

UN APPROCCIO ROBUSTO ALLA VERIFICA D'IPOTESI BASATO SULLA FUNZIONE DI VEROSIMIGLIANZA PESATA

Claudio Agostinelli

1. INTRODUZIONE

La verifica d'ipotesi basata sulle funzioni test rapporto di verosimiglianza di Neyman e Pearson (1928) e Wilks (1938) sono ampiamente utilizzate nell'inferenza parametrica. In generale, queste funzioni test hanno ottime proprietà asintotiche ma sono sensibili alla presenza di valori anomali nei dati. La presenza di contaminazione può comportare forti distorsioni al livello o alla potenza di questi test e quindi alterare la conclusione inferenziale. Per ovviare a questo problema molti autori hanno proposto test statistici con desiderabili proprietà di robustezza tra gli altri Huber (1965, 1981), Hampel et al. (1986), Ronchetti (1982), Heritier e Ronchetti (1994) e Markatou e He (1994). Una funzione test ha buone proprietà di robustezza se il suo livello e la sua potenza rimangono stabili in presenza di valori anomali nei dati, di valori cioè che non provengono dal modello specificato. Nella letteratura la misura della robustezza è fatta considerando la stabilità della funzione test per contaminazioni infinitesimali attraverso l'uso, ad esempio, della Funzione d'Influenza del livello e della potenza (Ronchetti, 1982) e valutando il valore di contaminazione più elevato che la funzione test può sopportare attraverso la funzione di punto di rottura del livello e della potenza. Simpson (1989) e He et al. (1990) impiegano la funzione di punto di rottura per caratterizzare le proprietà di robustezza di un test statistico. Simpson (1989) allo scopo di ottenere test statistici robusti e allo stesso tempo efficienti ha proposto una funzione test analoga ai test rapporto di verosimiglianza ma basate sulla distanza di Hellinger.

I test proposti in Heritier e Ronchetti (1994) sono derivati dagli stimatori di tipo M. Lo scopo è quello di ottenere dei test che siano robusti a bassi livelli di contaminazione. Come per i corrispondenti stimatori i test non risultano pienamente efficienti, rispetto ai test di verosimiglianza, quando non vi sono contaminazioni nei dati. Inoltre, il corrispondente test robusto del rapporto di verosimiglianza (*likelihood ratio-type*) ha una distribuzione asintotica più complessa di un χ^2 .

In questo lavoro presentiamo tre funzioni test, basate sulla funzione di verosimiglianza pesata analoghe a quelle della funzione di verosimiglianza (test di Wald, test del punteggio o *score* test e del rapporto di verosimiglianza) allo scopo di ottenere un'inferenza robusta e allo stesso tempo efficiente. Le tre funzioni test, sotto

condizioni non restrittive, hanno la stessa distribuzione asintotica delle versioni classiche, sono pienamente efficienti quando non vi sono contaminazioni nei dati e mantengono buone proprietà di robustezza ad elevati livelli di contaminazione.

Nella sezione 2 presentiamo brevemente l'approccio all'inferenza basata sulla funzione di verosimiglianza pesata nel caso di modelli con supporto continuo. Una discussione completa di questa metodologia può essere trovata in Markatou *et al* (1995, 1998). Alcune estensioni possono essere trovate in Agostinelli e Markatou (1998), Agostinelli (1997, 1998a, 1998b). Nella sezione 3 introduciamo le tre funzioni test, nella sezione 4 studiamo le proprietà asintotiche e di potenza contro alternative locali mentre le proprietà di robustezza sono studiate nella sezione 5 attraverso la funzione di punto di rottura. Nella sezione 6 sono riportati alcuni esempi e una simulazione Monte Carlo, mentre nell'appendice sono riportate le assunzioni e alcuni lemmi.

2. LA VEROSIMIGLIANZA PESATA

Consideriamo una variabile aleatoria reale X con funzione di ripartizione $F(x)$ e densità $f(x)$ rispetto alla misura di Lebesgue nello spazio campionario $\Omega \subseteq \mathcal{R}$. Una funzione di ripartizione appartenente alla famiglia parametrica $M = \{M(x; \theta) : \theta \in \Theta\}$ ed avente densità $m(x; \theta)$ rispetto alla misura di Lebesgue e $\Theta \in \mathcal{R}^r$ viene scelta come modello per descrivere la variabile aleatoria X . Nell'approccio classico basato sulla funzione di verosimiglianza si assume che la famiglia parametrica sia stata specificata in modo corretto, cioè che esista un valore del parametro, diciamo θ_0 , all'interno dello spazio parametrico Θ tale che $F(x) \equiv M(x; \theta_0)$. Diversamente, nell'approccio robusto, si assume che nella famiglia parametrica M esista un elemento M_{θ_0} tale che F stia in un suo "intorno" (Questo intorno può essere definito come modello ad errori grossolani o *gross error model*, si veda Huber, (1981) anche per altri tipi di intorni).

Consideriamo un campione di n osservazioni indipendenti $x = \{x_1, x_2, \dots, x_n\}$ dalla variabile aleatoria X . Posto $\mathbf{1}_{\omega}(x)$ la funzione indicatrice dell'insieme ω , definiamo $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty, x]}(x_i)$ la funzione di ripartizione empirica del campione \mathbf{x}

e $u(x; \theta) = \frac{d}{d\theta} \log m(x; \theta)$ la funzione punteggio che è la derivata prima della funzione di log-verosimiglianza. Lo stimatore di massima verosimiglianza $\hat{\theta}_n$ del parametro θ , nel caso di problemi regolari di stima (vedi le ipotesi b riportate in appendice) è definito come la soluzione dell'equazione $\sum_{i=1}^n u(x_i; \theta) = 0$.

Dato un punto x nello spazio campionario Ω possiamo costruire una funzione peso dipendente da x , dall'elemento della famiglia parametrica M indicizzato da θ e dall'ignota distribuzione F attraverso la funzione di ripartizione empirica \hat{F}_n , cioè $w(x; \theta, \hat{F}_n)$. Lo stimatore di verosimiglianza pesata $\hat{\theta}_w$ del parametro θ nel caso di problemi regolari di stima θ è definito come la soluzione dell'equazione:

$$\int w(x; \theta, \hat{F}_n) u(x; \theta) d\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n w(x_i; \theta, \hat{F}_n) u(x_i; \theta) = 0 \quad (1)$$

Lo stimatore di verosimiglianza pesata viene utilizzato per la costruzione di test robusti per le sue buone proprietà: esso gode delle stesse proprietà dello stimatore di verosimiglianza, sotto condizioni non restrittive, quando non vi è contaminazione nei dati; viceversa presenta proprietà di robustezza (in particolare possiede un punto di rottura positivo) in presenza di valori anomali. Inoltre, a differenza di altri stimatori robusti presentati in letteratura, la metodologia è applicabile per una ampia classe di inodelli.

Il tipo di funzione peso che viene usata deriva dal metodo di stima a minima distanza e si basa su una funzione chiamata residui di Pearson. Quest'ultima è una misura di discrepanza tra la forma della densità della variabile casuale $f(x)$ e la forma della densità del modello specificato $m(x; \theta)$. Essa è definita nel seguente modo

$$\delta(x; \theta, \hat{F}_n) = \frac{f^n(x) - m^n(x; \theta)}{m^n(x; \theta)} = \frac{f^n(x)}{m^n(x; \theta)} - 1$$

dove $f(x) = \int k(x; t, v) dF_t(t)$ è uno stimatore non parametrico della densità $f(x)$ basato sul campione x e sul nucleo $k(x; t, v)$ con parametro di liscio t , mentre $m(x; \theta) = \int k(x; t, v) dM(t; \theta)$ è una versione liscia della densità $m(x; \theta)$ basata sul medesimo stimatore del nucleo.

È importante notare che il codominio dei residui di Pearson δ $(-1, +\infty]$ essi sono nulli in quelle osservazioni per cui la densità attesa (misurata attraverso il modello ipotizzato) viene a coincidere con quella osservata (corrispondente al valore dello stimatore non parametrico).

La funzione peso è definita nel seguente modo.

$$w(x; \theta, \hat{F}_n) = \frac{A(\delta(x; \theta, \hat{F}_n)) + 1}{\delta(x; \theta, \hat{F}_n) + 1} \quad (2)$$

dove $A(\delta(x; \theta, \hat{F}_n))$ è una funzione dei residui di Pearson chiamata funzione di aggiustamento dei residui (*Residual Adjustment Function, RAF*) introdotta da Lindsay (1994). Alcuni esempi sono: la *KAF* di Hellinger $A_{HD}(\delta) = 2[\delta + 1 - |\delta|]$ e la *Negative Exponential Disparity* $A_{NED}(\delta) = 2 - (2 + \delta) \exp(-\delta)$. Il comportamento dei residui di Pearson è illustrato in Agostinelli (1998a) mentre per una discussione sul ruolo della funzione di aggiustamento dei residui, si veda Lindsay (1994), Markatou (1996) e Markatou *et al.* (1998); infine Cressie e Read (1984) per una classe di misure di divergenza da cui ottenere funzioni di aggiustamento dei residui. Si osservi che per alcune famiglie di aggiustamento dei residui e per particolari forme di contaminazione è possibile che i pesi definiti in (2) assumano valori negativi o superiori all'unità (in particolare quando la numerosità campionaria è piccola), così nella pratica per rendere gli algoritmi di calcolo più stabili è più conveniente utilizzare questa seconda definizione della funzione peso:

$$\omega(x; \theta, \hat{F}_n) = \min \left\{ 1, \frac{[A(\delta(x; \theta, \hat{F}_n)) + 1]^+}{\delta(x; \theta, \hat{F}_n) + 1} \right\}$$

dove $[\cdot]_+$ vale 0 se l'argomento è negativo e vale l'argomento stesso se l'argomento è positivo. In questo modo si ottiene una funzione che assume valori tra 0 e 1. In Markatou *et al.* (1995, 1998) vengono studiate le proprietà asintotiche dello stimatore di verosimiglianza pesata; essi dimostrano che sotto appropriate condizioni (riportate negli articoli appena citati) lo stimatore risulta consistente e asintoticamente normale.

Un aspetto pratico importante riguarda la scelta della forma del nucleo e della grandezza del parametro di liscio necessari per ottenere la stima $f(x)$ e la versione liscia del modello $m(x; 0)$. Mentre la forma del nucleo (entro certi limiti) influenza solo marginalmente i valori numerici delle stime, il parametro di liscio può avere importanti effetti sulle proprietà di robustezza dello stimatore. Brevemente, la scelta di un parametro di liscio di valore elevato produce una stima che non evidenzia la struttura della distribuzione dei dati e le stime che si ottengono sono in genere molto simili a quelle del metodo di massima verosimiglianza.

3. TRE FUNZIONI TEST BASATE SULLA VEROSIMIGLIANZA PESATA

Consideriamo per semplicità una variabile aleatoria X con distribuzione $F(x) = (1 - \varepsilon)M(x; \theta_T) + \varepsilon G(x)$ e $G(x)$ una qualsiasi funzione di ripartizione. Riprendendo la notazione introdotta nella sezione precedente, definiamo la matrice di informazione attesa di Fisher $i(\theta) = E_{\theta}(u(x; \theta); u(x; \theta))$. Una sua versione pesata chiamata matrice di informazione di Fisher pesata può essere definita come segue: $i_w(\theta) = E_F[\omega(x; \hat{\theta}_n, F)]$ (si veda Agostinelli, 1998a). Dato che il secondo fattore della i_w è generalmente sconosciuto consideriamo anche una versione per campioni finiti nella forma $i_w(\theta) = i(\theta) E_{\hat{F}_n}[\omega(x; \hat{\theta}_w, \hat{F}_n)]$.

Il nostro interesse è nel verificare il sistema di ipotesi:

$$H_0 : \theta_T \in \Theta_0$$

$$H_1 : \theta_T \in \Theta_1 \subseteq \Theta / \Theta_0 \quad (3)$$

Introduciamo ora le tre funzioni test. La funzione test basata sul rapporto di verosimiglianza pesato (*Weighted Likelihood Ratio test*) è definita a partire dalla seguente funzione:

$$\lambda_w(\theta) = -2E_{\hat{F}_n}[\omega(x; \hat{\theta}_w, \hat{F}_n)(l(x; \theta) - l(x; \hat{\theta}_w))] \quad (4)$$

la funzione del test di Wald pesato (*Weighted Wald test*) è definita come segue:

$$W_w(\theta) = (\theta - \hat{\theta}_w)' i_w(\hat{\theta}_w) (\theta - \hat{\theta}_w) = E_F[\omega(x; \hat{\theta}_w, F)] (\theta - \hat{\theta}_w)' i(\hat{\theta}_w) (\theta - \hat{\theta}_w) \quad (5)$$

mentre il test basato sulla funzione punteggio pesato (*Weighted Score test*) è:

$$\begin{aligned}
 S_w(\theta) &= E_{\hat{F}_n} [w(x; \hat{\theta}_w, \hat{F}_n) u(x; \theta)]' i_w(\theta)^{-1} E_{\hat{F}_n} [w(x; \hat{\theta}_w, \hat{F}_n) u(x; \theta)] \\
 &= \frac{1}{E_F[w(x; \hat{\theta}_w, F)]} E_{\hat{F}_n} [w(x; \hat{\theta}_w, \hat{F}_n) u(x; \theta)]' i(\theta)^{-1} E_{\hat{F}_n} [w(x; \hat{\theta}_w, \hat{F}_n) u(x; \theta)] \quad (6)
 \end{aligned}$$

Dalla funzione che definisce il test basato sulla funzione punteggio pesata, si noti, che diversamente dall'analogo test basato sulla funzione punteggio, non è sufficiente la stima sotto l'ipotesi nulla H_0 . Questo è uno svantaggio apparente se si pensa che una analisi dei dati comprende sempre la stima non vincolata.

Per semplicità le definizioni sono state date senza il fattore moltiplicativo n dimensione del campione. Per valutare la funzione test, sarà inoltre necessario utilizzare J_u invece di i_w . Mostreremo nel seguito che le proprietà asintotiche dei test rimangono inalterate sotto condizioni non restrittive. Infine con $\hat{\theta}_w$ intenderemo la soluzione dell'equazione di verosimiglianza pesata 1, oppure la stima ad un passo come definita in Agostinelli (1977) e Agostinelli e Markatou (1998).

L'equazione di verosimiglianza pesata 1 non corrisponde ad un problema di ottimizzazione, così per definire il valore della funzione test nel caso di ipotesi composte c/o in presenza di parametri di disturbo utilizziamo la soluzione derivante dalla seguente funzione di verosimiglianza pesata, con pesi calcolati in $\hat{\theta}_w$:

$$\hat{\theta}_0 = \arg \max_{\theta \in \Theta_0} \prod_{i=1}^n m(x_i; \theta)^{w(x_i; \hat{\theta}_w, F_n)} \quad (7)$$

Questo approccio è simile a quello classico (si confronti ad esempio Cox and Hinkley, 1974).

Consideriamo un semplice esempio per chiarire come le funzioni test vadano calcolate e come esse presentino analogie con le funzioni classiche basate sulla verosimiglianza: la famiglia parametrica normale con media ignota θ e varianza nota σ^2 . È noto che in questo caso i tre test basati sulla verosimiglianza coincidono. Semplici manipolazioni algebriche mostrano che lo stesso risultato vale anche per i tre test basati sulla verosimiglianza pesata, indipendentemente dalla presenza o meno del fattore contaminante. Infatti, sia \mathbf{I} : una funzione di distribuzione e come usuale $\hat{\theta}_w(F)$ la soluzione dell'equazione di verosimiglianza pesata $E_F[w(x; \theta, F) (\theta - x)] \sigma^{-2} = 0$ e quindi $E_F[w(x; \hat{\theta}_w, F)] \hat{\theta}_w(F) = E_F[w(x; \hat{\theta}_w, F) x]$ siccome la funzione punteggio $u(x; \theta)$ è $(x - \theta) \sigma^{-2}$. Il test di Wald pesato è definito come segue.

$$W_w(\theta, F) = E_F[w(x; \hat{\theta}_w, F)] (\theta - \hat{\theta}_w)^2 \sigma^{-2} \quad (8)$$

mentre il test basato sul rapporto di verosimiglianza pesato è:

$$\begin{aligned}
 \lambda_w(\theta, F) &= -2 E_F[w(x; \hat{\theta}_w, F) (\log m(x; \theta) - \log m(x; \hat{\theta}_w))] \\
 &= E_F[w(x; \hat{\theta}_w, F) ((x - \theta)^2 - (x - \hat{\theta}_w)^2)] \sigma^{-2} = W_w(\theta, F)
 \end{aligned}$$

ed infine il test della funzione punteggio pesata vale:

$$\begin{aligned}
 S_w(\theta, F) &= E_F[w(x; \hat{\theta}_w, F) (x - \theta) \sigma^{-2}]^2 / E_F[w(x; \hat{\theta}_w, F) \sigma^{-2}] \\
 &= \sigma^{-2} E_F[w(x; \hat{\theta}_w, F)]^2 (\hat{\theta}_w - \theta)^2 / E_F[w(x; \hat{\theta}_w, F)] = W_w(\theta, F)
 \end{aligned}$$

4. EQUIVALENZA ASINTOTICA

In questa sezione studiamo il comportamento asintotico dei tre test quando il modello è correttamente specificato. A questo scopo assumiamo che $\varepsilon=0$ così $F(x) \equiv M(x; \theta_T)$. Sia \hat{O} lo stimatore di massima verosimiglianza e \hat{O}_w lo stimatore di verosimiglianza pesata. In questo caso entrambi gli stimatori sono consistenti sotto appropriate condizioni. Se supponiamo $0 < i_w(\theta) \leq i(\theta) < +\infty$ (nel senso dell'ordinamento delle matrici) attraverso il lemma 2, possiamo mostrare l'asintotica equivalenza tra i test basati sulla verosimiglianza pesata e quelle basati sulla verosimiglianza. Posto $\lambda(\theta) = -2E_{\hat{F}_n} [l(x; \hat{O}) - l(x; \hat{\theta})]$ il test rapporto di verosimiglianza, $W(\theta) = (\hat{O} - \hat{\theta})' i(\theta) (\hat{O} - \hat{\theta})$ il test di Wald e $S(\theta) = u(\theta)' i(\theta)^{-1} u(\theta)$ il test di Rao o della funzione punteggio allora:

$$\begin{aligned} |\lambda(\theta) - \lambda(\hat{\theta})| &= 2 |E_{\hat{F}_n} [w(x; \hat{\theta}_w, \hat{F}_n) (l(x; \hat{O}) - l(x; \hat{\theta}_w)) - (l(x; \hat{O}) - l(x; \hat{\theta}))]| \\ &\leq 2 \sup_t |w(t; \hat{\theta}_w, \hat{F}_n) - 1| |E_{\hat{F}_n} [l(x; \hat{O}) - l(x; \hat{\theta})]| \\ &\quad + 2E_{\hat{F}_n} |l(x; \hat{\theta}) - l(x; \hat{\theta}_w)| \end{aligned} \quad (9)$$

quindi per il lemma 2 e il fatto che $E_{\hat{F}_n} |l(x; \hat{\theta}) - l(x; \hat{\theta}_w)|$ è asintoticamente zero per la consistenza di \hat{O} e \hat{O}_w , il test rapporto di verosimiglianza pesato è asintoticamente equivalente al test rapporto di verosimiglianza per $\theta = \theta_T$. Questo corrisponde all'equivalenza dei due test sotto l'ipotesi nulla. Per studiare il comportamento asintotico della potenza locale del test seguiamo Cox e Hinkley (1974) e poniamo $\theta = \theta_T + \varepsilon$. Usando il fatto che $l(x; \theta_T + \varepsilon) - l(x; \theta_T) = \varepsilon u(x; \theta_T) + o(\varepsilon)$ abbiamo:

$$\begin{aligned} &\leq 2 \sup_t |w(t; \hat{\theta}_w, \hat{F}_n) - 1| \{ |E_{\hat{F}_n} [l(x; \theta) - l(x; \theta_T)]| + |E_{\hat{F}_n} [l(x; \theta_T) - l(x; \hat{\theta})]| \} \\ &\quad + 2E_{\hat{F}_n} |l(x; \hat{\theta}) - l(x; \hat{\theta}_w)| \\ &= 2 \sup_t |w(t; \hat{\theta}_w, \hat{F}_n) - 1| \{ |\varepsilon| |E_{\hat{F}_n} [u(x; \theta_T)]| + |o(\varepsilon)| + |E_{\hat{F}_n} [l(x; \theta_T) - l(x; \hat{\theta})]| \} \\ &\quad + 2E_{\hat{F}_n} |l(x; \hat{\theta}) - l(x; \hat{\theta}_w)| \end{aligned} \quad (10)$$

Siccome $E_{M_T} |u(x; \theta_T)| \leq 1 + i(\theta_T)$, se $\varepsilon \rightarrow 0$ per $n \rightarrow \infty$, componente per componente, il test rapporto di verosimiglianza pesato e quello del rapporto di verosimiglianza hanno asintoticamente la stessa potenza locale. Consideriamo ora la differenza tra la funzione test di Wald pesata e la funzione del test di Wald:

$$\begin{aligned} |W_w(\theta) - W(\theta)| &= |(\theta - \hat{\theta}_w)' i_w(\hat{\theta}_w) (\theta - \hat{\theta}_w) - (\theta - \hat{\theta})' i(\hat{\theta}) (\theta - \hat{\theta})| \\ &\leq | \{ (\theta - \hat{\theta}_w)' i_w(\hat{\theta}_w) (\theta - \hat{\theta}_w) - (\theta - \hat{\theta})' i_w(\hat{\theta}_w) (\theta - \hat{\theta}) \} | \\ &\quad + |(\theta - \hat{\theta})' [i_w(\hat{\theta}_w) - i(\hat{\theta})] (\theta - \hat{\theta})| \end{aligned}$$

Posto $A(\theta) = \left| \{(\theta - \hat{\theta}_w)' i_w(\hat{\theta}_w) (\theta - \hat{\theta}_w) - (\theta - \hat{\theta})' i_w(\hat{\theta}_w) (\theta - \hat{\theta})\} \right|$ e $B(\theta) = \left| (\theta - \hat{\theta})' [i_w(\hat{\theta}_w) - i(\hat{\theta})] (\theta - \hat{\theta}) \right|$ abbiamo:

$$\begin{aligned} B(\theta) &= |(\theta - \hat{\theta})' [i_w(\hat{\theta}_w) + i(\hat{\theta}_w) - i(\hat{\theta}_w) - i(\hat{\theta})] (\theta - \hat{\theta})| \\ &\leq |(\theta - \hat{\theta})' [i_w(\hat{\theta}_w) - i(\hat{\theta}_w)] (\theta - \hat{\theta})| + |(\theta - \hat{\theta})' [i(\hat{\theta}_w) - i(\hat{\theta})] (\theta - \hat{\theta})| \end{aligned}$$

mentre il comportamento della matrice di informazione di Fisher pesata può essere studiato attraverso:

$$\begin{aligned} |j_w(\theta) - i(\theta)| &= |E_{M_T} [w(x; \hat{\theta}_w, M_T) - 1] i(\theta)| \\ &\leq \sup_t |w(t; \hat{\theta}_w, M_T) - 1| |i(\theta)| \end{aligned} \quad (11)$$

Per il lemma 2 e la limitatezza di $i(\theta_w)$ la 11 è asintoticamente zero (in probabilità) e $|i(\hat{\theta}_w) - i(\hat{\theta})|$ è asintoticamente zero (in probabilità) per la continuità di i rispetto al suo argomento θ e per la consistenza di $\hat{\theta}$ e $\hat{\theta}_w$. $A(\theta)$ può essere riformulata come segue:

$$\begin{aligned} A(\theta) &= |(\theta - \hat{\theta}_w)' i_w(\hat{\theta}_w) (\hat{\theta} - \hat{\theta}_w) + (\hat{\theta} - \hat{\theta}_w)' i_w(\hat{\theta}_w) (\theta - \hat{\theta})| \\ &\leq 2 |(\theta - \theta_T)' i_w(\hat{\theta}_w) (\hat{\theta} - \hat{\theta}_w)| + |(\theta_T - \hat{\theta}_w)' i_w(\hat{\theta}_w) (\hat{\theta} - \hat{\theta}_w)| \\ &\quad + |(\theta_T - \hat{\theta})' i_w(\hat{\theta}_w) (\hat{\theta} - \hat{\theta}_w)| \end{aligned}$$

Infine:

$$\begin{aligned} |W_w(\theta) - W(\theta)| &\leq 2 |(\theta - \theta_T)' i_w(\hat{\theta}_w) (\hat{\theta} - \hat{\theta}_w)| \\ &\quad + |(\theta_T - \hat{\theta}_w)' i_w(\hat{\theta}_w) (\hat{\theta} - \hat{\theta}_w)| + |(\theta_T - \hat{\theta}_w)' i_w(\hat{\theta}_w) (\hat{\theta} - \hat{\theta}_w)| \\ &\quad + \sup_t |w(t; \hat{\theta}_w, M_T) - 1| |(\theta - \hat{\theta})' i(\hat{\theta}_w) (\theta - \hat{\theta})| \\ &\quad + |(\theta - \hat{\theta})' [i(\hat{\theta}_w) - i(\hat{\theta})] (\theta - \hat{\theta})| \end{aligned}$$

Quindi per il lemma 2 il test di Wald pesato è asintoticamente equivalente al test di Wald per $\theta = \theta_T$. Questo corrisponde all'equivalenza sotto l'ipotesi nulla. Per studiare la potenza locale asintotica poniamo $\theta = \theta_T + \varepsilon$ dove si ha:

$$|\theta - \hat{\theta}| = |\theta_T - \hat{\theta} + \varepsilon| \leq |\theta_T - \hat{\theta}| + |\varepsilon| \quad (12)$$

così quando $\varepsilon \rightarrow 0$ per $n \rightarrow \infty$, si ottiene l'equivalenza componente per componente. Si noti che lo stesso risultato asintotico viene ottenuto utilizzando la matrice di informazione di Fisher pesata nella sua versione per campioni finiti, infatti:

$$\begin{aligned} |j_w(\theta) - i(\theta)| &= |E_{\hat{F}_n} [w(x; \hat{\theta}_w, \hat{F}_n) - 1] i(\theta)| \\ &\leq \sup_t |w(t; \hat{\theta}_w, \hat{F}_n) - 1| |i(\theta)| \end{aligned}$$

Infine, posto $u_w(\theta) = E_{\hat{F}_n} [w(x; \hat{\theta}_w, \hat{F}_n) u(x; \theta)]$ e considerando la differenza tra il test basato sulla funzione punteggio e la sua versione pesata si ottiene:

$$\begin{aligned} |S_w(\theta) - S(\theta)| &= |u_w(\theta)' i_w(\theta)^{-1} u_w(\theta) - u(\theta)' i(\theta)^{-1} u(\theta)| \\ &= \{u_w(\theta)' i_w(\theta)^{-1} u_w(\theta) - u(\theta)' i_w(\theta)^{-1} u(\theta)\} \\ &\quad + u(\theta)' [i_w(\theta)^{-1} - i(\theta)^{-1}] u(\theta) \end{aligned}$$

Posto $C(\theta) = (u_w(\theta)' i_w(\theta)^{-1} u_w(\theta) - u(\theta)' i_w(\theta)^{-1} u(\theta))$ e $D(\theta) = u(\theta)' [i_w(\theta)^{-1} - i(\theta)^{-1}] u(\theta)$ abbiamo:

$$C(\theta) = u_w(\theta)' i_w(\theta)^{-1} (u_w(\theta) - u(\theta)) + (u_w(\theta) - u(\theta))' i_w(\theta)^{-1} u(\theta)$$

inoltre

$$\begin{aligned} |u_w(\theta) - u(\theta)| &= |E_{\hat{F}_n} [w(x; \hat{\theta}_w, \hat{F}_n) - 1] u(x; \theta)| \\ &\leq \sup_t |w(t; \hat{\theta}_w, \hat{F}_n) - 1| E_{\hat{F}_n} |u(x; \theta)| \end{aligned}$$

Per $D(\theta)$ si noti che:

$$\begin{aligned} |i_w(\theta)^{-1} - i(\theta)^{-1}| &= \left| \left[\frac{1}{E_{M_T} [w(x; \hat{\theta}_w, M_T)]} - 1 \right] i(\theta)^{-1} \right| \\ &\leq \sup_t |w(t; \hat{\theta}_w, M_T) - 1| \frac{1}{E_{M_T} [w(x; \hat{\theta}_w, M_T)]} |i(\theta)^{-1}| \end{aligned}$$

Siccome $|u_w(\theta_j)|$ e $|u(\theta_j)|$ sono più piccoli o al più uguali a $E_{\hat{F}_n} |u(x; \theta_j)|$ e quest'ultima è asintoticamente limitata allora per $\theta = \theta_T$ si ottiene l'equivalenza asintotica sotto l'ipotesi nulla. Per studiare l'andamento della potenza locale del test della funzione punteggio pesata consideriamo lo sviluppo in serie di Taylor di $u(x; \theta_T + E) = u(x; \theta_T) + E u'(x; \theta_T) + o(\varepsilon)$ da cui si ottiene che $E_{\hat{F}_n} |u(x; \theta_T + E)| \leq E |u(x; \theta_T)| + |\varepsilon| E_{\hat{F}_n} |u'(x; \theta_T)| + |o(\varepsilon)|$ e asintoticamente i termini di destra sono limitati quando $\varepsilon \rightarrow 0$ per $n \rightarrow \infty$. Inoltre,

$$\begin{aligned} |i_w(\theta)^{-1} - i(\theta)^{-1}| &= \left| \left[\frac{1}{E_{\hat{F}_n} [w(x; \hat{\theta}_w, \hat{F}_n)]} - 1 \right] i(\theta)^{-1} \right| \\ &\leq \sup_t |w(t; \hat{\theta}_w, M_T) - 1| \frac{1}{E_{\hat{F}_n} [w(x; \hat{\theta}_w, \hat{F}_n)]} |i(\theta)^{-1}| \end{aligned}$$

allora le stesse proprietà asintotiche valgono anche per il test della funzione punteggio pesata con la matrice di informazione di Fisher pesata nella versione per campioni finiti.

Le considerazioni che sono state fatte implicano che i test basati sulla verosi-

miglianza pesata definiti come in 4, 5 e 6 hanno la stessa potenza asintotica per alternative parametriche locali come nei test basati sulla verosimiglianza con gli stessi valori critici asintotici quando il modello è correttamente specificato. Inoltre sotto le classiche assunzioni di regolarità **b**, **c** e **d** hanno la stessa distribuzione asintotica sotto l'ipotesi nulla H_0 , cioè una distribuzione Chi quadrato con un appropriato numero di gradi di libertà per un valore fissato del parametro di lisciamiento ν .

5. STUDIO DELLA ROBUSTEZZA ATTRAVERSO LA FUNZIONE DEL PUNTO DI ROTTURA

In questa sezione caratterizziamo le proprietà di robustezza delle tre funzioni test proposte. La Funzione d'Influenza dei tre test corrisponde a quella delle versioni classiche. Lo stesso risultato è ottenuto per lo stimatore di verosimiglianza pesata. Inoltre, è ben noto, (Beran, 1981 e Agostinelli, 1998a) che la Funzione d'Influenza non sempre è adatta a valutare le proprietà di robustezza. Infatti, tralasciando alcuni aspetti tecnici e posto che la definizione di Funzione d'Influenza (Hampel *et al.*, 1986) è:

$$IF(c; T, H) = \lim_{\varepsilon \rightarrow 0} \frac{T(H_\varepsilon) - T(H)}{\varepsilon}$$

dove $H_\varepsilon(x) = (1 - \varepsilon)H(x) + \varepsilon \mathbf{1}_c(x)$, Beran (1981) definisce la ε funzione d'influenza come segue:

$$\varepsilon IF(c; T, H) = \frac{T(H_\varepsilon) - T(H)}{\varepsilon}$$

Egli dimostra che per lo stimatore basato sulla distanza di Hellinger la convergenza della εIF alla funzione d'influenza (per $\varepsilon \rightarrow 0$) non è uniforme. Questo fatto implica che la forma della εIF possa risultare molto diversa da quella della funzione d'influenza. In particolare per la distanza di Hellinger si ha una εIF limitata (evidenziando quindi le proprietà di robustezza dello stimatore basato sulla distanza di Hellinger) ma una funzione d'influenza illimitata (evidenziando l'asintotica efficienza dello stimatore). Agostinelli (1998a) mette in luce che le proprietà di robustezza locali dello stimatore di verosimiglianza pesata dipendono sia dalla distribuzione da cui provengono i dati \mathbf{H} che dal modello parametrico specificato M . Si consideri la seguente definizione (non classica) di funzione di influenza:

$$IF_w(c; T, M, H) = \lim_{\varepsilon \rightarrow 0} \frac{T(H_\varepsilon, M) - T(H, M)}{\varepsilon}$$

dove H è in un intorno di M . Per lo stimatore di verosimiglianza pesata abbiamo:

$$IF_w(c; T, M, H) = D(M, H)^{-1} N(c, M, H)$$

dove (posto per semplicità di notazione $T = T(H, M)$):

$$\begin{aligned}
N(c, M, H) &= w(c; T, H) u(c; T) \\
&+ \int w'(\delta(x; T, H)) \frac{k(x; c, v)}{m(x; T)} u(x; T) dH(x) \\
&- \int w'(\delta(x; T, H)) (\delta(x; T, \mathbf{II}) + 1) u(x; T) dH(x) \\
&= w(c; T, H) u(c; T) \\
&+ \int [A'(\delta(x; T, H)) - w(\delta(x; T, \mathbf{II}))] \\
&\times \left[\frac{k(x; c, v)}{b^*(x)} - u^*(x; T) \right] u(x; T) dH(x) \\
D(H) &= \int w'(\delta(x; T, H)) (\delta(x; T, H) + 1) u^*(x; T) u(x; T) dH(x) \\
&- \int w(x; T, H) u'(x; T) dH(x) \\
&= \int [A'(\delta(x; T, H)) - w(\delta(x; T, \mathbf{II}))] u^*(x; T) u(x; T) dH(x) \\
&- \int w(\delta(x; T, H)) u'(x; T) dH(x)
\end{aligned}$$

Si noti che $IF_w(c; \mathbf{I}, \mathbf{II}, \mathbf{ii}) = IF(c; T, \mathbf{II})$ e che IF_u dipende da M attraverso il modello liscio m , i residui di Pearson \mathfrak{G} e la funzione punteggio u . Quando non è presente alcuna contaminazione nei dati, H coincide con M e la IF_u si riduce alla funzione di influenza dello stimatore di massima verosimiglianza mentre in presenza di contaminazione la IF_u può risultare assai differente (si veda gli esempi in Agostinelli, 1998a, pp 11-17).

Per questi motivi non si ritiene opportuno utilizzare la funzione d'influenza per studiare le proprietà di robustezza dei test ma piuttosto di concentrarsi nello studio del punto di rottura.

Il punto di rottura caratterizza la stabilità globale di uno stimatore. Ne esiste sia una versione asintotica (Hampel, 1971; Huber, 1981) e una per campioni finiti (Hodges, 1967; Donoho, 1982; Donoho e Huber, 1983). Cori iina descrizione sommaria si può dire che il punto di rottura di una funzione statistica è la più piccola frazione di contaminazione dei dati che può portare il valore della stima ad un arbitrario valore estremo. In accordo con questa descrizione si può dire che una statistica test raggiunge il punto di rottura se per un dato livello di contaminazione il suo valore p (p value) può essere portato al suo massimo e minimo valore. Si noti che quando il valore p raggiunge il suo massimo (rispettivamente minimo) la probabilità di commettere l'errore di primo tipo (rispettivamente l'errore di secondo tipo) risulta uguale a 1. Per semplificare la presentazione l'analisi è centrata sul funzionale associato alla funzione test. Quindi $u_w(\theta, F) = E_F[w(x; \hat{\theta}_w(F), F) u(x; \theta)]$ e $u_u(\theta, F) = E_F[w(x; \hat{\theta}_u(F), F)] i(\theta)$ allora il funzionale del test del rapporto di verosimiglianza pesata è $\lambda_{u_w}(\theta, F) = -2E_F[w(x; \hat{\theta}_w(F), F) (l(x; \theta) - l(x; \hat{\theta}_w(F)))]$, il funzionale del test di Wald pesato è $W_{u_w}(\theta, F) = (t) - \hat{\theta}_w(F)' i_{u_w}(\theta, F) (\theta - \hat{\theta}_w(F))$ e il funzionale del test della funzione punteggio pesata è $S_{u_u}(\theta, F) = u_u(\theta, F)' i_{u_u}(\theta, F) (\theta - \hat{\theta}_u(F))$.

In Markatou *et al.* (1998) è dimostrato che lo stimatore di verosimiglianza pesata $\hat{\theta}_w(F)$ ha punto di rottura positivo, diciamo E' . Questo significa (Hampel *et al.*, 1986) che deve esistere un insieme compatto, contenente il vero valore θ_T , che con probabilità uno contenga anche $\hat{\theta}_w(F)$. In altri termini, la distorsione dello stimatore $\hat{\theta}_w(F)$ è finita.

He *et al.* (1990) e Simpson (1989) hanno introdotto la funzione di rottura del livello e della potenza di un test. Introduciamo delle definizioni simili per studiare il comportamento delle funzioni test di verosimiglianza pesata. Sia $t(\theta, F)$ un funzionale generico con dominio F tutte le distribuzioni proprie definite su \mathcal{R}^d . Posto $t_{\max} = \sup_{F \in \mathbf{F}} \inf_{\theta \in \Theta_0} t(\theta, F)$ e $t_{\min} = \inf_{F \in \mathbf{F}} \inf_{\theta \in \Theta_0} t(\theta, F)$ allora la funzione di rottura del livello del test può essere definita nel seguente modo:

$$\varepsilon_0(M_{\theta_T}; t) = \inf \left\{ E : \sup_{G \in \mathbf{F}} \inf_{\theta \in \Theta_0} t(\theta, (1 - E)M_{\theta_T} + \varepsilon G) = t_{\min}; \theta_T \in O, \right\}$$

e la funzione di rottura della potenza del test può essere definita come segue:

$$\varepsilon_1(M_{\theta_T}; t) = \inf \left\{ E : \sup_{G \in \mathbf{F}} \inf_{\theta \in \Theta_0} t(\theta, (1 - E)M_{\theta_T} + \varepsilon G) = t_{\max}; \theta_T \in \Theta_1 \right\}$$

dove G è la distribuzione contaminata e F è assunto convesso. Empiricamente, E , è la più piccola frazione di contaminazione che può portare il valore p a s ed è di maggior interesse sotto l'ipotesi H , ($s=0, 1$). In generale, ε_0 è il livello di contaminazione per il quale il livello di significatività del test è 1 e E , è il livello di contaminazione per il quale la potenza del test è 0. Infatti, se $\theta_T \in \Theta / \Theta_0$ e $t_{\min} = 0$, allora $\varepsilon_1(M_{\theta_T}; t)$ è la più piccola frazione di contaminazione di M_{θ_T} che può portare il test ad essere inconsistente. In analogia con il punto di rottura di uno stimatore possiamo definire il punto di rottura della potenza del test (*powerbreakdown point*) $\varepsilon_1^*(t) = \sup_{\theta_T \in \Theta_1} \varepsilon_1(M_{\theta_T}, t)$ e il punto di rottura del livello del test (*levelbreakdown point*) $\varepsilon_0^*(t) = \sup_{\theta_T \in \Theta_0} \varepsilon_0(M_{\theta_T}, t)$.

Per studiare il comportamento delle funzioni di rottura dei test supponiamo che siano soddisfatte le seguenti condizioni:

Ipotesi a: Condizioni per lo studio del punto di rottura

a.1) $F = (1 - \varepsilon) M_{\theta_T} + \varepsilon G$

a.2) $0 < i(\theta) < \infty$ per tutti i $\theta \in O$

a.3) $\sup_{G \in \mathbf{F}} \|\theta_T - \hat{\theta}_w(F)\| \leq B_{\varepsilon^*} < \infty$ F-q.c. per tutti gli $E \leq E^*$, dove E^* è il punto di rottura di $\hat{\theta}_w(F)$ e B_{ε^*} è il corrispondente errore massimo.

Siccome la matrice di informazione di Fisher è definita positiva allora $W_{\min} = 0$ perché $W_w(\hat{\theta}_w(M_{\theta_T}), M_{\theta_T}) = 0$ e $\theta_w(M_{\theta_T}) = \theta_T$ per la consistenza di Fisher, mentre $W_{\max} = +\infty$.

Dalla disequaglianza triangolare e a.3, $\|\theta - \theta_T\| < \infty \Leftrightarrow \|0 - \hat{\theta}_w(F)\| < \infty$. Così W_{\max} è raggiunto se e soltanto se anche il test calcolato al vero valore raggiunge tale valore massimo per ogni $E \leq E^*$. Questo mostra che $\varepsilon_0(M_{\theta_T}, W_w) = E^*$.

Inoltre per ogni $\epsilon > 0$ e ogni $\theta \in \Theta_0$ tale che $\|\theta - \theta_T\| < B_\epsilon$, è possibile trovare una distribuzione G in modo che $\hat{\theta}_w(F) = \mathbf{0}$ e quindi per la corrispondente \mathbf{F} , $W_w(\theta, F) = 0 = W_{\min}$ e quindi $\epsilon_1(M_{\theta_T}, W_w) = 0$ per ogni θ_T abbastanza vicino a Θ_0 . Per ogni elemento dell'insieme $\Theta_T = \{\theta_T \mid \|\theta - \theta_T\| > B_\epsilon, \forall \theta \in \Theta_0\}$ si ha invece $\epsilon_1(M_{\theta_T}, W_w) = \mathbf{E}$. Quindi $\epsilon_0(W_w) = \mathbf{E}$ e se $\Theta_T \neq \emptyset$ allora $\mathbf{E}_1(W_w) = \mathbf{P}$, quindi la versione robusta del test di Wald ha lo stesso punto di rottura dello stimatore di massima verosimiglianza pesata.

Per studiare la funzione di rottura del test rapporto di verosimiglianza pesata e del test della funzione punteggio pesata, prima riscriviamo i test in una diversa forma. Usando lo sviluppo in serie di Taylor di $l(x; \theta) = \log m(x; \theta)$ in un intorno di $\hat{\theta}$, si ottiene:

$$\log m(x; \theta) = \log m(x; \hat{\theta}_w) + u(x; \hat{\theta}_w)(\theta - \hat{\theta}_w) + \frac{1}{2} \frac{\partial}{\partial \theta} u(x; \theta) \Big|_{\theta=\theta^*} (\theta - \hat{\theta}_w)^2$$

dove θ^* è tra θ e $\hat{\theta}_w$ allora:

$$\begin{aligned} \lambda_w(\theta, \mathbf{F}) &= -2\mathbf{E}_F \left[w(x; \hat{\theta}_w(F), \mathbf{E}) (u(x; \theta), (F)) (\mathbf{8} - \hat{\theta}_w(F)) \right. \\ &\quad \left. + \frac{1}{2} \frac{\partial}{\partial \theta} u(x; \theta) \Big|_{\theta=\theta^*} (\theta - \hat{\theta}_w)^2 \right] \\ &= -\mathbf{E}_F \left[w(x; \hat{\theta}_w(F), F) \frac{\partial}{\partial \theta} u(x; \theta) \Big|_{\theta=\theta^*} \right] (\theta - \hat{\theta}_w(F))^2 \end{aligned}$$

Empiricamente, la quantità $-\mathbf{E}_F [w(x; t), (F), F) u'(x; \theta) \Big|_{\theta=\theta}]$ potrebbe non essere strettamente positiva o limitata, in special modo quando $\mathbf{0}$ è lontano da $\hat{\theta}_w(F)$, perciò non è facile determinare i valori di L_{\dots} , e i_{\dots} , che dipendono dal tipo di contaminazione e dal valore di θ .

Per il fatto:

$$u(x; \theta) = u(x; \hat{\theta}_w) + \frac{\partial}{\partial \theta} u(x; \theta) \Big|_{\theta=\theta^{**}} (\theta - \hat{\theta}_w)$$

dove θ^{**} è tra θ e $\hat{\theta}_w$, si ottiene:

$$\begin{aligned} S_w(\theta, F) &= \mathbf{E}_F \left[w(x; \hat{\theta}_w(F), F) \frac{\partial}{\partial \theta} u(x; \theta) \Big|_{\theta=\theta^{**}} \right]^2 \\ &\quad \times \mathbf{E}_F [w(x; \hat{\theta}_w(F), F)]^{-1} i(\theta)^{-1} (\theta - \hat{\theta}_w(F))^2 \end{aligned} \quad (13)$$

Allora $S_{\min} = 0$ e $S_{\max} = +\infty$ ma diversamente dal test di Wald pesato W_w , il test della funzione punteggio pesata S_w può raggiungere il valore $+\infty$ o 0 attraverso il primo fattore della 13 e quindi $\epsilon_0(M_{\theta_T}, S_w) \leq \mathbf{E}$ e per la stessa ragione $\epsilon_1(M_{\theta_T}, S_w) \leq \mathbf{E}$.

Quando il modello specificato appartiene alla famiglia esponenziale uniparametrica con la parametrizzazione naturale cioè $m(x, \theta) = \exp[\theta s(x) - \tau(\theta)] h(x)$ e $u'(x; \theta) = -\tau''(\theta) = -i(\theta)$ allora

$$\begin{aligned}\lambda_w(\theta, F) &= E_F[w(x; \hat{\theta}_w(F), F)] i(\theta^*)(\theta - \hat{\theta}_w(F))^2 \\ &= W_w(\theta, F) \frac{i(\theta^*)}{i(\hat{\theta}_w(F))}\end{aligned}$$

$$\begin{aligned}S_w(\theta, F) &= E_F[w(x; \hat{\theta}_w(F), F)] \frac{i(\theta^{**})^2}{i(\theta)} (\theta - \hat{\theta}_w(F))^2 \\ &= W_w(\theta, F) \frac{i(\theta^{**})^2}{i(\theta) i(\hat{\theta}_w(F))}\end{aligned}$$

In questo caso, $\lambda_{\min} = 0$ e $A_{\max} = +\infty$ e le proprietà di $\lambda_w(\theta, F)$ e $S_w(\theta, F)$ sono le medesime del test $W_w(\mathbf{t}, F)$.

Diverso P in generale il comportamento dei test in presenza di parametri di disturbo. Per il test di Wald pesato, il punto di rottura della potenza non sarà inferiore al più piccolo punto di rottura degli stimatori coinvolti nel calcolo del test stesso, dato che un valore arbitrario dei parametri di disturbo può portare il valore atteso dei pesi a zero, mentre il punto di rottura del livello del test dipende soltanto dal punto di rottura degli stimatori del vettore dei parametri di interesse.

Questo comportamento non è condiviso, in generale, dal test del rapporto di verosimiglianza pesato e dal test della funzione punteggio pesata e quindi anche il punto di rottura del livello di questi test dipenderà dal punto di rottura degli stimatori dei parametri di disturbo coinvolti nel calcolo. Si noti che se il modello appartiene alla famiglia esponenziale con parametrizzazione naturale allora il punto di rottura del livello del test rapporto di verosimiglianza pesato e del test funzione punteggio pesata sarà lo stesso del test di Wald pesato.

6. ESEMPI E SIMULAZIONI

In questa sezione presentiamo alcuni esempi, teorici o riferiti a dati reali che illustrano le proprietà dei test di verosimiglianza pesata e alcuni confronti con altre funzioni test. La sezione si conclude con alcune simulazioni Monte Carlo.

Esempio 6 i. Verifica d'ipotesi sulla media della distribuzione normale (ripreso da Simpson, 1989).

Se $m(x; \theta) = \mathcal{N}(x, 1)$ (quasi certamente) è la densità normale con media 0 e varianza unitaria, allora il funzionale del test rapporto di verosimiglianza per $H_0: \theta_1 = 0$ contro $H_1: \theta_1 \neq 0$ è dato da $2\lambda(F) = \mu(F)^2$, dove $\mu(F)$ è la media della distribuzione F . È possibile mostrare che $\varepsilon_0(M_\theta, A) = \varepsilon_1(M_\theta, A) = \varepsilon_0(A) = \varepsilon_1(\lambda) = 0$. Da Simpson (1989) le proprietà di punto di rottura del test basato sulla distanza di Hellinger d è $\varepsilon_0(M_\theta, d) = \varepsilon_1(M_\theta, d) = 1$, mentre ad esempio $\varepsilon_1(M_\lambda, d) \geq .313$. Se si impiega lo stimatore ad un passo della verosimiglianza pesata con stime iniziali date dallo stimatore *Least Trimmed Squared* (con $\alpha = 0.5$) il test di Wald pesato possiede $\varepsilon_0^*(W_w) = \varepsilon_1(W_w) = 0.5$ e lo stesso risultato vale per gli altri due test.

Esempio 6.2. Verifica d'ipotesi sulla media della distribuzione normale con parametro di scala ignoto (ripreso da Simpson, 1989).

Sia $m(x, \theta)$ con $\theta = (p, \sigma)$ denoti la distribuzione normale $\mathcal{N}(\mu, \sigma^2)$. Il sistema d'ipotesi è $H_0 : p = 0$ contro $H_1 : p \neq 0$ con σ^2 non specificata. Il funzionale del test rapporto di verosimiglianza è pari a $2\lambda(F) = \log[1 + \mu(F)^2/v(F)]$ dove $v(F)$ è la varianza della distribuzione F . Seguendo Simpson (1989) $\varepsilon_0(M_\theta, \lambda) = 1$ mentre $\varepsilon_1(M_\theta, h) = 0$. Per il test basato sulla distanza di Hellinger si ha $\varepsilon_0(M_\theta, d) = 1$ e ad esempio, $\varepsilon_1(M_{(3,1)}, d) \geq .127$. Se si utilizza lo stimatore ad un passo della verosimiglianza pesata con stime iniziali date dallo stimatore Least Trimmed Squared (con $E^* = 0.5$) allora per il test di Wald pesato $\varepsilon_0^*(W_w) = \varepsilon_1^*(W_w) = 0.5$ e lo stesso risultato vale per gli altri due test.

Esempio 6.3. Regione di accettazione sotto contaminazione.

Supponiamo che la densità da cui le osservazioni sono campionate è una mistura tra due variabili normali del tipo $f(\mathbf{x}) = (1 - E) \mathcal{N}(0, 1) + E \mathcal{N}(8, 1)$ per un fissato valore E . Consideriamo una famiglia parametrica normale nella forma $M = \{\mathcal{N}(\mu, 1) : -\infty < \mu < \infty\}$ in cui la varianza è nota. In questo contesto i tre test di verosimiglianza coincidono (non solo asintoticamente), così pure, come abbiamo visto, i tre test basati sulla verosimiglianza pesata.

Nella figura 1 riportiamo il valore asintotico di tre funzioni test: il test del rapporto di verosimiglianza quando non vi è contaminazione, il test rapporto di verosimiglianza e il test del rapporto di verosimiglianza pesata considerando che per semplicità di calcolo il parametro di liscio v del nucleo sia tale che $v \rightarrow 0$. Il tratto continuo è riferito al test di verosimiglianza quando il livello di contaminazione E è nullo, esso è riportato per facilitare il confronto. Il tratto di punto e linea si riferisce al test basato sulla verosimiglianza pesata, mentre il tratto di linea è invece il test di verosimiglianza calcolato sotto la contaminazione. La linea retta rappresenta il valore di un chi-quadrato con un grado di libertà per un errore di primo tipo del 5%. A partire dall'angolo in alto a sinistra, le singole figure si riferiscono ai seguenti livelli di contaminazione: 5%, 10%, 20%, 30%, 40%, 50%. Come si vede facilmente, se l'ipotesi nulla è $H_0 : p = 0$ il test di verosimiglianza porta al rifiuto dell'ipotesi nulla per un livello di contaminazione superiore al 20%, per il fissato valore dell'errore di primo tipo al 5% (riportato come linea orizzontale).

Esempio 6.4. Telephone-Lines Faults data

Welch (1987) presenta dei dati derivanti da un esperimento in cui si vuole verificare se un metodo riduce il numero di guasti di una linea telefonica. Un approccio parametrico semplice è di utilizzare un modello con distribuzione normale con media μ e varianza σ^2 . Il test rapporto di verosimiglianza del sistema $H_0 : \mu = 0$ contro $H_1 : p > 0$ (non specificata) è equivalente al test t ad una coda.

Welch (1987) trova che questo test è altamente sensibile alla presenza o all'assenza della prima coppia di osservazioni. Simpson (1989) ha utilizzato questi dati per illustrare il test basato sulla distanza di Hellinger. Nella tabella 1 riportiamo i tre test basati sulla verosimiglianza pesata insieme con il test basato sulla verosimiglianza e il test basato sulla distanza di Hellinger presentato da Simpson (1989). Per i test basati sulla verosimiglianza pesata è stato utilizzato un nucleo corrispon-

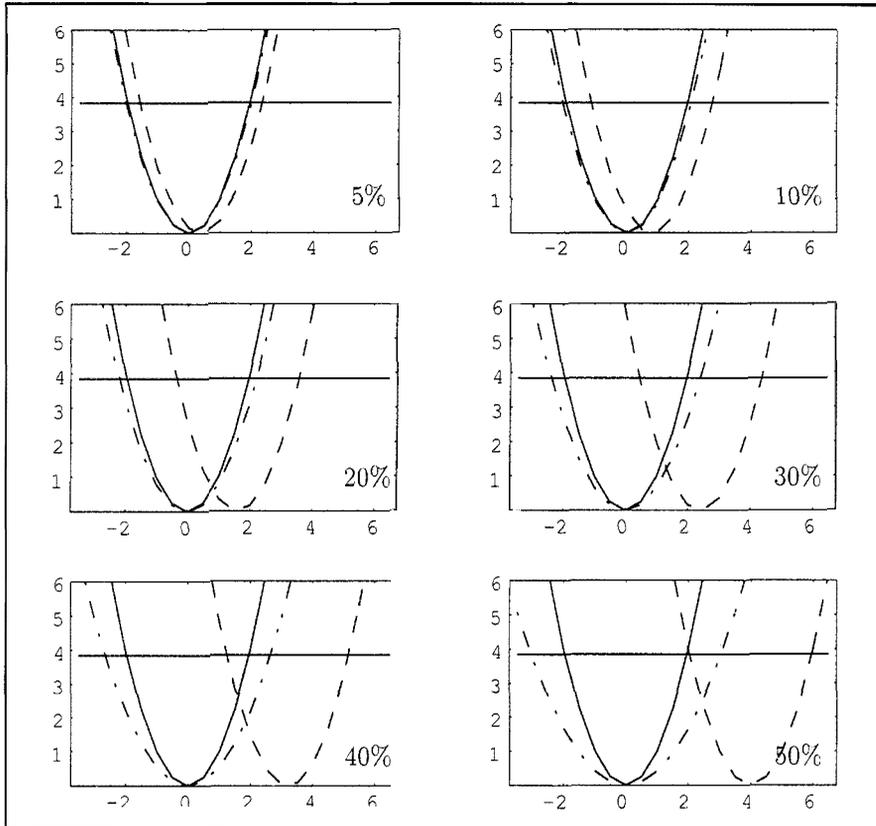


Figura 1 - Valore asintotico della funzione test di verosimiglianza (tratto a linea) e per il test di verosimiglianza pesata (tratto punto e linea) per un modello di posizione normale e diversi valori di contaminazione per $f(x) = (1 - \varepsilon) \mathcal{N}(0,1) + \varepsilon \mathcal{N}(8,1)$. La parabola a linea continua indica il valore asintotico del test di verosimiglianza per $f(x) = \mathcal{N}(0,1)$ ed è riportata come riferimento. La linea parallela all'asse delle ascisse corrisponde al valore del χ_1^2 con un errore di primo tipo del 5%.

dente alla densità normale con parametro di liscio v proporzionale al parametro di scala con $k = 0.0034$. I tre test di verosimiglianza pesata sono stati calcolati anche utilizzando un nucleo *biweight* nella forma $k(x) = (15/16) (1 - x^2)^2 \mathbf{1}_{[-1,1]}(x)$ e $v = 2.3 \sigma n^{-1/5} = 202.16$ dove $MAD/.6745$ è usato come stimatore di scala (Nota che il parametro di liscio è scelto con il criterio L, di Devroye and Györfi, 1985). Il nucleo *biweight* e lo stesso parametro di liscio sono stati utilizzati da Simpson (1989) per calcolare il test basato sulla distanza di Hellinger. In tabella 2 sono invece presentate le stesse funzioni test calcolate sui dati privi della prima osservazione (in questo caso Simpson, 1989, non riporta il valore delle stime dei parametri dello stimatore di Hellinger).

I tre test presentati si comportano bene, in particolare in entrambi i casi essi si comportano in maniera simile ai corrispondenti test di verosimiglianza calcolati sui dati privi della prima osservazione.

TABELLA 1

Telephone-Lines Faults data. Per ogni funzione test riportiamo il valore (prima colonna) e il corrispondente valore-p percentuale (seconda colonna)

	WL		ML		HD			
	normal	biweight						
σ_0	151.817	172.432		312.665		169.500		
$\hat{\mu}$	136.736	118.188		38.929		116.800		
$\hat{\sigma}$	102.170	127.723		310.232		144.600		
λ	3.102	0.421	2.827	0.711	0.468	32.388	2.740	0.850
W	3.886	0.094	3.328	0.272	0.469	32.324		
S	2.697	0.914	2.465	1.421	0.466	32.451		

TABELLA 2

Telephone-Lines Faults data. Per ogni funzione test riportiamo il valore (prima colonna) e il corrispondente valore-p percentuale (seconda colonna). La prima coppia di osservazioni è stata tolta

	WL		ML		IIL			
	normal	biweight						
σ_0	151.661	170.342		173.756		NA		
$\hat{\mu}$	136.865	117.892		117.923		NA		
$\hat{\sigma}$	101.958	127.534		127.614		NA		
λ	3.093	0.465	2.829	0.760	2.833	0.755	2.72	0.930
W	3.885	0.108	3.326	0.302	3.332	0.299		
S	2.692	0.980	2.490	1.422	2.447	1.538		

Alcune simulazioni Monte Carlo sono state svolte allo scopo di valutare il livello delle funzioni di punto di rottura in contesti abbastanza importanti. Sono state considerate due distribuzioni dei dati coinvolgenti il modello normale. La prima ha una contaminazione di tipo asimmetrico, attraverso la presenza di un nucleo di osservazioni distanziate dalle rimanenti $f_1(x) = (1 - \varepsilon) \mathcal{N}(0,1) + \varepsilon \mathcal{N}(8,1)$. La seconda presenta invece una contaminazione simmetrica, in cui la distribuzione ipotizzata ha code più corte, esemplificata dalla densità $f_2(x) = (1 - \varepsilon) \mathcal{N}(0,1) + \varepsilon \mathcal{N}(0,25)$. Da ognuna di queste due distribuzioni sono stati generati 5000 campioni di dimensione 80 per i seguenti livelli di contaminazione. $\varepsilon = \{0\%, 5\%, 10\%, 20\%, 30\%, 40\%\}$. Sono state considerate due funzioni di aggiustamento dei residui, quella basata sulla distanza di Hellinger ($A_{HD}(\delta) = 2[\sqrt{\delta} + 1 - 1]$) e la *Negative Exponential Disparity* ($A_{NED}(\delta) = 2 - (2 + \delta) \exp(-\delta)$). Infine per individuare le radici dell'equazione di stima della verosimiglianza pesata, si è utilizzato l'algoritmo di ricerca basato sul ricampionamento presentato in Markatou *et al.* (1998) con 100 ricampionamenti per ogni campione Monte Carlo. In presenza di più radici si è scelta la stima di verosimiglianza pesata attraverso il seguente criterio. Posto $\{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p\}$ le radici dell'equazione di verosimiglianza pesata, consideriamo la distanza $d(f^*, m)$ corrispondente alla RAF (*parallel disparity measure*) utilizzata e definiamo la stima come: $\hat{\theta}_w = \arg \min_{\hat{\theta}_i, i=1, \dots, p} d(f^*, m(\hat{\theta}_i))$. Nel caso della RAF di

Hellinger abbiamo, ad esempio, $d(f, m^*(\hat{\theta}_j)) = \int [f(x)^{1/2} - m(x, \hat{\theta}_j)^{1/2}]^2 dx$. Questo metodo utilizzato in Markatou *et al.* (1998) per la scelta della radice porta a stime con alto punto di rottura.

Utilizzando questo metodo di selezione si ottengono dei buoni risultati anche per contaminazioni piuttosto elevate. Nelle nostre simulazioni si sono individuate più radici per la contaminazione asimmetrica f_1 per livelli di contaminazioni superiori o uguali al 30%.

Abbiamo considerato la verifica d'ipotesi per il parametro di posizione μ con $H_0: \mu = 0$. In ogni tabella riportiamo nella prima parte i valori delle funzioni test quando il parametro di disturbo (\mathbf{O}) è considerato noto o ignoto. Si noti che nel caso di \mathbf{O} noto le tre funzioni test coincidono (sia per la verosimiglianza pesata che per la verosimiglianza). Oltre al valore del test riportiamo la stima della varianza Monte Carlo e i livelli stimati del test confrontati con quelli di un Chi-quadrato con un grado di libertà.

Nella seconda parte della tabella riportiamo le stesse quantità quando l'ipotesi è $H_1: \mu = 1$. I livelli stimati in questo caso sono una stima della potenza del test per una fissata ipotesi alternativa e un fissato valore nominale dell'errore di primo tipo. Abbiamo contrassegnato i test basati sulla funzione di verosimiglianza pesata con la sigla *WLT*, quelli di verosimiglianza con *MLT* e con *HRT* il test di Wald proposto in Heritier e Ronchetti (1994).

Nelle tabelle presentate riguardanti la densità f_1 si evidenzia un'ottimo comportamento dei test di verosimiglianza pesata. Risultati molto simili si ottengono per i test basati sulla funzione di aggiustamento dei residui *Negative Exponential Disparity*. In quest'ultimo caso già a un livello di contaminazione del 30% si comincia a notare un moderato degrado della qualità. Il comportamento di *HRT* è discreto solo per contaminazioni di piccola entità confermando quanto scritto da Heritier e Ronchetti (1994, p. 901): "*The bounded influence tests presented in this article are constructed to withstand small amounts of contamination. . . one should use these tests carefully when the expected amount of contamination exceeds 8-9%. A similar analysis for the Wald-type test shows that this test breaks down earlier.*".

La contaminazione di tipo f_2 è adatta a valutare l'impatto sul valore della funzione test di una stima non robusta dei parametri di disturbo e della sovrastima della varianza della stima del parametro di posizione dovuta alla presenza di code pesanti nel caso il parametro di scala sia considerato noto. In particolare si noti come nei test di verosimiglianza, nel caso del parametro di scala ignoto, al discreto andamento del livello si contrappone una diminuzione della potenza, viceversa nel caso del parametro di scala noto, accade il contrario, questo fatto è in accordo con quanto osservato in Tsou e Royall (1995). Per i test di verosimiglianza pesata lo stesso comportamento comincia a manifestarsi solo per elevati valori di contaminazione.

TABELLA 3

Confronto dei valori, del livello e della potenza tra i test di verosimiglianza pesata e i test di verosimiglianza per la densità f_1 e RAI: di Hellinger. 5000 replicazioni Monte Carlo

$\varepsilon = 0\%$

$H_0: \mu = 0$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (oignoto)	1.142	2.849	12.080	6.500	3.920	1.860
WLTest (onoto)	0.977	1.966	9.060	4.740	2.520	1.120
MLTest (oignoto)	1.003	2.187	10.100	5.120	2.740	1.320
MLTest (onoto)	0.967	1.943	9.480	4.740	2.480	1.020
HRTTest (σ ignoto)	1.331	3.893	13.571	6.933	4.149	2.605

$H_1: \mu = 1$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (oignoto)	88.383	645.572	100.000	100.000	100.000	100.000
WLTest (onoto)	75.899	293.924	100.000	100.000	100.000	100.000
MLTest (oignoto)	83.714	520.494	100.000	100.000	100.000	100.000
MLTest (onoto)	80.711	309.710	100.000	100.000	100.000	100.000
HRTTest (oignoto)	72.395	441.352	100.000	100.000	100.000	100.000

$\varepsilon = 5\%$

$H_0: \mu = 0$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (oignoto)	1.118	2.881	12.100	6.460	4.060	1.810
WLTest (onoto)	0.982	2.005	9.540	5.040	2.580	1.040
MLTest (σ ignoto)	3.487	3.481	62.200	36.460	18.900	5.780
MLTest (σ noto)	13.869	52.002	97.420	94.420	91.280	85.160
HRTTest (oignoto)	1.889	6.340	22.024	13.374	7.497	4.360

$H_1: \mu = 1$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (oignoto)	83.926	619.864	100.000	100.000	100.000	100.000
WLTest (onoto)	72.186	281.248	100.000	100.000	100.000	100.000
MLTest (oignoto)	7.440	7.729	98.160	92.500	80.620	57.420
MLTest (onoto)	29.613	112.390	100.000	99.980	99.820	99.620
HRTTest (oignoto)	53.959	304.842	100.000	100.000	100.000	99.923

$\varepsilon = 10\%$

$H_0: \mu = 0$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (oignoto)	1.141	2.835	12.020	6.900	3.960	1.780
WLTest (onoto)	0.981	1.999	9.620	5.080	2.360	1.080
MLTest (oignoto)	7.818	4.958	99.880	97.860	90.720	68.800
MLTest (onoto)	52.371	201.011	100.000	100.000	100.000	100.000
HRTTest (σ ignoto)	3.792	14.096	48.352	36.209	26.364	16.331

$H_1: \mu = 1$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (oignoto)	79.508	586.821	100.000	100.000	100.000	100.000
WLTest (onoto)	68.508	266.764	100.000	100.000	100.000	100.000
MLTest (σ ignoto)	0.614	0.314	0.540	0.000	0.000	0.000
MLTest (onoto)	4.115	14.021	55.880	42.620	31.040	20.060
HRTTest (oignoto)	36.495	187.288	99.953	99.950	99.859	99.802

TABELLA 4

Confronto dei valori, del livello e della potenza tra i test di verosimiglianza pesata e i test di verosimiglianza per h densità f_j e RAF di Hellinger. 5000 replicazioni Monte Carlo

$\epsilon = 20\%$

$H_0 : \mu = 0$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (σ ignoro)	1.273	4.150	13.540	8.060	5.080	3.100
WLTest (σ noto)	2.319	137.707	11.200	5.940	3.820	2.220
MLTest (σ ignoto)	18.454	7.868	100.000	100.000	100.000	100.000
MLTest (σ noto)	206.176	792.884	100.000	100.000	100.000	100.000
HRTTest (σ ignoto)	12.011	35.698	96.915	93.726	88.713	80.652

$H_1 : \mu = 1$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (σ ignoto)	69.370	581.370	98.620	98.620	98.620	98.620
WLTest (σ noto)	60.050	278.206	99.280	99.140	99.040	98.920
MLTest (σ ignoto)	2.678	0.946	14.260	12.200	1.800	0.140
MLTest (σ noto)	29.920	114.137	100.000	100.000	100.000	99.820
HRTTest (σ ignoto)	9.086	37.212	88.433	79.916	70.873	59.481

$\epsilon = 30\%$

$H_0 : \mu = 0$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (σ ignoto)	1.136	2.843	12.380	6.918	3.936	1.650
WLTest (σ noto)	0.972	1.929	9.986	4.457	2.200	0.955
MLTest (σ ignoto)	32.162	12.057	100.000	100.000	100.000	100.000
MLTest (σ noto)	462.380	1786.562	100.000	100.000	100.000	100.000
HRTTest (σ ignoto)	27.144	63.465	100.000	100.000	100.000	100.000

$H_1 : \mu = 1$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (σ ignoto)	61.644	432.567	100.000	100.000	100.000	100.000
WLTest (σ noto)	3.170	204.654	100.000	100.000	100.000	100.000
MLTest (σ ignoto)	10.998	3.346	100.000	100.000	100.000	99.620
MLTest (σ noto)	158.124	610.059	100.000	100.000	100.000	100.000
HRTTest (σ ignoto)	1.462	3.494	17.322	10.414	5.278	2.515

$\epsilon = 40\%$

$H_0 : \mu = 0$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (σ ignoto)	1.151	2.954	12.309	6.863	4.031	1.852
WLTest (σ noto)	0.984	1.874	9.804	5.773	1.743	0.871
MLTest (σ ignoto)	50.395	19.247	100.000	100.000	100.000	100.000
MLTest (σ noto)	820.985	3173.046	100.000	100.000	100.000	100.000
HRTTest (σ ignoto)	138.488	6664.038	100.000	100.000	100.000	100.000

$H_1 : \mu = 1$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (σ ignoto)	51.329	388.811	100.000	100.000	100.000	100.000
WLTest (σ noto)	-14.955	173.799	100.000	100.000	100.000	100.000
MLTest (σ ignoto)	23.862	7.282	100.000	100.000	100.000	100.000
MLTest (σ noto)	388.729	1501.786	100.000	100.000	100.000	100.000
HRTTest (σ ignoto)	64.656	2249.809	99.935	99.920	99.814	99.303

TAH I I A 5

Confronto dei valori, del livello e della potenza tra i test di verosimiglianza pesata e i test di verosimiglianza per la densità f_2 e RAF di Hellinger. 5000 replicazioni Monte Carlo.

$\varepsilon = 5\%$

$H_0 : \mu = 0$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTtest (o ignoto)	1.139	2.820	14.800	6.600	3.700	1.860
WLTtest (σ noto)	1.037	2.219	10.000	5.280	3.000	1.320
MLTest (o ignoto)	1.020	1.871	9.560	4.920	2.200	0.880
MLTest (σ noto)	2.136	8.544	26.140	17.820	12.380	7.700
HRTtest (σ ignoro)	1.355	3.940	15.440	8.480	5.360	2.880

$H_0 : \mu = 1$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTtest (σ ignoto)	81.145	608.839	100.000	100.000	100.000	100.000
WLTtest (o noto)	74.079	305.463	100.000	100.000	100.000	100.000
MLTest (σ ignoto)	43.449	427.067	100.000	100.000	99.940	99.700
MLTest (σ noto)	82.044	692.047	100.000	100.000	100.000	100.000
HRTtest (o ignoto)	68.208	413.854	100.000	100.000	100.000	100.000

$\varepsilon = 10\%$

$H_0 : \mu = 0$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTtest (o ignoto)	1.137	2.752	11.980	6.940	3.760	1.800
WLTtest (σ noto)	1.120	2.592	11.540	6.500	3.560	1.560
MLTest (o ignoto)	1.024	1.872	10.500	4.600	2.080	0.680
MLTest (o noto)	3.328	20.834	36.260	28.880	22.520	16.020
HRTtest (σ ignoto)	1.440	4.309	16.960	9.720	5.560	3.240

$H_0 : \mu = 1$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTtest (o ignoto)	73.917	563.022	100.000	100.000	100.000	100.000
WLTtest (o noto)	72.483	324.144	100.000	100.000	100.000	100.000
MLTest (o ignoto)	28.100	220.191	99.720	99.520	98.940	97.800
MLTest (o noto)	83.038	1079.664	100.000	100.000	100.000	99.980
HRTtest (o ignoto)	63.684	407.951	100.000	100.000	100.000	100.000

$\varepsilon = 20\%$

$H_0 : \mu = 0$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTtest (o ignoto)	1.154	2.712	12.480	6.760	3.640	1.820
WLTtest (o noto)	1.394	3.913	16.060	9.500	5.740	2.880
MLTest (o ignoto)	1.021	1.972	10.460	4.980	2.260	0.860
MLTest (σ noto)	5.700	63.670	49.400	41.340	34.600	28.460
HRTtest (σ ignoto)	1.612	5.196	19.980	12.200	7.240	3.960

$H_0 : \mu = 1$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTtest (o ignoto)	58.548	477.793	100.000	100.000	99.980	99.980
WLTtest (σ noto)	69.587	390.708	100.000	100.000	100.000	100.000
MLTest (o ignoto)	16.246	90.380	97.520	95.280	92.120	86.660
MLTest (o noto)	85.554	1874.988	99.920	99.780	99.720	99.620
HRTtest (σ ignoto)	54.141	379.548	99.960	99.960	99.960	99.960

TABELLA 6

Confronto dei valori, del livello e della potenza tra i test di verosimiglianza pesata e i test di verosimiglianza per la densità f_2 e RAF di Hellinger. 5000 replicazioni Monte Carlo

$\varepsilon = 30\%$

$H_0: \mu = 0$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (σ ignoto)	1.145	2.679	12.302	6.501	3.701	1.660
WLTest (σ noto)	2.019	8.704	24.145	16.543	11.302	6.861
MLTest (σ ignoto)	1.027	2.070	10.342	5.241	2.861	1.000
MLTest (σ noto)	8.086	130.057	57.271	49.830	43.849	36.767
HRTest (σ ignoto)	1.845	6.847	22.640	15.040	10.120	5.560

$H_0: \mu = 1$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (σ ignoto)	41.729	384.142	100.000	99.920	99.880	99.640
WLTest (σ noto)	68.279	559.619	100.000	100.000	100.000	100.000
MLTest (σ ignoto)	11.528	54.911	92.899	87.217	80.316	71.254
MLTest (σ noto)	87.701	2719.452	99.440	99.220	99.040	98.800
HRTest (σ ignoto)	44.604	351.851	99.920	99.920	99.920	99.920

$\varepsilon = 40\%$

$H_0: \mu = 0$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (σ ignoto)	1.175	2.967	12.703	7.113	3.887	2.084
WLTest (σ noto)	4.076	42.865	38.690	31.356	25.025	19.254
MLTest (σ ignoto)	1.011	2.125	10.078	5.189	2.745	1.042
MLTest (σ noto)	10.276	215.250	60.689	54.077	48.668	41.655
HRTest (σ ignoto)	2.209	10.422	26.974	19.502	13.299	7.989

$H_0: \mu = 1$

	Media	Varianza	10.0%	5.0%	2.5%	1.0%
WLTest (σ ignoto)	24.352	260.892	98.337	97.115	95.292	91.665
WLTest (σ noto)	70.510	1147.882	99.840	99.780	99.760	99.619
MLTest (σ ignoto)	9.038	39.430	86.696	78.622	70.427	58.686
MLTest (σ noto)	90.040	3458.312	98.638	98.357	97.996	97.696
HRTest (σ ignoto)	35.488	328.554	99.624	99.248	98.872	98.308

APPENDICE

Ipotesi b: Problema regolare di stitna

b.1) Lo spazio parametrico Θ è un intervallo aperto

b.2) Il supporto degli elementi della famiglia parametrica $M: \mathcal{A}_\theta = \{x: m(x; \theta) > 0\}$, θ indipendente dal valore del parametro θ , cioè $\mathcal{A}_\theta = \mathcal{A}$ per ogni $\theta \in \Theta$.

b.3) Per ogni $x \in \mathcal{A}$ la densità $m(x; \theta)$ è differenziabile tre volte rispetto a θ e la derivata terza C continua in θ .

b.4) L'integrale $\int dM(x; H)$ è differenziabile due volte sotto il segno di integrale.

b.5) L'informazione attesa di Fisher $i(\theta) = \int u(x; \theta)^2 dM(x; \theta)$ soddisfa alla condizione $0 < i(\theta) < \infty$ per tutti i valori di $\theta \in \Theta$.

b.6) Per ogni $\theta_T \in \mathbf{O}$ fissato esiste un numero positivo $c(\theta_T)$ e una funzione $B(x; \theta_T)$ tale che per ogni $x \in \mathcal{A}$ e $0 \in (\theta_T - c(\theta_T), \theta_T + c(\theta_T))$

$$\left| \frac{\partial^3}{\partial \theta^3} \log m(x; \theta) \right| \leq B(x; \theta) \quad (14)$$

e $\int B(x; \theta_T) dM(x; \theta_T) < \infty$.

Ipotesi c: Regolarità della funzione peso

c.1) $A(0) = 0$, $A'(0) = 1$ e $A''(\delta)$ è una funzione continua e limitata di δ .

c.2) $\frac{d}{d\delta} w(\delta)(\delta + 1)$ è una funzione limitata di δ .

Ipotesi d: Condizioni per la convergenza uniforme dei pesi

d.1) $k(x, t, \mathbf{v})$ è una densità ed è una funzione a variazione limitata e la larghezza della banda h è una costante positiva.

d.2) le osservazioni x_i ($i = 1, \dots, n$) formano un campione di osservazioni indipendenti estratte dalla distribuzione $M(\cdot; \theta_T)$ con densità $m(\cdot; \theta_T)$ (rispetto alla misura di Lebesgue) e $\theta_T \in \mathbf{O}$.

d.3) $\tilde{\theta}$ è uno stimatore consistente di θ_T .

d.4) $\sup \left| \frac{d}{d\theta} M(x; \theta) \right| < \infty$

Lemma 1. Sotto l'insieme di ipotesi d

$$\sup_x \left| \delta(x; \hat{\theta}_w(\hat{F}_n), \hat{F}_n) \right| \xrightarrow{p} 0 \quad (15)$$

Lemma 2. Sotto la validità delle ipotesi d e se $\sup_x \left| \frac{\partial}{\partial \theta} \delta \right| = C_\delta < \infty$ e $\sup_x \left| \frac{\partial^2}{\partial \theta^2} \delta \right| < \infty$ allora

$$\sup_x \left| w(x; \hat{\theta}_w(\hat{F}_n), \hat{F}_n) - 1 \right| \xrightarrow{p} 0 \quad (16)$$

Per le dimostrazioni si veda ad esempio Agostinelli (1998a).

RIFERIMENTI BIBLIOGRAFICI

- C. AGOSTINELLI (1997), *A one-step robust estimator based on the weighted likelihood methodology*, Working paper 1997.16, Dipartimento di Statistica, Università di Padova, Padova.
- C. AGOSTINELLI (1998a), *Inferenza statistica robusta basata sulla funzione di verosimiglianza pesata: alcuni sviluppi*, Tesi di Dottorato, Dipartimento di Statistica, Università di Padova.
- C. AGOSTINELLI (1998b), *Verosimiglianza pesata nel modello di regressione lineare, presentato alla XXXIX Riunione scientifica della Società Italiana di Statistica*, atti del convegno, Sorrento 1798.
- C. AGOSTINELLI, M. MARKATOU (1998), *A one-step robust estimator for regression based on the weighted likelihood reweighting scheme*, "Statistics & Probability Letters", 37, n. 4, pp. 341-350.
- I.V. BASAWA, H.L. KOUL (1988), *Large sample statistics based on quadratic dispersion*, "International Statistical Review", 56, pp. 199-219.
- A. BASU, B.G. LINDSAY (1994), *Minimum disparity estimation for continuous models: efficiency, distributions and robustness*, "Annals of the Institute of Statistical Mathematics", 46, pp. 683-705.
- R. BERAN (1981), *Efficient robust estimation in parametric models*, "Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete", 51, pp. 91-108.
- D.R. COX, D.V. HINKLEY (1974), *Theoretical statistics*, "Chapman & Hall", London.
- N. CRESSIE, T.R. READ (1984), *Multinomial goodness-of-fits tests*, "Journal of Royal Statistical Society", series B, 46, pp. 440-464.
- L. DEVROYE, I. GYORFI (1985), *Nonparametric density estimation: the L₁ view*, John Wiley, New York.
- D.L. DONOHO, P.J. HUBER (1983), *The notion of breakdown point*, in: P.J. BICKEL, K. DOKSUM, J.L. HODGES JR. (eds.), "Festschrift für Erich L. Lehmann", Wadsworth, Belmont, CA, pp. 157-184.
- F.R. HAMPEL (1971), *A general qualitative definition of robustness*, "Annals of Mathematical Statistics", 42, pp. 1887-1896.
- F.R. HAMPEL, E.M. RONCHETTI, P.J. ROUSSEEUW, W.A. STAHEL (1986), *Robust statistics: the approach based on influence functions*, John Wiley, New York.
- X. HE (1991), *A local breakdown property of robust tests in linear regression*, "Journal of Multivariate Analysis", 38, pp. 294-305.
- X. HE, D.G. SIMPSON, S.L. PORTNOY (1990), *Breakdown of robustness of tests*, "Journal of the American Statistical Association", 85, pp. 446-452.
- S. HERITIER, E.M. RONCHETTI (1994), *Robust bounded influence tests in general parametric models*, "Journal of the American Statistical Association", 89, pp. 897-904.
- P.J. HUBER (1965), *A robust version of the probability ratio test*, "The Annals of Mathematical Statistics", 36, pp. 1753-1758.
- P.J. HUBER (1981), *Robust statistics*, John Wiley, New York.
- B.G. LINDSAY (1994), *Efficiency versus robustness: the case for minimum Hellinger distance and related methods*, "Annals of Statistics", 22, pp. 1018-1114.
- M. MARKATOU (1996), *Robust statistical inference: weighted likelihoods or usual M-estimation?*, "Communications in Statistics, Theory and Methods", 25, pp. 2597-2613.
- M. MARKATOU, T.P. HETTMANSPERGER (1990), *Robust bounded-influence tests in linear models*, "Journal of the American Statistical Association", 85, pp. 187-190.
- M. MARKATOU, X. HE (1994), *Bounded influence and high breakdown point testing procedures in linear models*, "Journal of the American Statistical Association", 89, pp. 543-549.

- M. MARKATOU, A. BASU (1995), *Weighted likelihood estimating equations and robust tests of hypotheses*, Technical Report, Department of Statistics, Columbia University, New York.
- LI. MARKATOU, A. BASU, B.G. LINDSAY (1995), *Weighted likelihood estimating equations: the continuous case*, Technical Report, Department of Statistics, Columbia University, New York.
- M. MARKATOU, A. BASU, B.G. LINDSAY (1998), *Weighted likelihood estimating equations with a bootstrap root search*, "Journal of the American Statistical Association", 93, pp. 740-750.
- J. NEYMAN, E.S. PEARSON (1928), *On the use and interpretation of certain test criteria for purposes of statistical inference*, "Biometrika", series A, 20, pp. 175-240.
- B.L.S. PRAKASA RAO (1983), *Nonparametric functional estimation*, Academic Press, New York.
- E.M. RONCHETTI (1982), *Robust testing in linear models: the infinitesimal approach*, Ph.D. Thesis, ETH, Zurich.
- D.G. SIMPSON (1989), *Hellinger deviate tests: efficiency, breakdown points, and examples*, "Journal of the American Statistical Association", 84, pp. 107-113.
- T.S. TSOU, R.M. ROYALL (1995), *Robust likelihoods*, "Journal of the American Statistical Association", 90, pp. 316-320.
- W.J. WELCH (1987), *Re-randomizing the median in matched-pairs designs*, "Biometrika", 74, pp. 609-614.
- S.S. WILKS (1938), *The large-sample distribution of the likelihood ratio for testing composite hypothesis*, "The Annals of Mathematical Statistics", 9, pp. 60-62.

RIASSUNTO

Un approccio robusto alla verifica d'ipotesi basato sulla funzione di verosimiglianza pesata

Una versione pesata delle funzioni test basate sul rapporto di verosimiglianza C proposta per l'inferenza parametrica. Quando il modello parametrico è specificato correttamente le funzioni test qui introdotte sono asintoticamente equivalenti alle corrispondenti funzioni test basate sulle quantità di verosimiglianza. Le proprietà di robustezza di queste funzioni test sono studiate attraverso la funzione di punto di rottura e comparate con quelle dei test di verosimiglianza. Alcuni esempi e simulazioni illustrano le proprietà dei test introdotti.

SUMMARY

Robust testing hypotheses via Weighted Likelihood function

Weighted version of the Likelihood Ratio, Wald and score tests are proposed for parametric inference. If the parametric model is correct, the Weighted Likelihood tests are asymptotically equivalent to the corresponding Likelihood tests. Breakdown properties of the Weighted Likelihood tests and Likelihood tests are compared. Some comparison with the Hellinger tests (Simpson, 1989) and other robust tests (Heritier e Ronchetti, 1994) are presented too. Examples and Simulations examine the relative performance of the tests in finite samples and real data context.