni di regolarità lo stimatore di massima verosimiglianza è consistente. Sotto ulteriori condizioni, per θ parametro p -dimensionale, valgono i seguenti risultati asintotici, sotto θ .

$$l_*(\theta) \dot{\sim} N_p(0, i(\theta)), \quad (1.2)$$

$$(\hat{\theta} - \theta) \dot{\sim} N_p(0, i(\theta)^{-1}), \quad (1.3)$$

inoltre

$$W_e(\theta) = (\hat{\theta} - \theta)^\top i(\theta) (\hat{\theta} - \theta) \dot{\sim} \chi_p^2, \quad (1.4)$$

$$W_u(\theta) = l_*(\theta)^\top i(\theta)^{-1} l_*(\theta) \dot{\sim} \chi_p^2, \quad (1.5)$$

$$W(\theta) = 2\{l(\hat{\theta}) - l(\theta)\} \dot{\sim} \chi_p^2. \quad (1.6)$$

Nel seguito quando si studierà l'andamento del test di Wald in funzione di $\hat{\theta}$, si userà la notazione $W_e(\theta) = W_e(\theta; \hat{\theta})$.

Le quantità $W_u(\theta)$, $W_e(\theta)$ e $W(\theta)$ sono asintoticamente equivalenti e identificano le tre quantità pivotali approssimate basate sulla verosimiglianza, indicate come quantità di Wald, *score* e del rapporto di verosimiglianza. Queste quantità pivotali sono ampiamente utilizzate per costruire regioni di confidenza e test statistici. Nei test Wald e *score* $i(\theta)$ può essere sostituita con stime consistenti quali $i(\hat{\theta})$ e $j(\hat{\theta})$.

Nel seguito si sarà interessati alle versioni di queste quantità per test e regioni di confidenza per un sottoinsieme di componenti di θ . Se $\theta = (\tau, \zeta)$ con τ blocco di p_τ componenti di θ , si possono suddividere le quantità $\hat{\theta}$, $l_*(\theta)$, $i(\theta)$ e $j(\theta)$ nei blocchi di componenti corrispondenti: $\hat{\theta} = (\hat{\tau}, \hat{\zeta})$, $l_*(\theta)^\top = (l_\tau(\theta)^\top, l_\zeta(\theta)^\top)$,

$$j(\theta) = j(\tau, \zeta) = \begin{pmatrix} j_{\tau\tau} & j_{\tau\zeta} \\ j_{\zeta\tau} & j_{\zeta\zeta} \end{pmatrix} \quad i(\theta) = i(\tau, \zeta) = \begin{pmatrix} i_{\tau\tau} & i_{\tau\zeta} \\ i_{\zeta\tau} & i_{\zeta\zeta} \end{pmatrix}. \quad (1.7)$$

Con notazione simile si definiscono le inverse a blocchi

$$j(\theta)^{-1} = j(\tau, \zeta)^{-1} = \begin{pmatrix} j^{\tau\tau} & j^{\tau\zeta} \\ j^{\zeta\tau} & j^{\zeta\zeta} \end{pmatrix} \quad i(\theta)^{-1} = i(\tau, \zeta)^{-1} = \begin{pmatrix} i^{\tau\tau} & i^{\tau\zeta} \\ i^{\zeta\tau} & i^{\zeta\zeta} \end{pmatrix}. \quad (1.8)$$

Valgono le seguenti relazioni per le inverse di matrici a blocchi (analoghe anche per $j(\theta)^{-1}$)

$$\begin{aligned} i^{\tau\tau} &= (i_{\tau\tau} - i_{\tau\zeta} i_{\zeta\zeta}^{-1} i_{\zeta\tau})^{-1} \\ i^{\tau\zeta} &= -i^{\tau\tau} i_{\tau\zeta} i_{\zeta\zeta}^{-1} \\ i^{\zeta\tau} &= -i^{\zeta\zeta} i_{\zeta\tau} i_{\tau\tau}^{-1} \\ i^{\zeta\zeta} &= (i_{\zeta\zeta} - i_{\zeta\tau} i_{\tau\tau}^{-1} i_{\tau\zeta})^{-1}. \end{aligned} \quad (1.9)$$

Per l'inferenza sul parametro p_τ -dimensionale τ , vi sono risultati analoghi a quelli per l'inferenza globale su θ . Si indichi con $\hat{\theta}_\tau$ la stima di massima verosimiglianza di θ nel sottomodulo con τ fissato, $\hat{\theta}_\tau = (\tau, \hat{\zeta}_\tau)$, ottenuta tramite l'equazione di verosimiglianza $l_\zeta(\tau, \zeta) = 0$. Definite queste quantità, si hanno i seguenti risultati asintotici

$$(\hat{\tau} - \tau) \sim N_{p_\tau}(0, i^{\tau\tau}(\theta)) \quad (1.10)$$

$$l_\tau(\hat{\theta}_\tau) \sim N_{p_\tau}(0, i^{\tau\tau}(\theta)^{-1}) \quad (1.11)$$

che portano alle seguenti quantità pivotali

$$W_{eP}(\tau) = (\hat{\tau} - \tau)^\top (i^{\tau\tau}(\theta))^{-1} (\hat{\tau} - \tau) \sim \chi_{p_\tau}^2, \quad (1.12)$$

$$W_{uP}(\tau) = l_\tau(\hat{\theta}_\tau)^\top i^{\tau\tau}(\theta_\tau) l_\tau(\hat{\theta}_\tau) \sim \chi_{p_\tau}^2, \quad (1.13)$$

$$W_P = 2\{l(\hat{\theta}) - l(\hat{\theta}_\tau)\} \sim \chi_{p_\tau}^2. \quad (1.14)$$

Anche in questo caso, nella definizione di W_{eP} e W_{uP} spesso si sostituisce $i^{\tau\tau}(\theta)$ con $i^{\tau\tau}(\cdot)$, calcolata in $\hat{\theta}$ o $\hat{\theta}_\tau$, o con $j^{\tau\tau}(\hat{\theta})$. Solitamente è utilizzabile anche $j^{\tau\tau}(\hat{\theta}_\tau)$, che però, in qualche caso può non essere definita positiva, portando a valori negativi della statistica test (si veda Freedman, 2007).

1.5 Riparametrizzazioni

La funzione di verosimiglianza e il suo logaritmo non dipendono dalla parametrizzazione scelta per \mathcal{F} . Sia $\psi = \psi(\theta)$, con $\psi(\cdot)$ funzione biunivoca e regolare, ossia infinitamente derivabile con inversa infinitamente derivabile, definita da $\Theta \subseteq \mathbb{R}^p$ in $\Psi \subseteq \mathbb{R}^p$, una parametrizzazione alternativa del modello. Dal momento che sia θ che $\psi(\theta)$ individuano lo stesso elemento di \mathcal{F} , si ha

$$L^\Psi(\psi) = L^\Theta(\theta(\psi)), \quad (1.15)$$

$$l^\Psi(\psi) = l^\Theta(\theta(\psi)). \quad (1.16)$$

Vale inoltre la proprietà di equivarianza rispetto alla parametrizzazione della stima di massima verosimiglianza, ossia $\hat{\psi} = \psi(\hat{\theta})$ e $\hat{\theta} = \theta(\hat{\psi})$

Inoltre anche le altre quantità di verosimiglianza, come la funzione score e le matrici di informazione osservata e attesa variano secondo schemi regolari. Indicati con ψ^a, ψ^b, \dots ($a, b = 1, \dots, p$) le generiche componenti di ψ e θ^r, θ^s le generiche componenti di θ , la funzione score vale

$$l_a^\Psi = \frac{\partial l^\Psi(\psi)}{\partial \psi^a} = \sum_{r=1}^p l_r \theta_a^r, \quad (1.17)$$

con $l_r = l_r(\theta(\psi))$ e $\theta_a^r = (\partial \theta^r(\psi) / \partial \psi^a)$.

Per le derivate parziali seconde, e quindi per gli elementi della matrice di informazione osservata, si ha

$$l_{ab}^{\Psi} = \frac{\partial^2 l^{\Psi}(\psi)}{\partial \psi^a \partial \psi^b} = \sum_{r,s=1}^p l_{rs} \theta_a^r \theta_b^s + \sum_{r=1}^p l_r \theta_{ab}^r, \quad (1.18)$$

con $\theta_{ab}^r = (\partial^2 \theta^r(\psi) / \partial \psi^a \partial \psi^b)$.

Per l'informazione attesa nella nuova parametrizzazione, posto i_{rs} l'elemento di posizione (r, s) della matrice di informazione attesa nella parametrizzazione θ , vale

$$i_{ab}^{\Psi} = \sum_{r,s=1}^p i_{rs} \theta_a^r \theta_b^s. \quad (1.19)$$

Nel caso monparametrico il tutto si semplifica e si ha

$$l_{*}^{\Psi}(\psi) = l_{*}^{\Theta}(\theta(\psi))[\theta'(\psi)], \quad (1.20)$$

$$i^{\Psi}(\psi) = i^{\Theta}(\theta(\psi))[\theta'(\psi)]^2, \quad (1.21)$$

$$j^{\Psi}(\psi) = j^{\Theta}(\theta(\psi))[\theta'(\psi)]^2 + [l_{*}^{\Theta}(\theta(\psi))]^2 \theta''(\psi). \quad (1.22)$$

Quando (1.18) e (1.22) sono valutate in $\hat{\psi}$, il secondo addendo si annulla.

1.5.1 Invarianza dei test rispetto alla parametrizzazione

Data la proprietà di invarianza della funzione di verosimiglianza rispetto alla parametrizzazione, viene richiesta questa caratteristica anche alle statistiche test utilizzate per valutare ipotesi in parametrizzazioni differenti, che identificano però lo stesso elemento del modello statistico. Le statistiche test di tipo score e rapporto di verosimiglianza sono invarianti rispetto alla parametrizzazione, a differenza del test di Wald, il quale non è invariante rispetto alla parametrizzazione. Nel caso monparametrico, il test di Wald, considerata una parametrizzazione $\psi = \psi(\theta)$ e i problemi di verifica di ipotesi equivalenti $H_0 : \theta = \theta_0$ e $H_0 : \psi = \psi_0 = \psi(\theta_0)$, il test, nella parametrizzazione θ , assume la forma

$$W_e^{\Theta}(\theta_0; \hat{\theta}) = (\hat{\theta} - \theta_0)^2 i^{\Theta}(\hat{\theta}), \quad (1.23)$$

mentre nella parametrizzazione ψ , tenuto conto del principio di equivarianza della stima di massima verosimiglianza e dell'equazione (1.21), vale

$$W_e^\Psi(\psi_0; \psi(\hat{\theta})) = (\psi(\hat{\theta}) - \psi(\theta_0))^{2i^\Theta(\hat{\theta})} \left(\frac{d\theta(\psi)}{d\psi} \Big|_{\psi=\psi(\hat{\theta})} \right)^2. \quad (1.24)$$

In generale il valore delle statistiche test non coincide.

1.6 Famiglie esponenziali e modelli lineari generalizzati (MLG)

1.6.1 Famiglie esponenziali di ordine uno

La famiglia parametrica

$$\mathcal{F}_{en}^1 = \{p(y; \theta) = \exp\{\theta y - K(\theta)\} p_0(y), y \in \mathcal{Y}, \theta \in \Theta\}, \quad (1.25)$$

con $K(\theta) = \log \left(\int_{\mathcal{Y}} e^{\theta y} p_0(y) d\mu \right)$, $\Theta = \{\theta \in \mathbb{R} : K(\theta) < +\infty\}$ è detta famiglia esponenziale naturale di ordine uno generata da $p_0(y)$. Θ è detto spazio parametrico naturale, θ è il parametro naturale, μ è una misura dominante (di Lebesgue o di tipo contatore). La funzione generatrice dei momenti è

$$M_Y(t) = \exp\{K(\theta + t) - K(\theta)\}. \quad (1.26)$$

La funzione generatrice dei cumulanti, definita come $K_Y(t) = \log(M_Y(t))$ è pari a

$$K_Y(t) = K(\theta + t) - K(\theta). \quad (1.27)$$

Con la funzione generatrice dei cumulanti si ricavano immediatamente media e varianza:

$$E_\theta(Y) = k_1(Y) = \frac{d}{dt} K_Y(t) \Big|_{t=0} = \frac{dK(\theta)}{d\theta} = K'(\theta) = \mu. \quad (1.28)$$

$$Var_\theta(Y) = k_2(Y) = \frac{d^2}{dt^2} K_Y(t) \Big|_{t=0} = \frac{d^2 K(\theta)}{d\theta^2} = K''(\theta) = v(\mu). \quad (1.29)$$

Per questo motivo la funzione $K(\cdot)$ è detta funzione generatrice dei cumulanti. Le quantità di verosimiglianza principali in caso di osservazioni y_i ,

$i = 1, \dots, n$, i.i.d. sono

$$l(\theta) = \sum_{i=1}^n y_i \theta - nK(\theta), \quad l_*(\theta) = \sum_{i=1}^n y_i - nK'(\theta), \quad j(\theta) = nK''(\theta) = i(\theta) \quad (1.30)$$

1.6.2 Modelli lineari generalizzati (MLG)

I modelli lineari generalizzati estendono la classe dei modelli lineari normali ad altre distribuzioni. Le componenti di un MLG sono tre: componente casuale, predittore lineare e funzione di legame. Per quanto riguarda la componente casuale, si ipotizza che le osservazioni Y_i , $i = 1, \dots, n$ della variabile risposta siano indipendenti e appartengano a una famiglia di dispersione esponenziale. Tuttavia la trattazione si focalizzerà sulle famiglie esponenziali, principalmente binomiale e Poisson. Restringeremo quindi l'introduzione ai MLG riguardanti solo questi due modelli. Possiamo individuare univocamente la distribuzione tramite la parametrizzazione con media e funzione di varianza, $(\mu_i, v(\mu_i))$, quindi $Y_i \sim DE_1(\mu_i, v(\mu_i))$. Le variabili esplicative sono introdotte nel predittore lineare definito come $\eta_i = \sum_{j=1}^k \beta_j x_{ij}$, $\beta = (\beta_1, \dots, \beta_k)$ vettore di parametri k -dimensionale. La funzione di legame, invertibile e nota, mette in relazione la media della risposta con il predittore lineare. Sia $g(\cdot)$ la funzione di legame, allora $g(\mu_i) = \eta_i$.

Le quantità di verosimiglianza di interesse sono

$$l(\beta) = \sum_{i=1}^n [y_i \theta_i - K(\theta_i)], \quad \text{con } \theta_i = \theta(\mu_i) = \theta(g^{-1}(\eta_i)). \quad (1.31)$$

$$l_r = \frac{\partial l(\beta)}{\partial \beta_r} = \sum_{i=1}^n \left[y_i \frac{\partial \theta_i}{\partial \mu_i} - \frac{\partial K(\theta_i)}{\partial \beta_r} \right] = \sum_{i=1}^n (y_i - \mu_i) \frac{\partial \theta_i}{\partial \beta_r}, \quad (1.32)$$

dove la seconda uguaglianza si ottiene in quanto $\mu_i = g^{-1}(\eta_i)$, $\theta_i = \theta(g^{-1}(\eta_i))$, $\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{K''(\theta_i)} \Big|_{\theta_i=\theta(\mu_i)} = \frac{1}{v(\mu_i)}$ e quindi $\frac{\partial \theta_i}{\partial \beta_r} = \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_r} = \frac{1}{v(\mu_i)} \frac{1}{g'(\mu_i)} x_{ir}$. Le equazioni di verosimiglianza risultano

$$l_r = \sum_{i=1}^n \frac{(y_i - \mu_i)}{v(\mu_i)} \frac{\partial \mu_i}{\partial \beta_r}. \quad (1.33)$$

Le stime di massima verosimiglianza non sempre esistono finite e non sempre le medie stimate appartengono alla parte interna dello spazio delle medie $\mathcal{M} = \left\{ \mu \in \mathbb{R} : \mu_i = g^{-1} \left(\sum_{j=1}^k \beta_j x_{ij} \right), \beta \in \mathbb{R}^k \right\}$. Una spiegazione dettagliata si può trovare in Wedderburn (1976). Per il seguito della discussione saranno rilevanti solamente i risultati relativi ai modelli Poisson e binomiale e alcune loro funzioni di legame notevoli, di cui si discuterà più approfonditamente in seguito. In particolare, nel modello di Poisson si ha che:

- con il legame canonico $g(\mu_i) = \log(\mu_i)$ le stime di β non sono sempre finite,
- nella parametrizzazione $g(\mu_i) = \mu_i^\alpha$, $0 < \alpha \leq 1$ le stime di μ_i non sempre appartengono alla parte interna dello spazio delle medie.

Per quanto riguarda il modello binomiale si ha che:

- con la funzione di legame identità $g(\pi_i) = \pi_i$ e con il legame $g(\pi_i) = \arcsin \sqrt{\pi_i}$ le stime di π_i non sempre appartengono alla parte interna dello spazio delle medie,
- utilizzando le funzioni di legame: canonica $g(\pi_i) = \log[\pi_i/(1 - \pi_i)]$, probit $g(\pi_i) = \Phi^{-1}(\pi_i)$ (con $\Phi(\cdot)$ funzione di ripartizione della normale standard), log-log $g(\pi_i) = -\log[-\log(\pi_i)]$, log-log complementare $g(\pi_i) = \log[-\log(1 - \pi_i)]$ le stime di β non sono sempre finite.

Le altre quantità di verosimiglianza di interesse sono la matrice di informazione osservata e attesa. Si ha

$$j_{rs} = -l_{rs} = \sum_{i=1}^n \left[\frac{\partial \mu_i}{\partial \beta_s} \frac{\partial \theta_i}{\partial \beta_r} - (y_i - \mu_i) \frac{\partial^2 \theta_i}{\partial \beta_r \partial \beta_s} \right] \quad (1.34)$$

$$i_{rs} = \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta_s} \frac{\partial \theta_i}{\partial \beta_r} = \sum_{i=1}^n \frac{x_{ir} x_{is}}{(g'(\mu_i))^2 v(\mu_i)} \quad (1.35)$$

che può essere riscritta in forma matriciale come $i(\beta) = X^\top W X$, con $W = \text{diag}(1/[(g'(\mu_i))^2 v(\mu_i)])$ e X matrice del disegno (con righe x_i corrispondenti alle osservazioni delle variabili esplicative sull'unità statistica i).

Nel caso si utilizzi il legame canonico $\theta_i(\mu_i) = g(\mu_i)$ si ottengono semplificazioni notevoli, dal momento che $\frac{\partial \theta_i}{\partial \beta_r} = x_{ir}$ e $\frac{\partial^2 \theta_i}{\partial \beta_r \partial \beta_s} = 0$. Le componenti del vettore score sono

$$l_r = \sum_{i=1}^n (y_i - \mu_i) x_{ir} \quad (1.36)$$

e le matrici di informazione attesa e osservata coincidono:

$$j_{rs} = \sum_{i=1}^n \frac{\partial \mu_i}{\partial \beta_s} \frac{\partial \theta_i}{\partial \beta_r} = \sum_{i=1}^n v(\mu_i) x_{ir} x_{is} = i_{rs} \quad (1.37)$$

Si ha quindi che la matrice $j(\beta) = i(\beta)$ può essere riscritta in forma matriciale come $X^\top W X$, $W = \text{diag}(v(\mu_i))$.

Capitolo 2

Anomalie del test di Wald

2.1 Comportamento anomalo nel modello di regressione logistica

È stato mostrato da Hauck e Donner (1977) che il test Wald per la valutazione di un'ipotesi nulla per un singolo coefficiente nel modello di regressione logistica presenta un comportamento anomalo. In particolare si ha che la statistica Wald:

- 1) non è monotona crescente al distanziarsi della stima di massima verosimiglianza dal valore ipotizzato dall'ipotesi nulla.
- 2) la potenza del test Wald decresce fino a raggiungere il livello di significatività per alternative distanti dal valore nullo.

La struttura del problema che si analizza è la seguente. Siano Y_1, \dots, Y_n variabili risposta binarie indipendenti, a ognuna delle quali sia associato il vettore $x_i = (x_{i1}, \dots, x_{ik})$ di osservazioni di variabili concomitanti, con $x_{i1} = 1$ per ogni unità statistica e quindi β_1 parametro di intercetta. Si ponga la funzione di legame uguale al legame canonico (logit nel caso di osservazioni bernoulliane)

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = \sum_{j=1}^k \beta_j x_{ij}. \quad (2.1)$$

Si consideri dunque il problema di verifica di ipotesi $H_0 : \beta_k = \beta_k^0$ contro $H_1 : \beta_k \neq \beta_k^0$. Sia $\hat{\beta}_k$ la stima di massima verosimiglianza del parametro β_k e sia $i(\hat{\beta})^{-1}$ la matrice inversa dell'informazione attesa (coincidente con l'informazione osservata, dato l'utilizzo del legame canonico) valutata nella stima di massima verosimiglianza $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_k)$, e sia $i^{kk}(\hat{\beta})$ il suo elemento di posto (k, k) , il quale è anche la stima della varianza di $\hat{\beta}_k$. Il test di Wald definito dalla (1.12), assume la forma

$$W_e(\beta_k; \hat{\beta}) \Big|_{\beta_k = \beta_k^0} = \frac{(\hat{\beta}_k - \beta_k^0)^2}{i^{kk}(\hat{\beta})} \quad (2.2)$$

con distribuzione nulla approssimata χ_1^2 . Hauck e Donner (1977) mostrano che (2.2) tende a zero quando $\hat{\beta}_k$ tende a più o meno infinito, data una dimensione campionaria n finita e fissati i valori di $\hat{\beta}_1, \dots, \hat{\beta}_{k-1}$. Convienne ripercorrere i passi principali della dimostrazione. Si valuti prima la statistica Wald definita in (1.10) con informazione attesa calcolata in β , asintoticamente equivalente a (2.2). Si verifica, seguendo che la potenza del test decresce al livello di significatività (equivalente a dire che la distribuzione del test coincide con la distribuzione nulla) per alternative sempre più distanti dall'ipotesi nulla. Dato $i^{kk}(\beta)$ l'elemento (k, k) di $i(\beta)^{-1}$ valutata nei veri e ignoti valori dei parametri $\beta_j, j = 1, \dots, k$ si ha

$$\frac{\hat{\beta}_k - \beta_k}{(i^{kk}(\beta))^{\frac{1}{2}}} \xrightarrow{d} N(0, 1), \quad (2.3)$$

per $n \rightarrow \infty$. Sottraendo $\frac{\beta_k^0}{(i^{kk}(\beta))^{\frac{1}{2}}}$ e addizionando $\frac{\beta_k^0}{(i^{kk}(\beta))^{\frac{1}{2}}}$ si ha (sempre per $n \rightarrow \infty$):

$$Z = \frac{\hat{\beta}_k - \beta_k^0}{(i^{kk}(\beta))^{\frac{1}{2}}} \sim N\left(\frac{(\beta_k - \beta_k^0)}{(i^{kk}(\beta))^{\frac{1}{2}}}, 1\right). \quad (2.4)$$

Il quadrato della statistica (2.4) ha distribuzione chi-quadro non centrale, con parametro di non centralità ζ , $\chi_1^2(\zeta)$. Il parametro ζ è finito se $\beta_k - \beta_k^0 = \frac{\delta}{\sqrt{n}}$, ed è pari al quadrato della media della normale con varianza unitaria considerata, ossia

$$\zeta = \left(\frac{\beta_k - \beta_k^0}{(i^{kk}(\beta))^{1/2}} \right)^2 = \frac{(\beta_k - \beta_k^0)^2}{i^{kk}(\beta)} \quad (2.5)$$

e

$$Z^2 = \frac{(\hat{\beta}_k - \beta_k^0)^2}{(i^{kk}(\beta))} \quad \dot{\sim} \quad \chi_1^2(\zeta). \quad (2.6)$$

Sotto H_0 , la distribuzione asintotica di (2.4) è una $N(0, 1)$ e di conseguenza la distribuzione del suo quadrato un χ_1^2 , con parametro di non centralità 0. La quantità calcolata in (2.2), sotto β_k , è asintoticamente equivalente al quadrato di Z definita in (2.4), in quanto la matrice di informazione attesa calcolata nella stima di massima verosimiglianza converge alla matrice di informazione attesa calcolata nel vero e ignoto valore del parametro. Sapendo che la matrice di informazione attesa è simmetrica e definita positiva, in quanto matrice di covarianza del vettore *score*, partizionando $i(\beta)$ in blocchi di dimensione $(k-1)$ e 1, si può ricavare l'elemento (k, k) della sua inversa facendo riferimento alle formule (1.9) per le inverse di matrici a blocchi. Si ha quindi, per la (2.5), ed essendo i_{11}^{-1} definita positiva, la disuguaglianza

$$\zeta = (\beta_k - \beta_k^0)^2 (i_{kk} - i_{1k}^\top i_{11}^{-1} i_{1k}) \leq (\beta_k - \beta_k^0)^2 i_{kk}(\beta). \quad (2.7)$$

Sappiamo per la (1.35) che $i(\beta) = (X^\top W X)$, con $W = \text{diag}(v(\mu_1), \dots, v(\mu_n))$ dato il legame canonico, e la funzione di varianza per il modello di regressione logistica è $v(\mu_i) = v(\pi_i) = \pi_i(1 - \pi_i)$. Di conseguenza si ottiene che $i_{kk}(\beta) = \sum_{i=1}^n \pi_i(1 - \pi_i)x_{ik}^2$. Nella parametrizzazione naturale, evidenziando quindi gli elementi di β , l'ultimo termine della (2.7), che prende in considerazione $i_{kk}(\beta)$ vale

$$(\beta_k - \beta_k^0)^2 \sum_{i=1}^n \left[\frac{e^{\sum_{j=1}^{k-1} \beta_j x_{ij} + \beta_k x_{ik}}}{(1 + e^{\sum_{j=1}^{k-1} \beta_j x_{ij} + \beta_k x_{ik}})^2} x_{ik}^2 \right]. \quad (2.8)$$

Ogni termine della sommatoria (2.8), per $\beta_k \rightarrow \pm\infty$, ossia mano a mano che il vero valore del parametro si allontana dal valore fissato dall'ipotesi nulla, tende a zero. Di conseguenza, essendo $0 \leq \zeta \leq \sum_{i=1}^n (\beta_k - \beta_k^0)^2 \pi_i(1 - \pi_i)x_{ik}^2$, dal momento che l'ultimo termine tende a zero, si ha che anche ζ tende a zero. Ciò implica che la distribuzione di Z^2 , sotto β_k , per β_k che sia allontana da β_k^0 , tende a un χ_1^2 , ossia alla distribuzione nulla. Pertanto la potenza del test decresce al livello di significatività. La medesima dimostrazione può essere riprodotta anche per dimostrare che (2.2) $\rightarrow 0$ se $\hat{\beta}_k \rightarrow \pm\infty$, per una data dimensione campionaria n e per valori fissati di $\hat{\beta}_1, \dots, \hat{\beta}_{k-1}$.

Il problema di non monotonicità di Z^2 , e quindi $W_e(\beta_k; \hat{\beta}) \Big|_{\beta_k = \beta_k^0}$, è dovuto al fatto che il numeratore $(\beta_k - \beta_k^0)^2$ cresce al distanziarsi del vero valore del

parametro dal valore sotto H_0 , ma non velocemente quanto il denominatore $i^{kk}(\beta)$, espressione della curvatura della log-verosimiglianza. Infatti per $\beta_k \rightarrow \pm\infty$, la funzione di log-verosimiglianza diventa praticamente piatta, portando a valori di $i^{kk}(\beta)$ molto elevati.

2.2 Il test di Wald in famiglie esponenziali di ordine uno

Nel seguito si riportano i risultati di Væth (1985), che analizza in quali condizioni il test di Wald presenta il problema di non monotonicità nel caso di famiglie esponenziali di ordine uno. Siano Y_i , $i = 1, \dots, n$ variabili casuali indipendenti con distribuzione in una famiglia esponenziale di ordine uno, con θ parametro naturale, definita in (1.25). Si consideri la parametrizzazione seguente

$$\omega^\delta = \omega^\delta(\theta) = \int_0^\theta K''(s)^\delta ds \quad (2.9)$$

con $K(\cdot)$ funzione generatrice dei cumulanti definita nel paragrafo 1.6.1, $0 \leq \delta \leq 1$ (per approfondimenti si veda Barndorff-Nielsen, 1978, pagine 176-181). Non si faranno distinzioni tra parametrizzazioni che sono trasformazioni affini delle altre, in quanto la statistica Wald è invariante rispetto a questo tipo di trasformazioni. Al variare di δ si ottengono varie parametrizzazioni, di cui alcune presentano peculiari proprietà. In particolare:

- $\delta = 1$, $\omega^1 = K'(\theta)$ è la parametrizzazione con la media.
- $\delta = \frac{2}{3}$: normalizza la distribuzione dello stimatore di massima verosimiglianza $\hat{\omega}$ rendendo la asimmetria approssimativamente 0.
- $\delta = \frac{1}{2}$: questa parametrizzazione stabilizza la varianza di $\hat{\omega}$, essendo l'informazione di Fisher costante, in quanto $\frac{d^2 \log p}{d\omega^2}(\hat{\omega})$ è costante.
- $\delta = 0$: $\omega^0 = \theta$ è la parametrizzazione canonica.

Il caso di campionamento casuale semplice da popolazione omogenea

Il problema di non monotonicità del test di Wald si pone anche nel caso più semplice in cui le osservazioni y_i , $i = 1, \dots, n$, siano i.i.d. . Siano y_1, \dots, y_n

un campione casuale da una famiglia esponenziale definita in (1.25). Si consideri il problema di verificare, senza perdita di generalità, $H_0 : \theta = 0$ contro $H_1 : \theta \neq 0$, equivalente nella parametrizzazione ω^δ (2.9), se si mantiene 0 come estremo inferiore dell'integrale, a $H_0 : \omega^\delta = 0$ contro $H_1 : \omega^\delta \neq 0$ (tuttavia, anche nel seguito, si analizzeranno traslazioni di ω^δ , in quanto più spesso utilizzate). Ricordando (1.21), (1.22) e (1.30), risulta indifferente usare l'informazione attesa o l'informazione osservata, sia nella parametrizzazione θ che nella parametrizzazione ω^δ . Serve ora valutare il test Wald nella parametrizzazione ω^δ (2.9):

$$W_e^\delta(0; \hat{\omega}^\delta) = (\hat{\omega}^\delta)^2 i^\Omega(\hat{\omega}^\delta). \quad (2.10)$$

Serve quindi calcolare $i^\Omega(\omega^\delta)$. Ricordando (1.21) ed essendo $\theta(\omega^\delta)$ funzione inversa di $\omega^\delta(\theta)$, si ha quindi,

$$[\theta'(\omega^\delta)] = \frac{d\theta}{d\omega^\delta} = \frac{1}{\frac{d\omega^\delta}{d\theta}} = \frac{1}{K''(\theta)^\delta}. \quad (2.11)$$

Posto $\hat{\theta}$ soluzione dell'equazione di verosimiglianza

$$\sum_{i=1}^n y_i = nK'(\theta), \quad (2.12)$$

il test di Wald nella parametrizzazione ω^δ può essere scritto come

$$W_e^\delta(0; \hat{\omega}^\delta) = nK''(\hat{\theta}) \overbrace{\left(\frac{1}{K''(\hat{\theta})^\delta}\right)^2}^{i^\Omega(\hat{\omega}^\delta)} \overbrace{\left(\int_0^{\hat{\theta}} K''(s)^\delta ds\right)^2}^{(\hat{\omega}^\delta)^2 = [\omega^\delta(\hat{\theta})]^2}, \quad (2.13)$$

ossia come

$$W_e^\delta(0; \hat{\omega}^\delta) = n \left[K''(\hat{\theta})^{\frac{1}{2}-\delta} \int_0^{\hat{\theta}} K''(s)^\delta ds \right]^2. \quad (2.14)$$

Ottenuta questa forma generale del test, il problema è ora studiare il suo comportamento al variare della stima di massima verosimiglianza, ossia valutare se tenda o meno a zero, man mano che $\hat{\theta}$ si distanzia dal valore sotto l'ipotesi nulla, per esempio quando la stima di massima verosimiglianza si avvicina agli estremi dello spazio parametrico Θ . I seguenti risultati quindi valuteranno l'andamento di $W_e^\delta(0; \hat{\omega}^\delta)$ ipotizzando che $\hat{\theta}$ tenda a θ_l o θ_u ,

limite inferiore e superiore di Θ , senza addentrarsi nel fatto che la s.m.v $\hat{\theta}$ non possa assumere valori arbitrariamente vicini agli estremi di Θ nel caso di distribuzione discreta.

Valutando (2.14), il comportamento al limite di $W_e^\delta(0; \hat{\omega}^\delta)$ si rivela strettamente legato al comportamento di $K''(\theta)$ ossia la varianza dell'osservazione naturale e della statistica sufficiente. Si riportano i risultati di Væth (1985) riguardo all'analisi della (2.14). Nei seguenti casi il test di Wald si comporta in modo coerente, ossia non tende a zero:

- (a) $Var_\theta(Y)$ non tende a zero o a infinito quando θ si avvicina agli estremi di Θ .
- (b) Il test è basato sulla parametrizzazione $\delta = \frac{1}{2}$, che stabilizza la varianza, per cui il valore dell'usuale denominatore del test di Wald, $i(\hat{\theta})^{-1}$, è costante rispetto a $\hat{\theta}$. Questa parametrizzazione coincide con la parametrizzazione radice quadrata della media nel modello di Poisson $\sqrt{\lambda}$, mentre nel modello binomiale coincide con la parametrizzazione $\arcsin \sqrt{\pi}$.
- (c) $Var_\theta(Y)$ tende a zero quando θ si avvicina agli estremi di Θ e ci si basa su una parametrizzazione con $\delta > \frac{1}{2}$.
- (d) $Var_\theta(Y)$ non tende a zero quando θ si avvicina agli estremi di Θ e ci si basa su una parametrizzazione con $\delta < \frac{1}{2}$.

Altre condizioni connesse al comportamento sulle code della funzione di ripartizione di Y , $F_\theta(y)$ sono state studiate da Væth (1985), e riguardano i modelli gamma e binomiale negativo, con parametro di forma fissato.

Si analizzano ora due tra le principali e più utilizzate distribuzioni della famiglia esponenziale, Poisson e binomiale, e il comportamento nella parametrizzazione con la media ($\delta = 1$), nella parametrizzazione naturale ($\delta = 0$) e nella parametrizzazione che stabilizza la varianza ($\delta = \frac{1}{2}$).

Capitolo 3

Il modello di Poisson

3.1 Il caso di osservazioni omogenee

Si considera in primo luogo il caso di osservazioni Poisson indipendenti e identicamente distribuite con media λ , $Y_i \sim \mathcal{P}(\lambda)$, $i = 1, \dots, n$. Il parametro naturale è $\theta = \log(\lambda)$ e $K(\theta) = e^\theta$. La stima di massima verosimiglianza per θ è $\hat{\theta} = \log(\hat{\lambda}) = \log\left(\frac{t}{n}\right)$, con $t = \sum_{i=1}^n y_i$ statistica sufficiente. Il test di Wald assume la forma

$$W_e(\theta; \hat{\theta}) = ne^{\hat{\theta}}(\hat{\theta} - \theta)^2. \quad (3.1)$$

Si osserva quindi che per $\hat{\theta} \rightarrow +\infty$ il test si comporta in modo coerente in quanto sia la varianza della statistica sufficiente che il quadrato della stima di massima verosimiglianza tendono a più infinito. Tuttavia per $\hat{\theta} \rightarrow -\infty$ la statistica test tende a 0.

È necessario comunque tenere conto del fatto che lo stimatore non può assumere valori arbitrariamente piccoli data una dimensione campionaria finita n , dal momento che la distribuzione Poisson è discreta. Infatti il valore più piccolo che può assumere $\hat{\theta} = \log\left(\frac{t}{n}\right)$, se la statistica sufficiente non è sulla frontiera (ossia non assume valore 0), è quando $t = 1$. Si ha di conseguenza che $\hat{\theta} = \log\left(\frac{1}{n}\right)$ e, per il problema di verifica di $H_0 : \theta = 0$ contro $H_1 : \theta \neq 0$, il test $W_e(0; \hat{\theta})$ vale

$$W_e(0; \hat{\theta}) = n \left[\log\left(\frac{1}{n}\right) \right]^2 \frac{1}{n} = \left[\log\left(\frac{1}{n}\right) \right]^2. \quad (3.2)$$

Studiando il comportamento di questa quantità al variare di n , si ha che la quantità $W_e(0; \hat{\theta})$ è monotona crescente per $n \geq 1$ e quindi, man mano che la dimensione campionaria cresce, anche $W_e(0; \hat{\theta})$ cresce, e per $n \geq 7$, già viene giustamente rifiutata l'ipotesi nulla $\theta = 0$ al livello di significatività $\alpha = 0.05$.

Il comportamento del test di Wald, anche se non porta a conclusioni inferenziali errate già per n piccolo, non è comunque coerente. Considerando ancora al caso di osservazioni i.i.d., valutiamo l'andamento del test al variare di $\hat{\theta}$, ponendo sempre $H_0 : \theta = 0$ contro $H_1 : \theta \neq 0$. Il test (3.1) ha derivata rispetto a $\hat{\theta}$ pari a

$$\frac{dW_e(0; \hat{\theta})}{d\hat{\theta}} = n[2\hat{\theta}e^{\hat{\theta}} + \hat{\theta}^2e^{\hat{\theta}}] = ne^{\hat{\theta}}\hat{\theta}(2 + \hat{\theta}). \quad (3.3)$$

Così fattorizzata possiamo valutarne il segno e risulta

$$\frac{dW(0; \hat{\theta})}{d\hat{\theta}} \geq 0 \quad \text{per} \quad \hat{\theta} \leq -2 \quad \text{o} \quad \hat{\theta} \geq 0, \quad (3.4)$$

mentre altrove la derivata è negativa.

La non monotonicità del test all'allontanarsi da H_0 è presente solo nel test di Wald e non si presenta invece negli altri due test di verosimiglianza: il test log-rapporto di verosimiglianza (1.6) e il test *score* (1.5). Infatti, posto sempre $H_0 : \theta = 0$ i due test assumono la seguenti forme (si veda Pace e Salvani, 2001, pagina 205). Il test log-rapporto di verosimiglianza vale

$$W(0; \hat{\theta}) = 2n(\hat{\theta}e^{\hat{\theta}} - e^{\hat{\theta}} + 1). \quad (3.5)$$

Il test *score* con informazione attesa calcolata nel valore ipotizzato da H_0 ($\theta = 0$) vale

$$W_u^{i(0)}(0; \hat{\theta}) = n(e^{\hat{\theta}} - 1)^2, \quad (3.6)$$

oppure, con l'informazione attesa calcolata in $\hat{\theta}$,

$$W_u^{i(\hat{\theta})}(0; \hat{\theta}) = \frac{n(e^{\hat{\theta}} - 1)^2}{e^{\hat{\theta}}}. \quad (3.7)$$

La Figura 3.1 mostra l'andamento del valore assunto dai test di verosimiglianza al variare di $\hat{\theta}$. Per il grafico si è fissato n , numero di osservazioni,

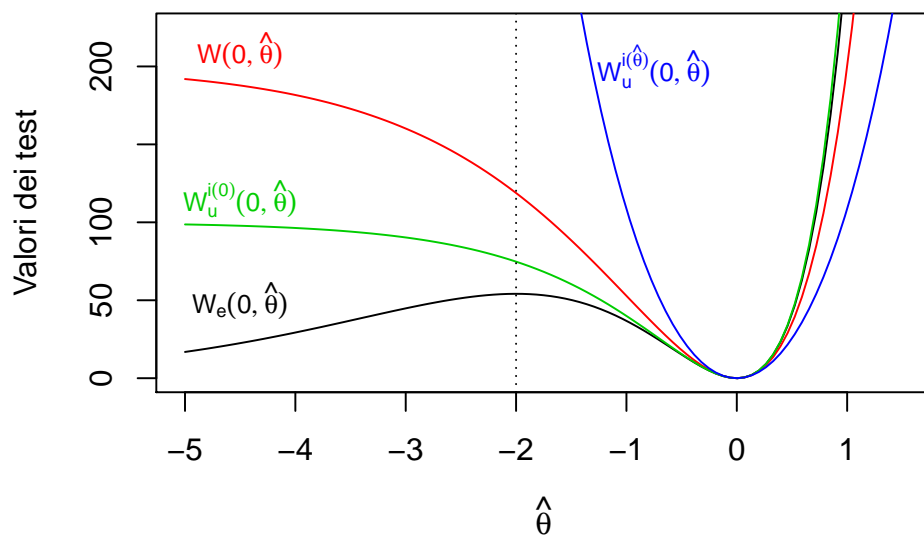


Figura 3.1: I test di verosimiglianza nel modello di Poisson al variare di $\hat{\theta}$. Wald (nero), log-rapporto di verosimiglianza (rosso), *score* con informazione attesa calcolata in $\theta = 0$ (verde), *score* con informazione osservata calcolata in $\hat{\theta}$ (blu).

a 100 e si è posto $H_0 : \theta = 0$ contro $H_1 : \theta \neq 0$. I test infatti assumono valore zero quando la stima di θ coincide con il valore sotto l'ipotesi nulla, ossia quando $\hat{\theta} = 0$. Nonostante i test si comportino in modo leggermente diverso, l'unico che presenta una chiara anomalia è il test Wald nella parametrizzazione canonica. Quanto ottenuto è in linea con il punto (f) di Væth (1985).

Posto θ vero valore del parametro, si può anche calcolare la probabilità che $\hat{\theta}$ appartenga alla regione di non monotonicità del test

$$Pr_{\theta}(\hat{\theta} \leq -2) = Pr_{\theta}\left(\log\left(\sum_{i=1}^n \frac{Y_i}{n}\right) \leq -2\right) = Pr_{\theta}\left(\sum_{i=1}^n Y_i \leq ne^{-2}\right). \quad (3.8)$$

Dal momento che la somma di Poisson i.i.d. è ancora una Poisson con media $(n\lambda) = (ne^{\theta})$, la (3.8) diventa

$$Pr_{\theta}\left(\mathcal{P}(ne^{\theta}) \leq ne^{-2}\right). \quad (3.9)$$

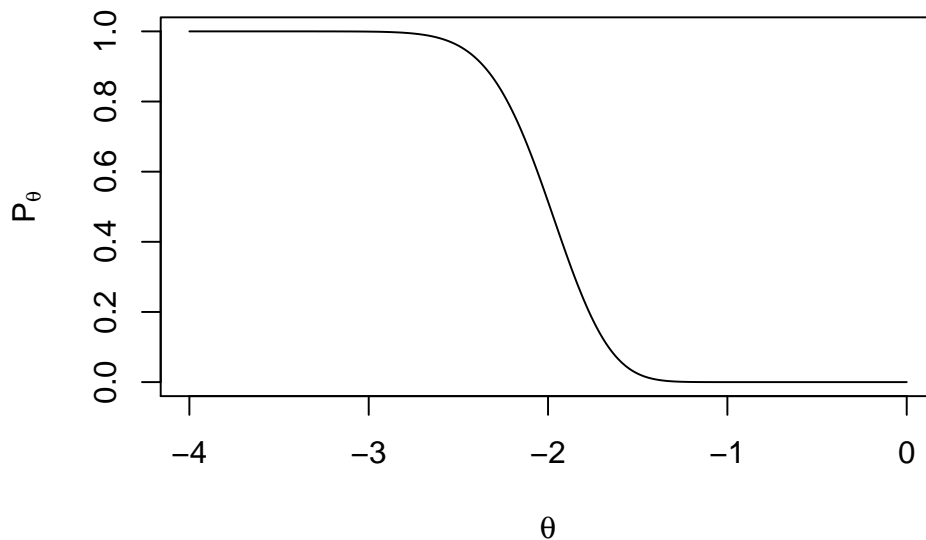


Figura 3.2: L'andamento di $Pr_{\theta}\left(\mathcal{P}(100e^{\theta}) \leq 100e^{-2}\right)$ al variare di θ .

L'andamento della probabilità di trovarsi nella regione di non monotonicità del test al variare di θ è descritta nella Figura 3.2 (dove si è posto

$n = 100$). Si può notare come per $\theta \leq -2.7$, la probabilità che $\hat{\theta}$ appartenga alla regione di non monotonicità sia praticamente 1.

Si verifica ora che le parametrizzazioni con $\delta = \frac{1}{2}$ e $\delta = 1$ non pongono questo tipo di problema. Infatti nel test di Wald con $\delta = 0$, quindi nella parametrizzazione con la media, si ha $\lambda = e^\theta$. Il test risulta $W_e^{\delta=0}(\lambda; \hat{\lambda}) = n \left[\frac{(\hat{\lambda} - \lambda)^2}{\hat{\lambda}} \right]$. Riscrivendolo in funzione di $\hat{\theta}$ si ottiene

$$W_e^{\delta=0}(\theta; \hat{\theta}) = \left[\frac{(e^{\hat{\theta}} - e^\theta)^2}{e^{\hat{\theta}}} \right]. \quad (3.10)$$

Per $\hat{\theta} \rightarrow -\infty$ il test tende a $+\infty$.

Nella parametrizzazione $\delta = \frac{1}{2}$, che nel caso della Poisson significa parametrizzazione con la radice quadrata della media, ossia $\omega = \sqrt{\lambda} = e^{\frac{\theta}{2}}$, il test di Wald assume la forma

$$W_e^{\delta=\frac{1}{2}}(\omega; \hat{\omega}) = 4n(\hat{\omega} - \omega)^2, \quad (3.11)$$

che riscritto in funzione di $\hat{\theta}$ diventa

$$W_e^{\delta=\frac{1}{2}}(\theta; \hat{\theta}) = 4n \left(e^{\frac{\hat{\theta}}{2}} - e^{\frac{\theta}{2}} \right)^2 \quad (3.12)$$

La Figura 3.3 mostra il test di Wald nelle tre parametrizzazioni prese in esame.

La Figura 3.4 mostra l'andamento dei test al variare di δ . Le curve sono state ottenute a partire da (2.14), che nel caso di osservazioni Poisson, per verificare $H_0 : \theta = 0$ contro $H_1 : \theta \neq 0$, assume la forma

$$W_e^\delta(0; \hat{\theta}) = \frac{n}{\delta^2} e^{\hat{\theta} - 2\delta\hat{\theta}} \left(\left[e^{\delta s} \right]_0^{\hat{\theta}} \right)^2 = \frac{n}{\delta^2} e^{\hat{\theta} - 2\delta\hat{\theta}} (e^{2\delta\hat{\theta}} + 1 - 2e^{\delta\hat{\theta}}). \quad (3.13)$$

La curva verde corrispondente alla parametrizzazione $\delta = \frac{1}{2}$ suddivide in due parti le curve: quelle con $\delta < \frac{1}{2}$ (nero, ciano, blu, magenta e arancione nel grafico), e con $\delta > \frac{1}{2}$ (viola, verde scuro e rosso nel grafico). Le curve corrispondenti al test di Wald con $\delta \in (0, 0.4)$ presentano il comportamento anomalo proprio della parametrizzazione canonica. Ciò è confermato dallo studio della derivata di (3.13). La derivata vale

$$\begin{aligned} \frac{dW_e^\delta(0; \hat{\theta})}{d\hat{\theta}} &= \frac{n}{\delta^2} \left[e^{\hat{\theta}} + e^{\hat{\theta} - 2\delta\hat{\theta}}(1 - 2\delta) - 2e^{\hat{\theta} - \delta\hat{\theta}}(1 - \delta) \right] \\ &= \frac{n}{\delta^2} e^{\hat{\theta}} \left[1 + e^{-2\delta\hat{\theta}}(1 - 2\delta) - 2e^{-\delta\hat{\theta}}(1 - \delta) \right]. \end{aligned} \quad (3.14)$$

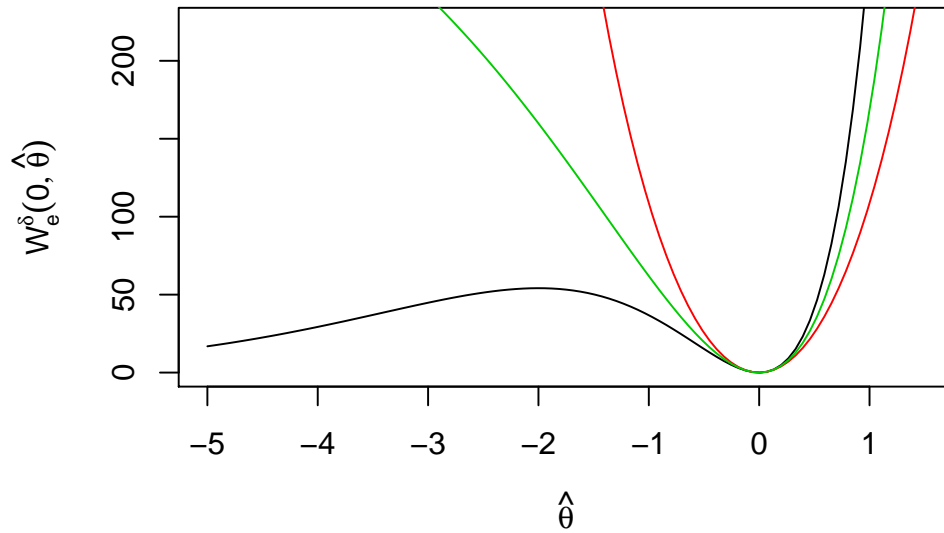


Figura 3.3: Il test di Wald al variare di $\hat{\theta}$. Parametrizzazione canonica (nero), parametrizzazione che stabilizza la varianza (verde), parametrizzazione con la media (rosso)

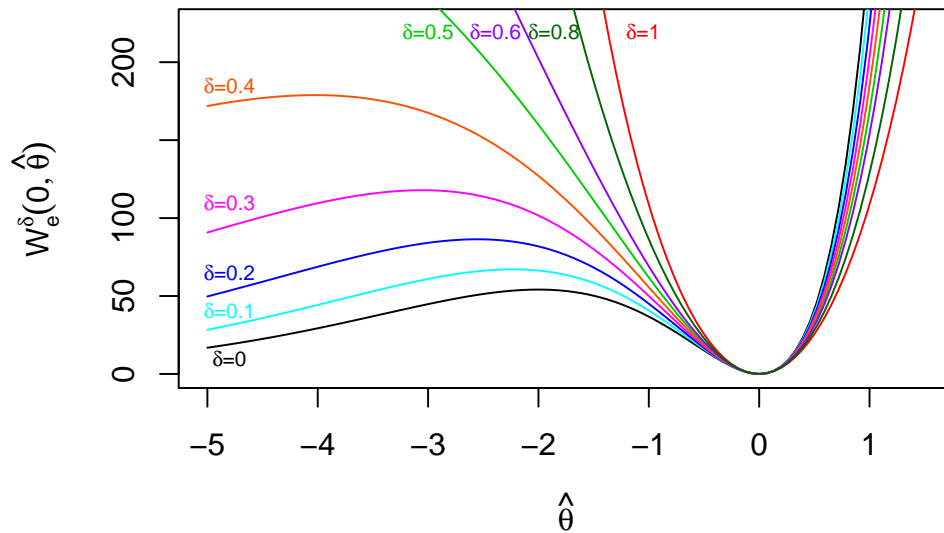


Figura 3.4: Il test di Wald al variare del valore di δ .

Nonostante si sia raccolto $e^{\hat{\theta}}$, la ricerca dei punti stazionari (oltre allo 0) non permette soluzione esplicita. Si è quindi ricorso a metodi numerici. Per valori di $\delta \geq \frac{1}{2}$ l'unico punto stazionario è lo zero.

```

massimo <- function(x,d)
{ 1+ exp(-2*d*x)*(1-2*d) - 2 * exp(-d*x)*(1-d) }
d.grid <- seq(0.01,0.49,length=100)
radici <- rep(NA,100)
for(i in 1:100){radici[i] = uniroot(massimo, c(-10,-1),d=d.grid[i])$root}
plot(radici,d.grid, type="l", xlab = expression(frac({d*W[e]^delta}{0,hat(
theta)),d*hat(theta))), ylab=expression(delta))

```

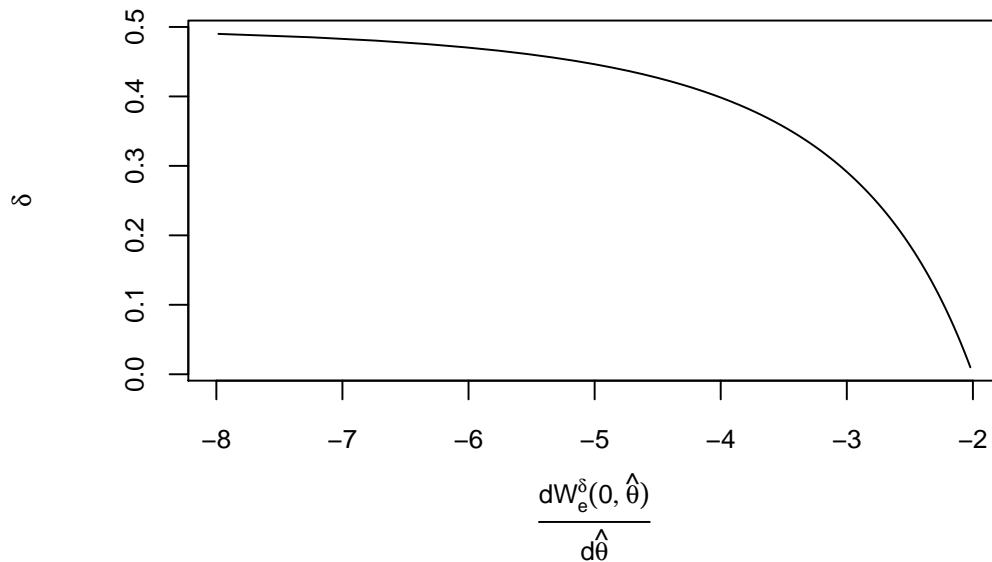


Figura 3.5: Il punto stazionario di (3.13) al variare di δ .

3.2 Modelli di regressione di Poisson

3.2.1 Legame canonico

Le problematiche sottolineate nel caso di osservazioni Poisson i.i.d. si riflettono anche sui modelli di regressione Poisson, in particolare quello con

legame canonico. In generale, secondo Væth (1985, pagina 208), il test di Wald applicato ai singoli parametri di regressione si comporterà in modo anomalo se la funzione di legame usata corrisponde, nel caso di osservazioni i.i.d., a una parametrizzazione che presenta il comportamento anomalo. Le funzioni di legame analizzate da Væth (1985) sono del tipo $\omega_i^\delta = \omega^\delta(\theta_i) = \sum_{j=1}^k \beta_j x_{ij}$. Nel seguito si mostrerà ciò nel modello di regressione di Poisson.

Siano y_i le osservazioni della variabile risposta, distribuita come una $\mathcal{P}(\lambda_i)$, e sia x_i una variabile esplicativa. Il primo modello analizzato utilizza la funzione di legame canonica, ossia $\theta_i(\lambda_i) = \theta_i = \eta_i$, con $\eta_i = \beta_1 + \beta_2 x_i$ predittore lineare. La statistica sufficiente bidimensionale è $s = (s_1, s_2) = (\sum_{i=1}^n y_i/n, \sum_{i=1}^n (y_i x_i)/n)$. Per una generica x_i , le quantità di verosimiglianza di interesse per valutare la nullità di β_2 sono

$$l(\beta) = \beta_1 s_1 + \beta_2 s_2 - e^{\beta_1} \sum_{i=1}^n e^{\beta_2 x_i}, \quad (3.15)$$

$$l_*(\beta) = (l_1(\beta), l_2(\beta)) = \left(\beta_1 s_1 - e^{\beta_1} \sum_{i=1}^n e^{\beta_2 x_i}, s_2 - e^{\beta_1} \sum_{i=1}^n e^{\beta_2 x_i} x_i \right), \quad (3.16)$$

$$j(\beta) = \begin{pmatrix} e^{\beta_1} \sum_{i=1}^n e^{\beta_2 x_i} & e^{\beta_1} \sum_{i=1}^n e^{\beta_2 x_i} x_i \\ e^{\beta_1} \sum_{i=1}^n e^{\beta_2 x_i} x_i & e^{\beta_1} \sum_{i=1}^n e^{\beta_2 x_i} x_i^2 \end{pmatrix}. \quad (3.17)$$

Facendo riferimento alle formule per le inverse di matrici (1.9)

$$\left[j^{22}(\beta) \right]^{-1} = e^{\beta_1} \left[\sum_{i=1}^n (e^{\beta_2 x_i} x_i^2) - \frac{\left(\sum_{i=1}^n e^{\beta_2 x_i} x_i \right)^2}{\sum_{i=1}^n e^{\beta_2 x_i}} \right]. \quad (3.18)$$

Per x quantitativa generica non è disponibile soluzione esplicita per il sistema di equazioni di verosimiglianza.

Si consideri quindi il problema più semplice di una popolazione composta da due gruppi con medie diverse. La variabile x è quindi una variabile indicatrice che assume valore 0 se l'unità i appartiene al primo gruppo, 1 se appartiene al secondo gruppo. Sia inoltre n_1 la numerosità delle osservazioni appartenenti al primo gruppo e n_2 la numerosità delle osservazioni appartenenti al secondo gruppo. Le stime di massima verosimiglianza sono $\hat{\beta}_1 = \log \left(\sum_{i=1}^n [y_i(1-x_i)]/n_1 \right)$, $\hat{\beta}_2 = \log \left(\sum_{i=1}^n (y_i x_i)/n_2 \right) - \hat{\beta}_1$. Il test

Wald per verificare $H_0 : \beta_2 = 0$ contro $H_1 : \beta_2 \neq 0$, assume la forma

$$W_{eP}(\beta_2; \hat{\beta}) \Big|_{\beta_2=0} = (\hat{\beta}_2)^2 \left[n_2 e^{\hat{\beta}_1 + \hat{\beta}_2} \left(1 - \frac{n_2 e^{\hat{\beta}_1 + \hat{\beta}_2}}{n_1 e^{\hat{\beta}_1} + n_2 e^{\hat{\beta}_1 + \hat{\beta}_2}} \right) \right]. \quad (3.19)$$

Il test può essere riscritto esplicitando $(e^{\hat{\beta}_1}, e^{\hat{\beta}_1 + \hat{\beta}_2}) = (\bar{y}_1, \bar{y}_2)$, funzione biunivoca della statistica sufficiente. Si ottiene

$$W_{eP}(\beta_2; \hat{\beta}) \Big|_{\beta_2=0} = \left(\log \frac{\bar{y}_2}{\bar{y}_1} \right)^2 \left[n_2 \bar{y}_2 \left(1 - \frac{n_2 \bar{y}_2}{n_1 \bar{y}_1 + n_2 \bar{y}_2} \right) \right]. \quad (3.20)$$

Il comportamento anomalo si può immediatamente notare, in quanto il limite per $\bar{y}_2 \rightarrow 0$ vale zero. Si vuole comunque valutare se il test di Wald è monotono crescente all'allontanarsi dall'ipotesi nulla. Possiamo riscrivere il test rispetto a $t = \bar{y}_2/\bar{y}_1$. Risulta

$$\begin{aligned} W_{eP}(\beta_2; \hat{\beta}) \Big|_{\beta_2=0} &= (\log t)^2 \bar{y}_1 \left[n_2 t \left(1 - \frac{n_2 t}{n_1 + n_2 t} \right) \right] \\ &= c (\log t)^2 \left[kt \left(1 - \frac{kt}{1 + kt} \right) \right], \end{aligned} \quad (3.21)$$

con $k = n_2/n_1$ e $c = n_1 \bar{y}_1$.

La derivata prima è

$$\frac{dW_{eP}(t)}{dt} = c \left[2 \frac{k \log t}{1 + kt} + (\log t)^2 \frac{k}{(1 + kt)^2} \right] = \frac{ck \log t}{(1 + kt)^2} [2 + 2kt + k \log t] \quad (3.22)$$

Si studia quindi il segno della derivata, tenendo conto che sia c che k sono positivi, il denominatore è sempre positivo, e che non si riesce ad ottenere una soluzione esplicita per l'equazione di annullamento del secondo fattore. Si pone quindi $k = 1$, equivalente a $n_1 = n_2$, e si ha

$$\frac{dW_{eP}(t)}{dt} > 0 \quad \text{per } 0 \leq x \leq 0.109 \quad \text{o} \quad x \geq 1 \quad (3.23)$$

mentre altrove la derivata è negativa.

L'uguaglianza delle medie (stimate) nelle due popolazioni si ha quando t assume il valore 1 e infatti il test di Wald assume valore zero. Per $t > 1$ il test cresce al crescere di t . Tuttavia non accade lo stesso per $t \leq 1$: il segno della derivata ci conferma che inizialmente il test cresce all'allontanarsi di t da 1, per poi, partendo da $t = 0.109$ e andando verso lo zero, decrescere verso 0.

3.2.2 Legame identità

Mentre nel caso di popolazione omogenea una riparametrizzazione porta a modelli stimati equivalenti, nell'ambito di regressione diverse funzioni di legame portano generalmente a modelli differenti, poiché si mette in relazione il parametro ω_i^δ , trasformazione non lineare di θ_i , con il predittore lineare η_i .

La funzione di legame identità ipotizza che la media della variabile risposta sia pari al predittore lineare, ossia $\lambda_i = \log(\theta_i) = \sum_{j=1}^k \beta_j x_{ij}$. Con una variabile esplicativa, si ha che $\lambda_i = \beta_1 + \beta_2 x_i$. Le quantità di verosimiglianza di interesse sono

$$l_*(\beta) = \left(l_1(\beta), l_2(\beta) \right) = \left(\sum_{i=1}^n \frac{y_i}{\beta_1 + \beta_2 x_i} - n, \sum_{i=1}^n \frac{y_i x_i}{\beta_1 + \beta_2 x_i} - \sum_{i=1}^n x_i \right), \quad (3.24)$$

$$j(\beta_1, \beta_2) = \begin{pmatrix} \sum_{i=1}^n \frac{y_i}{(\beta_1 + \beta_2 x_i)^2} & \sum_{i=1}^n \frac{y_i x_i}{(\beta_1 + \beta_2 x_i)^2} \\ \sum_{i=1}^n \frac{y_i x_i}{(\beta_1 + \beta_2 x_i)^2} & \sum_{i=1}^n \frac{y_i x_i^2}{(\beta_1 + \beta_2 x_i)^2} \end{pmatrix}. \quad (3.25)$$

Vincolandoci al semplice caso di due gruppi preso in esame precedentemente, l'informazione osservata si semplifica in

$$j(\beta_1, \beta_2) = \begin{pmatrix} \frac{\sum_{i=1}^n y_i (1 - x_i)}{\beta_1^2} + \frac{\sum_{i=1}^n y_i x_i}{(\beta_1 + \beta_2)^2} & \frac{\sum_{i=1}^n y_i x_i}{(\beta_1 + \beta_2)^2} \\ \frac{\sum_{i=1}^n y_i x_i}{(\beta_1 + \beta_2)^2} & \frac{\sum_{i=1}^n y_i x_i}{(\beta_1 + \beta_2)^2} \end{pmatrix}. \quad (3.26)$$

Si ottiene

$$\left[j(\beta)^{22} \right]^{-1} = \frac{\sum_{i=1}^n y_i (1 - x_i) \sum_{i=1}^n y_i x_i}{(\beta_1 + \beta_2)^2 \sum_{i=1}^n y_i (1 - x_i) + \beta_1^2 \sum_{i=1}^n y_i x_i}. \quad (3.27)$$

e il sistema di equazioni di verosimiglianza è risolvibile esplicitamente, con le stime di massima verosimiglianza pari a $\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i (1 - x_i)}{n_1}$ e $\hat{\beta}_2 = \frac{\sum_{i=1}^n y_i x_i}{n_2} - \hat{\beta}_1$. Il test di Wald per $H_0 : \beta_2 = 0$ contro $H_1 : \beta_2 \neq 0$ è quindi

$$W_{eP}(\beta_2; \hat{\beta}) \Big|_{\beta_2=0} = (\hat{\beta}_2)^2 \left[j(\hat{\beta})^{22} \right]^{-1}. \quad (3.28)$$

Si può riscrivere il test in funzione di una trasformazione biunivoca della statistica sufficiente bidimensionale: $(\hat{\beta}_1, \hat{\beta}_1 + \hat{\beta}_2) = (\bar{y}_1, \bar{y}_2)$:

$$W_{eP}(\beta_2; \hat{\beta}) \Big|_{\beta_2=0} = (\bar{y}_2 - \bar{y}_1)^2 \frac{n_1 n_2 \bar{y}_1 \bar{y}_2}{n_1 \bar{y}_1 \bar{y}_2^2 + n_1 \bar{y}_2 \bar{y}_1^2}. \quad (3.29)$$

Con la stessa notazione precedente, ossia con $t = \bar{y}_2/\bar{y}_1$, $k = n_2/n_1$ e $c = n_1 \bar{y}_1$, possiamo riscriverlo come

$$W_{eP}(\beta_2^0) = c(t-1)^2 \frac{k}{t+k}. \quad (3.30)$$

La derivata è

$$\frac{dW_{eP}(t)}{dt} = ck \left[\frac{2(t-1)(t+k) - (t-1)^2}{(t+k)^2} \right] = \frac{ck[t^2 + 2tk - 2k - 1]}{(t+k)^2}. \quad (3.31)$$

Il segno della derivata è

$$\frac{dW_{eP}(t)}{dt} \geq 0 \quad \text{per } t \geq 1 \quad (3.32)$$

e negativo per $0 < t < 1$.

Quindi per qualsiasi valore di k , il test di Wald assume valore 0 in $t = 1$, corrispondente all'ipotesi nulla di uguaglianza delle medie, per poi crescere all'allontanarsi da $t = 1$.

3.2.3 Legame radice quadrata

L'ultima funzione di legame presa in considerazione è la funzione di legame radice quadrata, che corrisponde nel caso di popolazione omogenea alla parametrizzazione che stabilizza la varianza. Insieme alla funzione di legame canonica (logaritmo della media) e alla funzione della legame identità, costituisce le funzioni di legame di *default* in \mathbb{R} per la famiglia Poisson.

La funzione di legame radice quadrata mette in relazione la radice quadrata della media della risposta con il predittore lineare, ossia $\sqrt{\lambda_i} = \sum_{j=1}^k \beta_j x_{ij}$. Ancora con una sola variabile esplicativa,

$$j(\beta_1, \beta_2) = \left(\begin{array}{c} \sum_{i=1}^n \left[\frac{2y_i}{(\beta_1 + \beta_2 x_i)^2} + 2 \right] \\ \sum_{i=1}^n \left[\frac{2y_i x_i}{(\beta_1 + \beta_2 x_i)^2} + 2x_i \right] \end{array} \quad \begin{array}{c} \sum_{i=1}^n \left[\frac{2y_i x_i}{(\beta_1 + \beta_2 x_i)^2} + 2x_i \right] \\ \sum_{i=1}^n \left[\frac{2y_i x_i^2}{(\beta_1 + \beta_2 x_i)^2} + 2x_i^2 \right] \end{array} \right). \quad (3.33)$$

Si ha inoltre che nel caso di due popolazioni le stime di massima verosimiglianza sono $\hat{\beta}_1 = \sqrt{\bar{y}_1}$ e $\hat{\beta}_2 = \sqrt{\bar{y}_2} - \hat{\beta}_1$, e la matrice di informazione si semplifica in

$$j(\hat{\beta}_1, \hat{\beta}_2) = \begin{pmatrix} 4n & 4n_2 \\ 4n_2 & 4n_2 \end{pmatrix}. \quad (3.34)$$

Il test per $H_0 : \beta_2 = 0$ contro $H_1 : \beta_2 \neq 0$ ha quindi la forma

$$W_{eP}(\beta_2; \hat{\beta}) \Big|_{\beta_2=0} = \frac{(\sqrt{\bar{y}_2} - \sqrt{\bar{y}_1})^2 4n_2 n_1}{n} = 4\bar{y}_1 \left(\sqrt{\frac{\bar{y}_2}{\bar{y}_1}} - 1 \right)^2 \frac{n_2 n_1}{n}. \quad (3.35)$$

Riscrivendolo in funzione di $t = \bar{y}_2/\bar{y}_1$, $c = n_1\bar{y}_1$ e $k = n_2/n_1$

$$W_{eP}(\beta_2; \hat{\beta}) \Big|_{\beta_2=0} = \frac{4ck}{1+k} (\sqrt{t} - 1)^2 \quad (3.36)$$

Si nota che al tendere di t a zero, il test si comporta in modo coerente. La derivata rispetto a t è

$$\frac{dW_{eP}(t)}{dt} = \frac{4ck}{1+k} (\sqrt{t} - 1) \frac{1}{\sqrt{t}}. \quad (3.37)$$

Studiandone il segno si ottiene

$$\frac{dW_{eP}(t)}{dt} \geq 0 \quad \text{per} \quad t \geq 1, \quad (3.38)$$

mentre altrove la derivata è negativa.

Nonostante i modelli stimati in questo semplice esempio siano equivalenti, e le diverse funzioni di legame costituiscano una mera riparametrizzazione, si può affermare comunque che dei tre modelli che abbiamo analizzato, solo il modello con funzione di legame canonica presenta anomalie nel test di Wald per verificare la nullità del singolo parametro. La Figura 3.6 mostra l'andamento dei test per la verifica della nullità di β_2 con le tre diverse funzioni di legame. L'unico che presenta il comportamento anomalo è il test con funzione di legame canonica. È da sottolineare comunque il fatto che il problema si presenta quando le medie stimate λ_i tendono a zero, e per questi casi va valutato se non sia più appropriato un modello per dati di conteggio con molti zeri, come ad esempio il modello di Poisson con inflazione di zeri, o il modello binomiale negativo e binomiale negativo con inflazione di zeri.

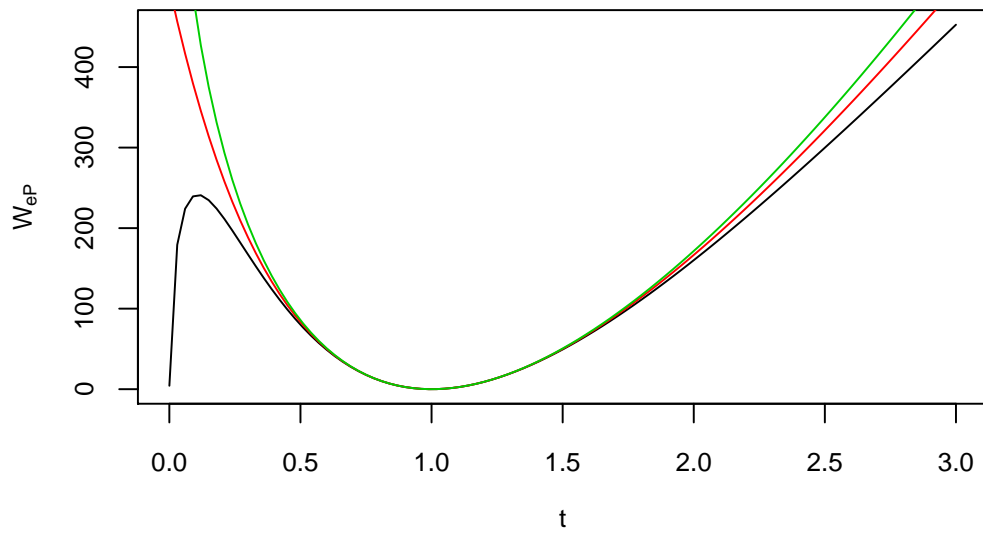


Figura 3.6: Il test Wald $W_{eP}(\beta_2; \hat{\beta}) \Big|_{\beta_2=0}$ al variare del valore di $t = \bar{y}_1/\bar{y}_2$. Legame canonico (nero), legame identità (rosso) e legame radice quadrata (verde).

Capitolo 4

Le principali funzioni di legame per dati binari

Nel paragrafo 2.2 si sono analizzate le parametrizzazioni per famiglie esponenziali di ordine uno del tipo ω^δ (2.9) e nel paragrafo 2.1 si è analizzato il legame canonico in ambito di regressione per dati binari. Tuttavia le funzioni di legame in ambito di regressione per dati binari più utilizzate, presenti di *default* su R, sono la funzione di legame probit, la funzione di legame log-log complementare e la funzione di legame log-log, non riconducibili alle parametrizzazioni ω^δ studiate da Væth (1985).

4.1 Il caso di osservazioni omogenee

Nel caso di osservazioni omogenee $Y_i \sim Bi(1, \pi)$, $i = 1, \dots, n$ indipendenti, la stima di massima verosimiglianza è $\hat{\pi} = \sum_{i=1}^n y_i/n$. Si consideri nel seguito la verifica di

$$H_0 : \pi = \pi_0 \quad \text{contro} \quad H_1 : \pi \neq \pi_0, \quad \pi_0 = 0.5. \quad (4.1)$$

4.1.1 Parametrizzazione probit

La parametrizzazione probit è $\psi = \Phi^{-1}(\pi)$, con $\Phi(\cdot)$ funzione di ripartizione della normale standard, $\psi \in \mathbb{R}$, e quindi $\pi = \Phi(\psi)$. Possiamo ottenere

il test di Wald nella nuova parametrizzazione facendo riferimento alla (1.21). Il test è

$$W_e^P(\psi; \hat{\psi}) = (\hat{\psi} - \psi)^2 i(\hat{\psi}) = (\hat{\psi})^2 \underbrace{\left[\frac{n}{\Phi(\hat{\psi})(1 - \Phi(\hat{\psi}))} \right]}_{i(\hat{\pi})=i(\pi(\hat{\psi}))} \underbrace{\left[\frac{\left(\frac{d\pi(\psi)}{d\psi} \Big|_{\psi=\hat{\psi}} \right)^2}{\left[\phi(\hat{\psi}) \right]^2} \right]}_{\left(\frac{d\pi(\psi)}{d\psi} \Big|_{\psi=\hat{\psi}} \right)^2}, \quad (4.2)$$

con $\phi(\cdot)$ funzione di densità della normale standard, e l'apice P per indicare la parametrizzazione probit.

La derivata di $W_e^P(\psi; \hat{\psi})$ rispetto a $\hat{\psi}$ è calcolata tramite la nota formula del derivazione di quoziente. La derivata del numeratore è

$$2n(\hat{\psi} - \psi)[\phi(\hat{\psi})]^2[1 - (\hat{\psi} - \psi)\hat{\psi}] \quad (4.3)$$

e la derivata del denominatore è

$$\phi(\hat{\psi}) - 2\Phi(\hat{\psi})\phi(\hat{\psi}) \quad (4.4)$$

Il problema della verifica di (4.1) equivale in questa parametrizzazione a $H_0 : \psi = \psi_0^P$ contro $H_1 : \psi \neq \psi_0^P$, $\psi_0^P = \Phi^{-1}(0.5) = 0$. Il grafico della funzione conferma la non monotonicità del test di Wald nella parametrizzazione probit. Non sono disponibili soluzioni esplicite per trovare i punti stazionari di $W_e^P(0; \hat{\psi})$, quindi si è ricorso a metodi numerici. I punti di massimo di $W_e^P(0; \hat{\psi})$ si hanno per $\hat{\psi} = \pm 1.575$ circa, ossia, con riferimento alla parametrizzazione con la media, quando $\hat{\pi} = 0.058$ e $\hat{\pi} = 0.942$. Per valori esterni a tale intervallo il test inizia a decrescere, nonostante l'evidenza contro l'ipotesi nulla aumenti.

Segue il codice R usato per i grafici e la ricerca numerica dei punti stazionari.

```
n<-100
psi.0.P<-0
wald.probit <- function(x)
{ n*((x-psi.0.P) * dnorm(x))^2 / (pnorm(x) - pnorm(x)^2) }
derivata.wald.probit <- function(x)
{ num <- ((x-psi.0.P)*dnorm(x))^2
  den <- pnorm(x)*(1- pnorm(x))
```

```

num.primo <- 2*(x-psi.0.P)*(dnorm(x)^2)*(1-(x-psi.0.P)*x)
den.primo <- dnorm(x) - 2*pnorm(x)*dnorm(x)
deriv <-n*(num.primo*den - num*den.primo)/(den^2)
return(deriv)}
curve(wald.probit(x), xlim = c(-5,5), ylim = c(-60,60),xlab = expression(
  hat(psi)), ylab="")
curve(derivata.wald.probit(x), add=T, col=2)
abline(h=0, lty=2)
x1<-uniroot(derivata.wald.probit, interval = c(-3,-0.5))
x2<-uniroot(derivata.wald.probit, interval = c(0.5,3))
abline(v=x1$root, lty=3)
abline(v=x2$root, lty=3)
abline(v=psi.0.P, lty=3)

```

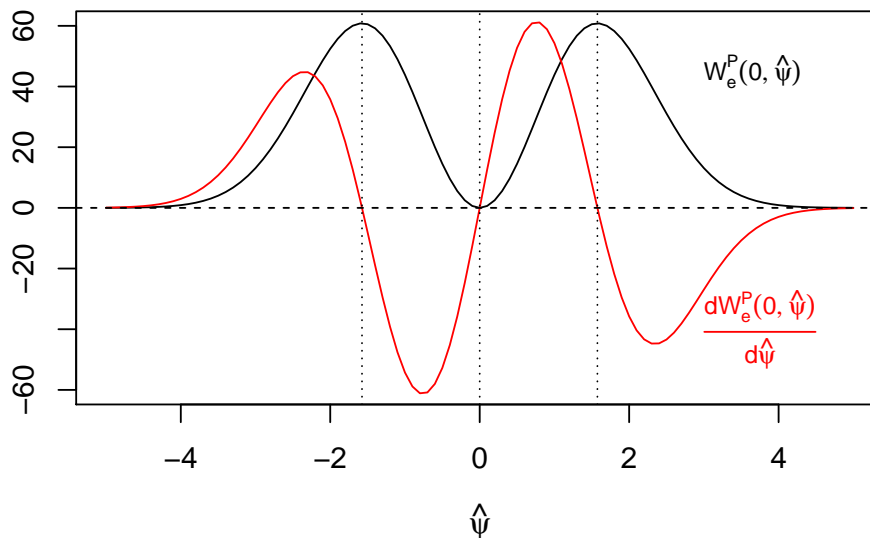


Figura 4.1: $W_e^P(0; \hat{\psi})$ e la sua derivata al variare di $\hat{\psi}$.

4.1.2 Parametrizzazione log-log complementare

La parametrizzazione è $\psi = \log[-\log(1 - \pi)]$, $\psi \in \mathbb{R}$, e quindi $\pi = 1 - \exp[-\exp(\psi)]$. Il test di Wald in questa parametrizzazione vale

$$W_e^C(\psi; \hat{\psi}) = (\hat{\psi} - \psi)^2 \frac{\overbrace{n}^{i(\hat{\pi})=i(\pi(\hat{\psi}))}}{(\exp[-\exp(\hat{\psi})] - \exp[-2\exp(\hat{\psi})])} \underbrace{\left(\frac{d\pi(\psi)}{d\psi} \Big|_{\psi=\hat{\psi}} \right)^2}_{\left(\exp[-\exp(\hat{\psi}) + \hat{\psi}] \right)^2}. \quad (4.5)$$

Per calcolare la derivata è necessario utilizzare la formula della derivata del quoziente. La derivata del numeratore è

$$2n(\hat{\psi} - \psi) \{ \exp[-\exp(\hat{\psi} + \hat{\psi})] \}^2 \{ 1 + (\hat{\psi} - \psi) [-\exp(\hat{\psi}) + 1] \}, \quad (4.6)$$

e la derivata del denominatore è

$$-\exp[-\exp(\hat{\psi}) + \psi] + 2\exp[-2\exp(\hat{\psi}) + \hat{\psi}]. \quad (4.7)$$

Il problema di verifica di (4.1) equivale a $H_0 : \psi = \psi_0$ contro $H_1 : \psi \neq \psi_0$, $\psi_0 = -0.3665$. Il grafico conferma la non monotonicità del test di Wald nella parametrizzazione log-log complementare. Per il calcolo delle radici della derivata non sono disponibili soluzioni esplicite e si è dunque ricorso a metodi numerici. I punti di massimo di $W_e^C(\psi_0; \hat{\psi})$ si hanno per $\hat{\psi} = -2.457$ e $\hat{\psi} = 1.155$. Per valori esterni a tale intervallo il test inizia a decrescere, nonostante $\hat{\psi}$ si allontani dal valore sotto l'ipotesi nulla. È inoltre interessante notare che, a differenza della parametrizzazione probit, questa parametrizzazione porti a un andamento del test di Wald asimmetrico rispetto a ψ_0 .

Seguono i comandi R per l'*output* grafico e la ricerca dei punti stazionari.

```
n<-100
psi.0.C <- - 0.3665
wald.cloglog <- function(x)
{ n * (x-psi.0)^2 * (( exp(-exp(x) + x))^2 / (exp(-exp(x)) - exp(-2*exp(x))
)) }
derivata.wald.cloglog <- function(x)
{ num <- ((x-psi.0)*exp(-exp(x)+x))^2
  den <- (exp(-exp(x)) - exp(-2*exp(x)))
  num.primo <- 2*(x-psi.0)*((exp(-exp(x) +x))^2)*(1+(x-psi.0)*(-exp(x)+1))
```



```

den.primo <- -exp(-exp(x) + x) + 2*exp(-2*exp(x) + x)
deriv <- -n*(num.primo*den - num*den.primo)/(den^2)
return(deriv) }
par(mar=c(4.5,3,1,2))
curve(wald.cloglog(x), xlim = c(-4,3), ylim=c(-130,110), xlab = expression
      (hat(psi)), ylab="")
curve(derivata.wald.cloglog(x), col=2, add=T)
abline(h=0, lty=2)
x1 <-uniroot(derivata.wald.cloglog, c(-4,-1))
x2 <-uniroot(derivata.wald.cloglog, c(0,3))
abline(v=x1$root, lty=3)
abline(v=x2$root, lty=3)
abline(v=psi.0, lty=3)

```

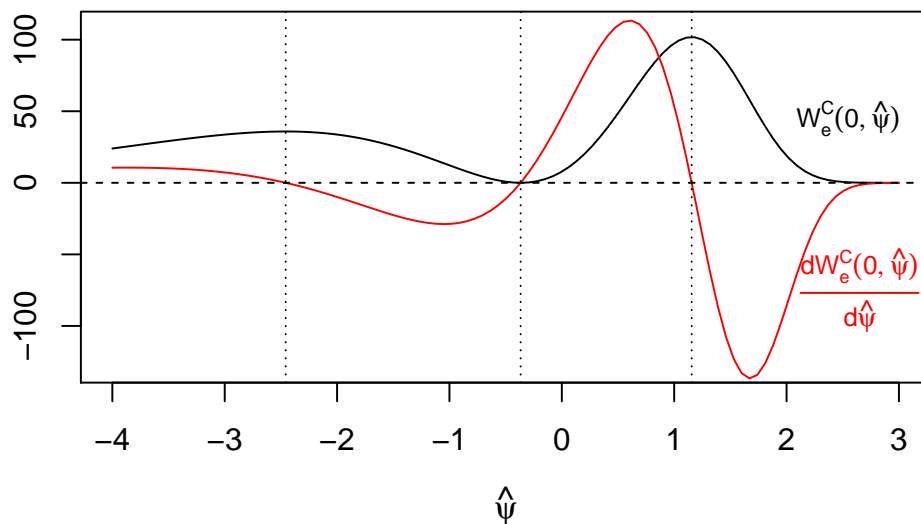


Figura 4.2: $W_e^C(-0.3665, \hat{\psi})$ e la sua derivata al variare di $\hat{\psi}$.

4.1.3 Parametrizzazione log-log

Questa parametrizzazione è $\psi = -\log[-\log(\pi)]$, $\psi \in \mathbb{R}$, e quindi $\pi = \exp[-\exp(-\psi)]$. Il test di Wald assume la forma

$$W_e^L(\psi; \hat{\psi}) = (\hat{\psi} - \psi)^2 \frac{\overbrace{\frac{i(\hat{\pi})=i(\pi(\hat{\psi}))}{n}}^{i(\hat{\pi})=i(\pi(\hat{\psi}))}}{(\exp[-\exp(-\hat{\psi})] - \exp[-2\exp(-\hat{\psi})])} \underbrace{\left(\frac{d\pi(\psi)}{d\psi} \Big|_{\psi=\hat{\psi}} \right)^2}_{\left(\frac{d\pi(\psi)}{d\psi} \Big|_{\psi=\hat{\psi}} \right)^2} \left(\exp[-\exp(-\hat{\psi})] - \hat{\psi} \right)^2 \quad (4.8)$$

Il problema di verifica di (4.1) equivale a $H_0 : \psi = \psi_0$ contro $H_1 : \psi \neq \psi_0$, $\psi_0 = 0.3665$. Si nota che il test è esattamente simmetrico al test nella parametrizzazione log-log complementare, dal momento che ψ_0^C nella parametrizzazione log-log complementare è esattamente $-\psi_0^L$ nella parametrizzazione log-log e $W_e^C(\psi_0^C, \psi) = W_e^L(-\psi_0^C, -\psi)$. Ciò è evidente anche dal confronto dei grafici in Figura 4.2 e in Figura 4.3. La simmetria ci permette di concludere che $W_e^L(\psi_0^L; \hat{\psi})$ presenta le medesime problematiche della parametrizzazione log-log complementare.

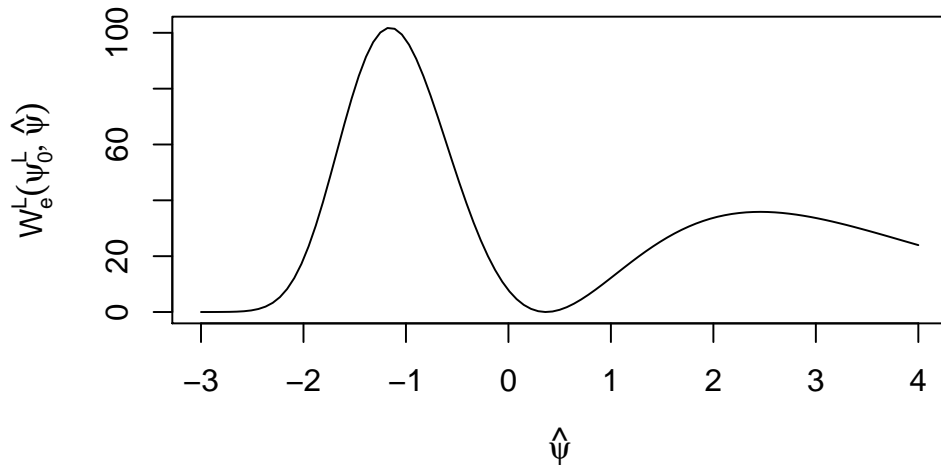


Figura 4.3: $W_e^L(\psi_0^L; \hat{\psi})$ al variare di $\hat{\psi}$.

4.1.4 Confronto tra le parametrizzazioni

Al fine di rendere confrontabili i test nelle tre parametrizzazioni prese in esame, si analizzeranno ora i test in funzione di $\hat{\pi}$, valutando anche il test nella parametrizzazione con la media π . I test valutano tutti (4.1) e risultano

$$W_e(\pi_0; \hat{\pi}) = \frac{n(\hat{\pi} - \pi_0)^2}{\hat{\pi}(1 - \hat{\pi})}, \quad (4.9)$$

$$W_e^P(\psi_0^P; \psi(\hat{\pi})) = \frac{n(\Phi^{-1}(\hat{\pi}) - \psi_0^P)^2}{\hat{\pi}(1 - \hat{\pi})} \left[\phi(\Phi^{-1}(\hat{\pi})) \right]^2, \quad (4.10)$$

$$W_e^C(\psi_0^C; \psi(\hat{\pi})) = \frac{n[\log(-\log(1 - \hat{\pi})) - \psi_0^C]^2}{\hat{\pi}(1 - \hat{\pi})} \left[-(1 - \hat{\pi}) \log(1 - \hat{\pi}) \right]^2, \quad (4.11)$$

$$W_e^L(\psi_0^L; \psi(\hat{\pi})) = \frac{n[-\log(-\log(\hat{\pi})) - \psi_0^L]^2}{\hat{\pi}(1 - \hat{\pi})} \left[-\hat{\pi} \log(\hat{\pi}) \right]^2. \quad (4.12)$$

La Figura 4.4 mostra i quattro test di Wald nelle parametrizzazioni prese in esame. Come già visto in precedenza, le parametrizzazioni usuali (log-log complementare, log-log e probit) presentano tutte il problema di non monotonicità del test Wald, insieme tra l'altro a quella canonica (come mostrato da Hauck e Donner (1977) e visto nel paragrafo 2.1). Le parametrizzazioni log-log e log-log complementare, a causa della loro non simmetria intorno a 0.5 (si veda Agresti, 2015, pagine 184-186) hanno inoltre la peculiarità di portare a un test di Wald non simmetrico, che tuttavia presenta, in entrambi i casi (dal momento che i test sono tra loro simmetrici) un comportamento anomalo.

Seguono i comandi per il confronto grafico dei test.

```
n<-100
pi.0 <- 0.5
psi.0.P<- 0
psi.0.C<- (-0.3665)
psi.0.L<- 0.3665
wald.media.pi <- function(x)
{ n * (x-pi.0)^2 / (x*(1-x)) }
```

```

wald.probit.pi <- function(x)
{ n*(qnorm(x)-psi.0.P)^2 * (dnorm(qnorm(x)))^2 /(x*(1-x)) }
wald.cloglog.pi <- function(x)
{ n*(log(-log(1-x)) - psi.0.C)^2 *(-(1-x)*log(1-x))^2 /(x*(1-x))}
wald.loglog.pi <- function(x)
{ n*(-log(-log(x)) - psi.0.L)^2 *(-x*log(x))^2 / (x*(1-x)) }
curve(wald.media.pi(x), xlim = c(0.00001,0.99999), ylim = c(0,125), xlab =
      expression(hat(pi)), ylab = expression(W[e]))
curve(wald.probit.pi(x),add=T, col=2)
curve(wald.cloglog.pi(x), add=T, col=3)
curve(wald.loglog.pi(x), add=T, col=4)

```

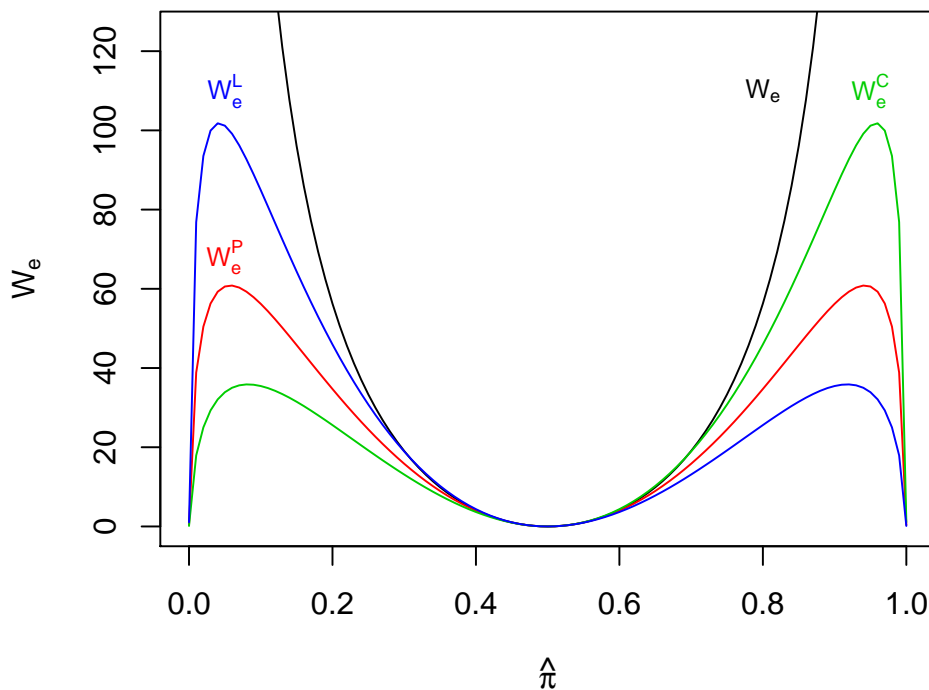


Figura 4.4: Confronto dei test di Wald al variare di $\hat{\pi}$. Parametrizzazione con la media (nero), parametrizzazione probit (rosso), parametrizzazione log-log complementare (verde), parametrizzazione log-log (blu).

4.2 Anomalie nel modello di regressione probit

Si è visto come nel caso di campionamento casuale semplice la parametrizzazione probit porti il test Wald a non essere monotono all'allontanarsi della stima di massima verosimiglianza dall'ipotesi nulla. Væth (1985), per la classe di parametrizzazioni individuata dalla (2.9), ha dimostrato che si riscontra il comportamento anomalo in ambito di modelli lineari generalizzati, riguardo ai test di nullità dei singoli parametri, quando nel corrispondente caso di campionamento casuale semplice si riscontra tale comportamento anomalo.

Dal momento che la parametrizzazione probit non appartiene a quelle indagate da Væth (1985), non è ovvio che si possa automaticamente estendere questo risultato alla funzione di legame probit in ambito di modelli lineari generalizzati. Nel seguito, seguendo il metodo illustrato in Hauck e Donner (1977), si verifica che il test di Wald per la nullità del singolo parametro non si comporta in modo coerente.

Il modello di regressione probit ipotizza $Y_i \sim Bi(1, \pi_i)$ indipendenti, $\eta_i = \sum_{j=1}^k \beta_j x_{ij}$ e $\pi_i = \Phi(\eta_i)$. Il test di Wald assume la forma

$$W_{eP}(\beta_k; \hat{\beta}) = \frac{(\hat{\beta}_k - \beta_k)^2}{i^{kk}(\hat{\beta})}. \quad (4.13)$$

Il problema di verificare la nullità di β_k si traduce nel verificare $H_0 : \beta_k = 0$ contro $H_1 : \beta_k \neq 0$. Suddividendo la matrice $i(\hat{\beta})$ in blocchi di dimensioni $(k-1)$ e 1 , così da isolare la parte relativa al parametro β_k e ricordando le formule (1.9) per le inverse di matrici a blocchi, si può affermare che $i_{kk}(\hat{\beta})$ è maggiore di $[i^{kk}(\hat{\beta})]^{-1}$.

Possiamo quindi affermare che la quantità in (4.13) è minore o uguale a

$$\hat{\beta}_k^2 i_{kk}(\hat{\beta}) \quad (4.14)$$

con $i_{kk}(\hat{\beta})$ definita, come in Agresti (2015, pagina 184), dalla seguente espressione

$$i_{kk}(\hat{\beta}) = \sum_{i=1}^n \frac{[\phi^2(\hat{\eta}_i) x_{ik}^2]}{\Phi(\hat{\eta}_i)[1 - \Phi(\hat{\eta}_i)]}. \quad (4.15)$$

Si vuole valutare se il test è monotono all'allontanarsi della stima di massima verosimiglianza $\hat{\beta}_k$ dall'ipotesi nulla.

Si dimostra invece che per $\hat{\beta}_k \rightarrow \pm\infty$ il test tende a zero. Infatti ogni elemento della sommatoria che compone la quantità (4.14) tende a zero, per $\hat{\beta}_1, \dots, \hat{\beta}_{k-1}$ fissati. Il limite per ogni singolo elemento della sommatoria vale

$$\lim_{\hat{\beta}_k \rightarrow \pm\infty} \frac{\hat{\beta}_k^2 \phi^2(\hat{\eta}_i) x_{ik}^2}{\Phi(\hat{\eta}_i)[1 - \Phi(\hat{\eta}_i)]} \quad (4.16)$$

Se x_{ik} vale 0, il termine è zero. Nel caso in cui $x_{ik} \neq 0$ si ottiene una forma indeterminata del tipo $\frac{0}{0}$. Applicando la regola di de l'Hôpital e ricordando che $d\hat{\eta}_i/d\hat{\beta}_k = x_{ik}$ e si ottiene

$$\lim_{\hat{\beta}_k \rightarrow \pm\infty} \frac{2[\hat{\beta}_k \phi^2(\hat{\eta}_i) + \hat{\beta}_k^2 \phi'(\hat{\eta}_i) x_{ik}] x_{ik}^2}{[x_{ik} - 2\Phi(\hat{\eta}_i) x_{ik}]}. \quad (4.17)$$

Semplificando si ottiene

$$\lim_{\hat{\beta}_k \rightarrow \pm\infty} \frac{2\hat{\beta}_k x_{ik} \phi(\hat{\eta}_i) + 2\hat{\beta}_k^2 x_{ik}^2 \phi'(\hat{\eta}_i)}{1 - 2\Phi(\hat{\eta}_i)}. \quad (4.18)$$

Il denominatore del limite è un valore finito pari a +1 o -1, a seconda che $\hat{\eta}_i$ tenda a $-\infty$ o $+\infty$ rispettivamente. Il numeratore invece, visto come somma di due addendi, tende a zero, in quanto ogni singolo addendo tende a zero. Infatti

$$\lim_{\hat{\beta}_k \rightarrow \pm\infty} 2\hat{\beta}_k \phi(\hat{\eta}_i) x_{ik} \quad (4.19)$$

tende a zero, a causa del termine $e^{-\frac{1}{2}\hat{\eta}_i^2}$ il quale tende a zero più velocemente di quanto $\hat{\beta}_k$ tenda a infinito. Analogamente il secondo termine del limite

$$\lim_{\hat{\beta}_k \rightarrow \pm\infty} 2\hat{\beta}_k^2 \phi'(\hat{\eta}_i) x_{ik}^2 \quad (4.20)$$

tende a zero a causa del fattore esponenziale contenuto in $\phi'(\hat{\eta}_i)$. La derivata della densità normale è infatti

$$\phi'(x) = -x\phi(x). \quad (4.21)$$

In conclusione, il test di Wald per la nullità di β_k , per $\hat{\beta}_k \rightarrow \pm\infty$, tende a 0. Il comportamento anomalo quindi, oltre che nel modello con osservazioni i.i.d. è presente anche nel modello di regressione probit.

Capitolo 5

Conclusioni

In questa relazione si è posta l'attenzione principalmente sul comportamento del test di Wald nei modelli binomiale e Poisson, sottolineandone i differenti comportamenti al variare della parametrizzazione e, in ambito di regressione, al variare della funzione di legame.

Si è mostrato che nel modello di Poisson, riprendendo i risultati già trovati da Væth (1985), sia in caso di osservazioni i.i.d. e parametrizzazione canonica, sia in ambito di regressione utilizzando la funzione di legame canonica, il test di Wald non è monotono al distanziarsi dall'ipotesi nulla quando le medie stimate sono vicine a zero.

Si è proceduto quindi ad analizzare un insieme di parametrizzazioni e funzioni di legame proprie del modello binomiale, sia in ambito di osservazioni omogenee, sia in ambito di regressione. In questo caso non ci si è limitati a riprendere risultati già dimostrati, ma si è proceduto a mostrare, in particolare, che le parametrizzazioni probit, log-log complementare e log-log, nel caso di osservazioni omogenee, presentano tutte il problema di non monotonicità del test di Wald. Si è inoltre dimostrato che il test di Wald per la nullità del singolo parametro di regressione, se si utilizza la funzione di legame probit, presenta il problema di non monotonicità.

La classe di modelli analizzati è comunque ristretta, e lascia spazio a numerosi ampliamenti dell'analisi del comportamento del test di Wald, ad esempio alle famiglie esponenziali multiparametriche, alle famiglie di dispersione esponenziale o a modelli normali non lineari.

Bibliografia

- Agresti, A. *Foundations of Linear and Generalized Linear Models*. Hoboken: Wiley, 2015.
- Barndorff-Nielsen, O. *Information and Exponential Families in Statistical Theory*. Chichester: Wiley, 1978.
- Freedman, D.A. «How can the score test be inconsistent?» In: *The American Statistician* 61.4 (2007), pp. 291–295.
- Hauck, W.W. e A. Donner. «Wald's test as applied to hypotheses in logit analysis». In: *Journal of the American Statistical Association* 72 (1977), pp. 851–853.
- Pace, L. e A. Salvani. *Introduzione alla statistica II Inferenza, Verosimiglianza, Modelli*. Milano: Cedam, 2001.
- *Teoria della Statistica*. Padova: Cedam, 1996.
- Væth, M. «On the use of Wald's test in exponential families». In: *International Statistical Review* 53, 2 (1985), pp. 199–214.
- Wedderburn, R. W. M. «On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models». In: *Biometrika* 63, 1 (1976), pp. 27–32.