

UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

Tesi di Laurea Triennale in Statistica e Tecnologie Informatiche



**Stima del parametro di forma della normale
asimmetrica tramite verosimiglianze profilo
penalizzate**

Estimator of the shape parameter of the Skew-normal by
penalized profile likelihoods

Relatore:

Chiar.ma Prof.ssa

Laura Ventura

Laureando:

Marco Serafin

573285

Anno Accademico 2010-2011

Indice

Introduzione	1
1 La normale asimmetrica	3
1.1 Il parametro di forma	6
1.2 I momenti	8
1.3 Aspetti statistici	9
2 Inferenza	13
2.1 Stima di α in 0	14
2.1.1 La parametrizzazione centrata	15
2.2 Monotonia di $l_p(\alpha)$	16
2.2.1 Soluzioni al problema della monotonia	18
2.2.2 Una verosimiglianza profilo penalizzata per α	20
3 Applicazioni	23
3.1 Studio di simulazione	25
3.1.1 Commenti	26
A Codice R	31
A.1 Algoritmo di massimizzazione	31
A.2 Codice simulazione	34
Bibliografia	39

Introduzione

La distribuzione normale, una delle distribuzioni più utilizzate in ambito statistico e non solo, come sappiamo si presta molto bene alla rappresentazione dell'andamento di diversi fenomeni. E' lo strumento più appropriato, infatti, per descrivere alcune variabili la cui distribuzione è simmetrica, con un singolo picco ed a forma di campana.

Esistono però situazioni in cui la distribuzione normale non risulta essere del tutto soddisfacente, in particolare quando l'insieme dei dati osservati presenta una distribuzione asimmetrica: nasce quindi l'esigenza di poter disporre di una famiglia di distribuzioni che sia più flessibile nel descrivere questi fenomeni.

La normale perde allora il ruolo centrale che siamo abituati ad attribuirle, e può essere vista come un caso limite di una famiglia di distribuzioni più generale: la famiglia della normale asimmetrica (Azzalini, 1985). Si tratta di una classe di distribuzioni la cui legge di probabilità è descritta non solo da parametri di posizione e di scala, che nella normale coincidono con la media e la varianza, ma anche da un terzo parametro che regola la forma della densità e che permette di trattare anche i casi in cui è evidente una certa asimmetria in distribuzione.

Nel Capitolo 1 viene descritta brevemente la distribuzione normale asimmetrica, evidenziandone le principali peculiarità e analogie con la distribuzione gaussiana. Si illustra quindi il comportamento del parametro di forma, per andare poi ad esplicitare i momenti e accennare i principali aspetti statistici che la caratterizzano.

Nel Capitolo 2 vengono introdotti i problemi inferenziali riguardanti la stima di massima verosimiglianza del parametro di forma α . Infatti, l'anda-

mento della funzione di log-verosimiglianza manifesta caratteristiche spiacevoli, in particolare in un intorno di $\alpha=0$. Si mostra poi come, utilizzando la verosimiglianza profilo per α , si possa avere a che fare con una funzione monotona crescente e che quindi sia difficoltoso individuare un punto di massimo. Vengono allora proposte delle possibili soluzioni per arginare queste complicazioni.

Il Capitolo 3 mostra un'esemplificazione su un data-set che è stato oggetto di diversi articoli in materia, caratterizzato dalle problematiche inferenziali illustrate in precedenza: il *frontier data-set*. Viene applicata una verosimiglianza profilo penalizzata al fine di ottenere una stima di massima verosimiglianza e gli intervalli di confidenza per il parametro di forma α . Vengono poi mostrati i risultati di uno studio di simulazione effettuato su campioni aventi numerosità differenti: da essi si ricavano le stime di massima verosimiglianza per α utilizzando tre diversi metodi a confronto, tra cui le penalizzazioni alla verosimiglianza profilo descritte nel Capitolo 2.

In Appendice, infine, viene riportato il codice R sviluppato per ottenere i risultati del Capitolo 3, ossia un algoritmo di ottimizzazione che lavora sulla verosimiglianza profilo penalizzata e la funzione per mezzo della quale si ottengono i risultati dello studio di simulazione.

Capitolo 1

La normale asimmetrica

L'intuizione che fosse opportuno rendere più elastica la distribuzione normale, al fine di poter sintetizzare anche fenomeni che presentano perturbazioni esterne, è da attribuire a Fernando de Helguero (1908):

Il compito della statistica [...] non consiste solo nel determinare la legge di dipendenza dei diversi valori ed esprimerla con pochi numeri, ma anche nel fornire un aiuto allo studioso che vuole cercare le cause della variazione.

...le curve teoriche studiate dal PEARSON e dall'EDGEWORTH [...] mentre danno con molta approssimazione la legge di variazione, a mio avviso sono difettose in quanto [...] nulla ci fanno sapere sulla legge di dipendenza, quasi nulla sulla relazione con la curva normale.

Io penso che miglior aiuto per lo studioso potrebbero essere delle equazioni che supponessero una perturbazione della variabilità normale per opera di cause esterne.

Dopo alcuni decenni, questo spunto viene concretizzato in un articolo redatto da Azzalini (1985), il quale introduce una classe di distribuzioni che include quella normale: la distribuzione normale asimmetrica, indicata con SN (dall'acronimo inglese *skew-normal*).

La distribuzione di probabilità normale asimmetrica è un caso più generico della ben nota distribuzione gaussiana. La generalizzazione avviene

per mezzo di un parametro aggiuntivo α , detto parametro di forma, il quale determina il livello di asimmetria di ogni specifica funzione di densità.

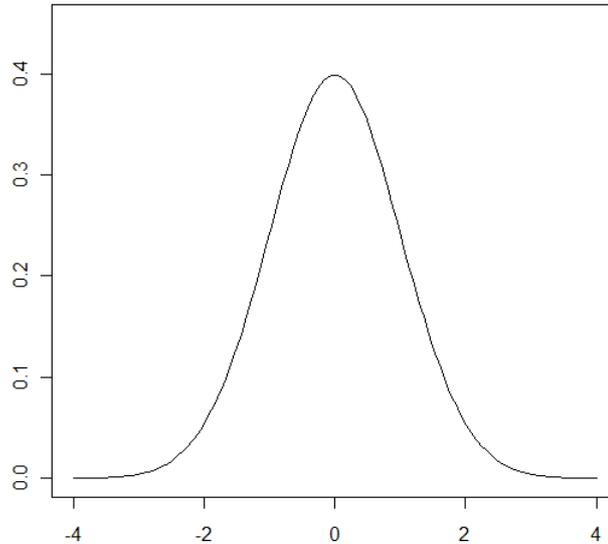


Figura 1.1: Funzione di densità SN con parametri $\mu=0$, $\sigma=1$, $\alpha=0$, equivalente ad una normale standard.

Se consideriamo una variabile casuale Z avente distribuzione normale asimmetrica, che d'ora in poi indichiamo con $Z \sim \text{SN}(\alpha)$, avremo quindi una densità esplicitata in una maniera che in parte rievoca quella della normale standard, opportunamente ponderata con una funzione che va ad esprimerne l'effetto di asimmetria della distribuzione stessa. Essa è data da

$$f(z) = 2 \phi(z) \Phi(\alpha z), \quad z \in \mathbb{R}, \quad (1.1)$$

dove α è il parametro di forma che regola l'intensità dell'asimmetria, mentre

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \text{e} \quad \Phi(\alpha z) = \int_{-\infty}^{\alpha z} \phi(t) dt$$

indicano, rispettivamente, la funzione di densità e la funzione di ripartizione di una normale standard.

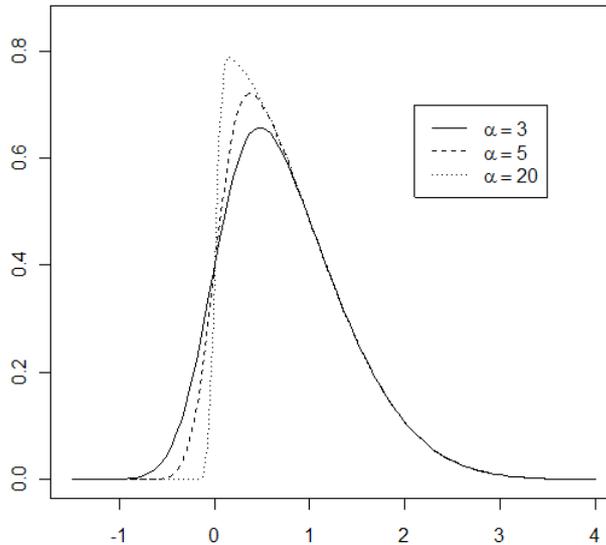


Figura 1.2: Funzione di densità SN al variare di $\alpha > 0$.

Il ruolo del parametro di forma e di scala, le due quantità chiave in un modello gaussiano, resta naturalmente inalterato; va però precisato che, essendo presente ora un'asimmetria, questi non coincidono più con la media e la varianza della distribuzione stessa. Andandoli a introdurre nel modello (1.1), possiamo definire la variabile $Y = \mu + \sigma Z$, avente densità

$$f(y; \mu, \sigma, \alpha) = \frac{2}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right) \Phi\left(\alpha \frac{y - \mu}{\sigma}\right), \quad y \in \mathbb{R}, \quad (1.2)$$

e scriveremo $Y \sim \text{SN}(\mu, \sigma, \alpha)$. Per quanto riguarda i domini dei parametri, abbiamo che $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$ e infine α può avere valori nell'intervallo $(-\infty, +\infty)$.

Da un punto di vista teorico, la classe normale asimmetrica ha il vantaggio di essere matematicamente trattabile e di godere di molte proprietà tipiche della distribuzione normale, tra cui:

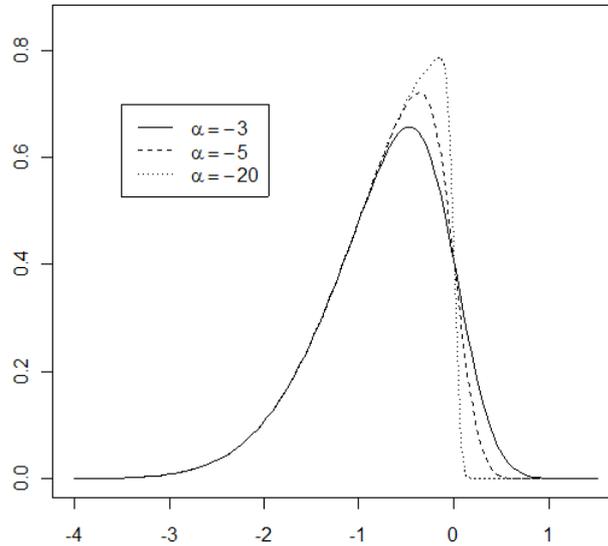
- (a) se $Z \sim \text{SN}(\alpha)$, allora $-Z \sim \text{SN}(-\alpha)$;
- (b) se $Z \sim \text{SN}(\alpha)$, allora $Z^2 \sim \chi_1^2$;
- (c) $1 - \Phi(-z; \alpha) = \Phi(z; -\alpha)$
- (d) la funzione di densità è fortemente unimodale; in particolare il logaritmo della densità è una funzione concava.

Un caso multivariato della normale asimmetrica è stato trattato dallo stesso Azzalini in collaborazione con altri studiosi: si vedano a tal proposito i lavori di Azzalini e Dalla Valle (1996), e quelli di Azzalini e Capitanio (1999).

1.1 Il parametro di forma

Prendendo come riferimento la gaussiana, è abbastanza intuitiva l'incidenza del parametro α sulla forma della funzione di densità. In particolare è interessante notare come:

- per $\alpha = 0$ si ottiene $Y \sim N(\mu, \sigma^2)$; i parametri di posizione μ e di scala σ coincidono allora con la media e la radice della varianza della distribuzione (si veda la Figura 1.1).
- al crescere in segno positivo del parametro di forma, emerge una sempre più evidente asimmetria positiva in distribuzione; in particolare avremo un addensamento delle osservazioni in corrispondenza dei valori più bassi, e una coda che si prolunga verso i valori più grandi (vedi Figura 1.2).
- al decrescere verso valori negativi di α si ottengano invece crescenti asimmetrie negative, caratterizzate da curve di densità che presentano code via via più pesanti verso i valori più bassi, e concentrano il grosso delle osservazioni in corrispondenza dei valori più grandi (Figura 1.3).

Figura 1.3: Funzione di densità SN al variare di $\alpha < 0$.

- per α tendente a infinito, si raggiunge come caso limite una distribuzione completamente asimmetrica, che converge alla così detta funzione di densità “semi-normale”: in Figura 1.4 sono illustrati i comportamenti della curva nel caso in cui il parametro di forma tende a $\pm \infty$.

E' importante sottolineare che, a differenza di altre distribuzioni proposte in letteratura, la famiglia SN consente di passare dal caso simmetrico al caso asimmetrico con continuità, modificando solo il valore di un parametro. E' doverosa comunque una certa attenzione nell'interpretazione dei parametri caratterizzanti la densità poichè, in maniera analoga a quanto accade per i parametri di posizione e di scala, il parametro di forma che regola l'asimmetria non coincide con il coefficiente di asimmetria della distribuzione, come avremo modo di approfondire in seguito.

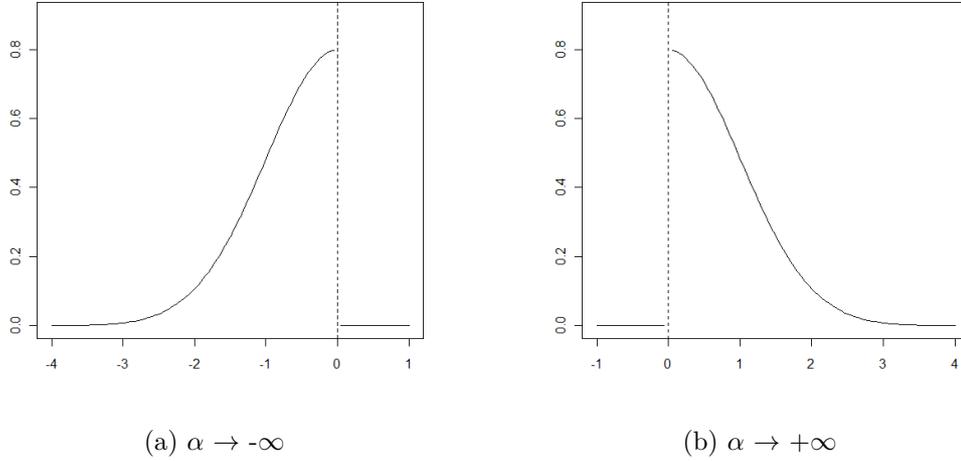


Figura 1.4: Casi limite della funzione di densità SN per $|\alpha| \rightarrow \infty$.

1.2 I momenti

In virtù della proprietà (b), i momenti di ordine pari di $Z \sim \text{SN}(\alpha)$ sono uguali a quelli di ordine pari di una normale standard. Per determinare i momenti dispari, invece, si sfrutta il seguente risultato:

Lemma 1. *Se Z è una variabile casuale $N(0,1)$, allora*

$$E\{\Phi(hZ+k)\} = \Phi\{k/\sqrt{(1+h^2)}\}$$

per h, k valori arbitrari $\in \mathbb{R}$.

Dal Lemma 1 segue che la funzione generatrice dei momenti della SN risulta

$$M(t) = 2 \exp(t^2/2) \Phi(\delta t), \quad (1.3)$$

dove il parametro δ è dato da

$$|\delta| = \sqrt{\frac{\pi}{2}} \frac{|\hat{\gamma}_1|^{2/3}}{|\hat{\gamma}_1|^{2/3} + ((4-\pi)/2)^{2/3}}$$

ed è legato al parametro di forma α tramite la relazione

$$\delta = \frac{\alpha}{\sqrt{1+\alpha^2}} \quad \Leftrightarrow \quad \alpha = \frac{\delta}{\sqrt{1-\delta^2}}.$$

A partire dalla (1.3), si possono definire tutti i momenti di una variabile casuale $SN(\alpha)$. Si trova

$$\begin{aligned}\mu_z &= E(Z) = \sqrt{\frac{2}{\pi}}\delta \\ \sigma_z^2 &= Var(Z) = 1 - \frac{2}{\pi}\delta^2 \\ \gamma_1 &= Asym(Z) = \frac{4 - \pi}{2} \text{sign}(\alpha) \left[\frac{\{E(Z)\}^2}{Var(Z)} \right]^{\frac{3}{2}} \\ \gamma_2 &= Kurt(Z) = 2(\pi - 3) \left[\frac{\{E(Z)\}^2}{Var(Z)} \right]^2\end{aligned}$$

che rappresentano, rispettivamente, media, varianza, coefficiente di asimmetria e di curtosi standardizzati. Si noti che, per costruzione, gli indici γ_1 e γ_2 hanno un campo di variazione limitato; in particolare, si ha $\gamma_1 \in (-0.995, 0.995)$ e $\gamma_2 \in (-0.869, 0.869)$. Di conseguenza, nel caso di campioni con marcata asimmetria o curtosi, è doverosa una certa accortezza nel trattare i dati.

1.3 Aspetti statistici

Consideriamo il modello descritto dalla funzione di densità (1.2). Dato un campione casuale semplice di numerosità n , la funzione di verosimiglianza per $\theta=(\mu,\sigma,\alpha)$ è data da

$$L(\theta; y) = L(\mu, \sigma, \alpha; y) = 2^n \sigma^{-n} \prod_{i=1}^n \phi\left(\frac{y_i - \mu}{\sigma}\right) \Phi\left(\alpha \frac{y_i - \mu}{\sigma}\right). \quad (1.4)$$

Derivando la (1.4) si ottengono le seguenti equazioni di verosimiglianza

$$\begin{cases} \sum_{i=1}^n z_i - \alpha \sum_{i=1}^n \beta_i = 0 \\ \sum_{i=1}^n z_i^2 - \alpha \sum_{i=1}^n \beta_i z_i - n = 0 \\ \sum_{i=1}^n \beta_i z_i = 0, \end{cases}$$

dove $\beta_i = \frac{\phi(\alpha z_i)}{\Phi(\alpha z_i)}$ e $z_i = \frac{y_i - \mu}{\sigma}$, $i=1, \dots, n$. Questo sistema di equazioni può essere semplificato come segue (Chiogna, 2005)

$$\begin{cases} \hat{\mu} = \bar{y} - \hat{\alpha} \hat{\sigma} \hat{\beta} \\ \hat{\sigma} = s^2 (1 - \hat{\alpha}^2 \hat{\beta}^2)^{-1} \\ \sum_{i=1}^n \beta_i z_i = 0, \end{cases}$$

dove \bar{y} e s^2 sono la media e la varianza campionaria, mentre $\hat{\beta} = \frac{1}{n} \sum_{i=1}^n \beta_i$.

Per quanto riguarda la stima di massima verosimiglianza $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\alpha})$ non è possibile individuare una espressione in forma chiusa; bisogna utilizzare una procedura di massimizzazione numerica in funzione delle osservazioni campionarie.

La matrice di informazione di Fisher per i tre parametri (μ, σ, α) è facilmente calcolabile, ed è data da (Azzalini, 1985)

$$i(\theta) = \begin{pmatrix} (1 + \alpha^2 a_0)/\sigma^2 & \left(E(Z) \frac{1+2\alpha^2}{1+\alpha^2} + \alpha^2 a_1 \right) / \sigma^2 & \left(\frac{\sqrt{2/\pi}}{(1+\alpha^2)^{3/2}} - \alpha a_1 \right) / \sigma \\ \left(E(Z) \frac{1+2\alpha^2}{1+\alpha^2} + \alpha^2 a_1 \right) / \sigma^2 & (2 + \alpha^2 a_2)/\sigma^2 & -\alpha a_2/\sigma \\ \left(\frac{\sqrt{2/\pi}}{(1+\alpha^2)^{3/2}} - \alpha a_1 \right) / \sigma & -\alpha a_2/\sigma & a_2 \end{pmatrix},$$

dove

$$a_k = a_k(\alpha) = E \left\{ Z^k \left(\frac{\phi(\alpha Z)}{\Phi(\alpha Z)} \right)^2 \right\}, \quad \text{per } k = 0, 1, 2;$$

tale quantità, in genere, viene valutata numericamente. Sono comunque state proposte delle approssimazioni per a_k : si veda l'approssimazione di Bayes e Branco (2007).

Una particolarità di $i(\theta)$ è che, per $\alpha \rightarrow 0$, essa diventa singolare e quindi non invertibile (Azzalini, 1985). Solo se la procedura di stima viene eseguita in condizioni di regolarità, è possibile utilizzare l'usuale risultato asintotico

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, i(\theta)^{-1}).$$

Nel modello che stiamo trattando, quindi, non è lecito l'utilizzo dei ri-

sultati sulla distribuzione asintotica delle stime di massima verosimiglianza; come vedremo in seguito, sarà necessario riparametrizzare il modello in modo da risolvere queste problematiche.

Capitolo 2

Inferenza

La distribuzione normale asimmetrica, come già accennato nel capitolo precedente, gode di buone proprietà, che le consentono di essere facilmente trattata dal punto di vista matematico. Dal lato pratico, tuttavia, subentrano diverse problematiche legate alle stime dei parametri che la caratterizzano, in particolare quelle relative allo stimatore del parametro di forma α .

Come sappiamo, per ottenere delle stime per θ a partire dai dati osservati possiamo avvalerci di due possibili metodi: il metodo dei momenti oppure il metodo della massima verosimiglianza.

Consideriamo dapprima il metodo dei momenti al fine di stimare il parametro di forma α . Ricordiamo che l'indice di asimmetria ha un forte vincolo, e cioè $|\gamma_1| < 0.995$. E' evidente, allora, che l'affidabilità di questo metodo viene meno ogni qualvolta il campione oggetto di studio presenta un coefficiente al di fuori del range di ammissibilità.

Il metodo della massima verosimiglianza, d'altra parte, non è esente da complicazioni. Vi sono, infatti, due problemi principali, dei quali uno è già stato accennato nel capitolo precedente:

1. per $\alpha = 0$, la funzione di verosimiglianza (1.4) presenta un punto di stazionarietà, e la matrice di informazione osservata risulta di conseguenza non invertibile;
2. nei così detti “dati di frontiera”, la verosimiglianza profilo per il parametro di forma mostra un andamento monotono che non permette di

ricavare la stima di massima verosimiglianza.

Esiste anche un supporto software per lavorare e fare inferenza su distribuzioni SN, per il programma statistico R. Si tratta della libreria “sn”, contenente un certo numero di funzioni inerenti le distribuzioni skew-normal e skew-t, che vanno dalla generazione di campioni alle stime di massima verosimiglianza.

Vediamo ora nel dettaglio i due punti 1. e 2. e alcune soluzioni proposte per ovviare questi problemi.

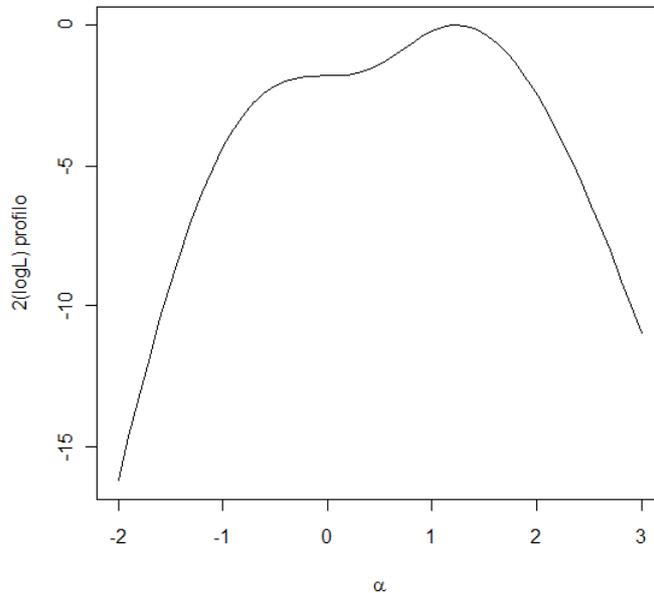


Figura 2.1: Dati AIS sul peso: log-verosimiglianza profilo normalizzata per α (*parametrizzazione diretta*).

2.1 Stima di α in 0

Da un punto di vista teorico, il problema della stima dei parametri e della singolarità della matrice di informazione di Fisher si presenta in corrispondenza di $\alpha=0$, per $n \rightarrow \infty$. In molte situazioni pratiche, tuttavia, si è

riscontrata questa particolarità anche in un intorno più o meno ampio dello zero, e per campioni di dimensioni anche modeste.

L'origine di queste difficoltà sta nel fatto che esiste un punto di stazionarietà in corrispondenza di $\alpha=0$, tale da non permettere alla curva della log-verosimiglianza profilo per α di manifestare l'andamento quadratico tipico di questo genere di funzioni: la log-verosimiglianza risulta essere, infatti, bimodale.

Per avere un'idea più chiara del comportamento della funzione di log-verosimiglianza profilo si veda la Figura 2.1, ottenuta facendo inferenza su un insieme di dati trattato in diversi articoli inerenti l'argomento della normale asimmetrica, ossia il data-set AIS (Australian Institute of Sport) reperibile nella libreria 'sn' del software statistico R. Si tratta delle realizzazioni di 13 variabili, rilevate su 202 atleti australiani. Il grafico fa riferimento allo studio della sola variabile peso corporeo in Kg, in quanto è la variabile in questo data-set che meglio si presta ad essere rappresentata con una densità normale asimmetrica.

2.1.1 La parametrizzazione centrata

Al fine di affrontare il primo dei problemi appena menzionati riguardo la stima dei parametri (μ, σ, α) , Azzalini (1985) propone una soluzione: una riparametrizzazione del modello normale asimmetrico.

Partendo dalla così-detta *parametrizzazione diretta* (indicata con DP dall'acronimo in lingua inglese) usata fino ad ora, per mezzo dell'identità

$$Y = \mu + \sigma Z = \xi + \omega \left(\frac{Z - E(Z)}{\sqrt{Var(Z)}} \right), \quad (2.1)$$

dove Z ha distribuzione $SN(\alpha)$ come descritto nella (1.1), si ottiene la *parametrizzazione centrata* (o CP), caratterizzata dai parametri (ξ, ω, γ_1) che ora coincidono con la media, la varianza e il coefficiente di asimmetria della distribuzione stessa. Come si vede dalla (2.1), il punto chiave della trasformazione del modello consiste nel considerare la variabile Z centrandola rispetto alla sua media e al suo scarto quadratico.

Possiamo ora esplicitare le espressioni dei momenti che si ottengono nel nuovo modello riparametrizzato:

$$\begin{aligned}\xi &= E(Y) = \mu + \sigma\mu_z \\ \omega^2 &= \text{Var}(Y) = \sigma^2(1 - \mu_z^2) \\ \gamma_1 &= \frac{E\{(Y - E(Y))^3\}}{\text{Var}(Y)^{3/2}} = \frac{4 - \pi}{2} \frac{\mu_z^3}{(1 - \mu_z^2)^{3/2}} \\ \gamma_2 &= \frac{E\{(Y - E(Y))^4\}}{\text{Var}(Y)^2} = 2(\pi - 3) \frac{\mu_z^4}{(1 - \mu_z^2)^2},\end{aligned}$$

dove μ_z è la media della variabile Z e γ_1 è l'indice di asimmetria, già ottenuti nel precedente calcolo dei momenti.

La *parametrizzazione centrata* offre quindi il vantaggio di fornire una più chiara interpretazione dei parametri caratterizzanti la distribuzione: come nel modello normale, ora ξ e ω^2 rappresentano la vera media e varianza della distribuzione, mentre il coefficiente di asimmetria γ_1 si ricollega in maniera più immediata al grado di asimmetria manifestato in distribuzione di quanto facesse il parametro α nella *parametrizzazione diretta*.

Dal punto di vista inferenziale la *parametrizzazione centrata* elimina le spiacevoli caratteristiche evidenziate prima riguardo la fase di stima: non è più presente, infatti, alcun punto di stazionarietà nella funzione di verosimiglianza per il nuovo parametro di forma γ_1 , come si vede ad esempio in Figura 2.2.

Anche il problema della singolarità della matrice di informazione attesa viene risolto tramite la *parametrizzazione centrata*, rendendo possibile l'utilizzo delle usuali metodiche dell'inferenza asintotica, come ad esempio la costruzione di intervalli di confidenza alla Wald.

2.2 Monotonia di $l_p(\alpha)$

Per quanto riguarda la monotonia della funzione di log-verosimiglianza profilo per il parametro di forma α , con la *parametrizzazione centrata*, come abbiamo visto, si risolvono le complicazioni che si presentano ogni qualvolta

si vuole stimare questo parametro, in tutte le circostanze riguardanti una variabile aleatoria distribuita come una normale asimmetrica.

Solo in alcuni casi, invece, si ha a che fare con campioni che, nonostante siano caratterizzati da un'asimmetria campionaria appartenente al range di ammissibilità $(-0.995, 0.995)$, manifestano una stima di massima verosimiglianza per α che va a cadere sulla frontiera dello spazio parametrico ammissibile; da cui il nome con cui vengono indicati questo tipo di dati, si parla infatti di “dati di frontiera”.

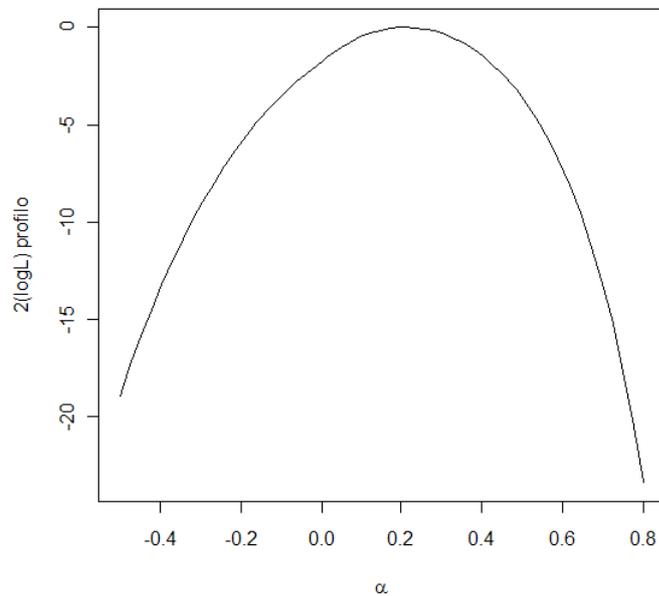


Figura 2.2: Dati AIS sul peso: log-verosimiglianza profilo normalizzata per γ_1 (parametrizzazione centrata).

In queste situazioni, andando a considerare la log-verosimiglianza profilo per il parametro di interesse, si nota una caratteristica spiacevole in fase di stima: l'andamento della funzione $l_p(\alpha)$ è di tipo monotono crescente. Questo preclude la possibilità di ottenere una stima finita per il parametro α .

Non è chiaro se questo comportamento sia imputabile in maniera intrinseca alla natura dei dati, anche se sembra manifestarsi con maggior frequen-

za per campioni di numerosità ridotta. Alcune spiegazioni formali possono essere:

- questo accade quando la stima di massima verosimiglianza per il parametro di posizione μ tende a identificarsi in un valore inferiore alla più piccola osservazione nel campione studiato (Sartori, 2006);
- supponiamo che tutte le osservazioni di un campione, derivanti da una variabile casuale $Y \sim SN(0, 1, \alpha)$, con $\alpha > 0$, siano positive: in questo caso, la log-verosimiglianza risulta essere una funzione crescente di α quando le osservazioni hanno tutte segno positivo. Quindi il metodo della massima verosimiglianza produce una stima tendente a infinito, considerando i dati come se provenissero da una mezza normale (Dalla Valle, 2004).

Sfortunatamente, però, questo problema sorge anche quando non tutte le osservazioni hanno segno positivo, ed è interessante notare che il raggiungimento della frontiera riguarda solo il parametro di forma e non quelli di posizione e scala (Pewsey, 2000).

Una esemplificazione di questa problematica può essere proposta andando a considerare un altro data-set, diverso da quello considerato in precedenza, anch'esso trattato in diversi articoli inerenti le difficoltà di stima tipiche della normale asimmetrica: si tratta del data-set “frontier”, anch'esso contenuto nella libreria “sn” di R. Si tratta di un campione ottenuto da una variabile aleatoria distribuita secondo la legge di probabilità $SN(0,1,5)$, che presenta la caratteristica di illustrare in maniera evidente la situazione fin qui descritta. A tal proposito si veda la Figura 2.3, che riporta l'andamento della funzione di log-verosimiglianza profilo per il parametro α .

2.2.1 Soluzioni al problema della monotonia

In merito a questo tipo di complicazione, sono stati proposti diversi metodi per rendere stimabile in maniera finita il parametro che regola la forma della distribuzione.

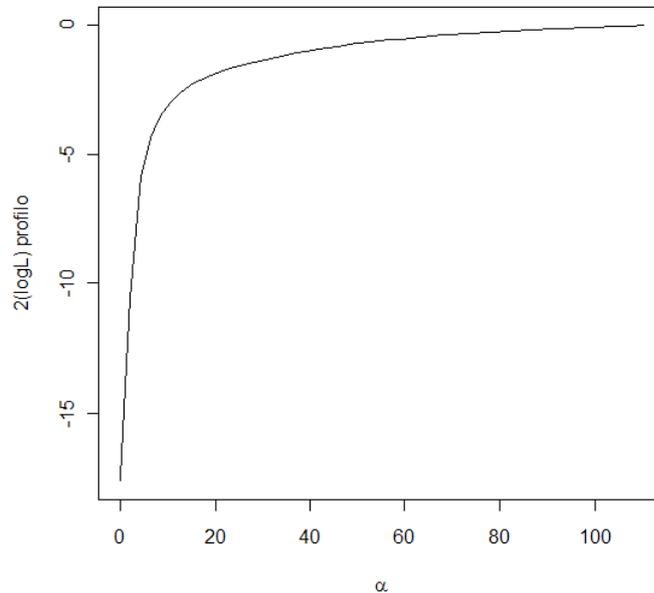


Figura 2.3: Dataset Frontier: log-verosimiglianza profilo normalizzata per α .

Una prima soluzione, proposta da Azzalini e Capitanio (1999), consiste nell'avviare l'algoritmo di ottimizzazione e farlo terminare quando si osserva che la log-verosimiglianza raggiunge un valore tale da non manifestare successive crescite significative della funzione stessa. Il valore così determinato costituisce la stima di massima verosimiglianza. Si tratta, tuttavia, di una metodica piuttosto arbitraria, in quanto non è definito in maniera univoca un criterio di arresto, e la natura dei dati potrebbe incidere molto sullo stimatore, fornendo stime poco attendibili.

Una soluzione alternativa consiste nell'agire sulle equazioni di verosimiglianza tramite una funzione penalizzante in α , grazie alla quale si passa da un contesto in cui la funzione punteggio non ha soluzioni, in quanto sempre positiva, a una situazione in cui esiste almeno un punto α^* tale da soddisfare l'uguaglianza a zero delle equazioni di verosimiglianza. Definiamo allora l'equazione di verosimiglianza profilo penalizzata come

$$U_P^*(\alpha) = l'_P(\alpha) + M(\alpha),$$

dove $l'_P(\alpha)$ è la funzione punteggio profilo per il parametro di forma, e $M(\alpha)$ è la correzione che permette di ottenere stime finite per il parametro di interesse. Se si desidera approfondire l'argomento, si veda il lavoro di Sartori (2006).

Un'ulteriore possibilità per affrontare le complicazioni relative alla monotonia, ed è quello su cui si concentra principalmente l'attenzione in questo documento, è stato avanzato da Ventura e Racugno (2011). La prossima sezione è dedicata alla descrizione di questo argomento.

2.2.2 Una verosimiglianza profilo penalizzata per α

La proposta è quella di considerare una verosimiglianza profilo penalizzata, grazie alla quale la funzione da massimizzare per ottenere la stima del parametro di interesse assume un andamento desiderabile. Andremo pertanto a considerare una pseudo-verosimiglianza per α , descritta come

$$L_{mp}^*(\alpha) = L_{mp}(\alpha)M(\alpha), \quad (2.2)$$

dove $M(\alpha)$ è il fattore di aggiustamento che modifica l'andamento della funzione. Per quanto riguarda $L_{mp}(\alpha)$, essa è la funzione di verosimiglianza profilo modificata per il parametro d'interesse, definita come (Barndorff-Nielsen, 1983)

$$L_{mp}(\alpha) = L_p(\alpha) \frac{1}{\hat{\sigma}_\alpha} |j_{\eta\eta}(\alpha, \hat{\eta}_\alpha)|^{-1/2},$$

dove $L_p(\alpha) = L(\alpha, \hat{\eta}_\alpha)$ è la verosimiglianza profilo per α , $\hat{\eta}_\alpha = (\hat{\mu}_\alpha, \hat{\sigma}_\alpha)$ è la stima vincolata di η per un valore di α fissato, mentre $j_{\eta\eta}(\alpha, \eta)$ è il blocco (η, η) della matrice di informazione osservata di Fisher.

Si tratta di un approccio a cavallo tra le teorie bayesiane e frequentiste, che comunque permette di applicare tutte le considerazioni inferenziali classiche che siamo abituati a considerare quando andiamo a stimare un parametro incognito in una distribuzione.

Il termine di correzione $M(\alpha)$ è dato da (Cabras et al., 2010):

$$M(\alpha) = \sqrt{\frac{a_2 A^2 [\pi(1 + a_0 \alpha^4) + \alpha^2(\pi(1 + a_0) - 4)] + 2\sqrt{2\pi} a_1 \alpha A^{3/2} - \pi a_1^2 \alpha^2 A^3 - 2}{\pi A^3 [2 + \alpha^2(2a_0 + a_2) + \alpha^4(a_0 a_2 - a_1^2)] - 2(\alpha + 2\alpha^3)^2 - 2\sqrt{2\pi} a_1 \alpha^3 \sqrt{A}(1 + 3\alpha^2 + 2\alpha^4)}}, \quad (2.3)$$

dove $A = 1 + \alpha^2$, mentre le quantità $a_i = a_i(\alpha)$, per $i=0,1,2$, sono così definite:

$$\begin{aligned} a_0(\alpha) &= \frac{2}{\pi} (1 + 8\alpha^2 \pi^{-2})^{-1/2} \\ a_1(\alpha) &= \frac{\alpha \sqrt{8\pi^3}}{(8\alpha^2 + \pi^2)(\alpha^2(8 + \pi^2) + \pi^2)^{1/2}} \\ a_2(\alpha) &= \frac{2}{\pi} (1 + 8\alpha^2 \pi^{-2})^{-3/2}. \end{aligned}$$

Grazie alla (2.3) il problema della monotonia, quindi, viene risolto: l'andamento della funzione di log-verosimiglianza profilo per α assume un comportamento che permette di individuare almeno un punto di massimo, dato dalla stima di massima verosimiglianza penalizzata per il parametro di forma.

E' importante notare, infine, che $M(\alpha)$ non dipende dai dati, e quindi nemmeno dagli incogniti parametri di posizione e di scala e dalle loro stime (Cabras et al., 2010).

Nel prossimo capitolo andremo ad applicare questi risultati ad un campione che presenta il problema della monotonia nella funzione di log-verosimiglianza, nella fase di stima per il parametro di forma α .

Capitolo 3

Applicazioni

Al fine di studiare il funzionamento della verosimiglianza profilo penalizzata, prenderemo in considerazione lo stesso data-set con cui venivano introdotti i problemi sulla monotonia: il data-set “frontier”.

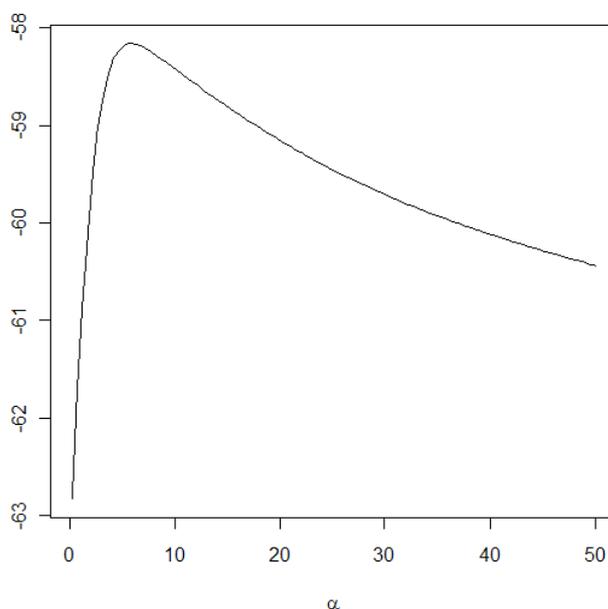


Figura 3.1: Data-set Frontier: verosimiglianza profilo penalizzata per α .

Nel precedente capitolo veniva mostrato come fosse problematica in que-

sto caso l'identificazione di una stima finita per α , dato che la funzione di verosimiglianza profilo per questo parametro mostra un andamento monotono crescente. Ora, invece, utilizzeremo la (2.2), ottenendo una funzione che ammette punto di massimo, cui coinciderà la stima di massima verosimiglianza per α : si veda a tal proposito la Figura 3.1.

Utilizzando un algoritmo di massimizzazione sulla verosimiglianza profilo penalizzata per il parametro di forma, se consideriamo il data-set "frontier" si ottiene un valore per lo stimatore di massima verosimiglianza pari a $\hat{\alpha} = 5.85207$ (Figura 3.2).

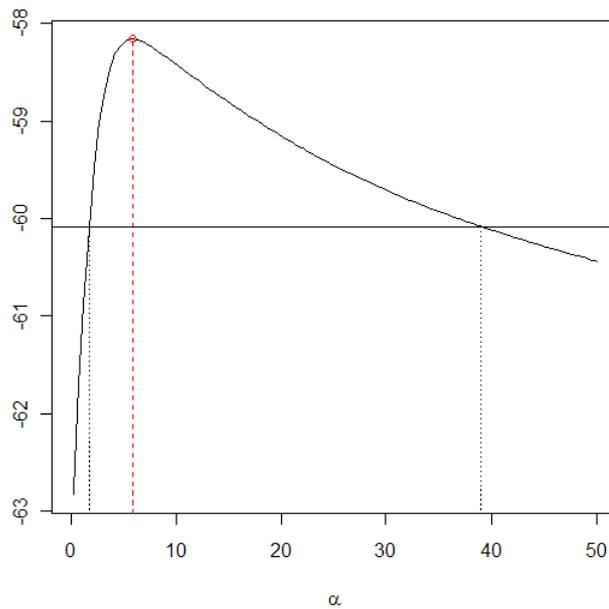


Figura 3.2: Data-set Frontier: stima e intervallo di confidenza al 95% per α .

Gli intervalli di confidenza si possono ottenere dal test log-rapporto di verosimiglianza, come

$$\begin{aligned}
 IC &= \{ \alpha \in \mathbb{R} : W_{mp}^*(\alpha) < \chi_{1,1-\alpha}^2 \} \\
 &= \{ \alpha \in \mathbb{R} : l_{mp}^*(\alpha) > l_{mp}^*(\hat{\alpha}^*) - \frac{1}{2} \chi_{1,1-\alpha}^2 \}
 \end{aligned}$$

Otteniamo un intervallo di confidenza al 95 % per il parametro di forma: come estremo inferiore è stato individuato il valore 1.725, mentre per quello superiore si ha 39.047.

3.1 Studio di simulazione

Si presentano adesso i risultati ottenuti tramite uno studio di simulazione basato su 5000 replicazioni, svolto al fine di studiare il comportamento dello stimatore per il parametro di forma della normale asimmetrica.

Tabella 3.1: Principali statistiche per le stime del parametro di forma, quando il vero α è 5, nei tre diversi metodi.

	media	bias	mediana	sd	MSE	% $\hat{\alpha} < +\infty$
$n = 30, \alpha = 5$						
$\hat{\alpha}$	5.345	0.345	4.399	4.209	17.838	70.38
$\hat{\alpha}_{VR}$	2.582	-2.418	2.245	1.670	8.636	100
$\hat{\alpha}_S$	4.032	-0.968	4.114	2.161	5.609	98.9
$n = 50, \alpha = 5$						
$\hat{\alpha}$	6.492	1.492	5.093	5.163	28.883	85.66
$\hat{\alpha}_{VR}$	3.564	-1.436	3.207	2.015	6.125	100
$\hat{\alpha}_S$	5.067	0.067	4.499	2.687	7.225	98.76
$n = 100, \alpha = 5$						
$\hat{\alpha}$	6.478	1.478	5.309	4.785	25.084	97.46
$\hat{\alpha}_{VR}$	4.510	-0.490	4.117	2.183	5.006	100
$\hat{\alpha}_S$	5.604	0.604	4.818	3.161	10.359	99.66
$n = 200, \alpha = 5$						
$\hat{\alpha}$	5.511	0.511	5.145	2.217	5.176	99.9
$\hat{\alpha}_{VR}$	4.707	-0.293	4.525	1.358	1.929	100
$\hat{\alpha}_S$	5.174	0.174	4.880	1.876	3.548	99.96

Vengono studiate, in particolare, le distribuzioni degli stimatori di α , ottenuti con tre metodi di stima differenti:

- il primo è quello proposto da Azzalini (1985), che ricava le stime di massima verosimiglianza per i tre parametri della normale asimmetrica tramite le operazioni inferenziali classiche;
- il secondo è quello proposto da Ventura e Racugno (2011), che va a penalizzare la funzione di verosimiglianza profilo per α ;
- il terzo è una versione preliminare della correzione avanzata da Sartori (2006), che agisce sulle equazioni di verosimiglianza.

In campioni di numerosità variabile, sono state effettuate 5000 repliche Monte Carlo di una funzione in R, che genera un campione casuale di dimensione a piacere e che produce le stime mediante le tre diverse possibilità. Dai vettori contenenti le stime, poi, vengono prodotte alcune statistiche allo scopo di studiare il comportamento dello stimatore di massima verosimiglianza per α in relazione al vero valore che assume il parametro. Si riportano, in particolare, la media, il bias, la mediana, lo scarto quadratico medio e l'errore quadratico medio (MSE)¹.

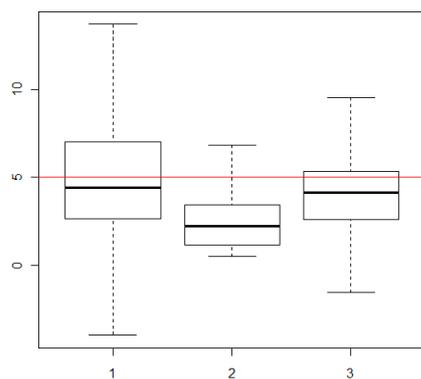
Nella Tabella 3.1 troviamo i risultati ottenuti da una prima simulazione, eseguita con $\alpha = 5$, per campioni aventi numerosità pari a $n = 30, 50, 100, 200$. In Figura 3.3, invece, vengono rappresentati i boxplot per le tre diverse modalità di stima considerate (la linea rossa nel grafico indica il vero valore del parametro di forma).

Per avere una idea più ampia di come lavorino i tre differenti stimatori, viene condotto un secondo studio di simulazione, andando a modificare il vero valore del parametro di forma. Ponendo $\alpha = 10$, si ottengono così i risultati riportati in Tabella 3.2 e, al solito, una rappresentazione grafica delle stime di massima verosimiglianza per i tre diversi stimatori (Figura 3.4).

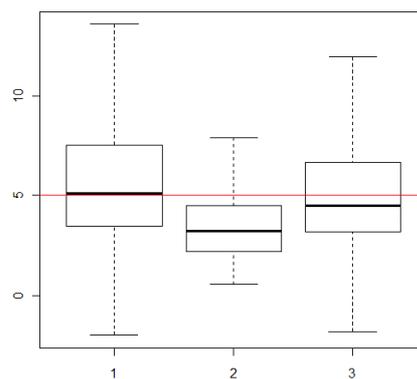
3.1.1 Commenti

Durante lo svolgimento degli studi di simulazione, si è notato che, utilizzando la verosimiglianza genuina che porta ad $\hat{\alpha}$, c'è la possibilità di ottenere stime per il parametro di forma che vanno ad infinito; questa caratteristica

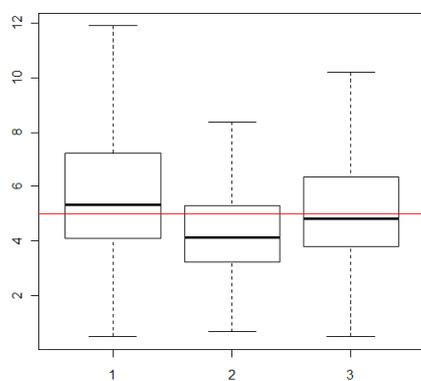
¹Per $\hat{\alpha}$ e $\hat{\alpha}_S$, solo quando la stima ottenuta è minore di infinito



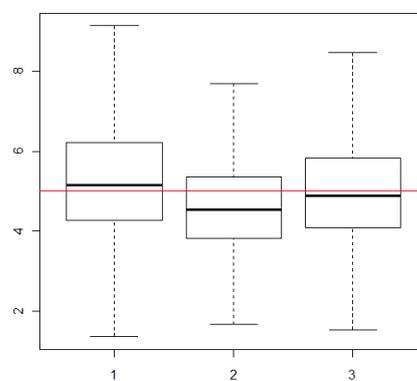
(a) $n = 30$



(b) $n = 50$



(c) $n = 100$



(d) $n = 200$

Figura 3.3: Boxplot, al variare della numerosità campionaria, per le stime di α , quando il vero valore del parametro è 5.

si presenta soprattutto quando si ha a che fare con numerosità campionarie modeste, e si realizza sempre meno quando si trattano campioni via via più ampi. Si tratta di un comportamento a cui eravamo preparati, in forza di quanto discusso nel Capitolo 2.

Tabella 3.2: Principali statistiche per le stime del parametro di forma, quando il vero α è 10, nei tre diversi metodi.

	media	bias	mediana	sd	MSE	% $\hat{\alpha} < +\infty$
$n = 30, \alpha = 10$						
$\hat{\alpha}$	7.444	-2.556	6.597	5.078	32.322	48.04
$\hat{\alpha}_{VR}$	3.489	-6.511	3.159	2.139	46.966	100
$\hat{\alpha}_S$	5.334	-4.666	5.324	2.360	27.337	98.16
$n = 50, \alpha = 10$						
$\hat{\alpha}$	10.792	0.792	9.075	6.738	46.030	63.72
$\hat{\alpha}_{VR}$	5.602	-4.398	5.107	2.970	28.165	100
$\hat{\alpha}_S$	7.798	-2.202	7.637	3.206	15.125	97.48
$n = 100, \alpha = 10$						
$\hat{\alpha}$	12.816	2.816	10.313	8.754	84.572	85.16
$\hat{\alpha}_{VR}$	8.130	-1.870	7.245	4.064	20.011	100
$\hat{\alpha}_S$	10.283	0.283	8.975	5.045	25.535	97.34
$n = 200, \alpha = 10$						
$\hat{\alpha}$	12.673	2.673	10.647	8.088	72.562	97.2
$\hat{\alpha}_{VR}$	9.440	-0.560	8.622	3.868	15.275	100
$\hat{\alpha}_S$	11.009	1.009	9.636	5.283	28.925	99.1

Molto più raramente, in particolare solo quando il coefficiente di asimmetria di un certo campione risulta essere in prossimità della soglia del proprio campo di variazione ($|\gamma_1| < 0.995$), anche lo stimatore di Sartori, $\hat{\alpha}_S$, fornisce stime che tendono ad infinito per il parametro di forma.

L'utilizzo della verosimiglianza profilo penalizzata che porta alla $\hat{\alpha}_{VR}$ (Ventura e Racugno, 2011), invece, garantisce l'individuazione di una stima sempre finita; la penalizzazione praticata, infatti, fa sì che le possibili

difficoltà di stima si ripercuotano in maniera meno evidente sullo stimatore, rispetto alle altre due metodiche di stima.

Osservando i risultati ottenuti negli studi di simulazione, emerge immediatamente che il comportamento dei tre stimatori migliora con l'aumentare della numerosità campionaria, sia ponendo $\alpha = 5$, sia con $\alpha = 10$. Infatti la media delle stime, per ognuna delle tre soluzioni, all'aumentare di n si avvicina sempre più al vero valore del parametro, riducendo anche la propria deviazione standard; come si nota, in particolare nel caso del parametro di forma pari a 10, si raggiunge una distorsione ragionevole nel parametro solo con campioni aventi almeno 100 unità.

E' anche interessante notare che, mediamente, $\hat{\alpha}$ tende a sovrastimare il parametro, evidenziando scostamenti in media positivi in campioni di piccole dimensioni, che tendono via via verso lo zero all'aumentare della numerosità campionaria. Il contrario accade per gli altri due stimatori, orientati invece verso valori che sottostimano il parametro di forma.

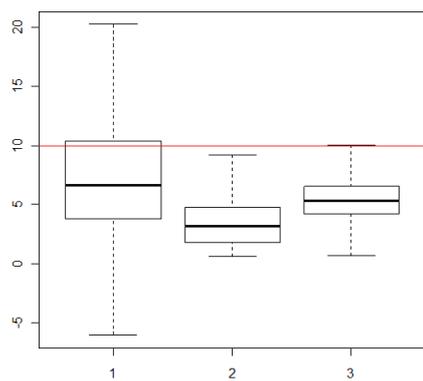
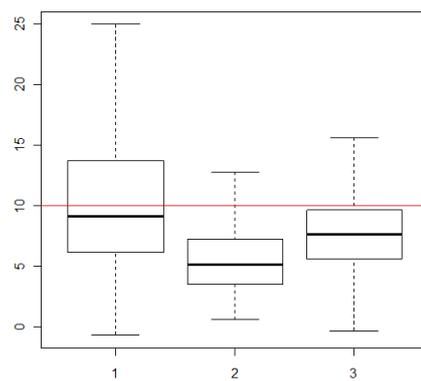
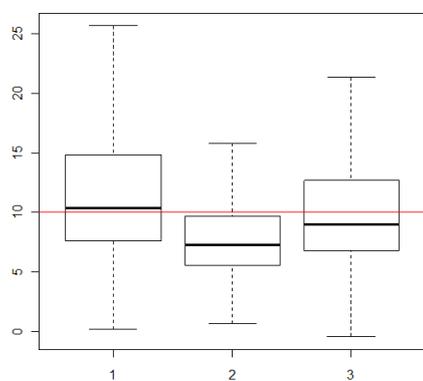
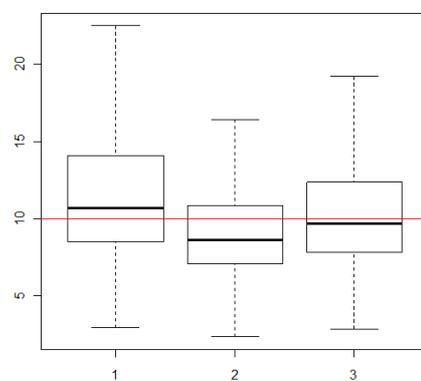
(a) $n = 30$ (b) $n = 50$ (c) $n = 100$ (d) $n = 200$

Figura 3.4: Boxplot, al variare della numerosità campionaria, per le stime di α , quando il vero valore del parametro è 10.

Appendice A

Codice R

Di seguito viene riportato il codice R sviluppato per ottenere le stime di massima verosimiglianza nei 'dati di frontiera', e i comandi utilizzati per lo studio di simulazione.

A.1 Algoritmo di massimizzazione

Per l'utilizzo del seguente codice R, è necessario caricare le librerie "sn" e "numDeriv"; la prima per poter utilizzare alcune funzioni che riguardano la normale asimmetrica, la seconda per effettuare il calcolo numerico della matrice di informazione osservata, necessaria per la stima di α da ottenere tramite la (2.2).

- **Funzione 1**, calcola la penalizzazione $M(\alpha)$ definita nella (2.3):

```
m.alpha <- function(x,mu,sigma,lambda){  
  
  a0 <- (2/pi)*(1+8*lambda^2*pi^(-2))^(1/2)  
  
  a1 <- (lambda*sqrt(8*pi^3))/((8*lambda^2+pi^2)*(lambda^2*(8+pi^2)  
    +pi^2)^(1/2))  
  
  a2 <- (2/pi)*(1+8*lambda^2*pi^(-2))^(3/2)  
  
  A <- 1+lambda^2
```

```

pi.greco.star <- sqrt((a2*A^2*(pi*(1+a0*lambda^4)+lambda^2*(pi*
  (1+a0)-4))+2*sqrt(2*pi)*a1*lambda*A^(3/2)-pi*a1^2*lambda^2*
  A^3-2)/(pi*A^3*(2+lambda^2*(2*a0+a2)+lambda^4*(a0*a2-a1^2))
  -2*(lambda+2*lambda^3)^2-2*sqrt(2*pi)*a1*lambda^3*sqrt(A)*
  (1+3*lambda^2+2*lambda^4)))

val <- list(ll = sum(dsn(x,mu,sigma,lambda,log=TRUE)),
  M.alpha = log(pi.greco.star))

val

}

```

- **Funzione 2**, calcola il determinante di $j_{\eta\eta}(\alpha, \eta)$:

```

j.theta <- function(x,mu,sigma,lambda) {

  -0.5*log(det(hessian(function(par) -sum(dsn(x,par[1],par[2],
    lambda,log=TRUE)),c(mu,sigma))))

}

```

- **Funzione 3**, raggruppa le funzioni 1 e 2 e ricava la stima di massima verosimiglianza per il parametro di forma utilizzando la $L_{mp}^*(\alpha)$, definita dalla (2.2):

```

lmp.star <- function(x,lambda) {

  lp <- function(mu,sigma) sum(dsn(x,mu,sigma,lambda,log=TRUE))

  ris.lp <- nlminb(c(-0.106,log(1.234)),function(par)
    -lp(par[1],exp(par[2])))

  ris.lp$par[2] <- exp(ris.lp$par[2])

}

```

```

c1 <- m.alpha(x,ris.lp$par[1],ris.lp$par[2],lambda)

c2 <- j.theta(x,ris.lp$par[1],ris.lp$par[2],lambda)

c1$ll+c1$M.alpha+c2-log(ris.lp$par[2])

}

```

- **Stima di massima varosimiglianza per α , intervalli di confidenza e grafico sul data-set “frontier”:**

```

maxll <- lmp.star(frontier,nlminb(1,function(par)
  -lmp.star(frontier,par))$par)

lmp.star <- Vectorize(lmp.star,"lambda")

lambda.hat <- nlminb(1,function(par) -lmp.star(frontier,par))$par

curve(lmp.star(frontier,x),xlim=c(0.2,50),xlab=expression(alpha),
  ylab="")

points(lambda.hat,maxll,col=2)

abline(v=lambda.hat,ylim=c(-55,maxll),lty="dashed")

lev <- -0.5*qchisq(0.95,1)+maxll

abline(h=lev)

IC.L <- uniroot(function(lambda) lmp.star(frontier,lambda)+
  0.5*qchisq(0.95,1)-maxll,lower=0,upper=10)$root

IC.U <- uniroot(function(lambda) lmp.star(frontier,lambda)+
  0.5*qchisq(0.95,1)-maxll,lower=30,upper=50)$root

c(IC.L,IC.U)

segments(IC.L,-100,IC.L,lmp.star(frontier,IC.L),lty="dotted")

```

```
segments(IC.U,-100,IC.U,lmp.star(frontier,IC.U),lty="dotted")
```

A.2 Codice simulazione

Per effettuare lo studio di simulazione, si utilizzano le funzioni sviluppate per ricavare la stima di massima verosimiglianza per il parametro di forma α , viste nel paragrafo precedente; di conseguenza, bisogna caricare tutte le librerie di R necessarie.

```
simula <- function(B,n,mu,sigma,lambda){  
  
  l.hat <- double(B)  
  
  lVR.hat <- double(B)  
  
  lS.hat <- double(B)  
  
  y <- double(n)  
  
  for(i in 1:B){  
  
    y <- rsn(n,mu,sigma,lambda)  
  
    l.hat[i] <- cp.to.dp(sn.mle(y=y,plot.it=F)$cp)[3]  
  
    lVR.hat[i] <- nlminb(1,function(par) -lmp.star(y,par))$par  
  
    lS.hat[i] <- sn.mmle(y=y,plot.it=F)$dp[3]  
  
  }  
  
  matrix(c(l.hat,lVR.hat,lS.hat),ncol=3)  
}
```

Gli argomenti della funzione “simula” sono:

- B, numero di replicazioni Monte Carlo che si vogliono eseguire;

- n , numerosità campionaria dei campioni da generare;
- μ , parametro di posizione delle skew-normal;
- σ , parametro di scala;
- λ , parametro di forma.

La funzione restituisce una matrice $3 \times n$, dove la prima colonna contiene le B stime ottenute dalla funzione di verosimiglianza profilo classica per α , la seconda colonna contiene quelle ottenute utilizzando la verosimiglianza profilo penalizzata $L_{mp}^*(\alpha)$ definita nella (2.2), mentre la terza colonna contiene le stime che si ricavano con il metodo di Sartori (2006).

Elenco delle figure

1.1	Funzione di densità SN con parametri $\mu=0$, $\sigma=1$, $\alpha=0$, equivalente ad una normale standard.	4
1.2	Funzione di densità SN al variare di $\alpha > 0$	5
1.3	Funzione di densità SN al variare di $\alpha < 0$	7
1.4	Casi limite della funzione di densità SN per $ \alpha \rightarrow \infty$	8
2.1	Dati AIS sul peso: log-verosimiglianza profilo normalizzata per α (<i>parametrizzazione diretta</i>).	14
2.2	Dati AIS sul peso: log-verosimiglianza profilo normalizzata per γ_1 (<i>parametrizzazione centrata</i>).	17
2.3	Dataset Frontier: log-verosimiglianza profilo normalizzata per α	19
3.1	Data-set Frontier: verosimiglianza profilo penalizzata per α	23
3.2	Data-set Frontier: stima e intervallo di confidenza al 95% per α	24
3.3	Boxplot, al variare della numerosità campionaria, per le stime di α , quando il vero valore del parametro è 5.	27
3.4	Boxplot, al variare della numerosità campionaria, per le stime di α , quando il vero valore del parametro è 10.	30

Bibliografia

- [1] Arellano-Valle, R. & Azzalini, A. (2008). *The centred parametrization for the multivariate skew-normal distribution*. J. Multivariate Analysis, 99, 1362-1382.
- [2] Azzalini, A. (1985). *A class of distributions wich includes the normal ones*. Scand. J. Statist., 12, 171-178.
- [3] Azzalini, A. (1986). *Further result on a class of distributions wich includes the normal ones*. Statistica, XLVI, 171-178.
- [4] Azzalini, A. & Capitanio, A. (1999). *Statistical applications of the multivariate skew-normal distribution*. J. Roy. Statist. Soc. B, 61, 579-602.
- [5] Azzalini, A. & Dalla Valle, A. (1996). *The multivariate skew-normal distribution*. Biometrika, 83, 715-726.
- [6] Bayes, C.L. & Branco, M.D. (2007). *Bayesian inference for the skewness parameter of the scalar skew-normal distribution*. Brazilian Jour. of Prob. and Statistics, 21, 141-163.
- [7] Barndorff-Nielsen, O.E. (1983). *On a formula for the distribution of the maximum likelihood estimator*. Biometrika, 70, 343-365.
- [8] Cabras, S., Castellanos, M.E., Racugno, W., Ventura, L. (2010), *A matching prior for the shape parameter of the skew-normal distribution*. Presentato.

-
- [9] Chiogna, M. (2005). *A note on the asymptotic distribution of the maximum likelihood estimator for the scalar Skew-normal distribution*. Statist. Methods and Applications, 14, 331-341.
- [10] Dalla Valle, A. (2004), *The skew-normal distribution*. Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality, Genton, M. G., Ed., Chapman & Hall / CRC, Boca Raton, FL, pp. 3-24.
- [11] de Helguero, F. (1908). *Sulla rappresentazione analitica delle statistiche abnormali*. Atti del IV congresso internazionale dei matematici, III, 288-299.
- [12] Pewsey, A. (2000). *Problems of inference for Azzalini's skew-normal distribution*. Journal of Applied Statistics, 27(7), 859-870.
- [13] Sartori, N. (2006). *Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions*. Journal of Statistical Planning and Inference, 136, 4259-4275.
- [14] Ventura, L., Racugno, W. (2011), *An adjustment of the modified profile likelihood from matching priors*. Working Papers, 2011.2, Dipartimento di Scienze Statistiche, Università degli Studi di Padova.