



UNIVERSITÀ DEGLI STUDI DI PADOVA



FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA
IN STATISTICA E INFORMATICA

TESI DI LAUREA

INFERENZA MULTIMODELLO: APPROCCIO FREQUENTISTA

RELATORE: CHIAR.MA PROF.SSA ALESSANDRA SALVAN

LAUREANDO: RICCARDO DE BIN

ANNO ACCADEMICO 2006/2007

Indice

1	Inferenza multimodello	1
1.1	Introduzione	1
1.2	Notazione	3
1.3	Criteri di selezione del modello	6
2	Metodi di costruzione di stimatori	7
2.1	Metodi ottimali per famiglie posizione scala	7
2.1.1	Stimatori compromesso ottimi	8
2.1.2	Stimatori compromesso di massima verosimiglianza	10
2.1.3	Approssimazione di ordine superiore	13
2.2	Metodi basati su criteri selezione modello	13
3	Problemi di distribuzione e costruzione di regioni di confidenza	15
3.1	Distribuzione stimatori selezione modello	15
3.1.1	Espressione asintotica dell'AIC	19
3.1.2	Distribuzione degli stimatori	20
3.1.3	Differente parametrizzazione	21
3.1.4	Intervalli di confidenza	21
3.2	Distribuzione degli stimatori compromesso	24
4	Applicazioni	25
4.1	Stimatori basati sull'AIC per Weibull	25
4.1.1	Simulazione della distribuzione	30

4.1.2	Simulazione della copertura	44
4.2	Stimatori per famiglie di posizione e scala	62
4.2.1	Stimatore basato sulla normale e normale di ordine 4	62
4.2.2	Stimatore basato sulla normale e normale di ordine s	63
4.2.3	Stimatore basato sulla normale e distribuzione dei valori estremi	73
4.2.4	Stimatore basato sulla normale, normale di ordine 4 e distribuzione dei valori estremi	75
4.2.5	Studio dell'efficienza	75
	Conclusioni	79
	Appendice: comandi R	81
	Bibliografia	119

Capitolo 1

Inferenza multimodello

1.1 Introduzione

La procedura usuale nelle applicazioni per l'inferenza su una quantità d'interesse consiste nell'adattare diversi modelli ai dati. Successivamente, attraverso qualche metodo di selezione, si sceglie il modello migliore, e lo si utilizza come se fosse il vero modello. Questo metodo, in pratica, procede come se il modello selezionato alla fine fosse stato scelto a priori, e nel fare inferenza sul parametro d'interesse non tiene conto dell'ulteriore incertezza dovuta alla selezione del modello. La conseguenza di questo è una stima della varianza di stima troppo ottimistica, con inevitabili ripercussioni sulla costruzione degli intervalli di confidenza (che risulteranno in genere troppo 'stretti', ossia con un livello di copertura inferiore a quello nominale), sui test di ipotesi, ecc.

Un'alternativa è rappresentata dal non condizionarsi ad uno specifico modello, ma fare invece inferenza tenendo conto dell'informazione derivante dai vari modelli adattati. Soprattutto quando si è in una situazione in cui non esiste un modello chiaramente migliore degli altri, è rischioso fare inferenza condizionandosi al singolo modello, mentre è sicuramente preferibile avere stimatori 'lisciati' su vari modelli. Le procedure inferenziali improntate a quest'idea prendono il nome di 'inferenza multimodello' (*model averaging*).

L'inferenza multimodello è stata ampiamente trattata secondo l'approc-

cio bayesiano (si veda, ad esempio, Hoeting, Madigan, Raftery & Volinsky (1999)). Se sono note le probabilità a priori per una serie di possibili modelli, unitamente alla densità a priori per i parametri di ogni modello, allora la teoria bayesiana è in grado, in via di principio, di fornire la distribuzione a posteriori per ogni parametro di interesse. L'approccio, in ogni caso, rimane problematico. Innanzitutto, ci sono difficoltà associate al metodo, spesso *ad hoc*, con il quale vengono fissate le probabilità a priori, soprattutto quando i modelli da considerare sono numerosi. In secondo luogo, l'impiego di distribuzioni a priori di default può portare a mescolare assieme a priori in contrasto tra loro.

Sporadici risultano invece i contributi nell'ambito dell'approccio frequentista (Buckland, Burnham & Augustin (1997), Burnham & Anderson (2002), Hjort & Claeskens (2003)). Questo può essere dovuto a vari fattori, tra i quali la difficoltà nel descrivere adeguatamente la scelta di un buon modello, che coinvolge concetti non formalizzati, come guardare i residui o cercare una trasformazione adeguata, oppure la difficoltà a formalizzare le questioni relative agli stimatori post-selezione.

Il presente lavoro intende seguire la via frequentista all'inferenza multimodello. In particolare, dopo questo capitolo introduttivo, nel Capitolo 2 si presenta una rassegna sui metodi di costruzione degli stimatori basati su più di un modello, partendo da quelli compromesso per famiglie di posizione e scala. Da questi, si passerà ad una versione approssimata basata sugli stimatori di massima verosimiglianza, per la cui costruzione si sfrutta la formula di Laplace: in base all'ordine di approssimazione, si descriveranno due versioni di stimatore approssimato. Successivamente, si riporta la costruzione di stimatori multimodello basati su criteri di selezione del modello, in particolar modo il criterio di Akaike.

Nel Capitolo 3, viceversa, si riassumono i risultati asintotici sviluppati da Hjort & Claeskens (2003) che permettono di ottenere la distribuzione limite degli stimatori basati su criteri di selezione del modello, grazie alla quale si ottiene un'espressione analitica per gli intervalli di confidenza.

Nel Capitolo 4, infine, si adatta la teoria esposta nei precedenti capitoli ad esempi corredati da studi di simulazione. Si parte costruendo uno stimatore basato sul modello Weibull e sul modello esponenziale, e si verifica l'effettiva corrispondenza sotto entrambi i modelli tra la distribuzione teorica e la distribuzione empirica. Sempre su questo esempio si verifica l'effettiva copertura degli intervalli di confidenza costruiti grazie alla teoria esposta nei capitoli precedenti.

Successivamente, sempre nel Capitolo 4, si definiscono, basandoli su vari modelli parametrici, alcuni stimatori multimodello per famiglie di posizione e scala, per i quali si effettua anche uno studio della distribuzione. In particolare, per lo stimatore basato sui modelli normale e normale di ordine s , si fa il confronto con lo stimatore costruito utilizzando il criterio AIC basato sugli stessi due modelli. Infine, vengono confrontati i vari stimatori in termini di errore quadratico medio, stimato tramite simulazione sotto vari modelli con numerosità differenti.

1.2 Notazione

Si consideri per i dati y appartenenti allo spazio campionario \mathcal{Y} , un modello statistico parametrico \mathcal{F} con densità di probabilità $f_Y(y, \theta)$, con $\theta \in \Theta \subseteq \mathbb{R}^p$ per qualche p . La sintesi informativa dei dati y su θ è la **funzione di verosimiglianza**

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

dove $c(y) > 0$ è una costante di proporzionalità arbitraria.

Se i dati sono costituiti da n osservazioni, $y = (y_1, \dots, y_n)$ indipendenti, la verosimiglianza assume la forma

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

dove $f_{Y_i}(y_i; \theta)$, con $i = 1, \dots, n$, indica la funzione di densità di probabilità dell' i -esima osservazione. Da questo punto in poi, ove non sorgano ambiguità,

se le osservazioni sono identicamente distribuite tale funzione verrà indicata con $f(y_i; \theta)$.

Spesso è conveniente considerare il logaritmo naturale della verosimiglianza, ossia la **funzione di log-verosimiglianza**

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

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

Per n osservazioni indipendenti ed identicamente distribuite si ha

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

Assumendo che la log-verosimiglianza sia una funzione di θ *regolare*, ossia infinitamente derivabile, si richiamano le seguenti **quantità di verosimiglianza** con la notazione che si userà nel seguito.

- La **funzione score**, ossia il vettore delle derivate parziali prime di $l(\theta)$ rispetto alle componenti $\theta_1, \dots, \theta_p$ di θ sarà indicata da una scrittura del tipo

$$l_* = l_*(\theta; y) = (l_1, \dots, l_p)^\top,$$

dove $l_r = \partial l(\theta) / \partial \theta_r$, $r = 1, \dots, p$, indica la r -esima derivata parziale e θ_r la r -esima componente del vettore di parametri $\theta = (\theta_1, \dots, \theta_p)^\top$.

- Per indicare la matrice di **informazione osservata** si userà il simbolo

$$j = j(\theta) = \left[-\frac{\partial^2 l}{\partial \theta \partial \theta^\top} \right]$$

o, in notazione compatta, $j = [-l_{rs}]$, dove $l_{rs} = \partial^2 l(\theta) / (\partial \theta_r \partial \theta_s)$ rappresenta l'elemento di posto (r, s) nella matrice j .

- La matrice di **informazione attesa** di Fisher verrà indicata con

$$i = i(\theta) = E_\theta[j(\theta)],$$

ed i_{rs} sarà il suo elemento di posto (r, s) .

Con l'acronimo MSE si indicherà l'errore quadratico medio, ossia la quantità:

$$MSE_{\theta}[\hat{\theta}(Y)] = E_{\theta}[(\hat{\theta}(Y) - \theta)^2]$$

Nelle applicazioni considerate in questa tesi, infine, si utilizzeranno delle distribuzioni continue note, in particolare, per un'osservazione univariata y :

- Distribuzione normale: $N(\mu; \sigma^2)$, con densità

$$f(y; \mu, \sigma) = -\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma}(y - \mu)^2\right),$$

dove $\mu \in \mathfrak{R}$ e $\sigma > 0$;

- Distribuzione di Weibull: $Wei(\theta, \gamma)$, con densità

$$f(y; \theta, \gamma) = \begin{cases} \theta \gamma (\theta y)^{\gamma-1} \exp\left(-(\theta y)^{\gamma}\right) & \text{se } y > 0 \\ 0 & \text{altrove} \end{cases},$$

dove $\theta > 0$ e $\gamma > 0$;

- Distribuzione esponenziale: $Exp(\theta)$, con densità

$$f(y; \theta) = \begin{cases} \theta \exp(-\theta y) & \text{se } y > 0 \\ 0 & \text{altrove} \end{cases},$$

dove $\theta > 0$;

- Distribuzione normale di ordine s (*exponential power*): $EP(s)$, con densità

$$f(y; \mu, \sigma, s) = \frac{1}{2s^{1/s}\Gamma(1 + 1/s)\sigma} \exp\left(-\frac{1}{s\sigma^s}|y - \mu|^s\right),$$

dove $\mu \in \mathfrak{R}$, $\sigma > 0$ e $s > 0$;

- Distribuzione dei valori estremi: $ve(\mu, \sigma)$, con densità

$$f(y; \mu, \sigma) = \frac{1}{\sigma} \exp\left(\frac{y - \mu}{\sigma} - \exp\left(\frac{y - \mu}{\sigma}\right)\right),$$

dove $\mu \in \mathfrak{R}$ e $\sigma > 0$.

1.3 Criteri di selezione del modello

In questa tesi verrà affrontata la costruzione di stimatori basati su criteri di selezione del modello. In questo paragrafo si ricorda brevemente che cosa si intende per criterio di selezione del modello, con particolare attenzione al criterio di Akaike, che verrà ampiamente utilizzato nel seguito (si confronti Burnham & Anderson (2002, Capitolo 7))

L'obiettivo è fornire uno strumento in grado di confrontare l'adattamento di vari modelli ai dati, al fine di selezionare il migliore. 'Migliore' va inteso come il modello che più si avvicini al modello generatore dei dati stessi, evitando quindi modelli che tendano a sovra-adattarsi al particolare campione di dati, non riuscendo a cogliere l'aspetto generale degli stessi.

Da un lato, dunque, si vuole che il modello segua bene il campione, e quindi, in un approccio basato sulla verosimiglianza, si vuole massimizzare la log-verosimiglianza. Dall'altra parte, si desidera un modello meno vincolato possibile al particolare campione: per tenere conto di questo aspetto, si può introdurre una quantità che penalizzi un'eccessivo numero di parametri.

Seguendo questa logica, si arriva ad una famiglia di criteri riconducibile a funzioni obiettivo del tipo

$$IC = -2 \log L(\hat{\theta}) + \text{penalità}(p),$$

dove $\text{penalità}(p)$ rappresenta la penalizzazione assegnata ad un modello con p parametri.

La funzione di penalità necessita semplicemente di essere positiva ed essere funzione crescente di p .

In questa tesi si utilizzerà il criterio proposto da Akaike (1973), l'**AIC** (**Akaike Information Criterion**), il quale prevede come funzione di penalità la funzione $2p$:

$$AIC = -2 \log f_Y(y; \hat{\theta}) + 2p.$$

Capitolo 2

Metodi di costruzione di stimatori

In questo capitolo si analizza la costruzione di stimatori che permettono di sfruttare l'informazione proveniente da più di un modello. Si distingue tra metodi ottimali per famiglie di posizione e scala, e metodi basati sulla ponderazione con pesi definiti in termini di criteri di selezione del modello.

Nel primo caso si parlerà di 'stimatori compromesso', mentre nel secondo si parlerà di 'stimatori frequentisti basati su criteri di selezione del modello'.

2.1 Metodi ottimali per famiglie di posizione e scala

Gli stimatori compromesso ottimali per specifiche classi di modelli sono stimatori che si comportano bene in più di un possibile modello. Sfruttando le proprietà delle particolari famiglie, si ottiene uno stimatore migliore, che tiene conto, pesandola, dell'informazione ricavabile da ogni modello.

Easton (1991) ha sviluppato gli stimatori compromesso ottimi per le famiglie di posizione e scala, introdotti da Pregibon and Tukey (1981), Tukey (1981) e Tukey (1987), utilizzando i metodi di approssimazione di Laplace per gli integrali doppi che definiscono gli stimatori compromesso ottimi stessi.

A loro volta questi ultimi sono una generalizzazione degli stimatori di Pitman per ognuna delle densità considerate della famiglia di posizione e scala. Il lavoro di Easton (1991) porta ad una nuova classe di stimatori compromesso, chiamati **stimatori compromesso di massima verosimiglianza (SCMV)**, che non sono altro che una media pesata degli stimatori di massima verosimiglianza per le densità considerate, con i pesi che dipendono dalle funzioni di verosimiglianza.

2.1.1 Stimatori compromesso ottimi

Sia $y = (y_1, \dots, y_n)$ un campione proveniente da una variabile casuale continua Y appartenente ad una famiglia di posizione e scala, con densità

$$f(y; \mu, \sigma) = \frac{1}{\sigma} f_0\left(\frac{y - \mu}{\sigma}\right),$$

dove $y \in \mathbb{R}$, $\mu \in \mathbb{R}$ e $\sigma > 0$. Il punto di partenza è lo stimatore introdotto da Pitman (1938). Esso è lo stimatore del parametro di posizione con errore quadratico medio minimo tra gli stimatori equivarianti rispetto a trasformazioni di posizione e scala.

Sia

$$I_f[h(\mu, \sigma)] = \int_{-\infty}^{+\infty} \int_0^{+\infty} h(\mu, \sigma) \frac{1}{\sigma^{n+1}} \prod_{j=1}^n \frac{1}{\sigma} f_0\left(\frac{y_j - \mu}{\sigma}\right) d\sigma d\mu.$$

In formula, lo stimatore di Pitman è

$$T_f(Y) = \frac{I_f\left[\frac{\mu}{\sigma^2}\right]}{I_f\left[\frac{\mu}{\sigma^2}\right]} = \frac{\int_{-\infty}^{+\infty} \int_0^{+\infty} \frac{\mu}{\sigma^{n+3}} \prod_{i=1}^n f\left(\frac{y_i - \mu}{\sigma}\right) d\sigma d\mu}{\int_{-\infty}^{+\infty} \int_0^{+\infty} \frac{1}{\sigma^{n+3}} \prod_{i=1}^n f\left(\frac{y_i - \mu}{\sigma}\right) d\sigma d\mu}. \quad (2.1)$$

Il passo successivo, operato da Pregibon & Tukey (1981), Tukey (1981) e Tukey (1987), è quello di creare degli stimatori che siano ottimali simultaneamente per più di una famiglia parametrica, in maniera che lo stimatore si comporti bene in un'ampia gamma di modelli, per cui sarà buono qualsiasi sia la vera distribuzione dei dati. I risultati sono presentati per sole due famiglie parametriche \mathcal{F}_1 e \mathcal{F}_2 , ma l'estensione a più famiglie è immediata.

L'obbiettivo, fondamentalmente, è trovare il giusto compromesso tra l'efficienza dello stimatore per le due famiglie parametriche \mathcal{F}_1 e \mathcal{F}_2 (con densità standard, rispettivamente, f_1 ed f_2) simultaneamente, dato che è impossibile aumentare l'efficienza per una senza diminuire quella per l'altra.

In particolare, per ogni $\varphi \in [0, 1]$, lo stimatore sarà quel $T(Y)$ che minimizza la funzione

$$\varphi \frac{\text{var}_{f_1}[T(Y)] - \text{var}_{f_1}[T_{f_1}(Y)]}{\text{var}_{f_1}[T_{f_1}(Y)]} + (1 - \varphi) \frac{\text{var}_{f_2}[T(Y)] - \text{var}_{f_2}[T_{f_2}(Y)]}{\text{var}_{f_2}[T_{f_2}(Y)]}, \quad (2.2)$$

dove:

- $T_{f_i}(Y)$ è lo stimatore di Pitman per la famiglia parametrica \mathcal{F}_i ;
- $\text{var}_{f_i}[\cdot]$ è la varianza calcolata rispetto alla distribuzione con densità f_i ;

con la quantità

$$\frac{\text{var}_{f_i}[T(Y)] - \text{var}_{f_i}[T_{f_i}(Y)]}{\text{var}_{f_i}[T_{f_i}(Y)]}$$

che può essere considerata come un 'surplus relativo di varianza'.

Attraverso operazioni di minimizzazione, si arriva a ricavare lo stimatore cercato, chiamato, proprio perché deve essere ottimale sotto entrambe le famiglie parametriche, **stimatore bi-ottimale**. Esso è

$$T(y, \varphi) = \frac{\varphi \alpha_1(y) T_{f_1}(y) + (1 - \varphi) \alpha_2(y) T_{f_2}(y)}{\varphi \alpha_1(y) + (1 - \varphi) \alpha_2(y)}, \quad (2.3)$$

dove

$$\alpha_i(y) = \frac{I_{f_i}(1/\sigma^2)}{\text{var}_{f_i}[T_{f_i}(Y)]}. \quad (2.4)$$

Perciò gli stimatori bi-ottimali sono una combinazione convessa degli stimatori di Pitman per le due famiglie parametriche considerate, dove i pesi nella combinazione convessa dipendono sia da φ sia dal campione y . Il valore di φ deve essere determinato attraverso i metodi Monte Carlo.

Per garantire l'ottimalità dello stimatore sotto entrambe le famiglie parametriche, non necessariamente si deve utilizzare la (2.2). È infatti sufficiente basarsi su una qualsiasi funzione che, minimizzata, porti alla scelta di uno

stimatore che abbia le caratteristiche desiderate sotto entrambe le famiglie parametriche.

Utilizzando la funzione

$$\max_{i=1,2} \frac{I[1/\sigma^2](T(Y) - T_{f_i}(y))^2}{\text{var}_{f_i}[T_{f_i}(Y)]},$$

si ricava, sempre minimizzando, lo stimatore, chiamato **stimatore bi-efficace**,

$$T_{bieff}(y) = \frac{\alpha'_1(y)T_{f_1}(y) + \alpha'_2(y)T_{f_2}(y)}{\alpha'_1(y) + \alpha'_2(y)}, \quad (2.5)$$

dove

$$\alpha'_i(y) = \left[\frac{I_{f_i}(1/\sigma^2)}{\text{var}_{f_i}[T_{f_i}(Y)]} \right]^{1/2}. \quad (2.6)$$

Quest'ultimo ha il vantaggio che il valore del parametro φ non deve essere conosciuto in anticipo.

2.1.2 Stimatori compromesso di massima verosimiglianza

Gli stimatori fin qui visti presentano lo svantaggio di un certo onere computazionale, dovuto agli integrali multipli presenti. Easton (1991) introduce una nuova tipologia di stimatori, gli stimatori compromesso di massima verosimiglianza, ottenuti applicando le tecniche di approssimazione di Laplace agli integrali che definiscono gli stimatori compromesso ottimi. Questi nuovi stimatori sono medie pesate degli stimatori di massima verosimiglianza per le due densità considerate, con i pesi che dipendono dal campione attraverso le due funzioni di verosimiglianza.

In particolare gli integrali doppi da considerare sono:

$$I_{f_i}[1/\sigma^2] = \int_{-\infty}^{+\infty} \int_0^{+\infty} \frac{1}{\sigma^3} \prod_{j=1}^n \frac{1}{\sigma} f\left(\frac{y_j - \mu}{\sigma}\right) d\sigma d\mu$$

e

$$I_{f_i}[\mu/\sigma^2] = \int_{-\infty}^{+\infty} \int_0^{+\infty} \frac{\mu}{\sigma^3} \prod_{j=1}^n \frac{1}{\sigma} f\left(\frac{y_j - \mu}{\sigma}\right) d\sigma d\mu.$$

Si prenda in esame l'ultimo, tenendo comunque conto che per l'altro si procede allo stesso modo. Si può riscrivere l'equazione come

$$I_{f_i}[\mu/\sigma^2] = \int_{-\infty}^{+\infty} \int_0^{+\infty} \frac{\mu}{\sigma^3} \exp[n\bar{l}_{f_i}(y; \mu, \sigma)] d\sigma d\mu,$$

dove

$$\bar{l}_{f_i}(y; \mu, \sigma) = \frac{\sum_{j=1}^n \log \frac{f_i((y_j - \mu)/\sigma)}{\sigma}}{n}.$$

Quindi la $\bar{l}_{f_i}(y; \mu, \sigma)$ non è che la funzione di log-verosimiglianza per (μ, σ) sotto il modello con famiglia parametrica \mathcal{F}_i divisa per la numerosità campionaria n .

Si supponga che, per n grande, ogni $\bar{l}_{f_i}(y; \mu, \sigma)$, come funzione di μ ed σ , abbia un unico massimo globale $M_{f_i}(y)$. Questo massimo si ottiene quando μ ed σ sono uguali agli stimatori di massima verosimiglianza dei parametri di posizione e scala $\hat{\mu}_{f_i}$ ed $\hat{\sigma}_{f_i}$ per la famiglia parametrica \mathcal{F}_i . Quindi $M_{f_i}(y) = \bar{l}_{f_i}(y; \hat{\mu}_{f_i}, \hat{\sigma}_{f_i})$. Si può dunque riscrivere il nostro integrale come

$$I_{f_i}[\mu/\sigma^2] = \exp(nM_{f_i}(y)) \int_{-\infty}^{+\infty} \int_{\hat{\sigma}_{f_i}}^{+\infty} \frac{\mu - \hat{\mu}_{f_i}}{(\sigma + \hat{\sigma}_{f_i})^3} \exp(ng_{f_i}(y; \mu, \sigma)) d\sigma d\mu,$$

dove $g_{f_i}(y; \mu, \sigma) = \bar{l}_{f_i}(y; \mu + \hat{\mu}_{f_i}, \sigma + \hat{\sigma}_{f_i}) - M_{f_i}(y)$.

Con questa formulazione, si può applicare il metodo di approssimazione di Laplace, che porta ad ottenere

$$I_{f_i}[\mu/\sigma^2] \approx \exp(nM_{f_i}(y)) \frac{\hat{\mu}_{f_i}}{\hat{\sigma}_{f_i}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp(na^T H_{f_i}(y)a/2) d\sigma d\mu, \quad (2.7)$$

dove:

- $a^T = (\mu - \hat{\mu}_{f_i}, \sigma - \hat{\sigma}_{f_i})$
- $H_{f_i}(y)$ è la matrice 2×2 delle derivate parziali seconde di $g_{f_i}(y; t, s)$ valutate in $t = s = 0$, ossia non è altro che la matrice di informazione osservata per la famiglia parametrica \mathcal{F}_i .

L'integrale presente nella (2.7) può essere facilmente calcolato notando la somiglianza con l'integrazione di una densità di una normale bivariata. Integrando, si ha

$$I_{f_i}[\mu/\sigma^2] \approx \frac{2\pi\hat{\mu}_{f_i} \exp(nM_{f_i}(y))}{n\hat{\sigma}_{f_i}^3 |\det(H_{f_i}(y))|^{1/2}}.$$

Similmente, l'altro integrale, $I_{f_i}[1/\sigma^2]$, risulta

$$I_{f_i}[\mu/\sigma^2] \approx \frac{2\pi \exp(nM_{f_i}(y))}{n\hat{\sigma}_{f_i}^3 |\det(H_{f_i}(y))|^{1/2}}.$$

A questo punto si possono sfruttare questi calcoli per ottenere un'approssimazione dello stimatore di Pitman, sostituendo nella (2.1):

$$T_{f_i} = \frac{I_{f_i}[\mu/\sigma^2]}{I_{f_i}[1/\sigma^2]} \approx \hat{\mu}_{f_i}.$$

Perciò lo stimatore di Pitman per la densità f_i è approssimato al 1° ordine dal corrispondente stimatore di massima verosimiglianza. L'approssimazione per i pesi nella formula, ad esempio, dello stimatore bi-ottimale (2.4) è sviluppato nella stessa maniera:

$$\alpha_i(y) = \frac{I_{f_i}[1/\sigma^2]}{\text{var}_{f_i}[T_{f_i}(Y)]} \approx \frac{2\pi \exp(nM_{f_i}(y))}{n\hat{\sigma}_{f_i}^3 |\det(H_{f_i}(y))|^{1/2} \text{var}_{f_i}[T_{f_i}(Y)]}.$$

Lo stimatore bi-ottimale può essere approssimato sostituendo le ultime approssimazioni nelle (2.3) e (2.4). Per cui

$$T(y, \varphi) \approx \frac{\varphi \tilde{\alpha}_1(y) \hat{\mu}_{f_1} + (1 - \varphi) \tilde{\alpha}_2(y) \hat{\mu}_{f_2}}{\varphi \tilde{\alpha}_1(y) + (1 - \varphi) \tilde{\alpha}_2(y)}, \quad (2.8)$$

dove

$$\tilde{\alpha}_i(y) = \frac{\exp(nM_{f_i}(y))}{\hat{\sigma}_{f_i}^3 |\det(H_{f_i}(y))|^{1/2} \text{var}_{f_i}[T_{f_i}(Y)]}. \quad (2.9)$$

Questo stimatore è chiamato **SCMV stile bi-ottimale**.

Allo stesso modo, si ricava lo **stimatore SCMV stile bi-efficace**:

$$T_{bief}(y) \approx \frac{\tilde{\alpha}'_1(y) \hat{\mu}_{f_1} + \tilde{\alpha}'_2(y) \hat{\mu}_{f_2}}{\tilde{\alpha}'_1(y) + \tilde{\alpha}'_2(y)}, \quad (2.10)$$

dove

$$\tilde{\alpha}'_i(y) = \left[\frac{\exp(nM_{f_i}(y))}{\hat{\sigma}_{f_i}^3 |\det(H_{f_i}(y))|^{1/2} \text{var}_{f_i}[T_{f_i}(Y)]} \right]^{1/2}. \quad (2.11)$$

2.1.3 Approssimazione di ordine superiore

Gli stimatori introdotti da Easton (1991) sono stati perfezionati da Ventura (1998), che ha sviluppato gli integrali doppi includendo il termine di ordine $O(n^{-1})$ nella formula di Laplace ed ottenendo così approssimazioni di secondo ordine. Lo stimatore preso in considerazione è quello SCMV stile bi-efficace (2.10), ritenuto da Easton (1991) stesso quello computazionalmente più semplice. Lo stimatore diventa

$$T_{app}(y) = \frac{\hat{\alpha}_1 T_{\mu_1} + \hat{\alpha}_2 T_{\mu_2}}{\hat{\alpha}_1 + \hat{\alpha}_2}, \quad (2.12)$$

con i pesi

$$\hat{\alpha}_i = \left[\frac{\exp(nM_{f_i}(y))}{\hat{\sigma}_{f_i}^3 |det(H_{f_i}(y))|^{1/2} var_{f_i}[T_{f_i}(Y)]} \right]^{1/2} \{1 + D_{f_i} + o(n^{-2})\}, \quad (2.13)$$

dove D_{f_i} è il termine di correzione di ordine $O(n^{-1})$.

Nella (2.12), inoltre, sono da specificare i T_{μ_i} , che non sono altro che gli stimatori di Pitman anch'essi calcolati includendo il termine di ordine $O(n^{-1})$ nella formula di Laplace. Ossia,

$$T_{\mu_i} = T_{f_i} + \frac{1}{2} \left(\sum_{q,r,s} H_{qr}^{-1} H_{s1}^{-1} \hat{l}_{qrs} - 6 \frac{H_{12}^{-1}}{\hat{\sigma}_{f_i}} \right),$$

dove:

- H_{qr} è l'elemento in posizione q, r della matrice di informazione osservata;
- \hat{l}_{qrs} è la derivata terza della log-verosimiglianza rispetto alle componenti q, r ed s del parametro, nel punto di massima verosimiglianza.

2.2 Metodi basati su criteri di selezione del modello

Questi tipi di stimatori nascono dall'idea di sfruttare per il calcolo della stima del parametro di interesse l'informazione proveniente da tutti i modelli,

pesandola con un'opportuna quantità che sia proporzionale ad un indice di bontà del modello.

L'idea è stata introdotta da Buckland, Burnham & Augustin (1997) e discussa ulteriormente da Burnham & Anderson (2002, Capitolo 4). Il metodo prevede di ordinare i modelli (si suppongano R differenti modelli) dal migliore al peggiore, basandosi su un criterio di selezione del modello. Una volta associati i modelli ad un corrispondente indice di bontà di adattamento, si procede relativizzando questi ultimi, in modo da ottenere dei pesi che indichino la plausibilità relativa di ogni modello. Poi, semplicemente, si calcola una media pesata degli stimatori relativi al parametro d'interesse

$$\hat{\mu} = \sum_{i=1}^R w_i \hat{\mu}_i, \quad (2.14)$$

dove $\hat{\mu}_i$ indica lo stimatore di μ per l' i -simo modello, e w_i è il peso assegnato al modello corrispondente.

I pesi w_i hanno come unica condizione quella che devono sommare ad 1 ($\sum_{i=1}^R w_i = 1$), e possono essere determinati in vari modi. Burnham & Anderson (2002, Capitolo 4) suggeriscono varie possibilità:

- metodo basato sui pesi di Akaike (AIC);
- metodo basato su procedure di tipo bootstrap.

Nel primo caso, semplicemente, si sfruttano i valori degli AIC di ciascun modello, relativizzati, per determinare il peso da assegnare ad ogni modello considerato. Una scelta possibile è quella presentata da Hjort & Claeskens (2003), che verrà utilizzata nel seguito:

$$w_i = \frac{\exp(\frac{1}{2}AIC_i)}{\sum_{i=1}^R \exp(\frac{1}{2}AIC_i)}. \quad (2.15)$$

Il calcolo dei pesi attraverso la procedura bootstrap, invece, prevede di utilizzare come pesi w_i le stime delle frequenze di selezione del modello ($\hat{\pi}_i$) ottenute col metodo bootstrap (per approfondire questo metodo, si veda Burnham & Anderson (2002), Capitolo 8).

Capitolo 3

Problemi di distribuzione e costruzione di regioni di confidenza

3.1 Distribuzione degli stimatori basati su criteri di selezione del modello

Hjort e Claeskens (2003) hanno sviluppato una teoria asintotica per la distribuzione degli stimatori introdotti da Buckland, Burnham & Augustin (1997) e discussi ulteriormente da Burnham & Anderson (2002, Capitolo 4).

Le condizioni sotto cui è sviluppata questa teoria prevedono la presenza di modelli annidati, in cui il vero modello sia incluso nel modello più ampio (questo almeno inizialmente, poi si vedrà come, con un minimo aggiustamento, si potrà tralasciare questa condizione).

Formalmente, si parte da un campione di osservazioni indipendenti y_1, \dots, y_n provenienti da una variabile casuale con funzione di ripartizione F . L'inferenza è fatta su un certo parametro d'interesse $\mu = \mu(F)$.

Come detto, inoltre, sono adattati vari modelli annidati, che partono da un modello ridotto $f(y, \theta)$ per arrivare al modello completo $f(y, \theta, \gamma)$, dove θ

è un vettore p -dimensionale di parametri presenti in tutti i modelli, mentre γ è un vettore q -dimensionale, che vale γ_0 nel modello ridotto (γ_0 indica un vettore di valori fissi e noti) e presenta più valori incogniti a mano a mano che ci si allontana dal modello ridotto verso quello completo. C'è, quindi, una varietà di stimatori da considerare, che vanno da $\hat{\mu}_{esteso} = \mu(\hat{\theta}_{esteso}, \hat{\gamma}_{esteso})$, usando gli stimatori di massima verosimiglianza nel modello completo, a $\hat{\mu}_{ridotto} = \mu(\hat{\theta}_{ridotto}, \gamma_0)$. Se si indica con $S \in \{1, \dots, q\}$ il sottoinsieme di indici delle componenti di γ ignote, lo stimatore generale ha la forma

$$\hat{\mu}_S = \mu(\hat{\theta}_S, \hat{\gamma}_S, \gamma_{0,S^C}),$$

con S^C il complementare di S .

L'obiettivo, in Hjort e Claeskens (2003) è quindi vedere come si comportano gli stimatori $\hat{\mu}_S$, ed, in particolare, la versione mediata di essi, nel caso:

$$f_{vero}(y) = f_n(y) = f(y, \theta_0, \gamma_0 + \delta/\sqrt{n}), \quad (3.1)$$

dove $\delta_1, \dots, \delta_q$ indicano lo scostamento in direzione $1, \dots, q$.

Innanzitutto si consideri la funzione punteggio del modello esteso

$$\begin{pmatrix} U(y) \\ V(y) \end{pmatrix} = \begin{pmatrix} \partial \log f(y, \theta_0, \gamma_0) / \partial \theta \\ \partial \log f(y, \theta_0, \gamma_0) / \partial \gamma \end{pmatrix},$$

dove U è un vettore p -dimensionale e V un vettore q -dimensionale. La matrice $(p+q) \times (p+q)$ di covarianza di (U, V) , sotto il modello nullo, è

$$i_{esteso} = \begin{pmatrix} i_{00} & i_{01} \\ i_{10} & i_{11} \end{pmatrix},$$

con inversa

$$i_{esteso}^{-1} = \begin{pmatrix} i^{00} & i^{01} \\ i^{10} & i^{11} \end{pmatrix};$$

in particolare, si indichi con K la quantità

$$K = i^{11} = (i_{11} - i_{10}i_{00}^{-1}i_{01})^{-1}. \quad (3.2)$$

Si stanno considerando i modelli indicizzati dal sottoinsieme S di $\{1, \dots, q\}$. Si ponga π_S la matrice di proiezione di $v = (v_1, \dots, v_q)^t$ sul sottovettore $\pi_S v = v_S$ di componenti v_j con $j \in S$. Quindi, π_S ha dimensione $|S| \times q$, dove $|S|$ è la dimensione di S . Per $V_S = \pi_S V$ si ha:

$$i_S = \text{var}_0 \begin{pmatrix} U(Y) \\ V_S(Y) \end{pmatrix} = \begin{pmatrix} i_{00} & i_{01,S} \\ i_{10,S} & i_{11,S} \end{pmatrix} = \begin{pmatrix} i_{00} & i_{01}\pi_S^t \\ \pi_S i_{10} & \pi_S i_{11,S}\pi_S^t \end{pmatrix}.$$

È necessario conoscere anche la sua inversa, che è

$$i_S^{-1} = \begin{pmatrix} i^{00,S} & i^{01,S} \\ i^{10,S} & i^{11,S} \end{pmatrix},$$

dove:

- $i^{11,S} = (\pi_S K^{-1} \pi_S^t)^{-1} = K_S;$
- $i^{01,S} = -i_{00}^{-1} i_{01} \pi_S^t K_S;$
- $i^{00,S} = i_{00}^{-1} + i_{00}^{-1} i_{01} \pi_S^t K_S \pi_S i_{01} i_{00}^{-1}.$

Lemma 1 (*Hjort & Claeskens, 2003*). Si consideri $\bar{U}_n = n^{-1} \sum_{i=1}^n U(Y_i)$ e $\bar{V}_n = n^{-1} \sum_{i=1}^n V(Y_i)$. Sotto la (3.1), vale:

$$\begin{pmatrix} \sqrt{n} \bar{U}_n \\ \sqrt{n} \bar{V}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} i_{01} \delta \\ i_{11} \delta \end{pmatrix} + \begin{pmatrix} M \\ N \end{pmatrix},$$

dove

$$\begin{pmatrix} M \\ N \end{pmatrix} \sim N_{p+q}(0, i_{esteso}).$$

Nei sotto-modelli, denotando con $(\hat{\theta}_S, \hat{\gamma}_S)$ lo stimatore di massima verosimiglianza nel modello che include γ_j parametri per $j \in S$, vale:

Lemma 2 (*Hjort & Claeskens, 2003*). Sotto la (3.1), vale:

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_S - \theta_0) \\ \sqrt{n}(\hat{\gamma}_S - \gamma_{0,S}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} C_S \\ D_S \end{pmatrix} = i_S^{-1} \begin{pmatrix} i_{01} \delta + M \\ \pi_S i_{11} \delta + N_S \end{pmatrix} \\ \sim N_{p+|S|} \left(i_S^{-1} \begin{pmatrix} i_{01} \\ \pi_S i_{11} \end{pmatrix} \delta, i_S^{-1} \right).$$

Prima di procedere, è necessario definire alcune quantità utili nel seguito. Innanzitutto si indichi con W la quantità:

$$W = i^{10}M + i^{11}N = K(N - i_{10}i_{00}^{-1}M). \quad (3.3)$$

Qui $M \sim N_p(0, i_{00})$, e non è difficile stabilire che M e W sono stocasticamente indipendenti, con $W \sim N_q(0, K)$. Una conseguenza dell'ultimo lemma, è che $\hat{\delta}_S = \sqrt{n}(\hat{\gamma}_S - \gamma_{0,S})$ tende in distribuzione a $D_S = K_S \pi_S K^{-1}(\delta + W)$. In particolare,

$$D_n = \hat{\delta}_{esteso} = \sqrt{n}(\hat{\gamma}_{esteso} - \gamma_0) \xrightarrow{d} D = \delta + W \sim N_q(\delta, K). \quad (3.4)$$

Inoltre, si definiscano

$$H_S = K^{-1/2} \pi_S^t K_S \pi_S K^{-1/2} \quad (3.5)$$

e

$$\omega = i_{10}i_{00}^{-1} \frac{\partial \mu}{\partial \theta} - \frac{\partial \mu}{\partial \gamma}, \quad (3.6)$$

dove le derivate parziali sono state ottenute sotto il modello (θ_0, γ_0) . È possibile notare come ω sia determinato per lo specifico parametro d'interesse μ . H_S , invece, è una matrice $q \times q$ di proiezione, quindi simmetrica ed idempotente, ed è ortogonale a $I - H_S$.

Lemma 3 (*Hjort & Claeskens, 2003*). *Si assuma che $\mu(\theta, \gamma)$ abbia derivate parziali continue nelle vicinanze di (θ_0, γ_0) . Allora lo stimatore di massima verosimiglianza di μ nel modello S ha distribuzione limite nella forma*

$$\sqrt{n}(\hat{\mu}_S - \mu_{vero}) \xrightarrow{d} \Lambda_S = \left(\frac{\partial \mu}{\partial \theta}\right)^t i_{00}^{-1} M + \omega^t (\delta - K^{1/2} H_S K^{-1/2} D),$$

dove le derivate parziali sono calcolate sotto il modello nullo (θ_0, γ_0) . Quindi,

$$\Lambda_S \sim N\left(\omega^t (I - K^{1/2} H_S K^{-1/2}) \delta, \left(\frac{\partial \mu}{\partial \theta}\right)^t i_{00}^{-1} \frac{\partial \mu}{\partial \theta} + \omega^t K^{1/2} H_S K^{1/2} \omega\right).$$

3.1.1 Espressione asintotica dell'AIC

Per sfruttare gli AIC come peso per gli stimatori multimodello, è necessario trovare una sua formulazione utilizzabile nelle formule asintotiche. Il punto di partenza è la formula del criterio di informazione di Akaike:

$$AIC_{n,S} = 2 \sum_{i=1}^n \log f(Y_i, \hat{\theta}_S, \hat{\gamma}_S, \gamma_{0,S^c}) - 2|S|,$$

dove ovviamente $|S|$ è il numero di elementi in S . Per capire il comportamento del criterio nel nostro caso, è utile partire dalla statistica rapporto di verosimiglianza, sviluppata al secondo ordine:

$$\begin{aligned} G_{n,S} &= 2 \sum_{i=1}^n \log \frac{f(Y_i, \hat{\theta}_S, \hat{\gamma}_S, \gamma_{0,S^c})}{f(Y_i, \theta_0, \gamma_0)} \\ &\doteq n \begin{pmatrix} \bar{U}_n \\ \bar{V}_{n,S} \end{pmatrix}^t i_S^{-1} \begin{pmatrix} \bar{U}_n \\ \bar{V}_{n,S} \end{pmatrix} \\ &\xrightarrow{d} \begin{pmatrix} i_{01}\delta + M \\ \pi_S i_{11}\delta + N_S \end{pmatrix}^t i_S^{-1} \begin{pmatrix} i_{01}\delta + M \\ \pi_S i_{11}\delta + N_S \end{pmatrix}, \end{aligned}$$

che è un chi-quadrato non centrale con $p + |S|$ gradi di libertà. Inoltre:

$$\begin{aligned} G_{n,S} - G_{n,\emptyset} &= n(\bar{V}_{n,S} - i_{10,S} i_{00}^{-1} \bar{U}_n)^t \times i^{11,S} (\bar{V}_{n,S} - i_{10,S} i_{00}^{-1} \bar{U}_n) \\ &\xrightarrow{d} (K_S^{-1}\delta + N_S - i_{10,S} i_{00}^{-1} M)^t \times K_S (K_S^{-1}\delta + N_S - i_{10,S} i_{00}^{-1} M), \end{aligned}$$

che è un $\chi_{|S|}^2(\delta^t K_S^{-1} \delta)$.

Combinando le cose, $\hat{\delta}_S = \sqrt{n}(\hat{\gamma}_S - \gamma_{0,S})$ è al più $o_p(1)$ distante da $\sqrt{n}K_S\pi_S(\bar{V}_n - i_{10,S}i_{00}^{-1}\bar{U}_n)$, e, similmente, $D_n \doteq \sqrt{n}K(\bar{V}_n - i_{10,S}i_{00}^{-1}\bar{U}_n)$, che implica $\hat{\delta}_S \doteq K_S\pi_S K^{-1}D_n$. La conseguenza importante è che $\hat{\gamma}_S$, lo stimatore del modello del sottoinsieme S , può essere espresso come una funzione di $\hat{\gamma}_{esteso}$. Inoltre, ne consegue anche che il criterio AIC può essere espresso in termini di D_n , infatti:

$$\begin{aligned} AIC_{n,S} &= G_{n,S} - G_{n,\emptyset} - 2|S| \\ &= D_n^t K^{-1/2} H_S K^{-1/2} D_n - 2|S| + o_p(1). \end{aligned} \tag{3.7}$$

3.1.2 Distribuzione degli stimatori

Si è quindi giunti ad avere le varie quantità d'interesse come determinazioni di $D_n = \hat{\delta}_{esteso}$. Nello specifico ci si sta riferendo a $\hat{\delta}_S$ (3.4) ed al criterio AIC (3.7). Si richiamino i pesi w_i descritti nella (2.15). È utile ora rappresentarli esplicitando la loro dipendenza da S e da D_n , in modo da riscrivere lo stimatore (2.14) come

$$\hat{\mu} = \sum_S c(S|D_n) \hat{\mu}_S. \quad (3.8)$$

Ovviamente la somma delle funzioni peso $c(S|d)$ mantiene la proprietà di essere uguale ad 1 per ogni d , altrimenti lo stimatore non sarebbe consistente.

Per generalizzare lo stimatore (3.8), utilizzando l' H_S definita nella (3.5), è necessario definire le quantità

$$G(d) = K^{-1/2} \left\{ \sum_S c(S|d) H_S \right\} K^{1/2}$$

e

$$\hat{\delta}(D) = G(D)^t D = K^{1/2} \left\{ \sum_S c(S|d) H_S \right\} K^{-1/2} D. \quad (3.9)$$

Si noti come $G(d)$ sia una matrice $q \times q$ di funzioni in $d = (d_1, \dots, d_q)^t$, e $\hat{\delta}(D)$ sia uno stimatore di δ basato su D_n . Si ricordi come $D_n = \hat{\delta}_{esteso}$ e tenda a $D \sim N_q(\delta, K)$ (vedi (3.4)).

Teorema 1 *Sotto l'ipotesi che le funzioni peso $c(S|d)$ sommino ad 1 per ogni z e che abbiano un numero al più infinito numerabile di punti di discontinuità, $\sqrt{n}(\hat{\mu} - \mu_{vero})$ sotto le assunzioni della (3.1) tende in distribuzione a*

$$\Lambda = \sum_S c(S|D) \Lambda_S = \left(\frac{\partial \mu}{\partial \theta} \right)^t i_{00}^{-1} M + \omega^t \{ \delta - \hat{\delta}(D) \}.$$

La media e la varianza sono rispettivamente

$$\omega^t \{ \delta - E[\hat{\delta}(D)] \}$$

e

$$\tau_0^2 + \omega^t var[\hat{\delta}(D)] \omega,$$

con errore quadratico medio $MSE[\Lambda] = \tau_0^2 + R(\delta)$, nella quale

$$R(\delta) = E[\omega^t \hat{\delta} - \omega^t \delta]^2 = \omega^t E[\{\hat{\delta}(D) - \delta\} \{\delta(D) - \delta\}] \omega,$$

dove:

$$\begin{aligned} - \tau_0^2 &= \left(\frac{\partial \mu}{\partial \theta} \right)^t i_{00}^{-1} \frac{\partial \mu}{\partial \theta}; \\ - \omega &= i_{10} i_{00}^{-1} \frac{\partial \mu}{\partial \theta} - \frac{\partial \mu}{\partial \gamma}. \end{aligned}$$

3.1.3 Differente parametrizzazione

Le ultime formule sono state viste basandosi sulla quantità D_n definita dalla (3.4) e sulla sua forma limite $D \sim N_q(\delta, K)$. È forse conveniente, almeno dal punto di vista interpretativo, ricondursi alla quantità

$$Z_n = \hat{K}^{-1/2} D_n = \hat{K}^{-1/2} \sqrt{n} (\hat{\gamma}_{esteso} - \gamma_0), \quad (3.10)$$

ed alla sua forma limite

$$Z = K^{-1/2} D \sim N_q(a, I),$$

dove $a = K^{-1/2} \delta$. Si noti, inoltre, come $Z_n \xrightarrow{d} N_q(a, I)$; con \hat{K} si è indicata una stima di K , consistente almeno sotto il modello nullo $\gamma = \gamma_0$.

Anche i pesi $c(S|D_n)$ dello stimatore in questione possono essere riscritti in funzione di Z_n invece che di D_n . Per ogni stimatore

$$\hat{\mu} = \sum_S \bar{c}(S|Z_n) \hat{\mu}_S$$

Infine, con questa notazione,

$$AIC_{n,S} = Z_n^t H_S Z_n - 2|S| + o_p(1).$$

3.1.4 Intervalli di confidenza

Si è già accennato nell'introduzione come il tradizionale uso della selezione del modello nelle procedure di inferenza conduca ad intervalli di confidenza troppo stretti, la cui probabilità di copertura non è quella dichiarata.

La teoria asintotica sviluppata da Hjort & Claeskens (2003) permette di definire intervalli di confidenza per μ con livello di copertura asintotico $1 - \alpha$ sotto tutti i modelli considerati.

Per fare questo si riparte dal metodo suggerito da Buckland, Burnham & Augustin (1997) e ripreso da Burnham & Anderson (2002, Capitolo 4), che prevede l'uso dei pesi $c(S|D_n)$ basati sul criterio AIC. Questo metodo porta ad intervalli del tipo $\hat{\mu} \pm u \widehat{se}_n$, dove u è l'appropriato quantile della normale, mentre \widehat{se}_n è lo scarto quadratico medio stimato attraverso la formula

$$\widehat{se}_n = \sum_S c(S|D_n) (\hat{\tau}_S^2/n + \hat{b}_S^2)^{1/2},$$

dove:

- $\hat{\tau}_S$ è uno stimatore consistente di $\tau_S = (\tau_0^2 + \omega^t K^{1/2} H_S) K^{1/2} \omega)^{1/2}$;
- $\hat{b}_S = \hat{\mu}_S - \hat{\mu}$.

L'obbiettivo è quindi studiare la probabilità di copertura $p_n = Pr\{-u \leq B_n \leq u\}$, dove

$$B_n = \frac{\hat{\mu} - \mu_{vero}}{\widehat{se}_n}.$$

La variabile B_n ha una distribuzione limite ben definita, dato che

- $\sqrt{n} \widehat{se}_n \xrightarrow{d} \widehat{se} = \sum_S c(S|D) \{\tau_S^2 + (\Lambda_S - \Lambda)^2\}^{1/2}$;
- $\sqrt{n}(\hat{\mu} - \mu_{vero}) \xrightarrow{d} \Lambda$.

Inoltre $\Lambda_S - \Lambda = \omega^t \{\hat{\delta}(D) - K^{1/2} H_S K^{-1/2} D\}$. Quindi

$$B_n \xrightarrow{d} B = \frac{\Lambda}{\widehat{se}} = \frac{\Lambda_0 + \omega^t \{\delta - \hat{\delta}(D)\}}{\sum_S c(S|D) \{\tau_S^2 + [\omega^t \{\hat{\delta}(D) - K^{1/2} H_S K^{-1/2} D\}]^2\}^{1/2}}$$

dove $\Lambda_0 = (\partial\mu/\partial\theta)^t i_{00}^{-1} M$.

La variabile B è una normale, data D , ma chiaramente non standardizzata quando pesata sulla distribuzione di D , e neppure centrata sullo 0, quindi la probabilità di copertura p_n è asintoticamente diversa dal livello nominale.

Per illustrare questo, poniamoci nel caso $q = 1$, in cui lo stimatore in questione è semplicemente

$$\hat{\mu} = \{1 - W(Z_n)\}\hat{\mu}_{ridotto} + W(Z_n)\hat{\mu}_{esteso},$$

per il quale:

- $\Lambda_\emptyset = \Lambda_0 + \omega K^{1/2}a;$
- $\Lambda_{esteso} = \Lambda_0 + \omega K^{1/2}(a - Z).$

Quindi in questo caso, $B = \delta/\widehat{se}$ assume la forma:

$$\frac{\Lambda_0 + \omega K^{1/2}\{a - W(Z)Z\}}{\{1 - W(Z)\}\{\tau_0^2 + \omega^2 K W(Z)^2 Z^2\}^{\frac{1}{2}} + W(Z)\{\tau_0^2 + \omega^2 K + \omega^2 K\{1 - W(Z)\}^2 Z^2\}^{\frac{1}{2}}}$$

dove $\Lambda_0 \sim N(0, \tau_0^2)$ indipendente da $Z \sim N(a, q)$. La probabilità di copertura asintotica può essere calcolata tramite integrazione con metodi numerici come $p(a) = \int Pr\{-u \leq B \leq u|z\}\phi(z - a) dz$.

Altrimenti, è possibile considerare le quantità:

$$\begin{aligned} inf_n &= \hat{\mu} - \hat{\omega}^t\{D_n - \hat{\delta}(D_n)\}/\sqrt{n} - u\hat{\kappa}/\sqrt{n}, \\ sup_n &= \hat{\mu} - \hat{\omega}^t\{D_n - \hat{\delta}(D_n)\}/\sqrt{n} + u\hat{\kappa}/\sqrt{n}, \end{aligned} \quad (3.11)$$

dove $\hat{\omega}$ e $\hat{\kappa}$ sono stimatori consistenti rispettivamente di ω e $\kappa = \tau_{esteso} = (\tau_0^2 + \omega^t K \omega)^{1/2}$ ed u come prima indica il quantile della normale. Si noti come la probabilità di copertura $p_n = Pr\{inf_n \leq \mu_{vero} \leq sup_n\}$ sia equivalente a $Pr\{-u \leq T_n \leq u\}$, dove

$$T_n = [\sqrt{n}(\hat{\mu} - \mu_{vero} - \hat{\omega}^t\{D_n - \hat{\delta}(D_n)\})]/\hat{\kappa}.$$

Ma c'è una simultanea convergenza in distribuzione:

$$(\sqrt{n}(\hat{\mu} - \mu_{vero}), D_n) \xrightarrow{d} (\Lambda_0 + \omega^t\{\delta - \hat{\delta}(D)\}, D).$$

Da questo segue che

$$T_n \xrightarrow{d} (\{\Lambda_0 - \omega^t(\delta - D)\}/\kappa) \sim N(0, 1).$$

3.2 **Distribuzione degli stimatori compromesso**

I risultati sulla distribuzione asintotica di Hjort & Claeskens (2003) non si applicano agli stimatori compromesso, presentati nel Paragrafo 2.1. Se per la loro costruzione sono state sfruttate le proprietà delle famiglie di posizione e scala, nulla è possibile dire riguardo la loro distribuzione asintotica. Quest'ultima, infatti, dipende non soltanto dal parametro di interesse, ma anche dalle specifiche distribuzioni sotto cui sono generati i dati. Formalmente, rimanendo, come in precedenza, nel caso di uno stimatore compromesso basato su due densità, il modello statistico diventa:

$$\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2,$$

in cui:

- \mathcal{F}_1 ha densità $f_1(y; \theta)$ con $\theta \in \Theta$;
- \mathcal{F}_2 ha densità $f_2(y; \theta)$ con $\theta \in \Theta$;

dove con Θ si intende lo spazio parametrico per θ , comune ai due modelli.

Lo spazio parametrico, quindi, è $\{1, 2\} \times \Theta$.

È evidente come i dati provengano da un elemento di \mathcal{F} , che, però, è ignoto. Finora non è nota alcuna teoria che permetta di ottenere intervalli di confidenza con livello di copertura approssimato prefissato sotto entrambe le famiglie.

Queste considerazioni, però, sono valide nel caso in cui le due distribuzioni non risultino annidate: in caso contrario, infatti, è possibile riportarsi alla teoria esposta nel paragrafo precedente per gli stimatori basati su criteri di selezione del modello e ricavare tramite essa la distribuzione desiderata.

Capitolo 4

Applicazioni

4.1 Stimatori basati su criteri di selezione del modello per la distribuzione Weibull

Si consideri $q = 1$ e uno stimatore del tipo

$$\hat{\mu} = \{1 - W(Z_n)\}\hat{\mu}_{ridotto} + W(Z_n)\hat{\mu}_{esteso},$$

come alla fine del Paragrafo 3.1. Si desidera applicare la teoria esposta in quel capitolo ad un esempio. Si considerino, in particolare, un campione casuale semplice da una distribuzione $Wei(\theta, \gamma)$, con γ prossimo a $\gamma_0 = 1$. È evidente come si abbia un modello esteso, quello Weibull, che contiene come modello ridotto un esponenziale con media θ .

Precisamente, siano y_1, \dots, y_n realizzazioni di una variabile casuale $Y \sim Wei(\theta, \gamma)$, con γ prossimo a $\gamma_0 = 1$. Le principali quantità di verosimiglianza risultano:

$$\begin{aligned} L(\gamma, \theta; y) &= \theta\gamma(\theta y)^{\gamma-1} \exp\{-(\theta y)^\gamma\}, \\ l(\gamma, \theta; y) &= \log\left(\theta\gamma(\theta y)^{\gamma-1} \exp\{-(\theta y)^\gamma\}\right) \\ &= \log(\theta) + \log(\gamma) + (\gamma - 1)\log(\theta) + (\gamma - 1)\log(y) - (\theta y)^\gamma, \\ l_\gamma &= \frac{1}{\gamma} + \log(\theta) + \log(y) - \log(\theta\gamma)(\theta\gamma)^\gamma, \end{aligned}$$

$$\begin{aligned}
l_\theta &= \frac{1}{\theta} + \frac{\gamma - 1}{\theta} - \gamma(\theta y)^{\gamma-1} y \\
&= \frac{\gamma}{\theta} - \gamma(\theta y)^{\gamma-1} y, \\
l_{\gamma\gamma} &= -\frac{1}{\gamma^2} - \log^2(\theta y)(\theta y)^\gamma, \\
l_{\theta\theta} &= -\frac{1}{\theta^2} - \frac{\gamma - 1}{\theta^2} - \gamma(\gamma - 1)(\theta y)^{\gamma-2} y^2 \\
&= -\frac{\gamma}{\theta^2} - \gamma(\gamma - 1)(\theta y)^{\gamma-2} y^2, \\
l_{\theta\gamma} &= \frac{1}{\theta} - (\theta y)^{\gamma-1} y - \gamma y(\theta y)^{\gamma-1} \log(\theta y).
\end{aligned}$$

Prima di procedere con il calcolo dell'informazione attesa, è utile ricordare che

$$Z = (\theta Y)^\gamma \sim \text{Exp}(1). \quad (4.1)$$

Inoltre, si richiamino le funzioni $\Gamma(\cdot)$ e derivate:

$$\begin{aligned}
\Gamma(\alpha) &= \int_0^{+\infty} x^{\alpha-1} e^{-x} dx, \\
\Gamma'(\alpha) &= \int_0^{+\infty} x^{\alpha-1} \log x e^{-x} dx, \\
\Gamma^{(n)}(\alpha) &= \int_0^{+\infty} x^{\alpha-1} (\log x)^n e^{-x} dx,
\end{aligned}$$

e le funzioni digamma, trigamma, ecc.

$$\begin{aligned}
\Psi(\alpha) &= \frac{d}{d\alpha} \log \Gamma(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}, \\
\Psi'(\alpha) &= \frac{d}{d\alpha} \Psi(\alpha) = \frac{\Gamma''(\alpha)\Gamma(\alpha) - (\Gamma'(\alpha))^2}{(\Gamma(\alpha))^2} = \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} - \Psi(\alpha)^2.
\end{aligned}$$

Proseguendo con il calcolo della matrice di informazione attesa,

$$\begin{aligned}
i_{\theta\theta} &= \frac{\gamma}{\theta^2} + \frac{\gamma(\gamma - 1)}{\theta^2} E[(\theta Y)^\gamma] \\
&= \frac{\gamma}{\theta^2} + \frac{\gamma^2}{\theta^2} - \frac{\gamma}{\theta^2} = \frac{\gamma^2}{\theta^2},
\end{aligned}$$

dove $E_{\theta,\gamma}[(\theta Y)^\gamma] = 1$ per la (4.1).

Il valore atteso della derivata seconda fatta due volte rispetto a γ , invece

$$\begin{aligned}
i_{\gamma\gamma} &= \frac{1}{\gamma^2} + \frac{1}{\gamma^2} E[\log^2(\theta Y)^\gamma (\theta Y)^\gamma] \\
&= \frac{1}{\gamma^2} (1 + E[Z \log^2 Z]),
\end{aligned}$$

con Z definita sempre dalla (4.1). In particolare,

$$E[Z \log^2 Z] = \int_0^{+\infty} z \log^2 z e^{-z} dz = \Gamma''(2),$$

dove, riprendendo le definizioni di funzioni digamma e trigamma,

$$\begin{aligned} \Gamma''(2) &= \Gamma(2) \left(\Psi'(2) + \Psi(2)^2 \right) \\ &= 1 \left(\Psi'(2) + \Psi(2)^2 \right) \\ &= \Psi'(1) - 1 + \Psi(2)^2. \end{aligned}$$

Ritornando a $i_{\gamma\gamma}$:

$$\begin{aligned} i_{\gamma\gamma} &= \frac{1}{\gamma^2} \left(1 + \Psi'(1) - 1 + \Psi(2)^2 \right) \\ &= \frac{1}{\gamma^2} \left(\Psi'(1) + \Psi(2)^2 \right) \\ &= \frac{1}{\gamma^2} \left(\Psi'(1) + (1 + \Psi(1))^2 \right) \\ &= \frac{1}{\gamma^2} \left(\frac{\pi^2}{6} + (1 - r)^2 \right) \quad \text{con } -r = \Psi(1) \approx -0.5772. \end{aligned}$$

Infine,

$$i_{\theta\gamma} = \frac{1}{\theta} \left(1 + \Psi(1) \right) = \frac{1 - r}{\theta}.$$

Per cui, la matrice di informazione attesa cercata è

$$i = \begin{pmatrix} \frac{\gamma^2}{\theta^2} & \frac{(1 - r)}{\theta} \\ \frac{(1 - r)}{\theta} & \frac{c^2}{\gamma^2} \end{pmatrix},$$

dove $c = \pi^2/6 + (1 - r)^2$.

L'inversa della matrice, è, invece,

$$i^{-1} = \frac{1}{\pi^2/6} \begin{pmatrix} \frac{c^2\theta^2}{\gamma^2} & -(1 - r)\theta \\ -(1 - r)\theta & \theta^2 \end{pmatrix}.$$

Si ipotizzi ora di essere interessati a stimare la mediana

$$\mu = \frac{(\log 2)^{1/\gamma}}{\theta}. \quad (4.2)$$

Lo stimatore compromesso, quindi, prende la forma

$$\begin{aligned}\hat{\mu} &= \{1 - W(Z_n)\}\hat{\mu}_{rid} + W(Z_n)\hat{\mu}_{est} \\ &= \{1 - W(Z_n)\}\frac{\log 2}{\hat{\theta}_{rid}} + W(Z_n)\frac{(\log 2)^{1/\hat{\gamma}_{est}}}{\hat{\theta}_{est}},\end{aligned}$$

dove $Z_n = \frac{\sqrt{n}(\hat{\gamma}_{full} - 1)}{\hat{K}^{1/2}}$, facilmente ricavabile dalle formule (3.2), (3.4) e (3.10). In questo caso,

$$Z_n = \hat{K}^{-1/2}D_n = \frac{\sqrt{n}(\hat{\gamma}_{full} - 1)}{\hat{K}^{1/2}},$$

con

$$D_n = \hat{\delta}_{full} = (\hat{\gamma}_{full} - \gamma_0)\sqrt{n} = (\hat{\gamma}_{full} - 1)\sqrt{n}$$

e \hat{K} stima di $K = \hat{i}^{11} = \frac{6\gamma^2}{\pi^2}$.

Per calcolare la distribuzione asintotica, è necessario ricavare anche le quantità ω e τ_0 . La prima, partendo dalla (3.6), risulta

$$\begin{aligned}\omega &= i_{10}i_{00}^{-1}\frac{\partial\mu}{\partial\theta} - \frac{\partial\mu}{\partial\gamma} \\ &= \frac{1-r}{\theta}\frac{\theta^2}{\gamma^2}(-1)\frac{(\log 2)^{1/\gamma}}{\theta^2} - \frac{(\log 2)^{1/\gamma}\log(\log 2)}{\theta}\left(-\frac{1}{\gamma^2}\right) \\ &= \frac{(\log 2)^{1/\gamma}}{\gamma^2\theta}\left(-(1-r) + \log(\log 2)\right),\end{aligned}$$

che, per $\gamma = 1$, diventa

$$\omega = \frac{\log 2}{\theta}\left(-(1-r) + \log(\log 2)\right).$$

Per quanto riguarda τ_0 , invece, si ha

$$\begin{aligned}\tau_0^2 &= \left(\frac{\partial\mu}{\partial\theta}\right)^t i_{00}^{-1} \frac{\partial\mu}{\partial\theta} \\ &= (-1)\frac{(\log 2)^{1/\gamma}}{\theta^2}\frac{\theta^2}{\gamma^2}(-1)\frac{(\log 2)^{1/\gamma}}{\theta^2} \\ &= \frac{(\log 2)^{2/\gamma}}{\theta^2\gamma^2},\end{aligned}$$

da cui

$$\tau_0 = \frac{(\log 2)^{1/\gamma}}{\theta^\gamma},$$

che, per $\gamma = 1$, diventa

$$\tau_0 = \frac{\log 2}{\theta}.$$

Con queste quantità, è possibile calcolare la distribuzione asintotica di μ , che, per il Teorema 1 risulta

$$\Lambda = \{1 - W(Z_n)\}\Lambda_{rid} + W(Z_n)\Lambda_{est}.$$

Non rimane che calcolare le distribuzioni Λ_{rid} e Λ_{est} , ripartendo dalla formula del Lemma 3,

$$\Lambda_S = \left(\frac{\partial \mu}{\partial \theta}\right)^T i_{00}^{-1} M + \omega(\delta - K^{1/2} H_S K^{-1/2} D).$$

Il primo addendo, oltre ad essere indipendente dal modello sotto cui è calcolata, è facilmente ricavabile, ricordando che $M \sim N_p(0; i_{00})$. Infatti

$$\left(\frac{\partial \mu}{\partial \theta}\right)^T i_{00}^{-1} N(0, i_{00})$$

è una variabile casuale che si distribuisce come una normale con media

$$\left(\frac{\partial \mu}{\partial \theta}\right)^T i_{00}^{-1} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$

e varianza

$$\left(\frac{\partial \mu}{\partial \theta}\right)^T i_{00}^{-2} \left(\frac{\partial \mu}{\partial \theta}\right) i_{00} = \left(\frac{\partial \mu}{\partial \theta}\right)^T i_{00}^{-2} \left(\frac{\partial \mu}{\partial \theta}\right) i_{00} = \tau_0^2;$$

quindi,

$$\left(\frac{\partial \mu}{\partial \theta}\right)^T i_{00}^{-1} M = \Lambda_0 \sim N(0, \tau_0^2).$$

Viceversa, il secondo addendo cambia a seconda del modello sotto cui è calcolato. In particolare, sotto il modello Weibull, dato che $H_S = 1$, e che $D \sim N(\delta, K)$,

$$\Lambda_{est} = \Lambda_0 + \omega(\delta - K^{1/2} H_S K^{-1/2} D) = \omega(\delta - K^{1/2} Z),$$

con $Z \sim N\left(\frac{\delta}{K^{1/2}}, \frac{K}{(K^{1/2})^2}\right) \sim N(a, I)$, posto $a = \frac{\delta}{K^{1/2}}$. Raccogliendo $K^{1/2}$,

$$\Lambda_{est} = \Lambda_0 + \omega K^{1/2}(a - Z).$$

Sotto il modello esponenziale, invece, considerato che $H_S = 0$, la distribuzione è semplicemente

$$\Lambda_{rid} = \Lambda_0 + \omega K^{1/2}a.$$

In definitiva,

$$\begin{aligned}\Lambda_{rid} &= \Lambda_0 + \omega\delta, \\ \Lambda_{est} &= \Lambda_0 + \omega K^{1/2}(\delta K^{-1/2} - Z),\end{aligned}$$

da cui

$$\begin{aligned}\Lambda &= \{1 - W(Z)\}(\Lambda_0 + \omega\delta) + W(Z)(\Lambda_0 + \omega K^{1/2}(\delta K^{-1/2} - Z)) \\ &= \Lambda_0 + \omega K^{1/2}\{a - W(Z)Z\},\end{aligned}$$

dove $\Lambda_0 \sim N(0, \tau_0^2)$ e $Z \sim N(a, 1)$ sono variabili casuali indipendenti.

Il peso $W(Z)$ rappresenta la versione asintotica dei pesi $W(Z_n)$: esso vale quindi 0 quando si è sotto il modello ridotto, 1 sotto il modello completo. Nel momento in cui sia necessario utilizzare i pesi campionari, viceversa, si utilizza la formula (2.15):

$$\frac{\exp(\frac{1}{2}AIC_i)}{\sum_{S' \in \mathcal{S}} \exp(\frac{1}{2}AIC_i)},$$

che, ricordando la (3.7), è uguale a

$$\frac{\exp(\frac{1}{2}D_n^t K^{-1/2} H_S K^{-1/2} D_n - |S|)}{\sum_{S' \in \mathcal{S}} \exp(\frac{1}{2}D_n^t K^{-1/2} H'_S K^{-1/2} D_n - |S'|)} + o_p(1),$$

dove con \mathcal{S} si intende l'insieme di tutti i modelli utilizzati.

4.1.1 Simulazione della distribuzione

Per verificare empiricamente questi risultati, e per analizzare la velocità di convergenza, si può procedere con uno studio di simulazione, utilizzando un

ambiente statistico, come R (per questa tesi è stata utilizzata la versione 2.3.1).

In particolare, si generano realizzazioni pseudo-casuali da una distribuzione Weibull per analizzare ciò che accade sotto il modello esteso (in questo esempio, si è fissato $\gamma = 1.10$ e $\theta = 5$) o da una distribuzione esponenziale per un'analisi sotto il modello ridotto (qui si è considerato il tasso di guasto pari a 5). Si procede quindi con un'analisi che porti ad avere una distribuzione dello stimatore $\hat{\mu}$, confrontabile con quella teorica (Λ).

Innanzitutto, si proceda con un campione di numerosità ampia, in maniera da osservare il comportamento asintotico. Con un campione di 250 osservazioni, si ottengono i grafici presentati in Figura 4.1 (modello esteso) ed in Figura 4.2 (modello ridotto), nei quali la linea tratteggiata indica la distribuzione teorica della variabile Λ , mentre quella continua rappresenta la distribuzione empirica di $\hat{\mu}$.

Come si può notare, sotto il modello ridotto la distribuzione limite si discosta in maniera veramente minima da quella esatta, stimata attraverso le simulazioni. Viceversa, sotto il modello esteso vi è una qualche differenza, soprattutto al centro delle distribuzioni.

Sicuramente più interessante, in ogni caso, è osservare il comportamento in presenza di un numero inferiore di osservazioni. I grafici nelle figure 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8 riportano per varie numerosità (rispettivamente 10, 20, 50, sia sotto il modello esteso, sia sotto quello ridotto), le distribuzioni teoriche (linea tratteggiata) e le distribuzioni simulate (linea continua). Sotto il modello esteso, inoltre, è riportata la distribuzione teorica in cui i pesi asintotici $W(Z)$ sono sostituiti dai pesi campionari ($W(Z_n)$). Nei grafici sotto il modello ridotto, viceversa, le distribuzioni coi pesi campionari non sono riportate, dacché non si scostano in maniera significativa da quelle coi pesi asintotici.

Sotto il modello ridotto, la differenza tra la distribuzione ottenuta dalle simulazioni (10000 ripetizioni) non si discosta eccessivamente dalla distribuzione asintotica prevista dalla teoria già con numerosità non elevate. In

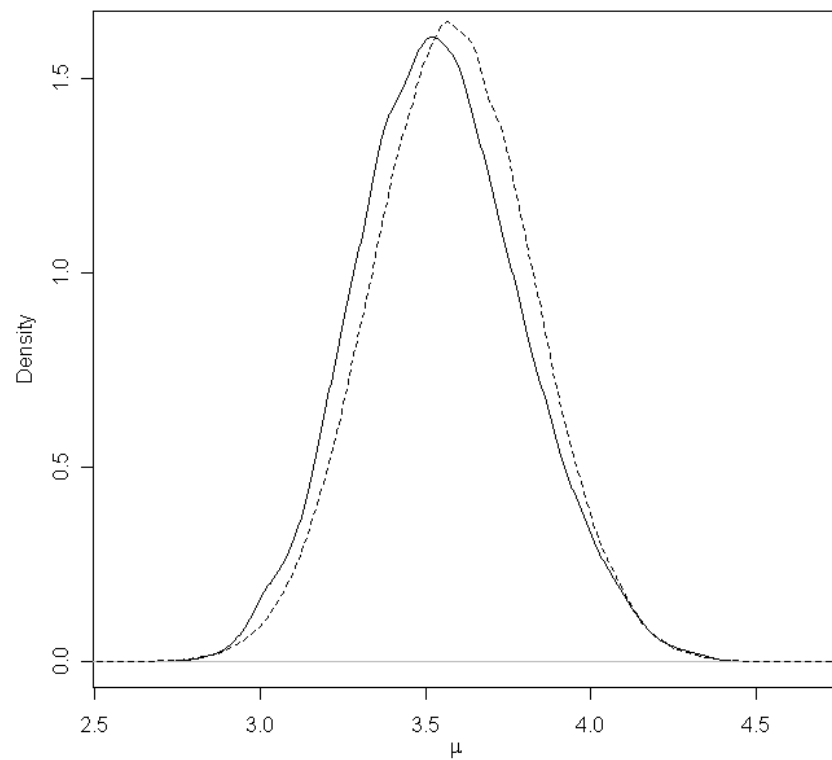


Figura 4.1: Confronto tra la distribuzione asintotica (linea tratteggiata) di $\hat{\mu}$, e la sua distribuzione simulata (linea continua) calcolata per 10000 repliche con campioni di numerosità 250, sotto il modello esteso $(Wei(1.10, 5))$.

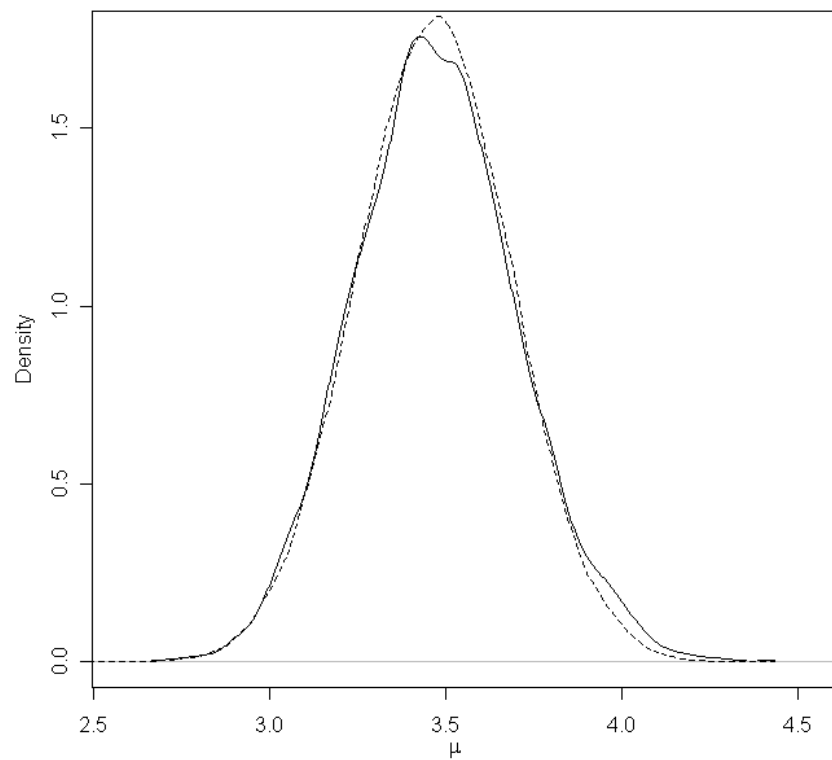


Figura 4.2: Confronto tra la distribuzione asintotica (linea tratteggiata) di $\hat{\mu}$, e la sua distribuzione simulata (linea continua) calcolata per 10000 repliche con campioni di numerosità 250, sotto il modello ridotto (esponenziale con tasso di guasto pari a 5).

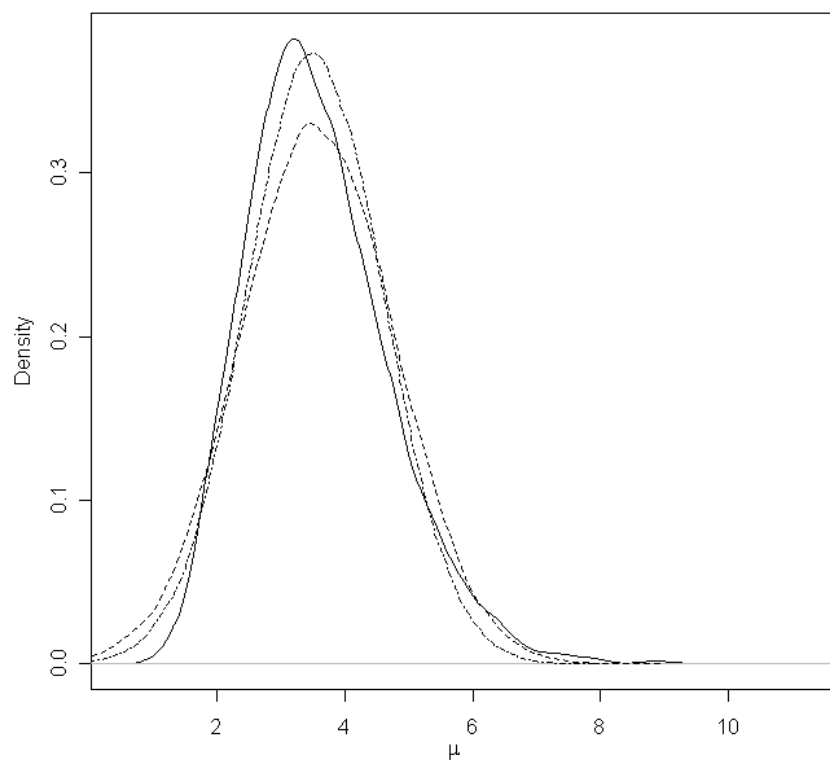


Figura 4.3: Confronto tra la distribuzione asintotica di $\hat{\mu}$, con pesi asintotici (linea tratteggiata) e pesi campionari (linea punto linea), e la sua distribuzione simulata (linea continua) calcolata per 10000 replicazioni con campioni di numerosità 10, sotto il modello esteso $(Wei(1.10, 5))$.

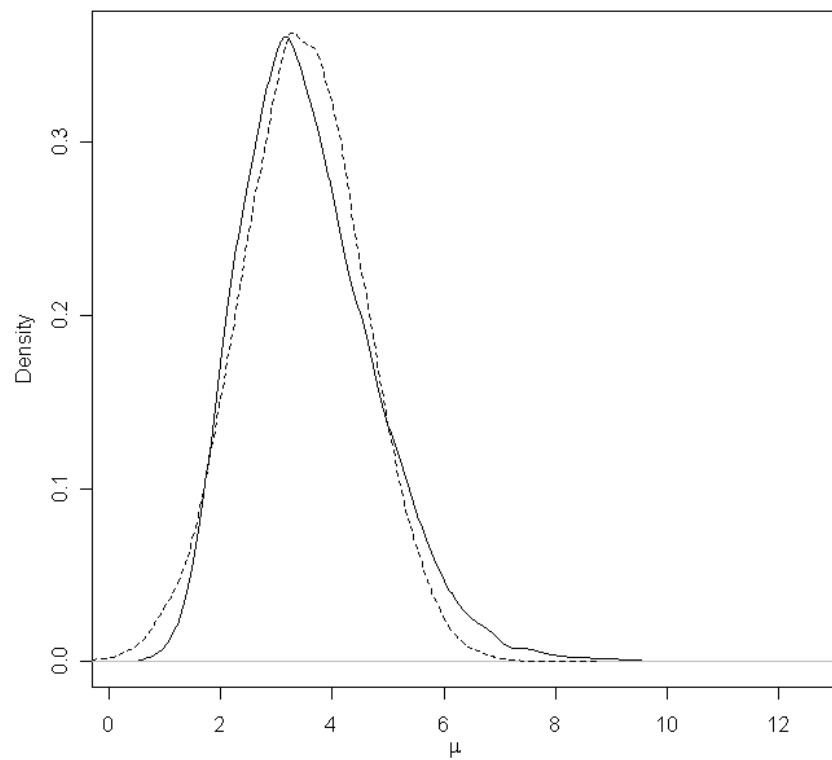


Figura 4.4: Confronto tra la distribuzione asintotica (linea tratteggiata) di $\hat{\mu}$, e la sua distribuzione simulata (linea continua) calcolata per 10000 repliche con campioni di numerosità 10, sotto il modello ridotto (esponenziale con tasso di guasto pari a 5).

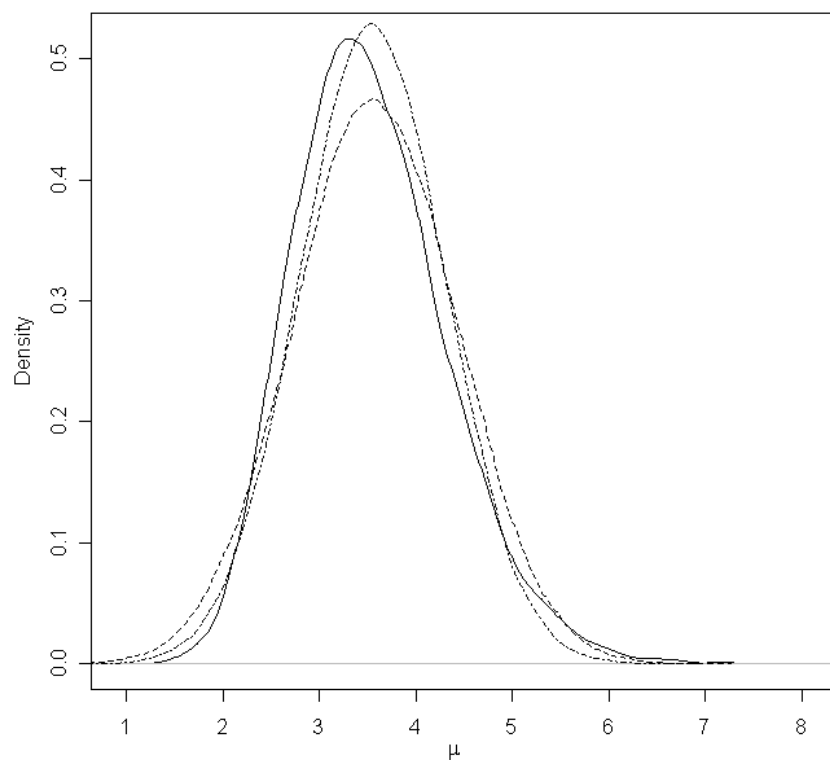


Figura 4.5: Confronto tra la distribuzione asintotica di $\hat{\mu}$, con pesi asintotici (linea tratteggiata) e pesi campionari (linea punto linea), e la sua distribuzione simulata (linea continua) calcolata per 10000 repliche con campioni di numerosità 20, sotto il modello esteso ($Wei(1.10, 5)$).

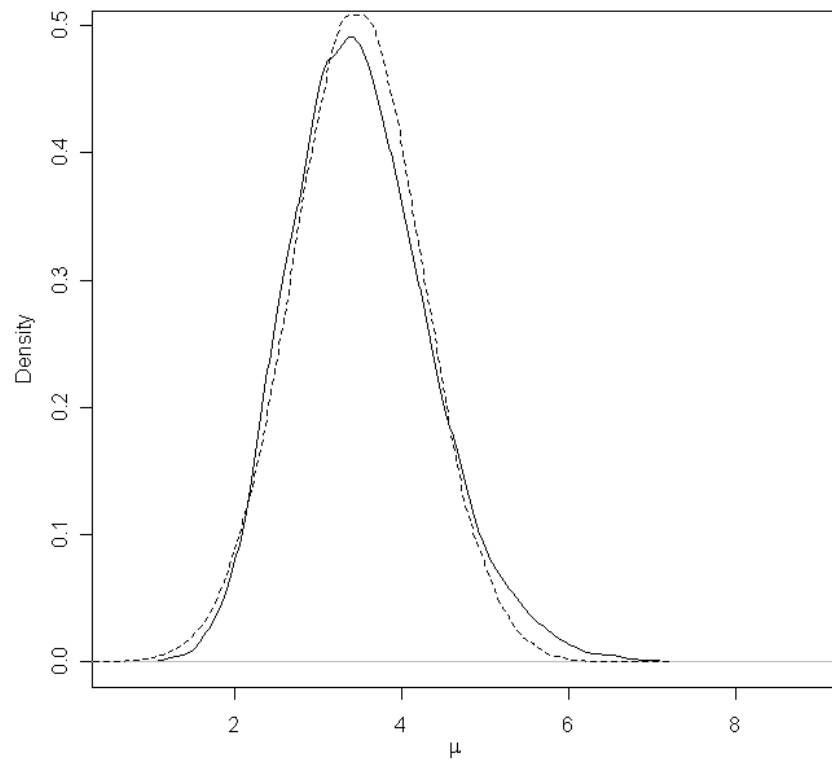


Figura 4.6: Confronto tra la distribuzione asintotica (linea tratteggiata) di $\hat{\mu}$, e la sua distribuzione simulata (linea continua) calcolata per 10000 repliche con campioni di numerosità 20, sotto il modello ridotto (esponenziale con tasso di guasto pari a 5).

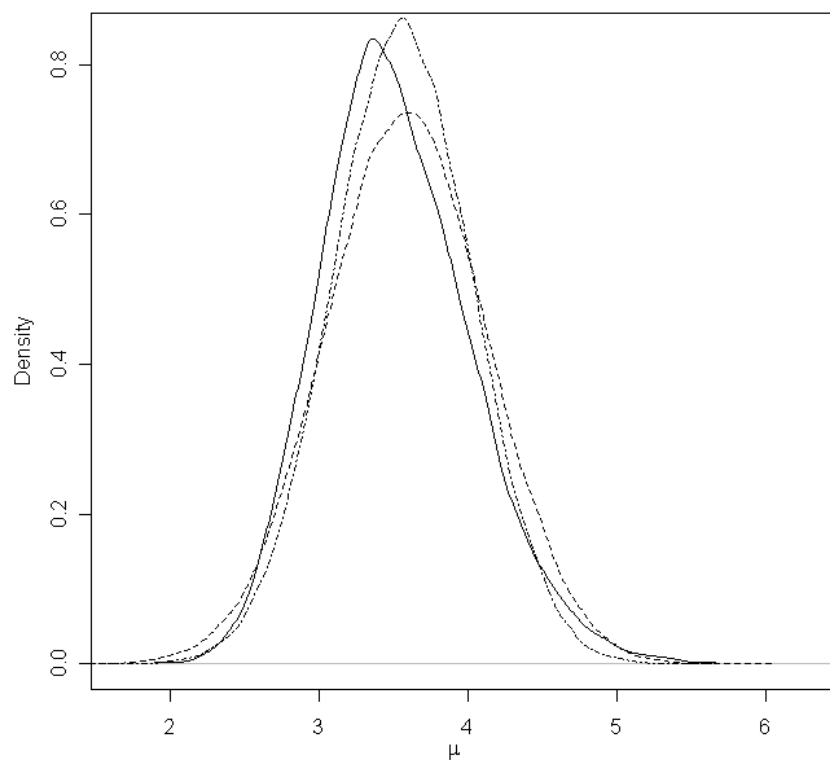


Figura 4.7: Confronto tra la distribuzione asintotica di $\hat{\mu}$, con pesi asintotici (linea tratteggiata) e pesi campionari (linea punto linea), e la sua distribuzione simulata (linea continua) calcolata per 10000 repliche con campioni di numerosità 50, sotto il modello esteso ($Wei(1.10, 5)$).

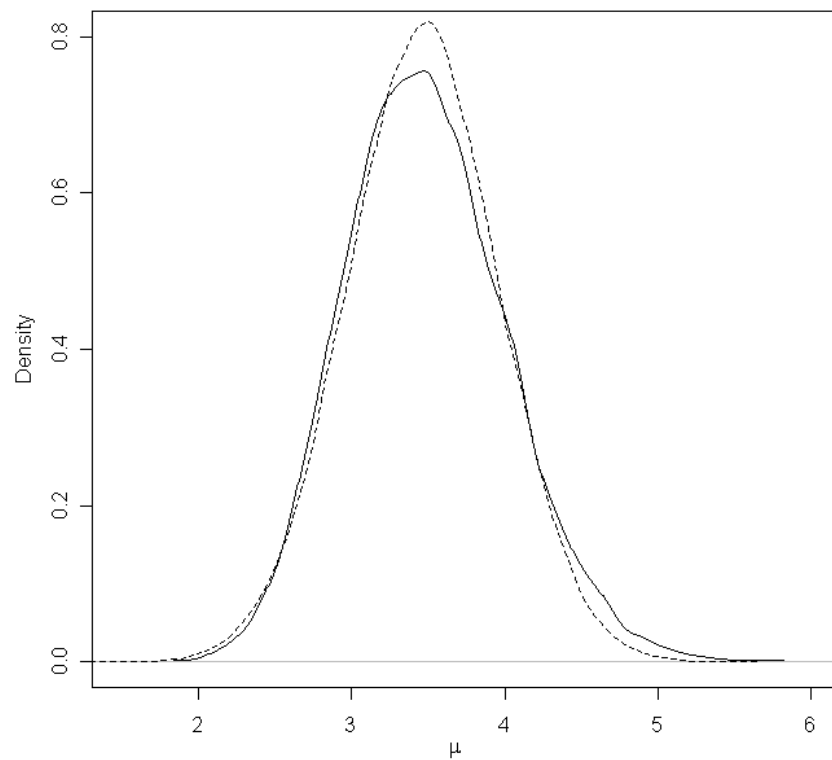


Figura 4.8: Confronto tra la distribuzione asintotica (linea tratteggiata) di $\hat{\mu}$, e la sua distribuzione simulata (linea continua) calcolata per 10000 repliche con campioni di numerosità 50, sotto il modello ridotto (esponenziale con tasso di guasto pari a 5).

realtà qualche scostamento eccessivo vi è, soprattutto nelle code, come si può osservare nei grafici di Figura 4.4, 4.6 e 4.8). Più in dettaglio, i valori della funzione di ripartizione, nominali ed effettivi, riportati in Tabella 4.1 ci confermano questa tendenza. Comunque, si può osservare che, all'aumentare della numerosità campionaria, gli scostamenti tra distribuzione asintotica ed empirica tendono a divenire sempre minori, fino, per campioni di 250 osservazioni, a divenire sufficientemente piccoli. Al centro della distribuzione, viceversa, i valori della funzione di ripartizione, nominali ed effettivi, risultano vicini già a numerosità campionarie basse.

Sotto il modello esteso, invece, la differenza tra distribuzione asintotica e distribuzione empirica è maggiore. Molta di questa differenza è dovuta alla lentezza dei pesi a raggiungere il livello asintotico, come dimostra la molto maggiore vicinanza della distribuzione simulata alla distribuzione teorica in cui i pesi siano calcolati su base campionaria (si vedano le Figure 4.3, 4.5, 4.7 e la Tabella 4.2). Con numerosità più elevata (Figura 4.1 e Tabella 4.3), la distanza tra distribuzione asintotica e distribuzione empirica cala leggermente, anche se persiste una differenza non trascurabile. In tutti i grafici delle distribuzioni calcolate sotto il modello esteso, comunque, le differenze più grosse sono al centro delle distribuzioni stesse, ossia l'esatto contrario di ciò che accadeva sotto il modello ridotto. Questo perché la moda delle distribuzioni asintotiche (sia con pesi asintotici, sia con pesi campionari) tende ad essere spostata leggermente a destra rispetto a quella della distribuzione empirica.

	numerosità campionaria			
percentile nominale	n=10	n=20	n=50	n=250
0.05	0.0239	0.0371	0.0465	0.0535
0.10	0.0787	0.0939	0.1023	0.1050
0.25	0.2491	0.2566	0.2628	0.2564
0.50	0.5109	0.5070	0.5048	0.5005
0.75	0.7277	0.7356	0.7381	0.7397
0.90	0.8552	0.8724	0.8776	0.8840
0.95	0.9082	0.9226	0.9266	0.9346

Tabella 4.1: *Valori effettivi della funzione di ripartizione delle distribuzioni dello stimatore compromesso basato su criteri di selezione del modello sotto il modello ridotto.*

valori funzione di ripartizione per n=10			
$\hat{\mu}$	empirica	pesi asint	pesi camp
1	0.00050	0.01645	0.00838
2	0.05460	0.09478	0.07316
3	0.32630	0.31610	0.30699
4	0.68230	0.63568	0.66703
5	0.89140	0.87924	0.91513
6	0.97080	0.97699	0.98976
7	0.99250	0.99765	0.99954

valori funzione di ripartizione per n=20			
$\hat{\mu}$	empirica	pesi asint	pesi camp
2	0.01010	0.03237	0.01946
2.5	0.07920	0.10389	0.08221
3	0.26180	0.24815	0.23313
3.5	0.51610	0.46473	0.47394
4	0.74120	0.68933	0.72493
4.5	0.88360	0.85874	0.89644
5	0.95500	0.95207	0.97332

valori funzione di ripartizione per n=50			
$\hat{\mu}$	empirica	pesi asint	pesi camp
2.2	0.00090	0.00518	0.00183
2.6	0.02040	0.03370	0.01970
3	0.14530	0.13984	0.11389
3.5	0.51690	0.43823	0.44892
4	0.83250	0.77953	0.82628
4.4	0.94870	0.93352	0.96371
4.8	0.98710	0.98823	0.99617

Tabella 4.2: Valori della funzione di ripartizione (empirica, asintotica, teorica con pesi campionari) sotto il modello esteso per vari valori dello stimatore basato su criteri di selezione del modello.

valori funzione di ripartizione per n=250		
$\hat{\mu}$	empirica	pesi asint
3	0.00980	0.00809
3.1	0.03290	0.02309
3.2	0.07800	0.05746
3.5	0.44270	0.36825
3.8	0.84650	0.81505
3.9	0.91740	0.90474
4	0.95950	0.95728

Tabella 4.3: *Valori della funzione di ripartizione (empirica ed asintotica) sotto il modello esteso per vari valori dello stimatore basato su criteri di selezione del modello.*

4.1.2 Simulazione della copertura

Un aspetto particolarmente interessante da verificare tramite simulazione è il livello di copertura effettivo degli intervalli di confidenza basati sugli stimatori compromesso. Si è già data in precedenza una formulazione esplicita per gli estremi degli intervalli di confidenza, attraverso le formule (3.11). Sostituendo all'interno di esse le quantità trovate nell'esempio, ed utilizzando come quantile $u = 1.645$, si ottengono intervalli di confidenza con livello nominale 0.90.

Come nel paragrafo precedente, è stato condotto uno studio di simulazione con diverse numerosità campionarie, sia sotto il modello ridotto, sia sotto quello esteso.

Si prenda prima in esame il modello ridotto, che, avendo un solo parametro, è più semplice da trattare. Infatti, è sufficiente generare un campione da un'esponenziale, il cui parametro vari a seconda dei diversi valori della mediana μ di interesse. Dalla (4.2) è immediato ricavare che il tasso di guasto dell'esponenziale è $\frac{\log 2}{\mu}$. Si procede quindi con il generare campioni per vari valori di μ (nei grafici presentati, si sono scelti 200 diversi valori tra 0.1 ed 8), e verificare la copertura effettiva.

Come si può verificare dai grafici (Figure 4.9, 4.10, 4.11 e 4.12), il livello di copertura effettivo non si discosta molto da quello nominale, nemmeno con numerosità piccole. Se con numerosità campionaria pari a 10 od a 20, effettivamente, il livello di copertura effettivo è leggermente inferiore a quello nominale, già con un campione di 50 elementi questa differenza cala decisamente, fino a praticamente annullarsi con una numerosità campionaria pari a 100.

Un altro metodo per costruire intervalli di confidenza, alternativo a quello asintotico, consiste nell'utilizzare una procedura **bootstrap**. Esso prevede di sfruttare la 'distribuzione bootstrap' dello stimatore per individuare il limite inferiore e quello superiore dell'intervallo di confidenza stesso ($\inf = \hat{\mu}_{(R+1)0.05}$, $\sup = \hat{\mu}_{(R+1)0.05}$, dove con R si intende il numero di ricampionamenti).

Anche in questo caso si può verificare il livello di copertura, procedendo

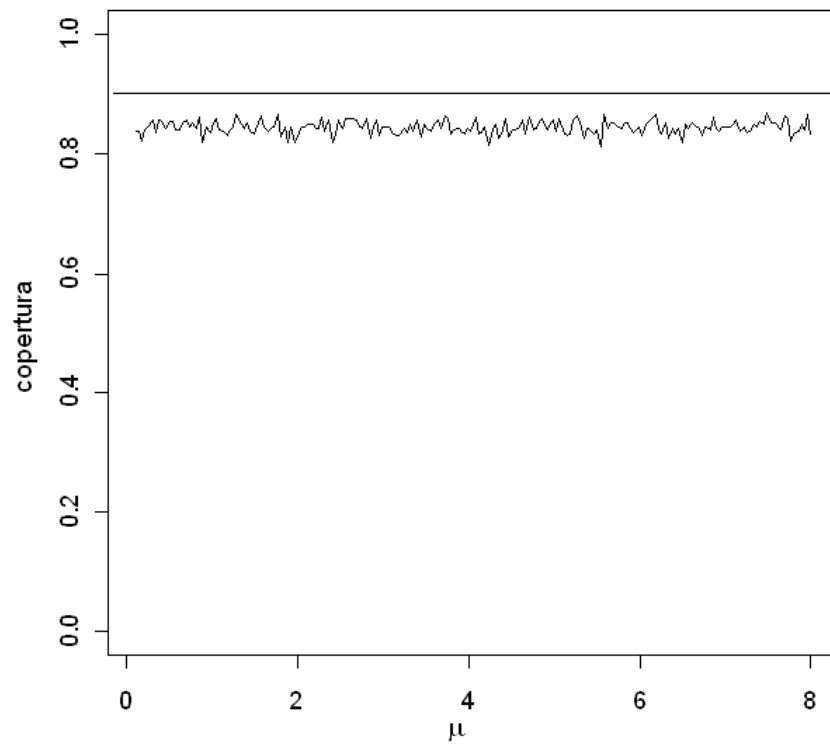


Figura 4.9: *Copertura simulata degli intervalli di confidenza calcolati tramite metodo asintotico, sotto il modello ridotto (esponenziale con tasso di guasto pari a 5) su 10000 ripetizioni di campioni di numerosità 10.*

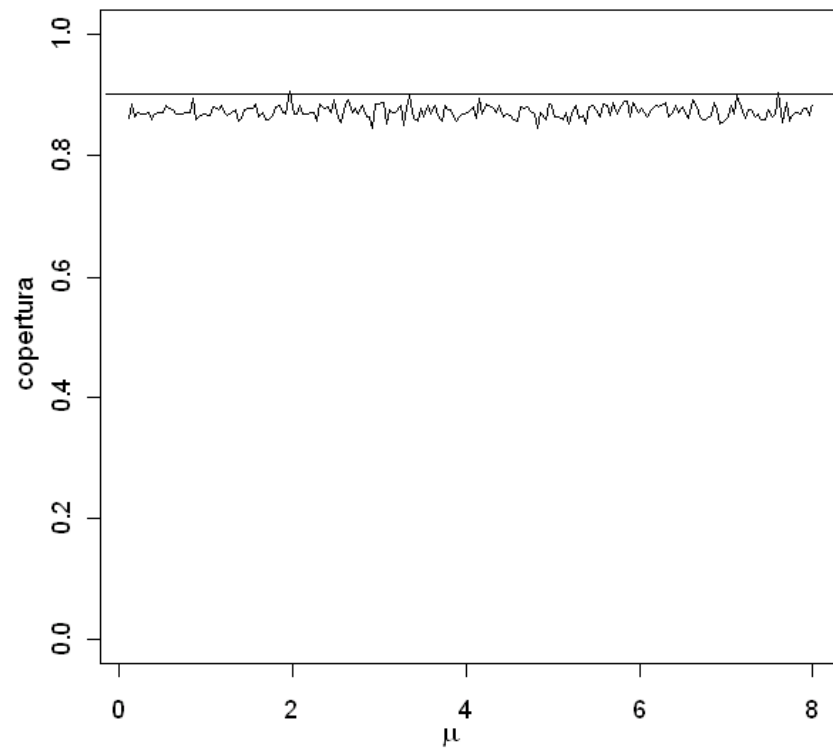


Figura 4.10: *Copertura simulata degli intervalli di confidenza calcolati tramite metodo asintotico, sotto il modello ridotto (esponenziale con tasso di guasto pari a 5) su 10000 ripetizioni di campioni di numerosità 20.*

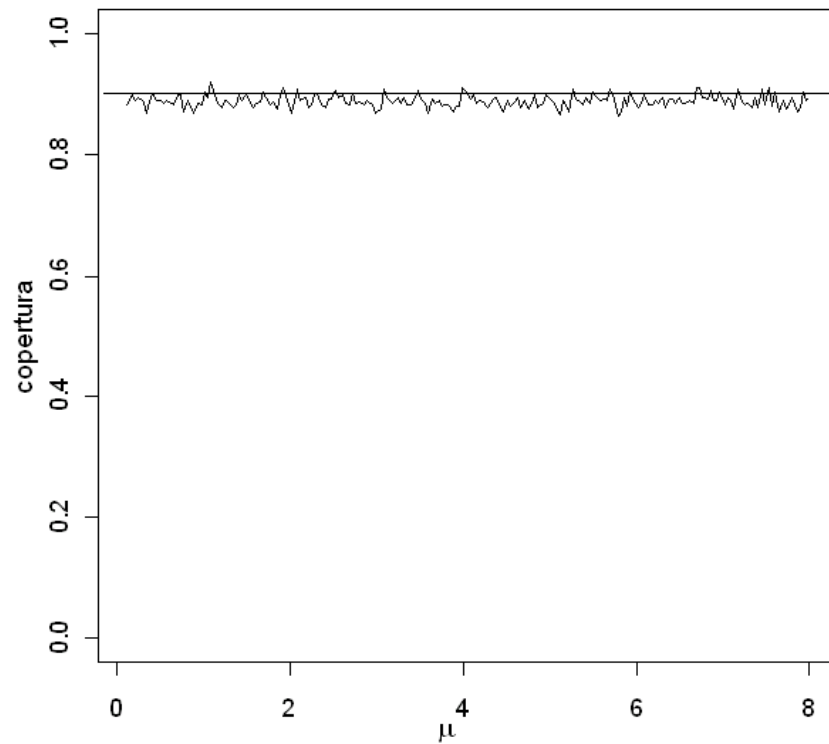


Figura 4.11: Copertura simulata degli intervalli di confidenza calcolati tramite metodo asintotico, sotto il modello ridotto (esponenziale con tasso di guasto pari a 5) su 10000 ripetizioni di campioni di numerosità 50.

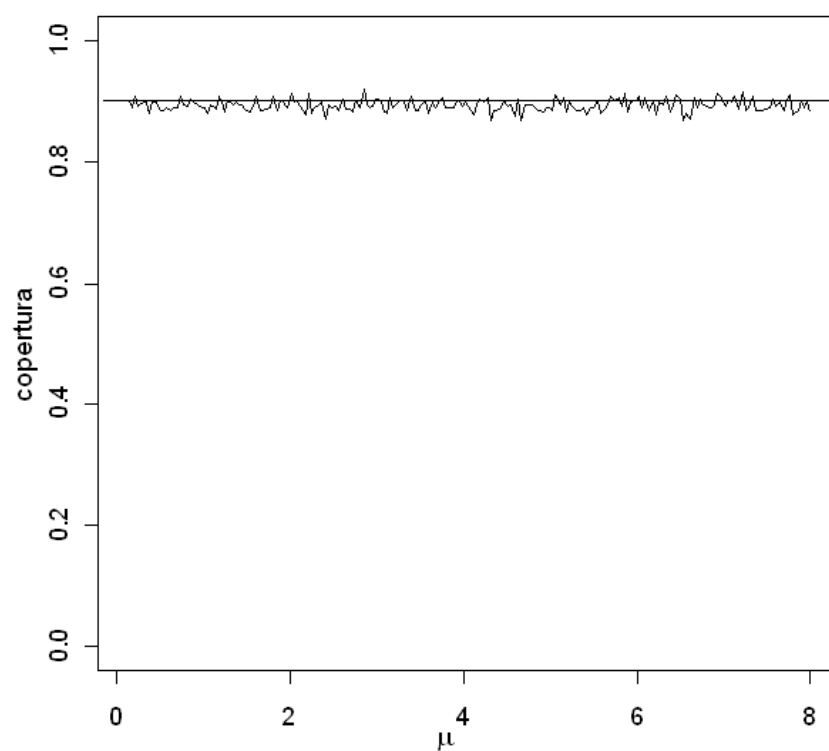


Figura 4.12: *Copertura simulata degli intervalli di confidenza calcolati tramite metodo asintotico, sotto il modello ridotto (esponenziale con tasso di guasto pari a 5) su 10000 ripetizioni di campioni di numerosità 100.*

nello stesso modo utilizzato in precedenza, ossia ripetendo varie prove, e calcolando la ‘percentuale di successi’.

I grafici ottenuti (Figure 4.13, 4.14 e 4.15), però, mostrano come con questo metodo si ottengano livelli di copertura effettivi abbastanza inferiori a quelli nominali. Il grafico in Figura 4.13, in particolare, mostra come con una numerosità campionaria pari a 10 elementi, il livello di copertura effettivo non superi mai lo 0.80, con un livello nominale di 0.90. Nemmeno con numerosità campionarie maggiori (rispettivamente 20 e 50 per le Figure 4.14 e 4.15) si ottengono risultati soddisfacenti, anche se per questi due grafici è stato necessario ridurre, a causa dell’onere computazionale, il numero di prove eseguite.

Effettivamente, comunque, i ricampionamenti bootstrap presentati nei grafici precedenti sono stati ottenuti con 100 ricampionamenti, sempre a causa dell’onere computazionale. Davison & Hinkley (1997, Capitolo 2) indicano 1000 come numero minimo di ricampionamenti per avere buoni risultati: limitando il numero di punti in cui calcolare la copertura, e limitando le ripetizioni, nel grafico presentato in Figura 4.16 si è provato ad utilizzare 1000 ricampionamenti, senza avere risultati significativi.

In realtà, il metodo bootstrap porta ad ottenere intervalli di confidenza con livelli tendenzialmente più bassi di quelli nominali. Davison & Hinkley (1997, Capitolo 2) suggeriscono un metodo che prevede il calcolo degli intervalli basandosi non più semplicemente sui dati del campione, ma sui dati studentizzati. Anche questo metodo, tuttavia, non si rivela soddisfacente, come si può notare nel grafico di Figura 4.17, in cui il livello effettivo di copertura, nonostante la studentizzazione, rimane eccessivamente basso.

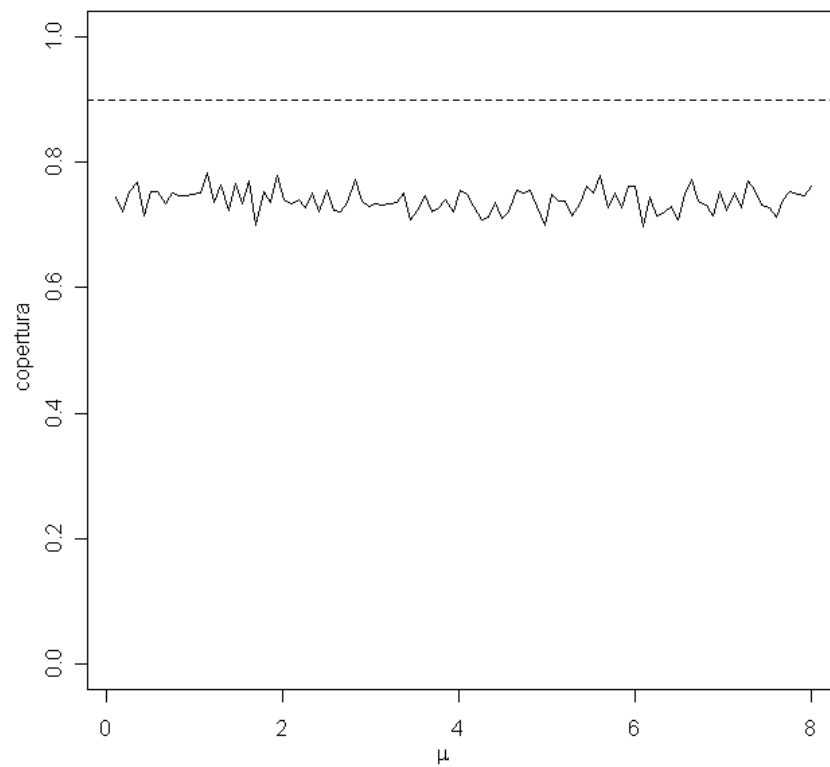


Figura 4.13: *Copertura simulata degli intervalli di confidenza calcolati tramite metodo bootstrap (100 ricampionamenti), sotto il modello ridotto (esponenziale con tasso di guasto pari a 5) su 500 ripetizioni di campioni di numerosità 10.*

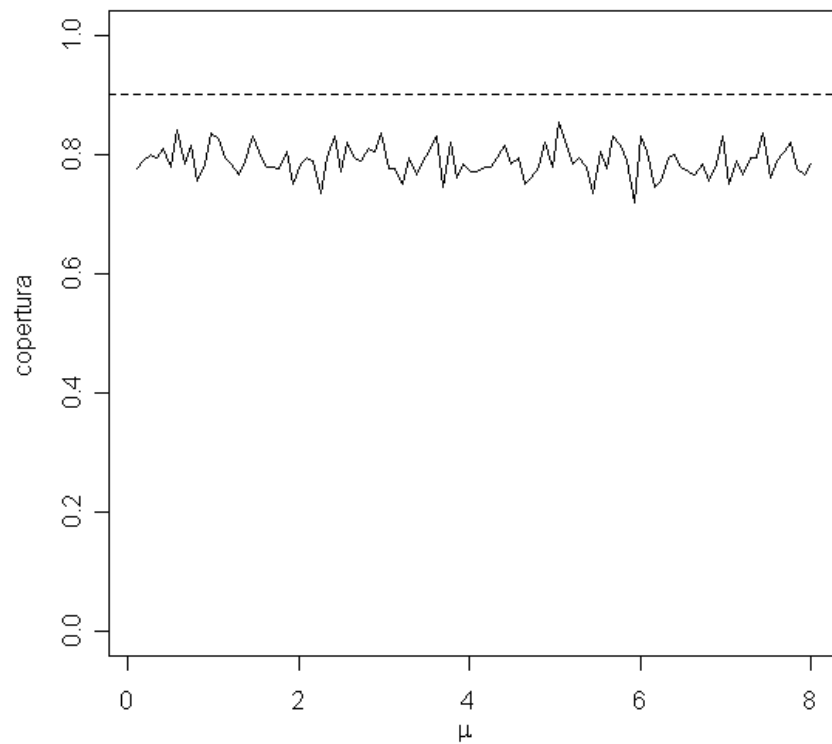


Figura 4.14: *Copertura simulata degli intervalli di confidenza calcolati tramite metodo bootstrap (100 ricampionamenti), sotto il modello ridotto (esponenziale con tasso di guasto pari a 5) su 200 ripetizioni di campioni di numerosità 20.*

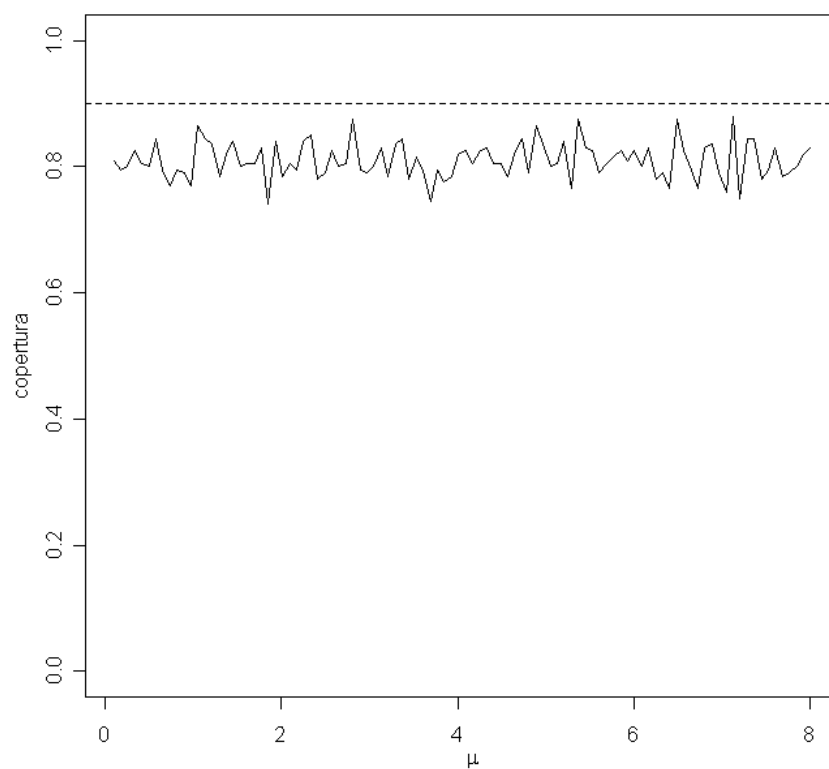


Figura 4.15: *Copertura simulata degli intervalli di confidenza calcolati tramite metodo bootstrap (100 ricampionamenti), sotto il modello ridotto (esponenziale con tasso di guasto pari a 5) su 200 ripetizioni di campioni di numerosità 50.*

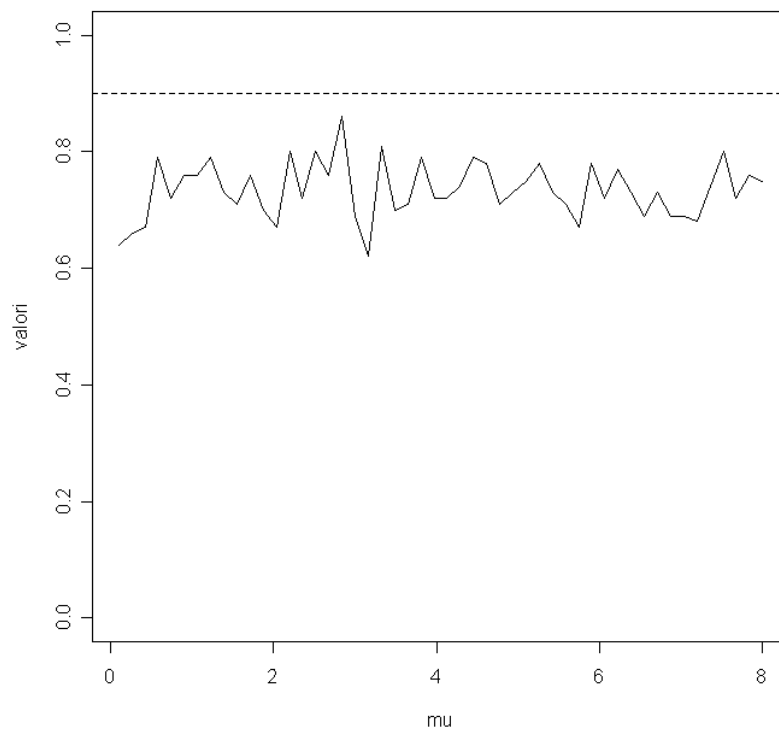


Figura 4.16: Copertura simulata degli intervalli di confidenza calcolati tramite metodo bootstrap (1000 ricampionamenti), sotto il modello ridotto (esponenziale con tasso di guasto pari a 5) su 100 ripetizioni di campioni di numerosità 10.

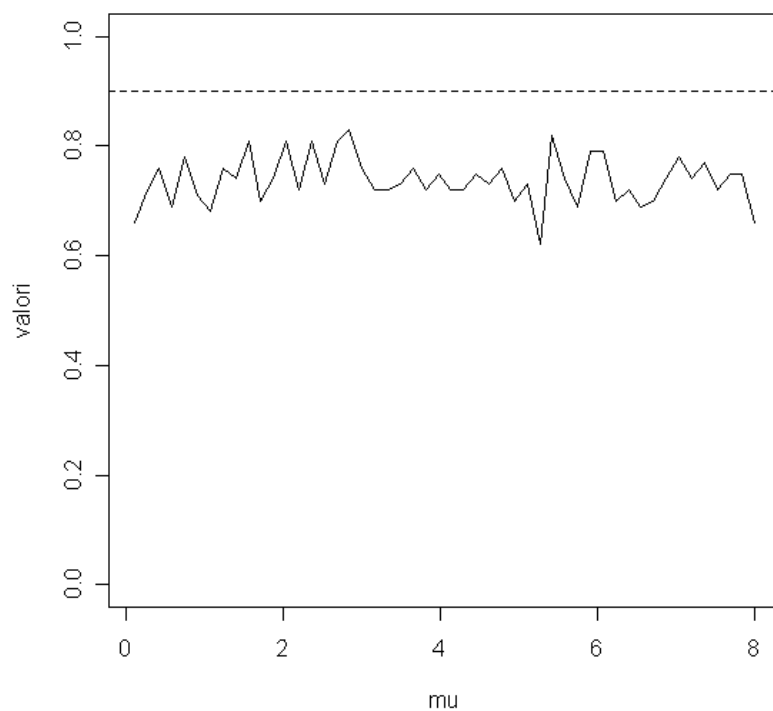


Figura 4.17: *Copertura simulata degli intervalli di confidenza calcolati tramite metodo bootstrap (100 ricampionamenti con campioni studentizzati), sotto il modello ridotto (esponenziale con tasso di guasto pari a 5) su 100 ripetizioni di campioni di numerosità 100.*

Ritornando agli intervalli di confidenza asintotici, si consideri ora la situazione sotto il modello esteso, tenendo conto che in questo caso, diversamente che sotto il modello ridotto, è necessario procedere con più cautela. Lo stimatore (4.2) della mediana, infatti, dipende in questo caso da due parametri, ed è quindi necessario fissarne uno per poi procedere per diversi valori dell'altro, in maniera da ottenere anche in questo caso vari valori di μ per cui verificare la copertura.

Si fissi inizialmente $\gamma = 1.10$, e si proceda simulando campioni casuali da una Weibull, col parametro di scala non fissato pari a $\frac{\log(2)^{(1/1.10)}}{\mu}$. I grafici che risultano, per differenti numerosità campionarie, sono quelli presentati nelle Figure 4.18, 4.19 e 4.20.

Il risultato è molto simile a quello ottenuto sotto il modello ridotto, con la copertura effettiva molto prossima a quella nominale. Anche in questo caso la differenza rintracciabile con una numerosità campionaria minima (Figura 4.18) cala significativamente all'aumentare della dimensione del campione, fino a divenire praticamente nulla con numerosità campionaria pari a 50 (Figura 4.20).

Si analizzi ora, invece, il comportamento se a rimanere fisso è il parametro θ (in questa simulazione è stato posto uguale a 0.2), ossia quando a variare è γ , che diventa pari a $(\log(\log(2)))/\log(0.2\mu)$.

Anche facendo variare γ , si ottengono risultati molto simili ai precedenti, per cui sono valide le considerazioni fatte in precedenza. La particolarità di questa situazione, invece, è la leggera diminuzione della copertura quando μ si avvicina allo 0 (Figure 4.21, 4.22 e 4.23). Questo è dovuto, probabilmente, al corrispondente andamento di γ , che, per valori di μ prossimi a 0, si avvicina anch'esso a 0, rendendo la distribuzione Weibull sempre più prossima ad una distribuzione 'degenere' in 0.

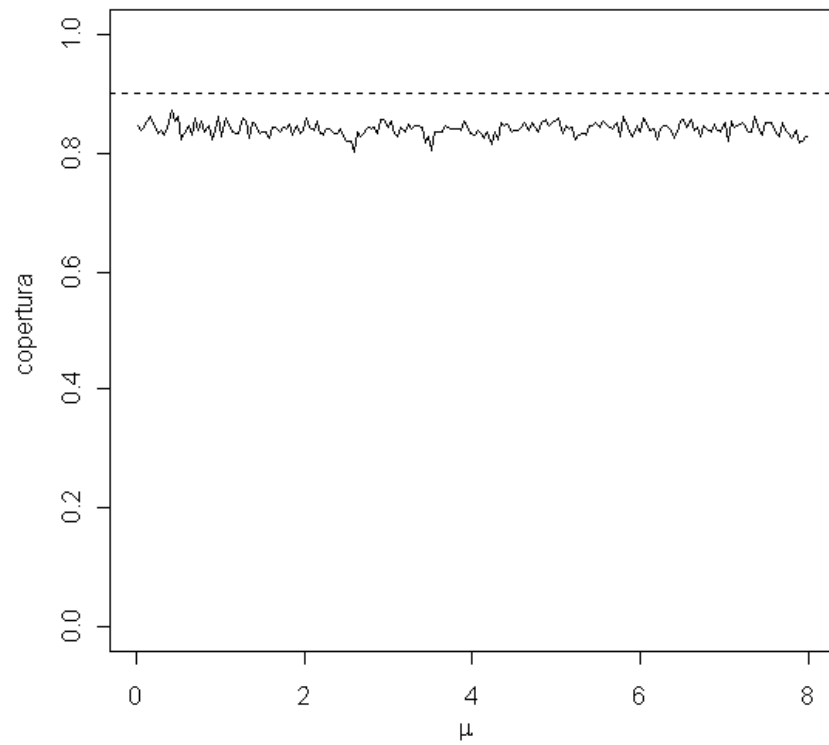


Figura 4.18: Copertura simulata degli intervalli di confidenza calcolati tramite metodo asintotico, sotto il modello esteso ($Wei(1.10, \frac{\log(2)^{(1/1.10)}}{\mu})$) su 10000 ripetizioni di campioni di numerosità 10.

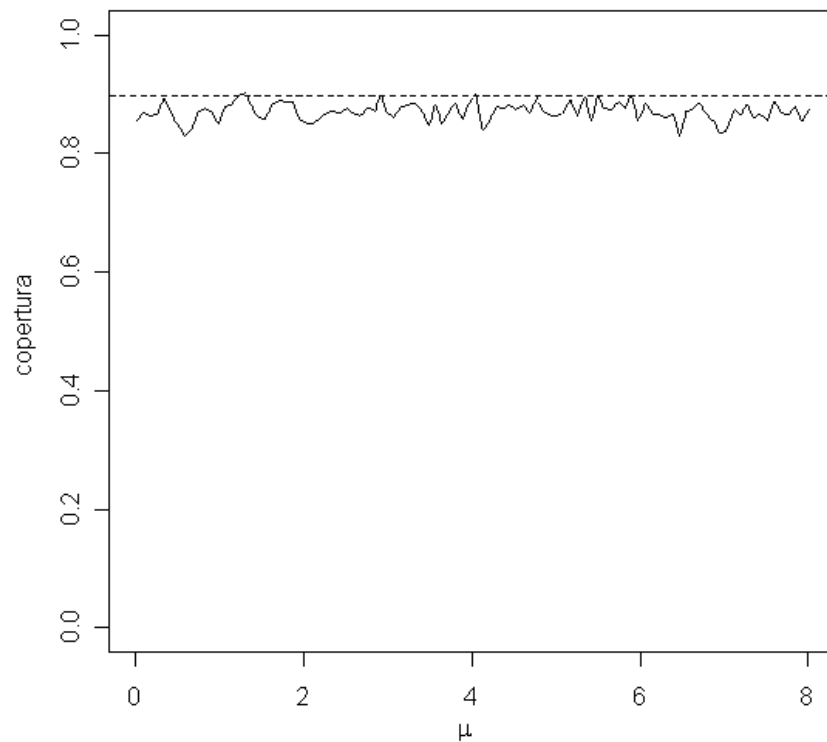


Figura 4.19: Copertura simulata degli intervalli di confidenza calcolati tramite metodo asintotico, sotto il modello esteso ($Wei(1.10, \frac{\log(2)^{(1/1.10)}}{\mu})$) su 10000 ripetizioni di campioni di numerosità 20.

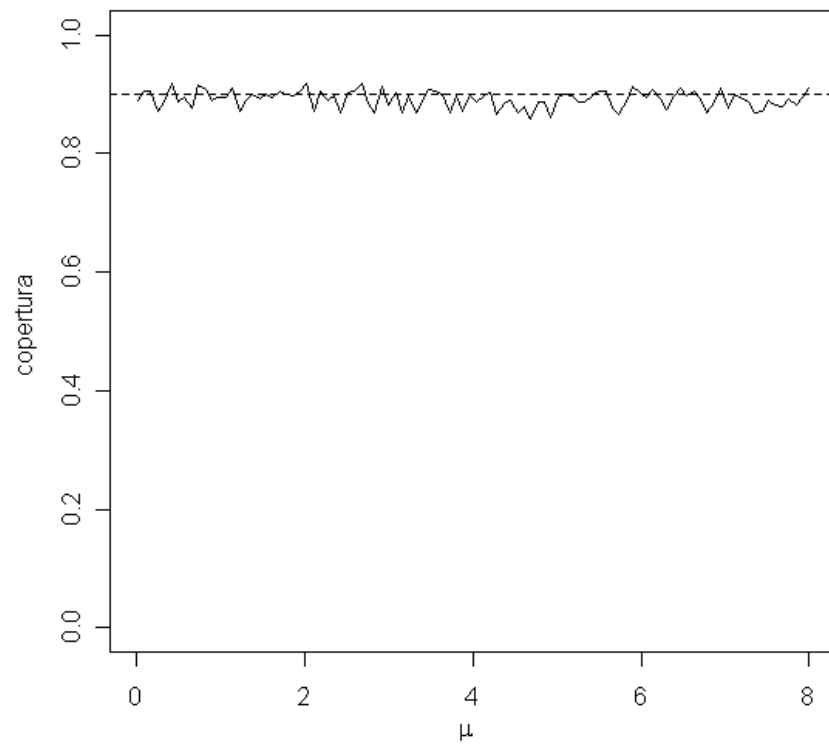


Figura 4.20: Copertura simulata degli intervalli di confidenza calcolati tramite metodo asintotico, sotto il modello esteso ($Wei(1.10, \frac{\log(2)^{(1/1.10)}}{\mu})$) su 10000 ripetizioni di campioni di numerosità 50.

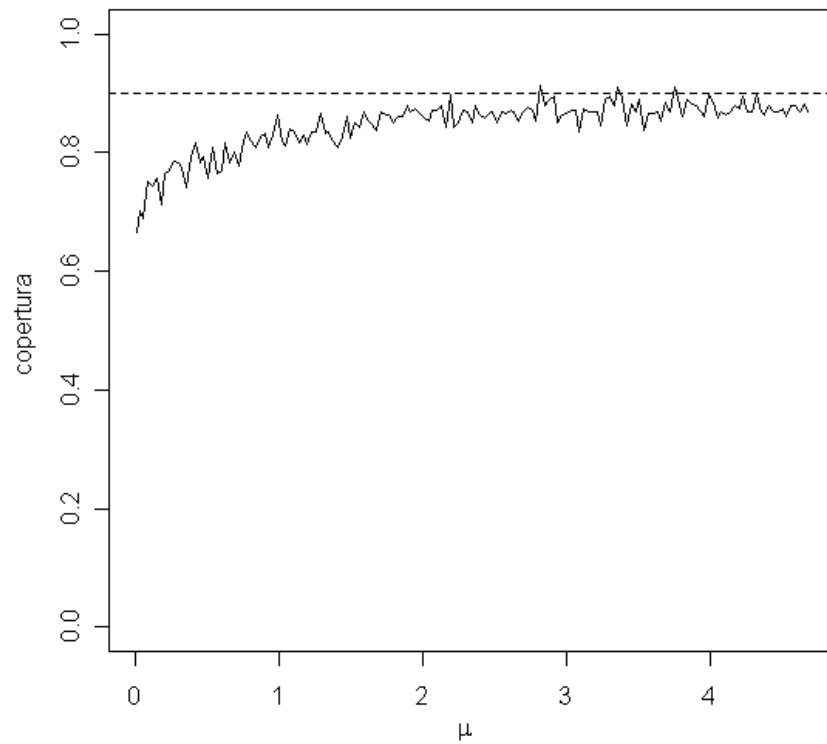


Figura 4.21: Copertura simulata degli intervalli di confidenza calcolati tramite metodo asintotico, sotto il modello esteso $(Wei((\log(\log(2))/\log(0.2\mu)), 5))$ su 10000 ripetizioni di campioni di numerosità 10.

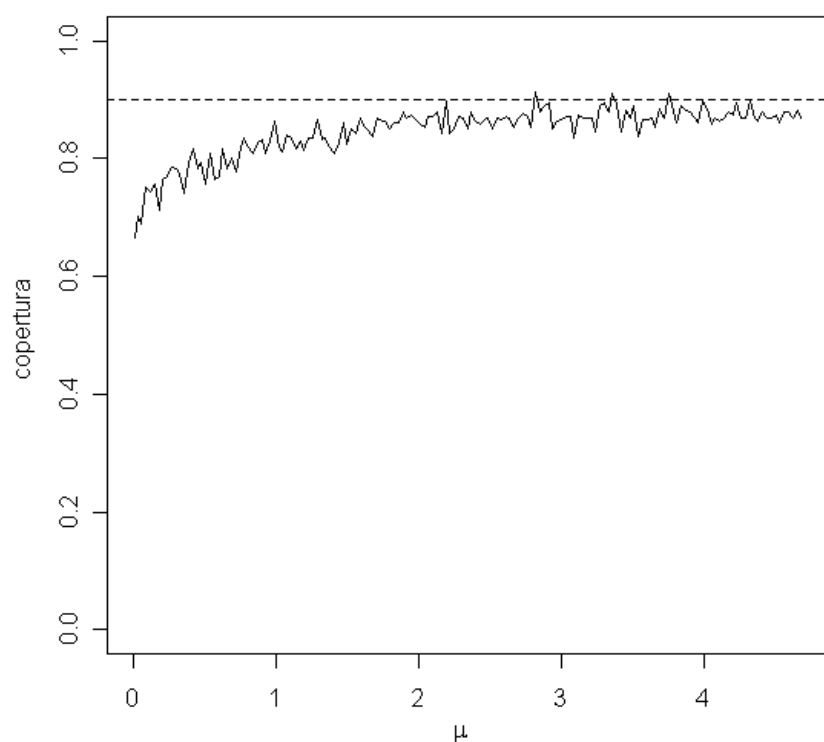


Figura 4.22: *Copertura simulata degli intervalli di confidenza calcolati tramite metodo asintotico, sotto il modello esteso $(Wei((\log(\log(2))/\log(0.2\mu)), 5))$ su 10000 ripetizioni di campioni di numerosità 20.*

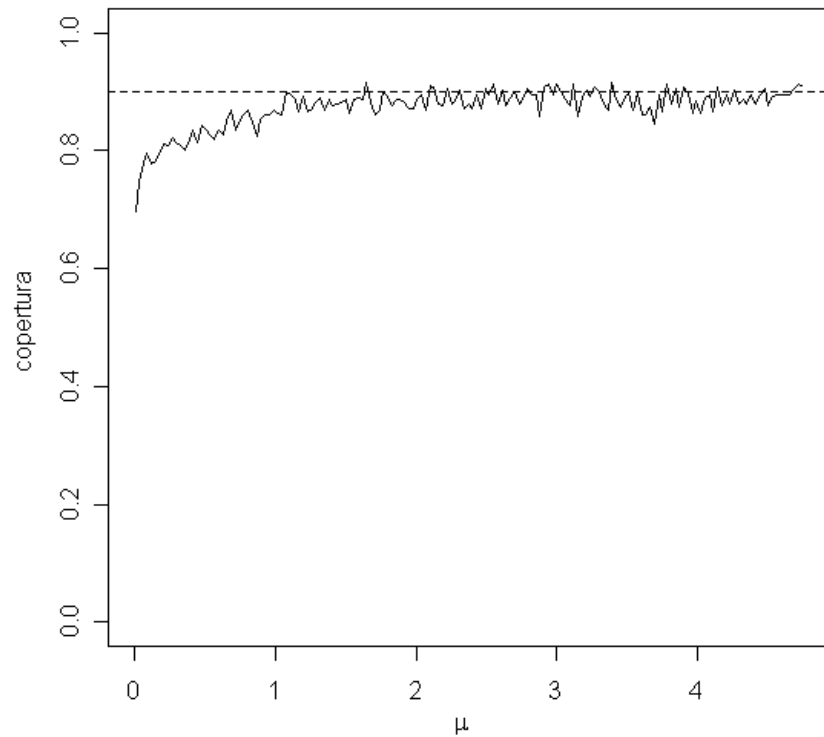


Figura 4.23: Copertura simulata degli intervalli di confidenza calcolati tramite metodo asintotico, sotto il modello esteso $(Wei((\log(\log(2))/\log(0.2\mu)), 5))$ su 10000 ripetizioni di campioni di numerosità 50.

4.2 Stimatori per famiglie di posizione e scala

4.2.1 Stimatore basato sulla normale e normale di ordine 4

Si desidera ora applicare la teoria esposta nel Paragrafo 2.1: si prenda un campione casuale semplice generato da una distribuzione ignota.

Si pensi, inizialmente, di stimare il parametro di posizione facendo riferimento a solo due distribuzioni, come presentato nella teoria. Nell'esempio, si prendano la distribuzione normale e la distribuzione normale di ordine 4 (EP(4)). Easton (1991) suggerisce che il miglior stimatore tra quelli proposti, è quello che è stato definito come 'stimatore SCMV stile biefficace' (si veda (2.10) e (2.11)). Esso si presenta nella forma

$$T_{bief}(y) \approx \frac{\tilde{\alpha}'_1(y)\hat{\mu}_{f_1} + \tilde{\alpha}'_2(y)\hat{\mu}_{f_2}}{\tilde{\alpha}'_1(y) + \tilde{\alpha}'_2(y)},$$

dove

$$\tilde{\alpha}'_i(y) = \left[\frac{\exp(nM_{f_i}(y))}{\hat{\sigma}_{f_i}^3 |det(H_{f_i}(y))|^{1/2} var_{f_i}[T_{f_i}(Y)]} \right]^{1/2}.$$

Per applicare la formula, quindi, è necessario conoscere, sotto entrambe le distribuzioni, le quantità:

- $\hat{\mu}$ e $\hat{\sigma}$, stime di massima verosimiglianza rispettivamente del parametro di posizione e scala;
- $\exp(nM_{f_i}(y))$, che non è altro che il valore della verosimiglianza in $(\hat{\mu}, \hat{\sigma})$;
- $|det(H_{f_i}(y))|^{1/2}$, la radice quadrata del determinante dell'informazione osservata;
- $var_{f_i}[T_{f_i}(Y)]$, la varianza dello stimatore di massima verosimiglianza per il parametro od una sua approssimazione.

Sotto la distribuzione normale, queste quantità sono note ed hanno una formulazione esplicita:

- $\hat{\mu} = \frac{\sum_{i=1}^n y_i}{n}$ e $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$;
- $\exp(nM_{f_i}(y)) = L(\hat{\mu}, \hat{\sigma}) = -n(1 + \log(2\pi))/2 - n \log(\hat{\sigma})$;
- $|det(H_{f_i}(y))|^{1/2} = \sqrt{2}/\hat{\sigma}^2$;
- $var_{f_i}[T_{f_i}(Y)] = 1/n$.

Per la distribuzione normale di ordine 4, invece, le corrispondenti quantità possono essere trovate per via numerica: sfruttando la procedura **nlm** dell'ambiente R, è agevole ricavare le stime di massima verosimiglianza, la matrice dell'informazione osservata e, attraverso esse, le entità necessarie (vedi Appendice). Ottenute queste quantità, sia per una distribuzione, sia per l'altra, è immediato ricavare i pesi, e quindi lo stimatore richiesto. Applicando quest'ultimo ad un campione pseudocasuale generato da una *t* di Student con 5 gradi di libertà, e ripetendo un numero sufficiente di volte l'operazione su campioni diversi, è possibile anche fornire una stima della distribuzione empirica dello stimatore stesso (vedi Figura 4.24).

4.2.2 Stimatore basato sulla normale e normale di ordine s

Se si volesse applicare a questo caso la teoria esposta nel Paragrafo 3.1, invece, è necessario trovare una famiglia di distribuzioni che includa come caso particolare un'altra. Volendo continuare ad usare la distribuzione normale, si può scegliere come modello esteso un modello con distribuzione normale di ordine s , con s libero: è evidente che per $s = 2$ si ritroverà il modello ridotto, ossia quello normale.

Si parta dalla funzione di densità della distribuzione normale di ordine s

$$f(y; \mu, \sigma, s) = \frac{1}{2s^{1/s}\Gamma(1 + 1/s)\sigma} \exp\left\{-\frac{|x - \mu|^s}{s\sigma^s}\right\},$$

dalla quale, grazie all'aiuto del ambiente R e del comando 'nlm', è possibile ricavare le stime di massima verosimiglianza per i tre parametri, unitamente

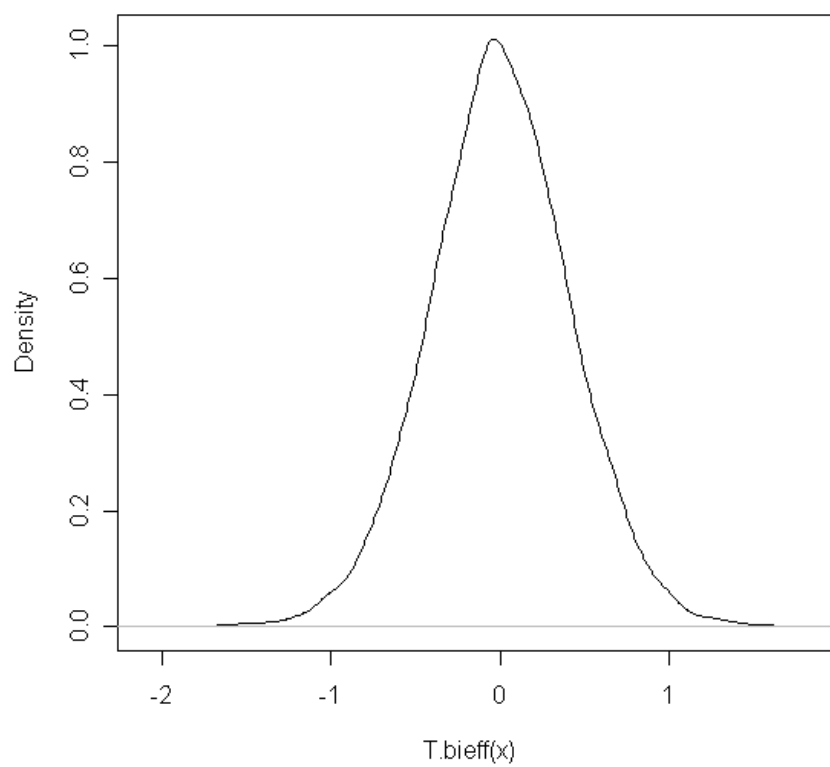


Figura 4.24: *Distribuzione simulata dello stimatore compromesso di massima verosimiglianza stile biefficace calcolata per 10000 repliche con campioni di numerosità 10 generati da una t di Student con 5 gradi di libertà.*

alla matrice di covarianza. In questo esempio, il parametro di disturbo θ è un parametro bidimensionale, in quanto $\theta = (\mu, \sigma)$, mentre il parametro d'interesse γ è il parametro s . Inoltre, dato che siamo interessati al parametro di posizione, lo stimatore cercato è semplicemente $\hat{\mu}$, da cui lo stimatore compromesso

$$\hat{\mu} = \{1 - W(Z_n)\}\hat{\mu}_{rid} + W(Z_n)\hat{\mu}_{est},$$

dove, anche in questo caso, $Z_n = \frac{\sqrt{n}(\hat{\gamma}_{full} - \gamma_0)}{\hat{K}^{1/2}}$, sempre ricavabile dalle formule (3.2), (3.4) e (3.10). In particolare, nell'esempio,

$$Z_n = \hat{K}^{-1/2}D_n = \frac{\sqrt{n}(\hat{\gamma}_{full} - 2)}{\hat{K}^{1/2}},$$

con

$$D_n = \hat{\delta}_{full} = (\hat{\gamma}_{full} - \gamma_0)\sqrt{n} = (\hat{\gamma}_{full} - 2)\sqrt{n}$$

e \hat{K} stima di $K = i^{11}$, ricavabile dalla matrice di covarianza ottenuta numericamente.

Fissato questo, è possibile andare alla ricerca delle quantità d'interesse per la costruzione della distribuzione asintotica dello stimatore stesso. In particolare,

$$\omega = i_{10}i_{00}^{-1}\frac{\partial\mu}{\partial\theta} - \frac{\partial\mu}{\partial\gamma} = \begin{pmatrix} i_{[1,1]} & i_{[1,2]} \end{pmatrix} \begin{pmatrix} i^{[1,1]} & i^{[1,2]} \\ i^{[2,1]} & i^{[2,2]} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} - 0,$$

e,

$$\tau_0^2 = \left(\frac{\partial\mu}{\partial\theta}\right)^t i_{00}^{-1} \frac{\partial\mu}{\partial\theta} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} i^{[1,1]} & i^{[1,2]} \\ i^{[2,1]} & i^{[2,2]} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = i^{[1,1]},$$

Ripetendo i calcoli svolti nel Paragrafo 4.1, richiamando il Teorema 1 ed il Lemma 3, si ottiene che la distribuzione asintotica risulta

$$\Lambda = \{1 - W(Z_n)\}\Lambda_{rid} + W(Z_n)\Lambda_{est}.$$

dove:

$$\Lambda_{rid} = \Lambda_0 + \omega K^{1/2}a,$$

$$\Lambda_{est} = \lambda_0 + \omega K^{1/2}(a - Z),$$

con $a = \frac{\delta}{K^{1/2}}$, mentre $\Lambda_0 \sim N(0, \tau_0^2)$ e $Z \sim N(a, 1)$ sono variabili casuali indipendenti.

Parallelamente è possibile calcolare lo stimatore compromesso per famiglie di posizione e scala. Esso è

$$T_{bief}(y) \approx \frac{\tilde{\alpha}'_{esteso}(y)\hat{\mu}_{esteso} + \tilde{\alpha}'_{ridotto}(y)\hat{\mu}_{ridotto}}{\tilde{\alpha}'_{esteso}(y) + \tilde{\alpha}'_{ridotto}(y)},$$

dove

$$\tilde{\alpha}'_i(y) = \left[\frac{\exp(nM_i(y))}{\hat{\sigma}_i^3 |\det(H_i(y))|^{1/2} \text{var}_i[T_i(Y)]} \right]^{1/2}$$

per $i \in \{esteso, ridotto\}$.

Anche in questo caso le quantità necessarie possono essere calcolate esplicitamente (per il modello ridotto) o in via numerica (per il modello esteso).

Si può quindi confrontare la distribuzione empirica di questo stimatore con Λ .

I grafici nelle Figure 4.25, 4.26 e 4.27 mostrano la distribuzione dello stimatore compromesso e dello stimatore basato sui pesi AIC sotto il modello ridotto (normale di media 2 e varianza 16), confrontati con la distribuzione asintotica di quest'ultimo, rispettivamente per campioni di numerosità 50, 100 e 250. Non è stato possibile provare con numerosità inferiore, per problemi nelle routine di ottimizzazione di R. I grafici nelle Figure 4.28, 4.29 e 4.30 mostrano anch'essi le distribuzioni dei due stimatori sopraccitati e la distribuzione asintotica, questa volta sotto il modello esteso (normale di ordine $s = 2.2$ con $\mu = 2$ e $\sigma = 4$)

Analizzando i grafici, sia sotto il modello ridotto, sia sotto quello esteso, è evidente come le distribuzioni dei due stimatori siano molto simili, e anche come la distribuzione teorica prevista con la teoria di Hjort & Claeskens (2003) bene approssimi la distribuzione empirica. Effettivamente la numerosità dei campioni è abbastanza elevata, ma le discrepanze tra le distribuzioni nei grafici sono particolarmente piccole. Questo ci dà ancora una volta una conferma della bontà delle approssimazioni viste nel Capitolo 2.

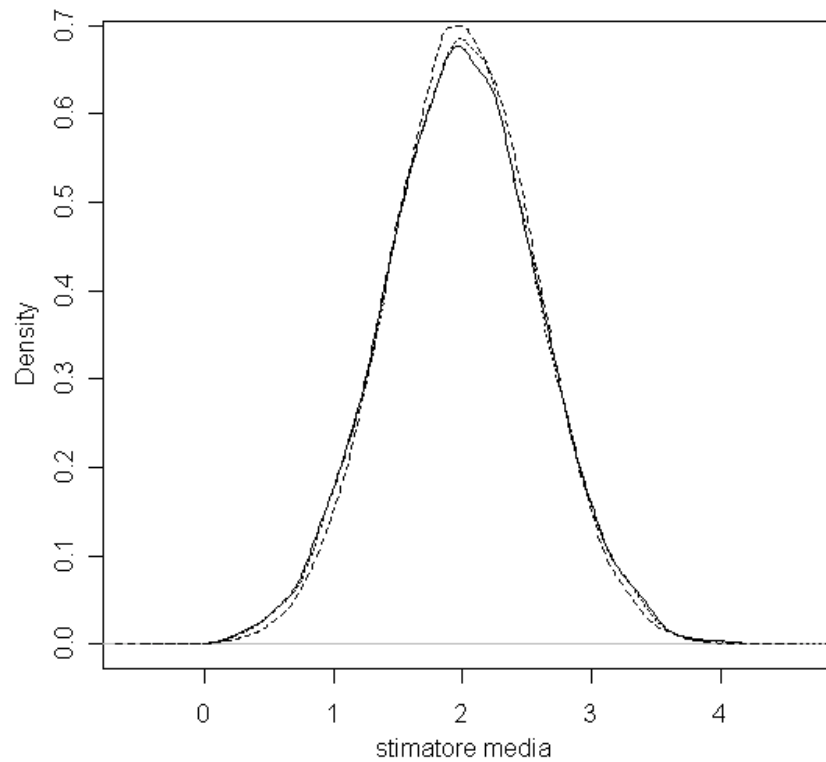


Figura 4.25: *Distribuzione degli stimatori per la media, costruiti per famiglie di posizione e scala (linea continua) e basati su criteri di selezione del modello (linea puntata), calcolati per 10000 replicazioni con campioni di numerosità 50, con i dati provenienti da una normale di media 2 e varianza 16; e confronto con la distribuzione asintotica (linea tratteggiata).*

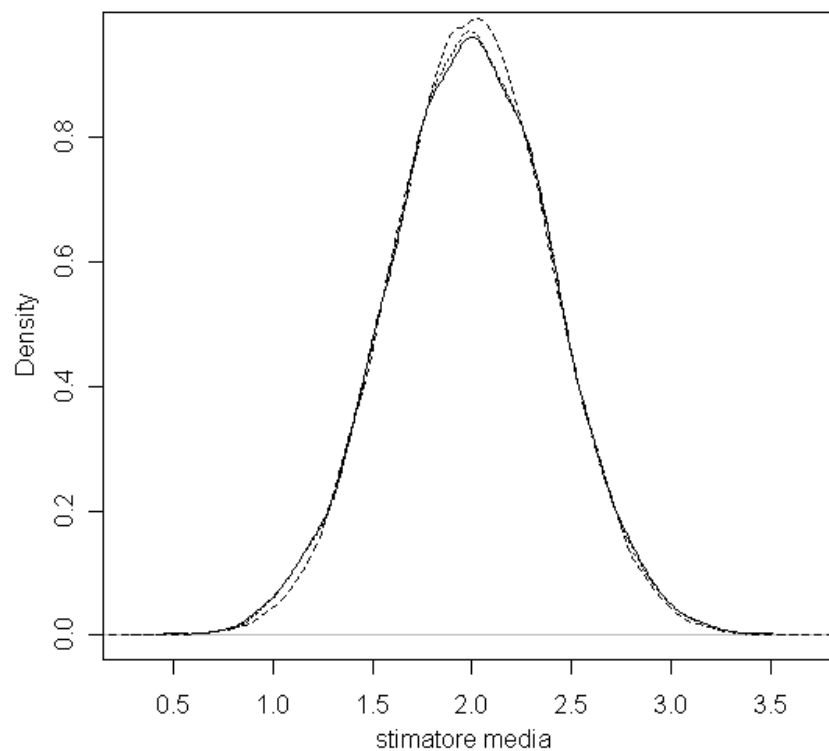


Figura 4.26: *Distribuzione degli stimatori per la media, costruiti per famiglie di posizione e scala (linea continua) e basati su criteri di selezione del modello (linea puntata), calcolati per 10000 repliche con campioni di numerosità 50, con i dati provenienti da una normale di media 2 e varianza 16; e confronto con la distribuzione asintotica (linea tratteggiata).*

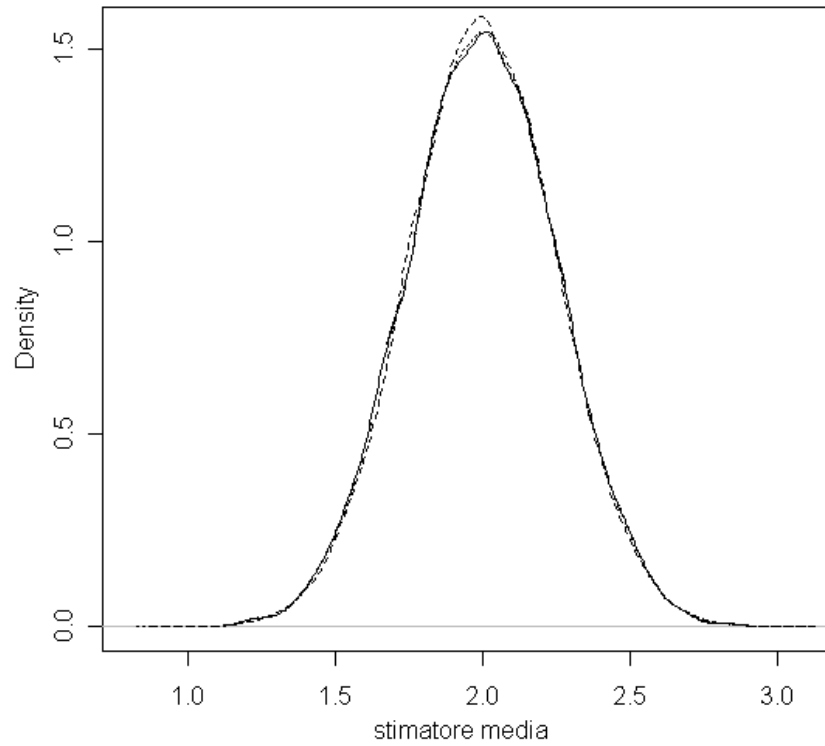


Figura 4.27: *Distribuzione degli stimatori per la media, costruiti per famiglie di posizione e scala (linea continua) e basati su criteri di selezione del modello (linea puntata), calcolati per 10000 replicazioni con campioni di numerosità 50, con i dati provenienti da una normale di media 2 e varianza 16; e confronto con la distribuzione asintotica (linea tratteggiata).*

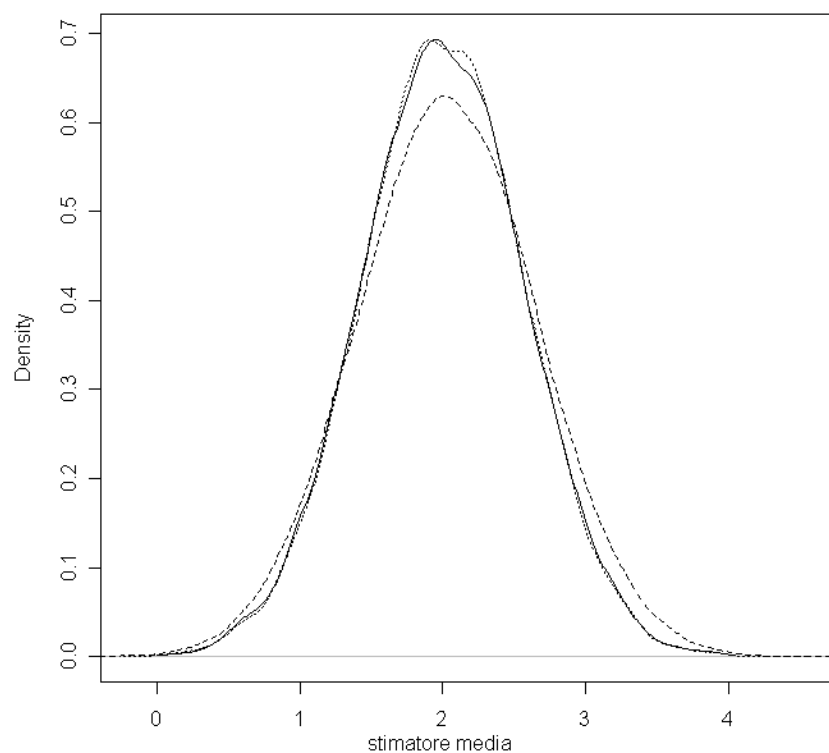


Figura 4.28: *Distribuzione degli stimatori per la media, costruiti per famiglie di posizione e scala (linea continua) e basati su criteri di selezione del modello (linea puntata), calcolati per 10000 repliche con campioni di numerosità 50, con i dati provenienti da una normale di ordine $s = 2.2$ con $\mu = 2$ e $\sigma = 4$; e confronto con la distribuzione asintotica (linea tratteggiata).*

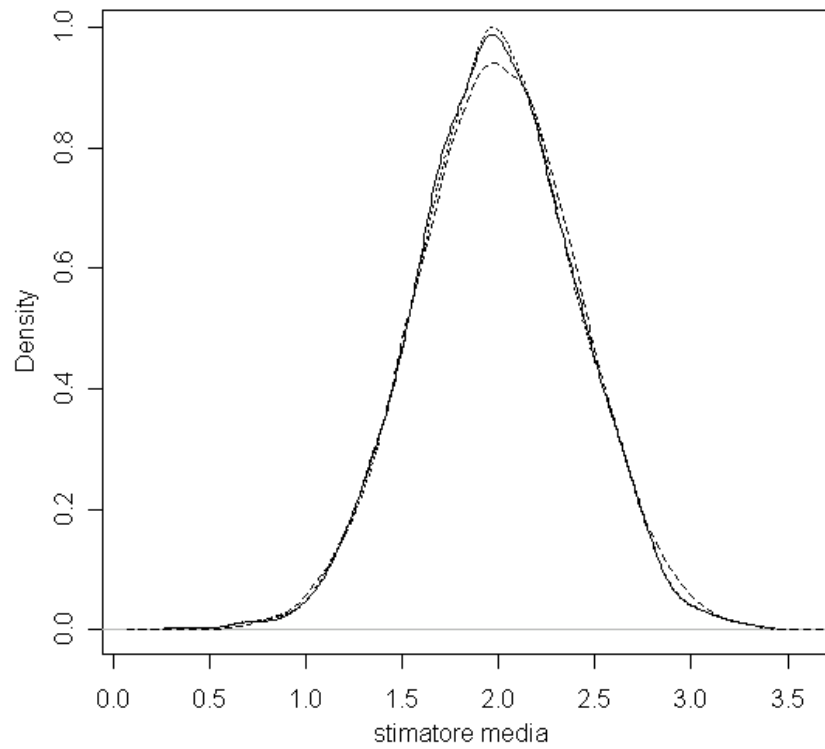


Figura 4.29: *Distribuzione degli stimatori per la media, costruiti per famiglie di posizione e scala (linea continua) e basati su criteri di selezione del modello (linea puntata), calcolati per 10000 repliche con campioni di numerosità 50, con i dati provenienti da una normale di ordine $s = 2.2$ con $\mu = 2$ e $\sigma = 4$; e confronto con la distribuzione asintotica (linea tratteggiata).*

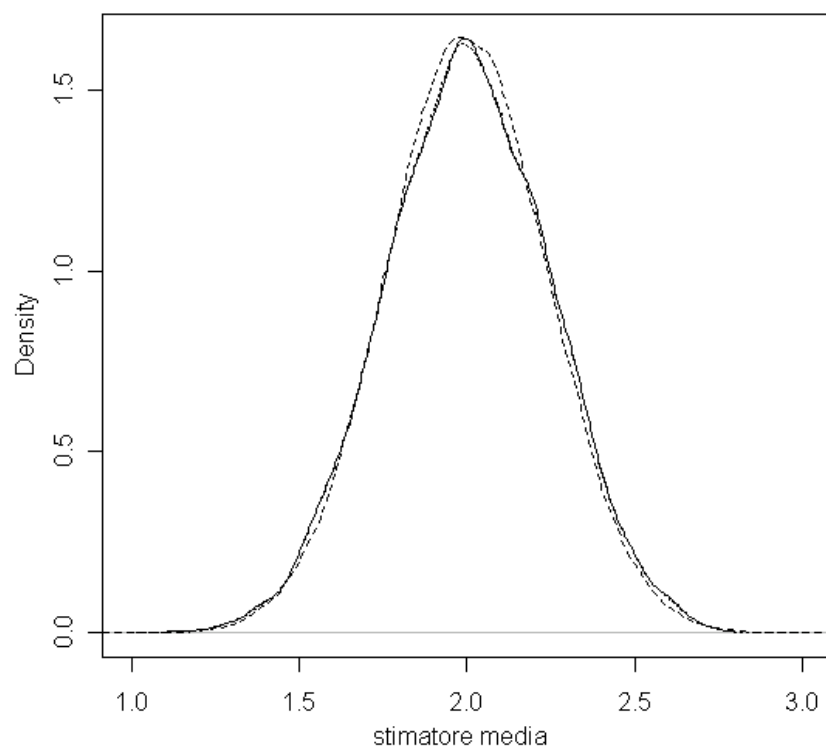


Figura 4.30: *Distribuzione degli stimatori per la media, costruiti per famiglie di posizione e scala (linea continua) e basati su criteri di selezione del modello (linea puntata), calcolati per 10000 repliche con campioni di numerosità 50, con i dati provenienti da una normale di ordine $s = 2.2$ con $\mu = 2$ e $\sigma = 4$; e confronto con la distribuzione asintotica (linea tratteggiata).*

4.2.3 Stimatore basato sulla normale e distribuzione dei valori estremi

Ritornando alla teoria esposta nel Paragrafo 2.1, una possibile alternativa alla distribuzione exponential (4) è utilizzare al suo posto la distribuzione dei valori estremi.

Il procedimento è lo stesso presentato nel Sottoparagrafo 4.2.1, quindi si fa riferimento allo ‘stimatore SCMV stile biefficace’ (formula (2.10))

$$T_{bieff}(y) \approx \frac{\tilde{\alpha}'_1(y)\hat{\mu}_{f_1} + \tilde{\alpha}'_2(y)\hat{\mu}_{f_2}}{\tilde{\alpha}'_1(y) + \tilde{\alpha}'_2(y)}$$

con pesi (vedi (2.11))

$$\tilde{\alpha}'_i(y) = \left[\frac{\exp(nM_{f_i}(y))}{\hat{\sigma}_{f_i}^3 |det(H_{f_i}(y))|^{1/2} var_{f_i}[T_{f_i}(Y)]} \right]^{1/2}.$$

Anche in questo caso le quantità:

- $\hat{\mu}$ e $\hat{\sigma}$, stime di massima verosimiglianza rispettivamente del parametro di posizione e scala;
- $\exp(nM_{f_i}(y))$, che non è altro che il valore della verosimiglianza in $(\hat{\mu}, \hat{\sigma})$;
- $|det(H_{f_i}(y))|^{1/2}$, la radice quadrata del determinante dell’informazione osservata;
- $var_{f_i}[T_{f_i}(Y)]$, la varianza dello stimatore di massima verosimiglianza per il parametro od una sua approssimazione;

possono essere ricavate esplicitamente o per via numerica.

La distribuzione per questo stimatore, calcolata su dati provenienti da una distribuzione t di Student con 5 gradi di libertà, è quella presentata nel grafico di Figura 4.31.

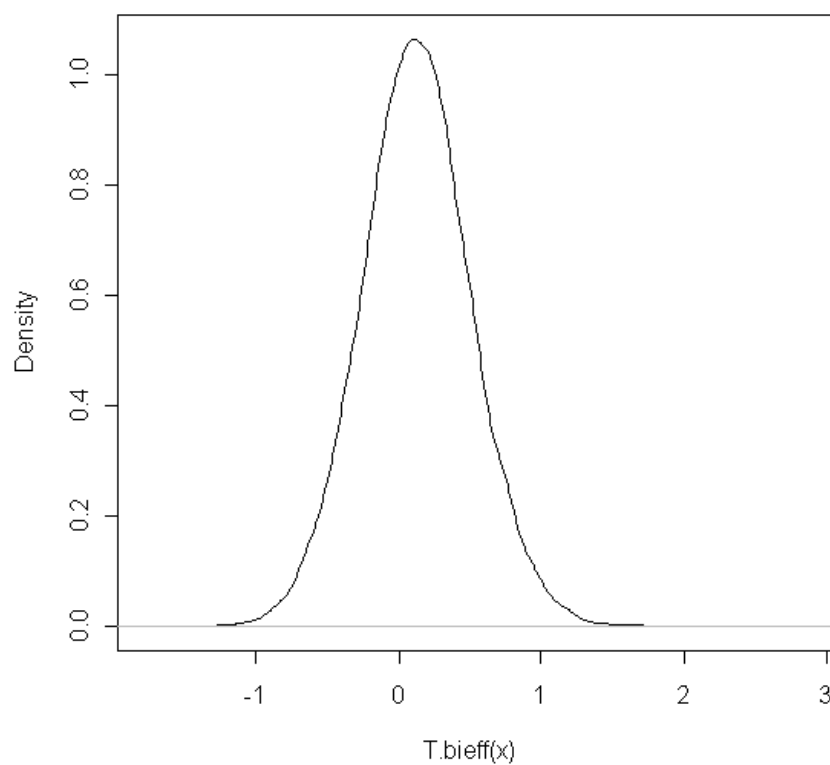


Figura 4.31: *Distribuzione simulata dello stimatore compromesso di massima verosimiglianza stile biefficace calcolata per 10000 repliche con campioni di numerosità 10 generati da una t di Student con 5 gradi di libertà.*

4.2.4 Stimatore basato sulla normale, normale di ordine 4 e distribuzione dei valori estremi

Come esempio finale è possibile provare ad utilizzare tutte tre le distribuzioni utilizzate precedentemente per ricavare informazioni sul parametro d'interesse.

L'estensione è immediata, lo stimatore diventa

$$T_{bief}(y) \approx \frac{\tilde{\alpha}'_{norm}(y)\hat{\mu}_{norm} + \tilde{\alpha}'_{ep(4)}(y)\hat{\mu}_{ep(4)} + \tilde{\alpha}'_{ve}(y)\hat{\mu}_{ve}}{\tilde{\alpha}_{norm}'(y) + \tilde{\alpha}'_{ep(4)}(y) + \tilde{\alpha}'_{ve}(y)}$$

dove

$$\tilde{\alpha}'_i(y) = \left[\frac{\exp(nM_i(y))}{\hat{\sigma}_i^3 |det(H_i(y))|^{1/2} var_i[T_i(Y)]} \right]^{1/2}$$

per $i \in \{norm, ep(4), ve\}$. Come negli esempi precedenti, le quantità necessarie si ricavano in maniera esplicita o per via numerica. La distribuzione empirica è quella di Figura (4.32)

4.2.5 Studio dell'efficienza

Per comparare gli stimatori presentati nei vari esempi, è possibile ricorrere ad un confronto dell'efficienza. Sono quindi stati stimati tramite simulazione, sotto vari modelli, gli errori quadratici medi per i vari stimatori.

Nella tabella 4.4 è riportato l'MSE sotto la normale, la normale di ordine 4 e la valori estremi per la media campionaria e per gli stimatori compromesso di massima verosimiglianza presentati negli esempi.

La tabella ci conferma la bontà degli stimatori compromesso di massima verosimiglianza. In particolare, si può notare come essi si avvicinino agli stimatori migliori (ad esempio, si confronti lo stimatore basato su normale ed EP(4) e lo stimatore media campionaria sotto la distribuzione normale standard), ed abbiano una maggiore efficienza quando la distribuzione è diversa (si vedano i valori dell'MSE per i due stimatori citati nel caso i dati provengano da una EP(4)). Si noti, infine, che gli stimatori ottenuti utilizzando come pesi i valori dell'AIC si comportino grossomodo alla stessa

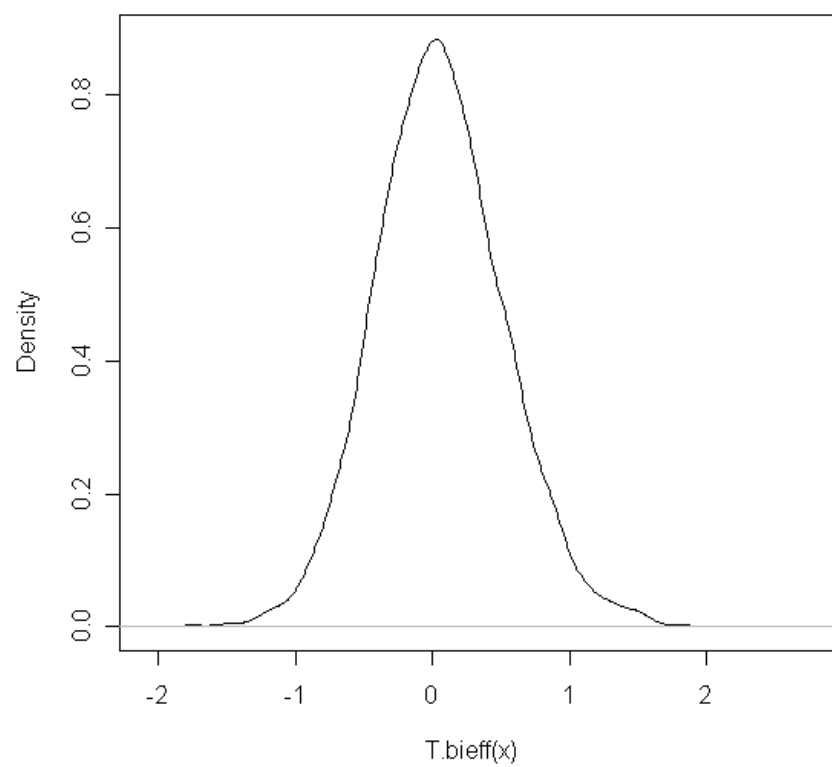


Figura 4.32: *Distribuzione simulata dello stimatore compromesso di massima verosimiglianza stile biefficace calcolata per 10000 repliche con campioni di numerosità 10 generati da una t di Student con 5 gradi di libertà.*

	n=5		
Stimatore	normale	EP(4)	ve
Media campionaria	0.1861980	0.1336952	0.6845050
SCMV norm + EP(4)	0.1911192	0.1246001	0.7295363
SCMV norm + ve	0.2102327	0.1677166	0.4702628
SCMV norm + EP(4) + ve	0.2016248	0.1483649	0.5166543
stim AIC norm + EP(4)	0.1865952	0.1321032	0.7083288
stim AIC norm + ve	0.2326470	0.1804954	0.3692942
stim AIC norm+EP(4)+ve	0.2306365	0.1779511	0.3745656
	n=10		
Stimatore	normale	EP(4)	ve
Media campionaria	0.1004879	0.0700564	0.5010312
SCMV norm + EP(4)	0.1025018	0.0644631	0.5438016
SCMV norm + ve	0.1150704	0.0898832	0.3075036
SCMV norm + EP(4) + ve	0.1099148	0.0786091	0.3423291
stim AIC norm + EP(4)	0.1004949	0.0700039	0.5019690
stim AIC norm + ve	0.1553820	0.1157596	0.1801763
stim AIC norm+EP(4)+ve	0.1552869	0.1156596	0.1803151

Tabella 4.4: *MSE per i vari stimatori (calcolato su 2000 ripetizioni) sotto la normale ($N(0,1)$), la EP(4) (con media 0 e varianza 1) e la ve(0,1).*

maniera, soprattutto nel momento in cui si utilizzano come differenti modelli quello normale e quello EP(4). Si dimostrano parecchio migliori, invece, quando uno di questi modelli è quello dei valori estremi ed effettivamente i dati provengano dalla distribuzione valori estremi (si noti l'ultima colonna della Tabella 4.4).

Conclusioni

In questa tesi è stata considerata una strategia inferenziale che prevede di utilizzare per l'inferenza l'informazione proveniente da vari modelli, invece di seguire il classico approccio della selezione del migliore modello.

Nel Capitolo 2 sono stati descritti i due metodi studiati in letteratura per la costruzione degli stimatori multimodello, ossia quelli compromesso per famiglie di posizione e scala (con una versione approssimata basata sugli stimatori di massima verosimiglianza) e quelli basati su criteri di selezione del modello, in particolar modo il criterio di Akaike.

Nel Capitolo 3, viceversa, è stata riportata una teoria asintotica che ha permesso di ottenere una distribuzione limite degli stimatori basati su criteri di selezione del modello e conseguentemente un'espressione analitica per gli intervalli di confidenza.

Nel Capitolo 4, infine, la teoria esposta nei precedenti capitoli è stata adattata ad esempi di simulazione, verificandone il comportamento.

L'esempio in cui si sono considerati il modello Weibull ed il modello esponenziale, che partiva da un suggerimento di Hjort & Claeskens (2003), ha prodotto risultati abbastanza buoni per quanto riguarda i livelli di copertura degli intervalli di confidenza, mentre ha lasciato qualche dubbio riguardo la capacità della distribuzione asintotica di adattarsi bene al comportamento dello stimatore, soprattutto quando il campione pseudocasuale proveniva dal modello esteso. Sotto quest'ultimo, infatti, c'è qualche problema al centro della distribuzione, con la moda effettiva spostata a sinistra rispetto alla moda prevista dalla distribuzione asintotica.

Nel secondo esempio in cui si è utilizzata la teoria asintotica, viceversa, i risultati sono stati più che soddisfacenti, sia sotto il modello ridotto (normale), sia sotto quello esteso ($EP(s)$). Questo esempio, inoltre, ha suggerito che, limitatamente al caso preso in esame, gli stimatori compromesso per famiglie di posizione e scala e gli stimatori basati sull'AIC non si comportano in maniera sostanzialmente differente. Questa analogia di comportamento è stata confermata anche dalle valutazioni di efficienza dei vari stimatori, con quelli compromesso che si comportano in maniera molto simile a quelli basati sull'AIC.

Alcuni sviluppi ulteriori potrebbero riguardare i seguenti aspetti.

- Un differente metodo, non esplorato nella tesi, per la definizione di stimatori multimodello potrebbe essere quello di combinare, anziché gli stimatori di interesse, le relative equazioni di stima. Ad esempio, si potrebbero utilizzare, per le note proprietà di robustezza, le equazioni di stima che definiscono stimatori di tipo M. Si considerino R modelli differenti, ed un campione di dati y proveniente da una distribuzione ignota. Per ogni modello, si ottiene una funzione di stima $q_i(\mu, y)$. Uno stimatore basato sulla combinazione di equazioni di stima, lo si chiami $\tilde{\mu}$, sarebbe quindi quel μ che risolve l'equazione di stima

$$q(\mu, y) = 0,$$

dove $q(\mu, y) = \sum_{i=1}^R w_i q_i(\mu, y)$, per appropriati valori di w_i . Con w_i , invece, si indica il peso dell' i -simo modello.

- Un diverso approccio potrebbe essere quello di utilizzare metodi semi-parametrici, seguendo l'approccio di Severini (1992). Sembra di interesse un confronto tra le due strategie inferenziali.
- La costruzione di stimatori basati su metodi di selezione del modello si è limitata qui ai pesi di Akaike. Resta da valutare il comportamento di stimatori basati su altri criteri di selezione, ad esempio il BIC.

Appendice: comandi R

Grafici delle distribuzioni sotto il modello esteso

```
stime<-NULL
pesi.est[i]<-NULL
pesi.rid[i]<-NULL

#####
# funzioni necessarie

logV <- function(ganna, theta, dati){
  sum(log(ganna)+log(theta)+(ganna-1)*log(theta)
    +(ganna-1)*log(dati)-(theta*dati)^ganna)
}

erre<-0.5772
J.10<-function(theta,erre=0.5772) { (1-erre)/theta }
J.00<-function(theta,ganna) { ganna^2/theta^2 }
omega<-function(theta,ganna,erre=0.5772){
  (log(2)^(1/ganna))*(erre-1 +log(log(2)))/((ganna^2)*theta)
}
tau.0<-function(theta,ganna){(log(2)^(1/ganna))/(theta*ganna)}
K<-function(ganna){ 6*(ganna^2)/(pi^2) }
```

```

Z.n<-function(ganna,K,n){ sqrt(n)*(ganna-1)/sqrt(K) }

#####

for(i in 1:10000)
{
  # genero campione casuale dal modello esteso, in questo caso,
  # di numerosità 10
  camp<-rweibull(10,1.10,5)
  # ottengo le stime di massima verosimiglianza per i parametri
  # tramite la funzione uniroot
  gannaS.hat<-uniroot(function(x) length(camp)/x+sum(log(camp))
    -length(camp)*sum(camp^x*log(camp))/sum(camp^x),
    c(1e-05,10))$root
  lambdaS.hat<-1/mean(camp^gannaS.hat)
  # riparametrizzo le stime di massima verosimiglianza
  ganna.hat<-gannaS.hat
  theta.hat<-lambdaS.hat^(1/gannaS.hat)
  parametri<-c(theta.hat,ganna.hat)
  # formula stimatore mu:((log(2))^(1/ganna.hat))/theta.hat
  # calcolo di mu col modello esteso
  mu.est<-((log(2))^(1/ganna.hat))/theta.hat
  # calcolo di mu col modello ridotto
  mu.rid<-mean(camp)*log(2)

  # calcolo i pesi
  AIC.w<-2*(logV(ganna.hat,theta.hat,camp)-1)
  AIC.e<-2*(logV(1,(1/mean(camp)),camp)-0)
  W_Z.est<-exp(AIC.w/2)/(exp(AIC.w/2)+exp(AIC.e/2))
  W_Z.rid<-exp(AIC.e/2)/(exp(AIC.w/2)+exp(AIC.e/2))
  # calcolo lo stimatore

```

```

mu<-W_Z.est*mu.est + W_Z.rid*mu.rid
# tengo memoria dello stimatore dell'i-sima ripetizione
stime[i]<-mu
pesi.est[i]<-W_Z.est
pesi.rid[i]<-W_Z.rid
}

# disegno la distribuzione empirica dello stimatore
plot(density(stime))

# simulo le variabili casuali necessarie per il calcolo della
# distribuzione asintotica
sim1<- rnorm(100000,0, tau.0(0.20,1.10))
sim2<- rnorm(100000,(0.10/sqrt(K(1.10))),1)
# calcolo la distribuzione asintotica
lambda.est<- ((log(2))^(1/1.10))/0.20 + (sim1
+omega(0.20,1.10)*sqrt(K(1.10))*(0.10/sqrt(K(1.10))
-sim2))/sqrt(length(camp))
lambda.rid<- ((log(2))^(1/1.10))/0.20
+ (sim1 + 0.10*omega(0.20, 1.10))/sqrt(length(camp))
lambda<-mean(pesi.est)*lambda.est + mean(pesi.rid)*lambda.rid

# aggiungo al grafico la distribuzione asintotica
# con pesi asintotici
lines(density(lambda.est),lty=2)
# e quella con pesi campionari
lines(density(lambda),lty=2)

```

Grafici delle distribuzioni sotto il modello ridotto

```
stime<-NULL

#####
# funzioni necessarie

logV <- function(ganna, theta, dati){
  sum(log(ganna)+log(theta)+(ganna-1)*log(theta)
    +(ganna-1)*log(dati)-(theta*dati)^ganna)
}

erre<-0.5772
J.10<-function(theta,erre=0.5772) { (1-erre)/theta }
J.00<-function(theta,ganna) { ganna^2/theta^2 }
omega<-function(theta,ganna,erre=0.5772){
  (log(2)^(1/ganna))*(erre-1 +log(log(2)))/((ganna^2)*theta)
}
tau.0<-function(theta,ganna){(log(2)^(1/ganna))/(theta*ganna)}
K<-function(ganna){ 6*(ganna^2)/(pi^2) }
Z.n<-function(ganna,K,n){ sqrt(n)*(ganna-1)/sqrt(K) }

#####
for(i in 1:10000)
{
  # genero campione casuale dal modello ridotto, in questo
  # caso, di numerosità 10
  camp<-rexp(10,0.20)
  # ottengo le stime di massima verosimiglianza per i
  # parametri tramite la funzione uniroot
```



```

gannaS.hat<-uniroot(function(x) length(camp)/x+sum(log(camp))
- length(camp)*sum(camp^x*log(camp))/sum(camp^x),
c(1e-05,10))$root
lambdaS.hat<-1/mean(camp^gannaS.hat)
# riparametrizzo le stime di massima verosimiglianza
ganna.hat<-gannaS.hat
theta.hat<-lambdaS.hat^(1/gannaS.hat)
parametri<-c(theta.hat,ganna.hat)
# formula stimatore mu:((log(2))^(1/ganna.hat))/theta.hat
# calcolo di mu col modello esteso
mu.est<-((log(2))^(1/ganna.hat))/theta.hat
# calcolo di mu col modello ridotto
mu.rid<-mean(camp)*log(2)

# calcolo i pesi
AIC.w<-2*(logV(ganna.hat,theta.hat,camp)-1)
AIC.e<-2*(logV(1,(1/mean(camp)),camp)-0)
W_Z.est<-exp(AIC.w/2)/(exp(AIC.w/2)+exp(AIC.e/2))
W_Z.rid<-exp(AIC.e/2)/(exp(AIC.w/2)+exp(AIC.e/2))
# calcolo lo stimatore
mu<-W_Z.est*mu.est + W_Z.rid*mu.rid
# tengo memoria dello stimatore dell'i-sima ripetizione
stime[i]<-mu
}

# disegno la distribuzione empirica dello stimatore
plot(density(stime))

# simulo le variabili casuali necessarie per il calcolo della
# distribuzione asintotica
sim1<- rnorm(100000,0, tau.0(0.20,1))

```

```
sim2<- rnorm(100000,(0/sqrt(K(1))),1)
# calcolo la distribuzione asintotica
lambda.est<- log(2)/0.20 + (sim1 - sim2))/sqrt(length(camp))
lambda.rid<- log(2)/0.20 + (sim1)/sqrt(length(camp))
lambda<-W_Z.est*lambda.est + W_Z.rid*lambda.rid

# aggiungo al grafico la distribuzione asintotica
lines(density(lambda),lty=2)
```

Grafici delle coperture asintotiche sotto il modello ridotto

```

coperturaHjort<-function(mu.0,delta,n,R)
{
  ### FUNZIONI NECESSARIE ###
  erre<-0.5772
  omega<-function(theta,ganna,erre=0.5772)
  {
    (log(2)^(1/ganna))*(erre-1+log(log(2)))/((ganna^2)*theta)
  }
  tau.0<-function(theta, ganna)
  {
    (log(2)^(1/ganna))/(theta*ganna)
  }
  K<-function(ganna)
  {
    6*(ganna^2)/(pi^2)
  }
  Z.n<-function(ganna,K,n)
  {
    sqrt(n)*(ganna-1)/sqrt(K)
  }
  D_n<-function(ganna,n)
  {
    sqrt(n)*(ganna-1)
  }

  cont<-0
  logV <- function(ganna, theta, dati)
  {

```

```

      sum(log(ganna)+log(theta)+(ganna-1)*log(theta)
        + (ganna-1)*log(dati)-(theta*dati)^ganna)
    }
### ITERAZIONI ###
for (i in 1:R)
{
  camp<-rexp(n,log(2)/mu.0)
  gannaS.hat<-uniroot(function(x) length(camp)/x+sum(log(camp))
    -length(camp)*sum(camp^x*log(camp))/sum(camp^x),
    c(1e-05,10))$root
  lambdaS.hat<-1/mean(camp^gannaS.hat)
  ganna.hat<-gannaS.hat
  theta.hat<-lambdaS.hat^(1/gannaS.hat)
  parametri<-c(theta.hat,ganna.hat)
  parametri
  mu.est<-((log(2))^(1/ganna.hat))/theta.hat
  mu.rid<-mean(camp)*log(2)
  AIC.w<-2*(logV(ganna.hat,theta.hat,camp)-1)
  AIC.e<-2*(logV(1,(1/mean(camp)),camp)-0)
  W_Z.est<-exp(AIC.w/2)/(exp(AIC.w/2)+exp(AIC.e/2))
  W_Z.rid<-exp(AIC.e/2)/(exp(AIC.w/2)+exp(AIC.e/2))
  mu.hat<-W_Z.est*mu.est + W_Z.rid*mu.rid
  quantile<-1.645
  D_n.hat<-D_n(ganna.hat,n)
  low<-mu.hat - omega(theta.hat,ganna.hat)
  *(D_n.hat - W_Z.est*D_n.hat)/sqrt(n)
  -(quantile*sqrt((tau.0(theta.hat,ganna.hat))^2
  +omega(theta.hat,ganna.hat)*K(ganna.hat)
  *omega(theta.hat,ganna.hat))/sqrt(n))
  up<-mu.hat + omega(theta.hat,ganna.hat)
  *(D_n.hat - W_Z.est*D_n.hat)/sqrt(n)

```

```

      +(quantile*sqrt((tau.0(theta.hat,ganna.hat))^2
      +omega(theta.hat,ganna.hat)*K(ganna.hat)
      *omega(theta.hat,ganna.hat))/sqrt(n))

      if (mu.0>low && mu.0 <up) cont<-cont+1
    }
    copertura<-cont/R
    copertura
  }

mu<-seq(0.1,8,length=200)
valori<-NULL
for (i in 1:200) valori[i]<-coperturaHjort(mu[i],0,10,1000)
plot(mu,valori,ylim=c(0,1),type='l')
abline(h=0.90,lty='dashed')

```

Grafici delle coperture bootstrap sotto il modello ridotto

```
library(boot)

coperturaHjort<-function(mu.0,delta,n,R)
{
  ### FUNZIONI NECESSARIE ###
  erre<-0.5772
  omega<-function(theta,ganna,erre=0.5772)
  {
    (log(2)^(1/ganna))*(erre-1+log(log(2)))/((ganna^2)*theta)
  }
  tau.0<-function(theta, ganna)
  {
    (log(2)^(1/ganna))/(theta*ganna)
  }
  K<-function(ganna)
  {
    6*(ganna^2)/(pi^2)
  }
  Z.n<-function(ganna,K,n)
  {
    sqrt(n)*(ganna-1)/sqrt(K)
  }
  D_n<-function(ganna,n)
  {
    sqrt(n)*(ganna-1)
  }

  cont<-0
```

```

logV <- function(ganna, theta, dati)
{
  sum(log(ganna)+log(theta)+(ganna-1)*log(theta)
    + (ganna-1)*log(dati)-(theta*dati)^ganna)
}

### ITERAZIONI ###
for (i in 1:R)
{
  camp<-rexp(n,log(2)/mu.0)
  logV2 <- function(param, dati)
  {
    sum(-log(param[1])-log(param[2])-(param[1]-1)*log(param[2])
      - (param[1]-1)*log(dati)+(param[2]*dati)^param[1])
  }
  vvv<-nlm(logV2, c(2,2), hessian=T, dati=camp)
  vvv$estimate
  ganna.hat<- vvv$estimate[1]
  theta.hat<-vvv$estimate[2]
  parametri<-c(theta.hat,ganna.hat)
  mu.est<-((log(2))^(1/ganna.hat))/theta.hat
  mu.rid<-mean(camp)*log(2)
  AIC.w<-2*(logV(ganna.hat,theta.hat,camp)-1)
  AIC.e<-2*(logV(1,(1/mean(camp)),camp)-0)
  W_Z.est<-exp(AIC.w/2)/(exp(AIC.w/2)+exp(AIC.e/2))
  W_Z.rid<-exp(AIC.e/2)/(exp(AIC.w/2)+exp(AIC.e/2))
  mu.hat<-W_Z.est*mu.est + W_Z.rid*mu.rid
  stimatore <- function(x,ind)
  {
    camp[ind]<-x[ind]
    vv2<-nlm(logV2, c(2,2), hessian=T, dati=camp)
    vv2$estimate
  }
}

```

```

    ganna.hat<- vv2$estimate[1]
    theta.hat<-vv2$estimate[2]
    mu.est<-((log(2))^(1/ganna.hat))/theta.hat
    mu.rid<-mean(camp[ind])*log(2)
    AIC.w<-2*(logV(ganna.hat,theta.hat,camp[ind])-1)
    AIC.e<-2*(logV(1,(1/mean(camp[ind])),camp[ind])-0)
    W_Z.est<-exp(AIC.w/2)/(exp(AIC.w/2)+exp(AIC.e/2))
    W_Z.rid<-exp(AIC.e/2)/(exp(AIC.w/2)+exp(AIC.e/2))
    mu<-W_Z.est*mu.est + W_Z.rid*mu.rid
    mu
  }
  # eseguo bootstrap
  b1<-boot(camp,stimatore,99)
  s<-sort(b1$t)
  low<-s[5]
  up<-s[95]
  if (mu.0>low & mu.0 <up) cont<-cont+1
}
  copertura<-cont/R
  copertura
}

mu<-seq(0.1,8,length=100)
valori<-NULL
for (i in 1:100) valori[i]<-coperturaHjort(mu[i],0,10,100)
plot(mu,valori,ylim=c(0,1),type='l')
abline(h=0.90,lty='dashed')

```


Grafici delle coperture asintotiche sotto il modello esteso (gamma fissato)

```

coperturaHjort<-function(mu.0,n,R)
{
  ### FUNZIONI NECESSARIE ###
  erre<-0.5772
  omega<-function(theta,ganna,erre=0.5772)
  {
    (log(2)^(1/ganna))*(erre-1+log(log(2)))/((ganna^2)*theta)
  }
  tau.0<-function(theta, ganna)
  {
    (log(2)^(1/ganna))/(theta*ganna)
  }
  K<-function(ganna)
  {
    6*(ganna^2)/(pi^2)
  }
  Z.n<-function(ganna,K,n)
  {
    sqrt(n)*(ganna-1)/sqrt(K)
  }
  D_n<-function(ganna,n)
  {
    sqrt(n)*(ganna-1)
  }

  cont<-0
  logV <- function(ganna, theta, dati)
  {

```

```

      sum(log(ganna)+log(theta)+(ganna-1)*log(theta)
      + (ganna-1)*log(dati)-(theta*dati)^ganna)
    }
### ITERAZIONI ###
for (i in 1:R)
{
  camp<-rweibull(n,1.10,mu.0/(log(2)^(1/1.10)))
  logV2 <- function(param, dati)
  {
    sum(-log(param[1])-log(param[2])
    -(param[1]-1)*log(param[2])
    -(param[1]-1)*log(dati)+(param[2]*dati)^param[1])
  }
  vvv<-nlm(logV2, c(2,2), hessian=T, dati=camp)
  vvv$estimate
  ganna.hat<- vvv$estimate[1]
  theta.hat<-vvv$estimate[2]
  parametri<-c(theta.hat,ganna.hat)
  parametri
  mu.est<-((log(2))^(1/ganna.hat))/theta.hat
  mu.rid<-mean(camp)*log(2)
  AIC.w<-2*(logV(ganna.hat,theta.hat,camp)-1)
  AIC.e<-2*(logV(1,(1/mean(camp)),camp)-0)
  W_Z.est<-exp(AIC.w/2)/(exp(AIC.w/2)+exp(AIC.e/2))
  W_Z.rid<-exp(AIC.e/2)/(exp(AIC.w/2)+exp(AIC.e/2))
  mu.hat<-W_Z.est*mu.est + W_Z.rid*mu.rid
  quantile<-1.645
  D_n.hat<-D_n(ganna.hat,n)
  low<-mu.hat - omega(theta.hat,ganna.hat)*(D_n.hat-W_Z.est
  *D_n.hat)/sqrt(n)-(quantile*sqrt((tau.0(theta.hat,ganna.hat))^2
  +omega(theta.hat,ganna.hat)*K(ganna.hat)

```

```

*omega(theta.hat,ganna.hat))/sqrt(n))
up<-mu.hat - omega(theta.hat,ganna.hat)*(D_n.hat-W_Z.est
*D_n.hat)/sqrt(n)+(quantile*sqrt((tau.0(theta.hat,ganna.hat))^2
+omega(theta.hat,ganna.hat)*K(ganna.hat)
*omega(theta.hat,ganna.hat))/sqrt(n))

  if (mu.0>low && mu.0 <up) cont<-cont+1
}
copertura<-cont/R
copertura
}

mu<-seq(0.01,8,length=100)
valori<-NULL
for (i in 1:100) valori[i]<-coperturaHjort(mu[i],50,500)
plot(mu,valori,ylim=c(0,1),type='l')
abline(h=0.90,lty='dashed')

```

Grafici delle coperture asintotiche sotto il modello esteso (theta fissato)

```

coperturaHjort<-function(mu.0,n,R)
{
  ### FUNZIONI NECESSARIE ###
  erre<-0.5772
  omega<-function(theta,ganna,erre=0.5772)
  {
    (log(2)^(1/ganna))*(erre-1+log(log(2)))/((ganna^2)*theta)
  }
  tau.0<-function(theta, ganna)
  {
    (log(2)^(1/ganna))/(theta*ganna)
  }
  K<-function(ganna)
  {
    6*(ganna^2)/(pi^2)
  }
  Z.n<-function(ganna,K,n)
  {
    sqrt(n)*(ganna-1)/sqrt(K)
  }
  D_n<-function(ganna,n)
  {
    sqrt(n)*(ganna-1)
  }

  cont<-0
  logV <- function(ganna, theta, dati)
  {

```

```

    sum(log(ganna)+log(theta)+(ganna-1)*log(theta)
    + (ganna-1)*log(dati)-(theta*dati)^ganna)
  }
### ITERAZIONI ###
for (j in 1:R)
{
  camp<-rweibull(n,(log(log(2))/log(0.2*mu.0)),5)
  gannaS.hat<-uniroot(function(x) length(camp)/x+sum(log(camp))
    - length(camp)*sum(camp^x * log(camp))/sum(camp^x),
    c(1e-05,10))$root
  lambdaS.hat<-1/mean(camp^gannaS.hat)
  ganna.hat<-gannaS.hat
  theta.hat<-lambdaS.hat^(1/gannaS.hat)
  parametri<-c(theta.hat,ganna.hat)
  parametri
  mu.est<-((log(2))^(1/ganna.hat))/theta.hat
  mu.rid<-mean(camp)*log(2)
  AIC.w<-2*(logV(ganna.hat,theta.hat,camp)-1)
  AIC.e<-2*(logV(1,(1/mean(camp)),camp)-0)
  W_Z.est<-exp(AIC.w/2)/(exp(AIC.w/2)+exp(AIC.e/2))
  W_Z.rid<-exp(AIC.e/2)/(exp(AIC.w/2)+exp(AIC.e/2))
  mu.hat<-W_Z.est*mu.est + W_Z.rid*mu.rid

  quantile<-1.645
  D_n.hat<-D_n(ganna.hat,n)
  low<-mu.hat - omega(theta.hat,ganna.hat)*(D_n.hat-W_Z.est
    *D_n.hat)/sqrt(n)-(quantile*sqrt((tau.0(theta.hat,ganna.hat))^2
    +omega(theta.hat,ganna.hat)*K(ganna.hat)
    *omega(theta.hat,ganna.hat))/sqrt(n))
  up<-mu.hat + omega(theta.hat,ganna.hat)*(D_n.hat-W_Z.est
    *D_n.hat)/sqrt(n)+(quantile*sqrt((tau.0(theta.hat,ganna.hat))^2

```

```
+omega(theta.hat,ganna.hat)*K(ganna.hat)
*omega(theta.hat,ganna.hat))/sqrt(n))

  if (mu.0>low && mu.0 <up) cont<-cont+1
}
copertura<-cont/R
copertura
}

mu<-seq(0.01,4.99,length=200)
valori<-NULL
for (i in 1:200) valori[i]<-coperturaHjort(mu[i],10,100)
plot(mu,valori,ylim=c(0,1),type='l')
abline(h=0.90,lty='dashed')
```

Calcolo e distribuzione stimatore per famiglie di posizione e scala (normale + normale di ordine 4)

```
stime<-NULL

for(j in 1:10000){
  camp<-rt(10,5)
  t.hat<-NULL
  s.hat<-NULL
  dH12<-NULL
  nM<-NULL
  var_T.f<-NULL

  ### SOTTO LA NORMALE ###
  t.hat[1]<-mean(camp)
  s.hat[1]<-sqrt((1/length(camp))*sum((camp-t.hat[1])^2))
  dH12[1]<-sqrt(2)/(s.hat[1]^2)
  nM[1]<- -length(camp)*(1+log(2*pi))/2-length(camp)*log(s.hat[1])
  var_T.f[1]<-1/length(camp)

  ### SOTTO LA EP(4) ###
  logV.ep4<-function(param,dati){
    length(camp)*log(param[2])
    + sum((abs(camp-param[1]))^4)/(4*(param[2]^4))
  }
  ep4.optim<-nlm(logV.ep4, c(2,2), hessian=T, dati=camp)
  if (det(ep4.optim$hessian)<0){
    ep4.optim<-nlm(logV.ep4, c(10,10), hessian=T, dati=camp)}
  if (det(ep4.optim$hessian)<0){
    ep4.optim<-nlm(logV.ep4, c(-10,10), hessian=T, dati=camp)}
```

```

t.hat[2]<-ep4.optim$estimate[1]
s.hat[2]<-ep4.optim$estimate[2]
dH12[2]<-sqrt(det(ep4.optim$hessian))
nM[2]<- - length(camp)*(log(2) + log(4^(1/4)) + log(gamma(5/4))
  + log(s.hat[2])) - sum(abs(camp-t.hat[2])^4)/(4*(s.hat[2]^4))
J<-solve(ep4.optim$hessian)
var_T.f[2]<-J[1,1]

### CALCOLO PESI SCMV BIEFFICACE ###
alpha<-NULL
for (i in 1:2){
  alpha[i]<-sqrt( exp(nM[i])/((s.hat[i]^3)*dH12[i]*var_T.f[i]))
}
nM[1]
nM[2]
alpha[1]
alpha[2]

### CALCOLO STIMATORE SCMV BIEFFICACE ###
t.bieff<-(alpha[1]*t.hat[1]+alpha[2]*t.hat[2])/(alpha[1]+alpha[2])
t.bieff
stime[j]<-t.bieff
}

### GRAFICO DELLA DISTRIBUZIONE ###
plot(density(stime),col=2)

```


Calcolo, distribuzione stimatore per famiglie di posizione e scala (normale + normale di ordine s), confronto con distribuzione stimatore compromesso e distribuzione asintotica sotto il modello ridotto

```
library(normalp)
stimeE<-NULL
stimeH<-NULL

for(j in 1:10000){
  camp<-rnorm(100,2,4)

  #####
  ### STIMATORE PER FAMIGLIE DI POSIZIONE E SCALA ###
  t.hat<-NULL
  s.hat<-NULL
  dH12<-NULL
  nM<-NULL
  var_T.f<-NULL

  ### SOTTO LA NORMALE ###
  t.hat[1]<-mean(camp)
  s.hat[1]<-sqrt((1/length(camp))*sum((camp-t.hat[1])^2))
  dH12[1]<-sqrt(2)/(s.hat[1]^2)
  nM[1]<- -length(camp)*(1+log(2*pi))/2
           - length(camp)*log(s.hat[1])
  var_T.f[1]<-1/length(camp)

  ### SOTTO LA EP(s) ###
```

```

logV.epS<-function(param,dati){
length(camp)*( log(param[3]^(1/param[3]))
+log(gamma(1+1/param[3]))+log(param[2])) )
+sum((abs(camp-param[1]))^param[3])/
(param[3]*(param[2]^param[3]))
}
epS.optim<-nlm(logV.epS, c(2,2,2), hessian=T, dati=camp)
if (det(epS.optim$hessian)<0){
vvv<-nlm(logV.epS, c(4,4,1), hessian=T, dati=camp)}
if (det(epS.optim$hessian)<0){
vvv<-nlm(logV.epS, c(-2,2,1), hessian=T, dati=camp)}
t.hat[2]<-epS.optim$estimate[1]
s.hat[2]<-epS.optim$estimate[2]
pot.hat<-epS.optim$estimate[3]
dH12[2]<-sqrt(det(epS.optim$hessian))
nM[2]<- -length(camp)*(log(pot.hat^(1/pot.hat))
+log(gamma(1+1/pot.hat))
+log(s.hat[2]))-sum((abs(camp-t.hat[2]))^pot.hat)
/(pot.hat*(s.hat[2]^pot.hat))

J<-solve(epS.optim$hessian)
var_T.f[2]<-J[1,1]

### CALCOLO PESI SCMV BIEFFICACE ###
alpha<-NULL
for (i in 1:2){
alpha[i]<-sqrt(exp(nM[i])/((s.hat[i]^3)*dH12[i]*var_T.f[i]))
if (alpha[i]=='NaN') alpha[i]<-10^-180
}
nM[1]
nM[2]

```

```

alpha[1]
alpha[2]

### CALCOLO STIMATORE SCMV BIEFFICACE ###
t.bieff<-(alpha[1]*t.hat[1]+alpha[2]*t.hat[2])/(alpha[1]+alpha[2])
stimeE[j]<-t.bieff

#####
### STIMATORE COMPROMESSO BASATO SU CRITERI
# DI SELEZIONE DEL MODELLO ###

mu.est<- epS.optim$estimate[1]
mu.rid<-mean(camp)

vett<-matrix(c(1,0),ncol=1)
vrig<-matrix(c(1,0),ncol=2)
J.10<-c(epS.optim$hessian[3,1],epS.optim$hessian[3,2])
J.00<-matrix(c(epS.optim$hessian[1,1],epS.optim$hessian[1,2],
epS.optim$hessian[2,1],epS.optim$hessian[2,2]),ncol=2,byrow=T)

### CALCOLO I PESI ###
AIC.w<- -2*(logV.epS(c(epS.optim$estimate[1],epS.optim$estimate[2],
epS.optim$estimate[3]),camp) + length(camp)*log(2)+1)
AIC.e<- -2*(logV.epS(c(epS.optim$estimate[1],epS.optim$estimate[2],
2),camp)+ length(camp)*log(2)+0)
W_Z.est<-exp(AIC.w/2)/(exp(AIC.w/2)+exp(AIC.e/2))
W_Z.rid<-exp(AIC.e/2)/(exp(AIC.w/2)+exp(AIC.e/2))
mu<-W_Z.est*mu.est + W_Z.rid*mu.rid
stimeH[j]<-mu
}

```

```
#####
```

```
### GRAFICI ###
```

```
### DISTRIBUZIONE STIMATORE PER FAM. POS. E SCALA ###
```

```
plot(density(stimeE))
```

```
### DISTRIBUZIONE STIMATORE BASATO SU CRIT. SEL. MODELLO ###
```

```
lines(density(stimeH),lty=3)
```

```
### DISTRIBUZIONE ASINOTITICA SOTTO IL MODELLO RIDOTTO ###
```

```
sim1<- rnorm(100000,0, 4)
```

```
sim2<- rnorm(100000,0,1)
```

```
lambda.rid<- 2 + (sim1)/sqrt(length(camp))
```

```
lines(density(lambda.rid),lty=2)
```

Calcolo, distribuzione stimatore per famiglie di posizione e scala (normale + normale di ordine s), confronto con distribuzione stimatore compromesso e distribuzione asintotica sotto il modello esteso

```
library(normalp)
stimeE<-NULL
stimeH<-NULL
omega<-NULL
tau.0<-NULL
K<-NULL

for(j in 1:1000){
  camp<-rnormp(100,2,4,2.1)

  #####
  ### STIMATORE PER FAMIGLIE DI POSIZIONE E SCALA ###
  t.hat<-NULL
  s.hat<-NULL
  dH12<-NULL
  nM<-NULL
  var_T.f<-NULL

  ### SOTTO LA NORMALE ###
  t.hat[1]<-mean(camp)
  s.hat[1]<-sqrt((1/length(camp))*sum((camp-t.hat[1])^2))
  dH12[1]<-sqrt(2)/(s.hat[1]^2)
  nM[1]<- -length(camp)*(1+log(2*pi))/2-length(camp)*log(s.hat[1])
  var_T.f[1]<-1/length(camp)
```

```

### SOTTO LA EP(s) ###
logV.epS<-function(param,dati){
length(camp)*(log(param[3]^(1/param[3]))
+log(gamma(1+1/param[3]))+log(param[2]))
+sum((abs(camp-param[1]))^param[3])
/(param[3]*(param[2]^param[3]))
}
epS.optim<-nlm(logV.epS, c(2,2,2), hessian=T, dati=camp)
if (det(epS.optim$hessian)<0){
  vvv<-nlm(logV.epS, c(4,4,1), hessian=T, dati=camp)}
if (det(epS.optim$hessian)<0){
  vvv<-nlm(logV.epS, c(-2,2,1), hessian=T, dati=camp)}
t.hat[2]<-epS.optim$estimate[1]
s.hat[2]<-epS.optim$estimate[2]
pot.hat<-epS.optim$estimate[3]
dH12[2]<-sqrt(det(epS.optim$hessian))
nM[2]<- -length(camp)*(log(pot.hat^(1/pot.hat))
+log(gamma(1+1/pot.hat))+log(s.hat[2]))
-sum((abs(camp-t.hat[2]))^pot.hat)
/(pot.hat*(s.hat[2]^pot.hat))

J<-solve(epS.optim$hessian)
var_T.f[2]<-J[1,1]

### CALCOLO PESI SCMV BIEFFICACE ###
alpha<-NULL
for (i in 1:2){
alpha[i]<-sqrt(exp(nM[i])/((s.hat[i]^3)*dH12[i]*var_T.f[i]))
if (alpha[i]=='NaN') alpha[i]<-10^-180
}

```

```

nM[1]
nM[2]
alpha[1]
alpha[2]

### CALCOLO STIMATORE SCMV BIEFFICACE ###
t.bieff<-(alpha[1]*t.hat[1]+alpha[2]*t.hat[2])/(alpha[1]+alpha[2])
stimeE[j]<-t.bieff

#####
### STIMATORE COMPROMESSO BASATO SU CRITERI
# DI SELEZIONE DEL MODELLO ###

mu.est<- epS.optim$estimate[1]
mu.rid<-mean(camp)

vett<-matrix(c(1,0),ncol=1)
vrig<-matrix(c(1,0),ncol=2)
J.10<-c(epS.optim$hessian[3,1],epS.optim$hessian[3,2])
J.00<-matrix(c(epS.optim$hessian[1,1],epS.optim$hessian[1,2],
  epS.optim$hessian[2,1],epS.optim$hessian[2,2]),ncol=2,byrow=T)
omega[j]<- J.10%*%J.00%*%vett
tau.0[j]<-vrig%*%(solve(J.00))%*%vett
K[j]<-solve(epS.optim$hessian)[3,3]

### CALCOLO I PESI ###
AIC.w<- -2*(logV.epS(c(epS.optim$estimate[1],epS.optim$estimate[2],
  epS.optim$estimate[3]),camp) + length(camp)*log(2)+1)
AIC.e<- -2*(logV.epS(c(epS.optim$estimate[1],epS.optim$estimate[2],
  2),camp)+ length(camp)*log(2)+0)
W_Z.est<-exp(AIC.w/2)/(exp(AIC.w/2)+exp(AIC.e/2))

```

```

W_Z.rid<-exp(AIC.e/2)/(exp(AIC.w/2)+exp(AIC.e/2))
mu<-W_Z.est*mu.est + W_Z.rid*mu.rid
stimeH[j]<-mu
}

#####
### GRAFICI ###

### DISTRIBUZIONE STIMATORE PER FAM. POS. E SCALA ###
plot(density(stimeE))
### DISTRIBUZIONE STIMATORE BASATO SU CRIT. SEL. MODELLO ###
lines(density(stimeH),lty=3)

### DISTRIBUZIONE ASINOTITICA SOTTO IL MODELLO ESTESO ###
sim1<- rnorm(100000,0, sqrt(mean(tau.0)*length(camp)))
sim2<- rnorm(100000,(0.1/sqrt(mean(K))),1)

lambda.est<- 2 + (sim1 + mean(omega)*sqrt(mean(K))
  *(0.1/sqrt(mean(K))-sim2))/sqrt(length(camp))
lambda.rid<- 2 + (sim1 + 0.1*mean(omega))/sqrt(length(camp))
lambda<-W_Z.est*lambda.est + W_Z.rid*lambda.rid

lines(density(lambda),lty=2)

```


Calcolo e distribuzione stimatore per famiglie di posizione e scala (normale + valori estremi)

```
stime<-NULL

for(j in 1:10000){
  camp<-rt(10,5)
  t.hat<-NULL
  s.hat<-NULL
  dH12<-NULL
  nM<-NULL
  var_T.f<-NULL

  ### SOTTO LA NORMALE ###
  t.hat[1]<-mean(camp)
  s.hat[1]<-sqrt((1/length(camp))*sum((camp-t.hat[1])^2))
  dH12[1]<-sqrt(2)/(s.hat[1]^2)
  nM[1]<- -length(camp)*(1+log(2*pi))/2-length(camp)*log(s.hat[1])
  var_T.f[1]<-1/length(camp)

  ### SOTTO LA VALORI ESTREMI ###
  logV.ve <- function(param, dati){
    length(camp)*log(param[2])-sum(camp-param[1])/param[2]
    + sum(exp((camp-param[1])/param[2]))
  }
  ve.optim<-optim(c(2,2), logV.ve, hessian=T, dati=camp)
  ve.optim$par
  t.hat[2]<-ve.optim$par[1]
  s.hat[2]<-ve.optim$par[2]
  dH12[2]<-sqrt(det(ve.optim$hessian))
  nM[2]<- -length(camp)*log(s.hat[2])+sum(camp-t.hat[2])/s.hat[2]
```

```
-sum(exp((camp-t.hat[2])/s.hat[2]))
J<- solve(ve.optim$hessian)
var_T.f[2]<-J[1,1]

### CALCOLO PESI SCMV BIEFFICACE ###
alpha<-NULL
for (i in 1:2){
alpha[i]<-sqrt( exp(nM[i])/((s.hat[i]^3)*dH12[i]*var_T.f[i]))
# /length(camp)
}

### CALCOLO STIMATORE SCMV BIEFFICACE ###
t.bieff<-(alpha[1]*t.hat[1]+alpha[2]*t.hat[2])/(alpha[1]+alpha[2])
stime[j]<-t.bieff
}

### GRAFICO DELLA DISTRIBUZIONE ###
plot(density(stime))
```

Calcolo e distribuzione stimatore per famiglie di posizione e scala (normale + normale di ordine 4 + valori estremi)

```
stime<-NULL

for(j in 1:10000){
  camp<-rt(10,5)
  t.hat<-NULL
  s.hat<-NULL
  dH12<-NULL
  nM<-NULL
  var_T.f<-NULL

  ### SOTTO LA NORMALE ###
  t.hat[1]<-mean(camp)
  s.hat[1]<-sqrt((1/length(camp))*sum((camp-t.hat[1])^2))
  dH12[1]<-sqrt(2)/(s.hat[1]^2)
  nM[1]<- -length(camp)*(1+log(2*pi))/2-length(camp)*log(s.hat[1])
  var_T.f[1]<-1/length(camp)

  ### SOTTO LA EP(4) ###
  logV.ep4<-function(param,dati){
    length(camp)*log(param[2])
    +sum((abs(camp-param[1]))^4)/(4*(param[2])^4)
  }
  ep4.optim<-nlm(logV.ep4, c(2,2), hessian=T, dati=camp)
  if (det(ep4.optim$hessian)<0){
    ep4.optim<-nlm(logV.ep4, c(10,10), hessian=T, dati=camp)}
  if (det(ep4.optim$hessian)<0){
    ep4.optim<-nlm(logV.ep4, c(-10,10), hessian=T, dati=camp)}
```

```

t.hat[2]<-ep4.optim$estimate[1]
s.hat[2]<-ep4.optim$estimate[2]
dH12[2]<-sqrt(det(ep4.optim$hessian))
nM[2]<- -length(camp)*(3*log(4)/4+log(gamma(5/4))+log(s.hat[2]))
      -sum(abs(camp-t.hat[2])^4)/(4*(s.hat[2])^4)
J<-solve(ep4.optim$hessian)
var_T.f[2]<-J[1,1]

```

```

### SOTTO LA VALORI ESTREMI ###

```

```

logV.ve <- function(param, dati){
  length(camp)*log(param[2])-sum(camp-param[1])/param[2]
  + sum(exp((camp-param[1])/param[2]))
}
ve.optim<-optim(c(2,2), logV.ve, hessian=T, dati=camp)
ve.optim$par
t.hat[3]<-ve.optim$par[1]
s.hat[3]<-ve.optim$par[2]
dH12[3]<-sqrt(det(ve.optim$hessian))
nM[3]<- -length(camp)*log(s.hat[3])+sum(camp-t.hat[3])/s.hat[3]
      -sum(exp((camp-t.hat[3])/s.hat[3]))
J<- solve(ve.optim$hessian)
var_T.f[3]<-J[1,1]

```

```

### CALCOLO PESI SCMV BIEFFICACE###

```

```

alpha<-NULL
for (i in 1:3){
alpha[i]<-sqrt( exp(nM[i])/((s.hat[i]^3)*dH12[i]*var_T.f[i]))
if (alpha[i]=='NaN') alpha[i]<-10^-180
}

```

```

### CALCOLO STIMATORE SCMV BIEFFICACE ###

```

```
t.bieff<- (alpha[1]*t.hat[1] + alpha[2]*t.hat[2]  
  + alpha[3]*t.hat[3])/(alpha[1] + alpha[2]+alpha[3])  
stime[j]<-t.bieff  
}
```

```
### GRAFICO DELLA DISTRIBUZIONE ###
```

```
plot(density(stime),col=2)
```

Calcolo dell'efficienza

```

racc.bieff_ne<-NULL
racc.bieff_nv<-NULL
racc.bieff_nev<-NULL
racc.num<-NULL
racc.AIC_ne<-NULL
racc.AIC_nv<-NULL
racc.AIC_nev<-NULL

### funzioni di log-verosimiglianza ###
logV.norm<- function(param, camp){
  length(camp)*(log(sqrt(2*pi))+log(param[2]))
  + sum((camp-mean(camp))^2)/(2*param[2]^2)
}
logV.ve <- function(param, camp){
  length(camp)*log(param[2])-sum(camp-param[1])/param[2]
  +sum(exp((camp-param[1])/param[2]))
}
logV.ep4<-function(param,dati){
  length(camp)*(log(param[2])+3*log(4)/4 + log(gamma(5/4)))
  + sum((abs(camp-param[1]))^4)/(4*(param[2])^4)
}

for(j in 1:2000){
  camp<-rnorm(5,0,1)

  t.hat<-NULL
  s.hat<-NULL
  dH12<-NULL
  nM<-NULL
  var_T.f<-NULL

```

```
### SOTTO LA NORMALE ###
```

```
t.hat[1]<-mean(camp)
s.hat[1]<-sqrt((1/length(camp))*sum((camp-t.hat[1])^2))
dH12[1]<-sqrt(2)/(s.hat[1]^2)
nM[1]<- -length(camp)*(1+log(2*pi))/2-length(camp)*log(s.hat[1])
var_T.f[1]<-1/length(camp)
```

```
### SOTTO LA EP(4) ###
```

```
ep4.optim<-nlm(logV.ep4, c(0,1), hessian=T, dati=camp)
if(det(ep4.optim$hessian) < 0){
  ep4.optim<-nlm(logV.ep4, c(0,10), hessian=T, dati=camp)}
if(det(ep4.optim$hessian) < 0){
  ep4.optim<-nlm(logV.ep4, c(-5,1), hessian=T, dati=camp)}
t.hat[2]<-ep4.optim$estimate[1]
s.hat[2]<-ep4.optim$estimate[2]
dH12[2]<-sqrt(det(ep4.optim$hessian))
nM[2]<- -length(camp)*(3*log(4)/4+log(gamma(5/4))+log(s.hat[2]))
  -sum(abs(camp-t.hat[2])^4)/(4*(s.hat[2])^4)
J<-solve(ep4.optim$hessian)
var_T.f[2]<-J[1,1]
```

```
### SOTTO LA VALORI ESTREMI ###
```

```
logV.ve <- function(param, dati){
  length(camp)*log(param[2])-sum(camp-param[1])/param[2]
  +sum(exp((camp-param[1])/param[2]))
}
ve.optim<-optim(c(0,1), logV.ve, hessian=T, dati=camp)
if(det(ve.optim$hessian) < 0){
  ve.optim<-nlm(logV.ve, c(0,10), hessian=T, dati=camp)}
if(det(ve.optim$hessian) < 0){
```

```

ve.optim<-nlm(logV.ve, c(-5,1), hessian=T, dati=camp)}
ve.optim$par
t.hat[3]<-ve.optim$par[1]
s.hat[3]<-ve.optim$par[2]
dH12[3]<-sqrt(det(ve.optim$hessian))
nM[3]<- -length(camp)*log(s.hat[3])+sum(camp-t.hat[3])/s.hat[3]
      -sum(exp((camp-t.hat[3])/s.hat[3]))
J<- solve(ve.optim$hessian)
var_T.f[3]<-J[1,1]

### CALCOLO PESI CMLE BIEFFICACE###
alpha<-NULL
for (i in 1:3){
alpha[i]<-sqrt( exp(nM[i])/((s.hat[i]^3)*dH12[i]*var_T.f[i]))
if (alpha[i]=='NaN') alpha[i]<-10^-180
}

### CALCOLO STIMATORI CMLE BIEFFICACI ###
t.bieff_ne<-(alpha[1]*t.hat[1]+alpha[2]*t.hat[2])/(alpha[1]+alpha[2])
racc.bieff_ne[j]<-(t.bieff_ne)^2
t.bieff_nv<-(alpha[1]*t.hat[1]+alpha[3]*t.hat[3])/(alpha[1]+alpha[3])
racc.bieff_nv[j]<-(t.bieff_nv)^2
t.bieff_nev<- (alpha[1]*t.hat[1]+alpha[2]*t.hat[2]
              + alpha[3]*t.hat[3])/(alpha[1]+alpha[2]+alpha[3])
racc.bieff_nev[j]<-(t.bieff_nev)^2

racc.num[j]<-mean(camp)^2

### CALCOLO STIMATORE COMPROMESSO CON PESI AIC ###
# calcolo i pesi
AIC.e<- -2*(logV.ep4(ep4.optim$estimate,camp)+length(camp)*log(2)+1)

```



```

AIC.n<- -2*(logV.norm(c(t.hat[1],s.hat[2]),camp)+1)
AIC.v<- -2*(logV.ve(c(t.hat[3],s.hat[3]),camp)+1)

# NORMALE + EP4
W_Z.uno<-exp(AIC.n/2)/(exp(AIC.n/2)+exp(AIC.e/2))
W_Z.due<-exp(AIC.e/2)/(exp(AIC.n/2)+exp(AIC.e/2))
mu<-W_Z.uno*t.hat[1] + W_Z.due*t.hat[2]
racc.AIC_ne[j]<-mu^2

# NORMALE + VE
W_Z.uno<-exp(AIC.n/2)/(exp(AIC.n/2)+exp(AIC.v/2))
W_Z.tre<-exp(AIC.v/2)/(exp(AIC.n/2)+exp(AIC.v/2))
mu<-W_Z.uno*t.hat[1] + W_Z.tre*t.hat[3]
racc.AIC_nv[j]<-mu^2

# NORMALE + EP4 + VE
W_Z.uno<-exp(AIC.n/2)/(exp(AIC.n/2)+exp(AIC.e/2)+exp(AIC.v/2))
W_Z.due<-exp(AIC.e/2)/(exp(AIC.n/2)+exp(AIC.e/2)+exp(AIC.v/2))
W_Z.tre<-exp(AIC.v/2)/(exp(AIC.n/2)+exp(AIC.e/2)+exp(AIC.v/2))
mu<-W_Z.uno*t.hat[1] + W_Z.due*t.hat[2] + W_Z.tre*t.hat[3]
racc.AIC_nev[j]<-mu^2
}

# MSE MEDIA
sum(racc.num)/length(racc.num)
# MSE STIM BIEFF NORM + EP(4)
sum(racc.bieff_ne)/length(racc.bieff_ne)
# MSE STIM BIEFF NORM + V.E.
sum(racc.bieff_nv)/length(racc.bieff_nv)
# MSE STIM BIEFF NORM + EP(4) + V.E.
sum(racc.bieff_nev)/length(racc.bieff_nev)

```

```
# MSE STIM PESI AIC NORM + EP(4)
sum(racc.AIC_ne)/length(racc.AIC_ne)
# MSE STIM PESI AIC NORM + VE
sum(racc.AIC_nv)/length(racc.AIC_nv)
# MSE STIM PESI AIC NORM + EP(4) + VE
sum(racc.AIC_nev)/length(racc.AIC_nev)
```

Bibliografia

- Buckland S. T., Burnham K. P. & Augustin N. H. (1997). Model selection: an integral part of inference. *Biometrics*, **53**, 603-618.
- Burnham K. P. & Anderson D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretical Approach*. 2a ed. Springer-Verlag, New York.
- Davison A. C. & Hinkley D.V. (1997). *Bootstrap Methods and their Applications*. Cambridge University Press, Cambridge.
- Easton G. S. (1991). Compromise maximum likelihood estimators for location. *Journal of the American Statistician Association*, **86**, 1051-1064.
- Hjort N. L. & Claeskens G. (2003). Frequentist model average estimators. *Journal of the American Statistician Association*, **98**, 879-899.
- Hoeting J. A., Madigan D., Raftery A. E. & Volinsky C. T. (1999). Bayesian model averaging: a tutorial. *Statistical Science*, **19**, 382-417.
- Pace L. & Salvan A. (1996). *Teoria della Statistica*. Cedam, Padova.
- Severini T.A. (1994). Nonparametric conditional inference for a location parameter. *Journal of the Royal Statistical Society*, **56**, 353-362.
- Ventura L. (1998). Higher-order approximations for Pitman estimators and for optimal compromise estimators. *Canadian Journal of Statistics*, **6**, 1, 49-55.