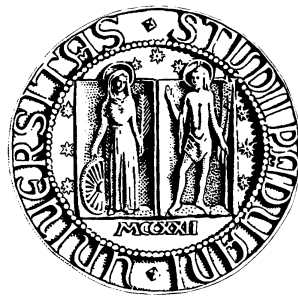


UNIVERSITÀ DEGLI STUDI DI PADOVA
FACOLTÀ DI SCIENZE STATISTICHE
TESI DI LAUREA SPECIALISTICA IN STATISTICA E INFORMATICA



ASPETTI STATISTICI NELLA
NORMALE ASIMMETRICA ESTESA

Relatore: Chiar.mo Prof. Adelchi Azzalini

Candidato: Antonio Canale

565574-STF

ANNO ACCADEMICO 2007–2008

*Ai miei nonni:
Otto, Sughì,
Piero e Teresina*

Indice

Introduzione	1
1 Le distribuzioni asimmetriche	5
1.1 Primi studi sulle distribuzioni asimmetriche	5
1.2 La distribuzione normale asimmetrica	7
1.3 Generazione	8
1.3.1 Troncamento di una normale bivariata	9
1.3.2 Convoluzione di normali	9
1.4 Normale asimmetrica multivariata	10
1.5 Momenti	10
1.6 Aspetti statistici	12
1.6.1 Verosimiglianza e matrice d'informazione	12
1.6.2 Parametrizzazione centrata	14
2 La distribuzione normale asimmetrica estesa	17
2.1 Generazione	18
2.1.1 Campioni selezionati	18
2.1.2 Distribuzione condizionata di una SN_2	19
2.2 Momenti	20
2.3 Comportamento della densità	21
2.3.1 Asimmetria	23

2.3.2	Curtosi	27
2.3.3	Comportamenti limite	27
2.3.4	Code a confronto	30
2.4	Problemi di identificabilità del modello	31
2.5	Caso multivariato	32
3	Aspetti statistici	33
3.1	Funzione di verosimiglianza e vettore punteggio	34
3.2	Matrice d'informazione	35
3.2.1	Calcolo della matrice d'informazione	35
3.2.2	Caratteristiche della matrice d'informazione	42
3.3	Stima di massima verosimiglianza	46
3.3.1	Peculiarità della verosimiglianza	46
3.3.2	Calcolo delle stime	49
3.4	Studio numerico della distribuzione delle stime	54
3.5	Una parametrizzazione alternativa	59
4	Regioni di tolleranza per la <i>ESN</i> doppia	67
4.1	Per la normale e la normale asimmetrica	67
4.2	Per la normale asimmetrica estesa	71
4.2.1	Casi limite: $\tau \rightarrow \pm\infty$	71
4.2.2	Casi per valori finiti di τ	75
	Conclusioni	83
A	Codice R	85
A.1	Calcolo della matrice d'informazione attesa	85
A.2	Calcolo delle stime di massima verosimiglianza	87
B	Ulteriori grafici e risultati	91
B.1	Distribuzione delle stime	91

B.2 Distribuzioni limite ESN_2	95
Riferimenti bibliografici	101

Elenco delle figure

1.1	Funzione di densità per una $SN(\alpha)$ per alcune scelte di α	8
1.2	Dati AIS: 2logverosimiglianza profilo relativa per α e per (ω, α)	12
1.3	Distribuzione empirica delle SMV. Parametrizzazione diretta.	14
1.4	Dati AIS: 2logverosimiglianza profilo relativa per γ_1 e per (σ^2, γ_1)	16
1.5	Distribuzione empirica delle SMV. Parametrizzazione centrata.	16
2.1	Densità per alcune scelte di $\tau \geq 0$	22
2.2	Densità per alcune scelte di $\tau \leq 0$	24
2.3	Densità per alcune scelte di $\tau < 0$	25
2.4	Indice di asimmetria γ_1 in funzione di τ per alcune scelte di α	26
2.5	Indice di asimmetria γ_1 in funzione di α per alcune scelte di $\tau \geq 0$	27
2.6	Indice di asimmetria γ_1 in funzione di α per alcune scelte di $\tau \leq 0$	28
2.7	Indice di curtosi γ_2 in funzione di α	28
2.8	Indice di curtosi γ_2 in funzione di τ	29
2.9	Densità a confronto	32
3.1	Determinante della matrice d'informazione attesa in funzione di α	42
3.2	Determinante della matrice d'informazione attesa in funzione di τ	44
3.3	Determinante della matrice d'informazione attesa in funzione di (α, τ)	45
3.4	Logverosimiglianza profilo per α in campioni simulati	48
3.5	Logverosimiglianza profilo per τ in campioni simulati	50
3.6	Logverosimiglianza profilo per τ in campioni simulati. Errori di stima.	51

3.7	Funzioni di densità con i parametri della tabella 3.1	53
3.8	Distribuzione delle stime, $n = 200$ (DP, massimizzazione diretta)	55
3.9	Distribuzione delle stime, $n = 200$ (DP, profilo per τ)	56
3.10	Distribuzione delle stime, $n = 200$ (DP, dettaglio fig. 3.8)	57
3.11	Distribuzione delle stime, $n = 200$ Istogrammi per ξ e α	58
3.12	Distribuzione delle stime, $n = 200$ (CP, massimizzazione diretta)	64
3.13	Distribuzione delle stime, $n = 200$ (CP, profilo per τ)	65
4.1	Valori di probabilità osservati e fissati	69
4.2	Valori osservati di h in funzione di α^*	70
4.3	Valori osservati di $1/\log(\exp(h/2) - 1)$ in funzione di α^*	70
4.4	Valori di probabilità osservati e fissati(<i>ESN</i> , approccio n.1)	74
4.5	Valori di probabilità osservati e fissati(<i>ESN</i> , approccio n.2, prova)	76
4.6	Valori osservati di h in funzione di α^* (<i>ESN</i>)	77
4.7	Valori osservati di k in funzione di τ	77
4.8	Valori osservati di h in funzione di α^* e τ (<i>ESN</i>)	78
4.9	Valori osservati di k in funzione di α^* per alcuni τ	79
4.10	Valori di probabilità osservati e fissati(<i>ESN</i> , approccio n.2)	80
B.1	Distribuzione delle stime, $n = 500$ (DP, massimizzazione diretta)	92
B.2	Distribuzione delle stime, $n = 500$ (DP, profilo per τ)	93
B.3	Distribuzione delle stime, $n = 500$ (DP, dettaglio fig. B.1)	94

Introduzione

La normale asimmetrica (Azzalini, 1985) è una famiglia di distribuzioni di probabilità che include al suo interno, come caso particolare, la distribuzione normale. Oltre agli usuali parametri di posizione e scala, questa distribuzione prevede l'introduzione di un terzo parametro che ne regola l'asimmetria. Essa gode di buone proprietà dal lato matematico-probabilistico e non mancano molti risultati utili in ambito inferenziale. Generalizzando a sua volta questo modello, con l'aggiunta di un quarto parametro, si può ottenere un'ulteriore famiglia di distribuzioni di probabilità, nota in letteratura come *normale asimmetrica estesa*, oggetto di questo lavoro di tesi.

A differenza della normale asimmetrica, per la normale asimmetrica estesa non sono noti alcuni importanti risultati inferenziali, prima fra tutti un'espressione per la matrice d'informazione attesa.

Nel capitolo 1 viene introdotta la normale asimmetrica, riportando alcuni dei risultati noti in letteratura quali i possibili fenomeni generatori, il calcolo dei momenti, la matrice d'informazione attesa e le sue caratteristiche e, infine, una parametrizzazione che risolve alcuni problemi in campo inferenziale.

Nel capitolo 2 viene presentata la normale asimmetrica estesa, sottolineandone di volta in volta le analogie con la normale asimmetrica. Dopo aver riportato alcuni dei possibili fenomeni generatori, è riportato il calcolo dei momenti. La densità è studiata per quanto riguarda la forma della curva al variare dei parametri e discutendo il suo comportamento per i casi limite dei parametri. Verrà inoltre introdotto il problema della possibile non identificabilità del modello.

Nel capitolo 3 è studiata la funzione di verosimiglianza e la matrice d'informazione attesa.

Si sviluppa il calcolo della matrice di informazione attesa, risultato non noto in letteratura. Il comportamento della matrice è studiato e confrontato con quello della distribuzione normale e normale asimmetrica. Si è analizzato inoltre il comportamento della verosimiglianza e delle stime di massima verosimiglianza per alcuni campioni simulati, evidenziando non pochi problemi nella stima dei parametri, causati dall'aggiunta del quarto parametro. In particolare la scarsa informatività della logverosimiglianza profilo per il quarto parametro, unita alla forte correlazione tra i primi tre parametri e il quarto, portano non di rado a stime distanti nello spazio parametrico dal vero valore dei parametri, ma con una densità sostanzialmente equivalente. Alla fine del capitolo viene proposta una parametrizzazione alternativa che risolve alcuni dei problemi evidenziati nel corso del capitolo.

Nel capitolo 4 si discute sulla possibilità di costruire delle regioni di assegnata probabilità. Dopo aver riportato alcuni risultati per la normale asimmetrica, si propongono due approcci per la normale asimmetrica estesa doppia. Si mostrerà come per questa distribuzione la costruzione di questo tipo di regioni non è semplice a causa della mancanza di una distribuzione di riferimento per la forma quadratica.

In appendice è riportato il codice R utilizzato per ottenere i principali risultati del capitolo 3.

Summary

The skew-normal (Azzalini, 1985) is a class of probability distributions which includes the normal ones as a special case. In addition to the usual position and scale parameters, this distribution has a third parameter which regulates skewness. It has good mathematical and probabilistic properties and there are a lot of useful inferential results too. Generalizing this model, with the addition of a fourth parameter, we can obtain another class of probability distributions, known as *extended skew-normal*, core of this thesis project.

Regardless of the skew-normal, for the extended skew-normal some important inferential results are still unknown, first of all an expression for the expected information matrix.

In chapter 1 the skew-normal is introduced. Some known results are given, such as the generation mechanism, moments, the expected information matrix and its characteristics and at last a parametrization which resolves some inferential problems.

In chapter 2 the extended skew-normal is presented remarking its similarity with the skew-normal. After explaining some of the possible generation mechanisms, the moments calculation is reported. The shape of the density is studied in function of the parameters, and the limit behaviour of it is shown. In this chapter also the problem of the non-identifiability of the model will be introduced.

In chapter 3 the likelihood function is studied and the expected information matrix is obtained. The expressions for the expected information matrix are derived and its behaviour is studied and compared with the skew-normal one. The behaviour of the likelihood and of the maximum likelihood estimation is also studied for some simulated samples, pointing out not rarely estimating problems. In particular we can see that the profile loglikelihood for the fourth parameter is very less informative for inference. This will lead not rarely to estimations very far from each other and from the real value of the parameter but with substantially equivalent density. At the end an alternative parametrization that settles some of the problems shown during the chapter will be discussed.

In chapter 4 the possibility of building regions of given probability for the extended skew-normal is discussed. Starting from known results for the skew-normal, two different approaches are discussed. It will be shown that building regions of given probability for this distributions is not simple due to the lack of a reference distribution for the quadratic form.

In the appendix we'll find the R code used to get the results of chapter 3.

Capitolo 1

Le distribuzioni asimmetriche

1.1 Primi studi sulle distribuzioni asimmetriche

Lo studio delle distribuzioni di probabilità parametriche nasce dall'esigenza per la Statistica di descrivere i fenomeni fisici, biologici, economici o sociali con delle espressioni che abbiano delle comode proprietà matematiche, una buona flessibilità di adattamento ai dati in esame facendo variare il valore dei parametri, e, da ultimo ma non meno importante, che siano una verosimile astrazione della realtà e che ben rappresentino il meccanismo generatore dei dati oggetto di studio.

I primi studi sulle distribuzioni di probabilità asimmetriche appaiono alla fine del diciannovesimo secolo. Edgeworth (1886) per primo esamina il problema di adattamento di distribuzioni asimmetriche ai dati. Qualche anno dopo Karl Pearson discute (Pearson (1893) e Pearson (1895)) la costruzione di densità asimmetriche dalla mistura di due curve normali.

Uno statistico italiano, Fernando de Helguero, al IV Congresso Internazionale dei Matematici di Roma del 1909, introduce una nuova famiglia di distribuzioni asimmetriche. A riguardo egli scrive:

Il compito della statistica nelle sue varie applicazioni alle scienze economiche e biologiche 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 e le loro modificazioni.

L'idea che de Helguero propone riguarda una modifica delle ditribuzioni di probabilità simmetriche tramite perturbazione per poter descrivere quei fenomeni in cui si osservano campioni selezionati.

[...]Le curve teoriche studiate dal PEARSON e dall'EDGEWORTH per la perequazione delle statistiche abnormali in materiale omogeneo, mentre dànno con molta approssimazione la legge di variazione (meglio della curva normale perchè ne sono delle generalizzazioni), a mio avviso sono difettose in quanto [...] nulla ci fanno sapere sulla legge di dipendenza, quasi nulla sulla relazione colla curva normale che pure deve essere considerata fondamentale

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

I moderni studi sulle distribuzioni asimmetriche, si sviluppano in seguito all'articolo di Azzalini (1985), in cui troviamo il seguente utile risultato.

Lemma 1.1. *Se f_0 è una funzione di densità di probabilità unidimensionale simmetrica in 0, G una funzione di ripartizione tale che G' esiste simmetrica in 0 e $\alpha \in \mathbb{R}$, allora*

$$f(z) = 2f_0(z)G\{\alpha z\} \quad (-\infty < z < \infty) \quad (1.1)$$

è una funzione di densità.

Il lemma descrive la costruzione di una funzione di densità asimmetrica a partire dalla *perturbazione* tramite αx di una funzione di densità simmetrica f_0 . Chiaramente la nuova famiglia di funzioni f include il caso di partenza, quando $\alpha = 0$.

Da quest'idea nasce un nutrito insieme di distribuzioni asimmetriche di cui la *normale asimmetrica*, che sarà brevemente trattata nel prossimo paragrafo, ne è il primo e il più chiaro esempio.

1.2 La distribuzione normale asimmetrica

Definiamo ora la distribuzione di probabilità *normale asimmetrica* facendo riferimento al Lemma 1.1.

Siano $f_0 = \phi$ e $G = \Phi$ la funzione di densità e di ripartizione di una normale standardizzata. Allora la densità

$$\phi(z; \alpha) = 2\phi(z)\Phi(\alpha z) \quad (-\infty < z < \infty) \quad (1.2)$$

è chiamata *normale asimmetrica* con parametro di forma α e scriveremo, per seguire l'usuale notazione in uso in letteratura, che $Z \sim SN(\alpha)$.

Generalizziamo l'espressione (1.2) introducendo dei parametri di posizione e scala. Sia $Z \sim SN(\alpha)$, $\xi \in \mathbb{R}$ e $\omega \in \mathbb{R}^+$. Allora, definita $Y = \xi + \omega Z$, la densità di Y sarà:

$$\phi(z; \xi, \omega, \alpha) = \frac{2}{\omega} \phi\left(\frac{y - \xi}{\omega}\right) \Phi\left(\alpha \frac{y - \xi}{\omega}\right) \quad (-\infty < y < \infty) \quad (1.3)$$

e scriveremo che $Y \sim SN(\xi, \omega^2, \alpha)$.

Proprietà

Si riportano in seguito alcune importanti proprietà introdotte in Azzalini (1985) in cui risultano evidenti i legami tra la distribuzione normale e la distribuzione normale asimmetrica.

1. Se $\alpha = 0$ ci si riconduce alla densità di una $N(0, 1)$;
2. se $X \sim SN(\alpha)$, allora $-X \sim SN(-\alpha)$;
3. se $\alpha \rightarrow \infty$ ci si riconduce alla mezza normale, $2\phi(x)$ per $z \geq 0$;
4. se $X \sim SN(\alpha)$, allora $Z^2 \sim \chi_1^2$;
5. per un fissato α la (1.2) è fortemente unimodale, ossia $\log(\phi(z; \alpha))$ è una funzione concava di z .

La distribuzione normale risulta quindi un caso interno a questa famiglia di distribuzioni.

Per dare un'idea della forma che la curva dell'equazione 1.2 assume si veda la Figura 1.1. Nel grafico sono state inserite le curve per alcune scelte del parametro α positivo. Per α negativi, come conseguenza della proprietà (2), avremo che il grafico della densità risulta speculare rispetto all'asse verticale.

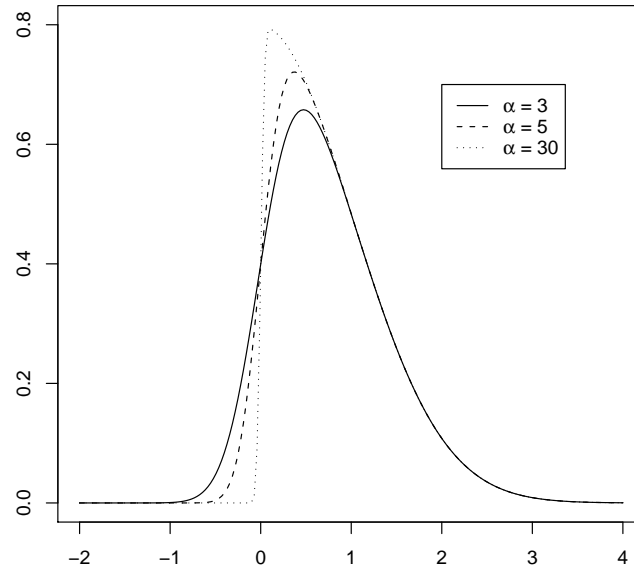


Figura 1.1: Funzione di densità per una $SN(\alpha)$ per alcune scelte di α

1.3 Generazione

La densità in (1.2) è stata introdotta a partire dalla densità normale standardizzata, applicando su di essa una perturbazione che da luogo all'asimmetria.

In questa sezione si vogliono descrivere alcuni meccanismi reali che danno luogo a questo tipo di distribuzione.

1.3.1 Troncamento di una normale bivariata

Si supponga che $\tilde{f}(v, z)$ sia una normale bivariata con vettore delle medie pari a $(0 \ 0)^T$ e matrice di varianza con diagonale unitaria e correlazione pari a ρ . Si supponga inoltre di osservare un campione censurato in una delle due marginali, ossia di osservare solamente (v, z) se $v \geq 0$. La densità della variabile che osserviamo sarà quindi:

$$f(v, z) = \begin{cases} 2\tilde{f}(v, z) & \text{per } v \geq 0; \\ 0, & \text{altrimenti.} \end{cases}$$

se marginalizziamo rispetto alla variabile z , integrando $f(v, z)$ sul dominio di v , otteniamo esattamente che:

$$Z \sim SN\left(\frac{\rho}{\sqrt{1-\rho^2}}\right).$$

In contesti di selezione del personale o di ammissione a scuole o enti di formazione in cui è previsto un test d'ingresso di tipo attitudinale, l'osservazione del punteggio ottenuto in test successivi, è evidentemente correlata al test d'ingresso. Questo tipo di meccanismo ben si presta ad essere modellato pensando al troncamento di una distribuzione bivariata come quello appena descritto. Si veda a riguardo Birnbaum (1950) e Arnold et al. (1993).

Si potrà estendere in seguito questa idea di campione selezionato, trattando la possibilità di effettuare il troncamento in un qualsiasi punto del dominio della funzione (cfr. cap. 2).

1.3.2 Convoluzione di normali

Un altro meccanismo di generazione di dati aventi distribuzione normale asimmetrica è presentato in Azzalini (1986).

Lemma 1.2. *Consideriamo due variabili casuali normali standard indipendenti, U_0 e U_1 , e una costante $|\delta| < 1$. Se definiamo: $Z = \delta|U_0| + \sqrt{1-\delta^2}U_1$, allora $Z \sim SN\left(\frac{\delta}{\sqrt{1-\delta^2}}\right)$.*

Questo meccanismo di generazione compare nei lavori di Weinstein (1964), Aigner (1964) e Andel et al. (1984).

1.4 Normale asimmetrica multivariata

Della normale asimmetrica, è stata studiata anche una versione multivariata. Si vedano a riguardo il lavori di Azzalini & Dalla Valle (1996) e Azzalini & Capitanio (1999). La distribuzione multivariata viene introdotta solamente per alcuni sviluppi del capitolo 2 e ne viene riportata solamente la funzione di densità.

Una variabile casuale d -dimensionale Z , ha distribuzione normale asimmetrica d -dimensionale, se ha una funzione di densità del tipo:

$$f_Z(z; \Omega_z, \alpha) = 2\phi_d(z; \Omega_z)\Phi(\alpha^T z) \quad (1.4)$$

dove α è un vettore d -dimensionale $\in \mathbb{R}^d$, $\phi(z; \Omega_z)$ è la densità di una variabile casuale normale d -dimensionale con vettore delle medie nullo e matrice di correlazione Ω_z calcolata in $z \in \mathbb{R}^d$ e $\Phi(\cdot)$ è la funzione di ripartizione di una normale standard. Diremo che $Z \sim SN_d(\Omega, \alpha)$.

Generalizzando l'espressione (1.4) con l'introduzione dei parametri di posizione e scala, abbiamo che $Y = \xi_d + \omega Z$ dove adesso ξ_d è un vettore d -dimensionale e $\omega = \text{diag}(\omega_1, \dots, \omega_d)$. La funzione di densità di Y sarà:

$$2\phi_d((y - \xi); \Omega)\Phi(\alpha^T \omega^{-1}(y - \xi)) \quad (1.5)$$

con $\Omega = \omega\Omega_z\omega$.

1.5 Momenti

La funzione generatrice dei momenti è facilmente ottenibile sfruttando alcuni risultati utili ben noti:

Lemma 1.3 (Completamento del quadrato). *Sia $\mathbf{A}_{k \times k}$ una matrice quadrata definita positiva.*

Allora:

$$\int_{\mathbb{R}} \frac{1}{(2\pi)^{k/2}} \exp \left\{ -\frac{1}{2}(y^T \mathbf{A}y - 2b^T y) \right\} dx = \frac{\exp \left\{ \frac{1}{2}(\mu^T \mathbf{A}\mu) \right\}}{|\mathbf{A}|^{1/2}}$$

Lemma 1.4. *Se $U \sim N(0, 1)$ e $h, k \in \mathbb{R}$ allora:*

$$E[\Phi(hU + k)] = \Phi\left(\frac{k}{\sqrt{1 + h^2}}\right).$$

Utilizzando questi due risultati otteniamo la funzione generatrice dei momenti:

$$M_Z(t) = 2 \exp\left(\frac{t^2}{2}\right) \Phi(\delta t) \quad (1.6)$$

dove $\delta = \alpha/\sqrt{1 + \alpha^2}$, e la funzione generatrice dei cumulanti:

$$K_Z(t) = \log(M_Z(t)) = \left(\frac{t^2}{2}\right) + \log(2\Phi(\delta t)). \quad (1.7)$$

Derivando la funzione generatrice dei cumulanti otteniamo:

$$\mathbb{E}[Z] = b\delta, \quad (1.8)$$

$$\mathbb{V}ar[Z] = 1 - (b\delta)^2, \quad (1.9)$$

$$\gamma_1[Z] = \frac{4 - \pi}{2} \text{sign}(\alpha) \left[\frac{\{\mathbb{E}[Z]\}^2}{\mathbb{V}ar[Z]} \right]^{3/2}, \quad (1.10)$$

$$\gamma_2[Z] = 2(\pi - 3) \left[\frac{\{\mathbb{E}[Z]\}^2}{\mathbb{V}ar[Z]} \right]^2, \quad (1.11)$$

dove $b = \sqrt{\frac{2}{\pi}}$ e γ_1, γ_2 sono il terzo e il quarto cumulante standardizzato, noti anche rispettivamente come coefficiente di asimmetria e di curtosi.

Si noti che gli indici γ_1 e γ_2 hanno per costruzione un campo di variazione limitato. In particolare $|\gamma_1| \leq 0.995$ circa, mentre $\gamma_2 \leq 0.869$. Questo implica un limite della normale asimmetrica nel rappresentare il comportamento in quanto ad asimmetria e curtosi di campioni di dati particolarmente asimmetrici o con pronunciata curtosi.

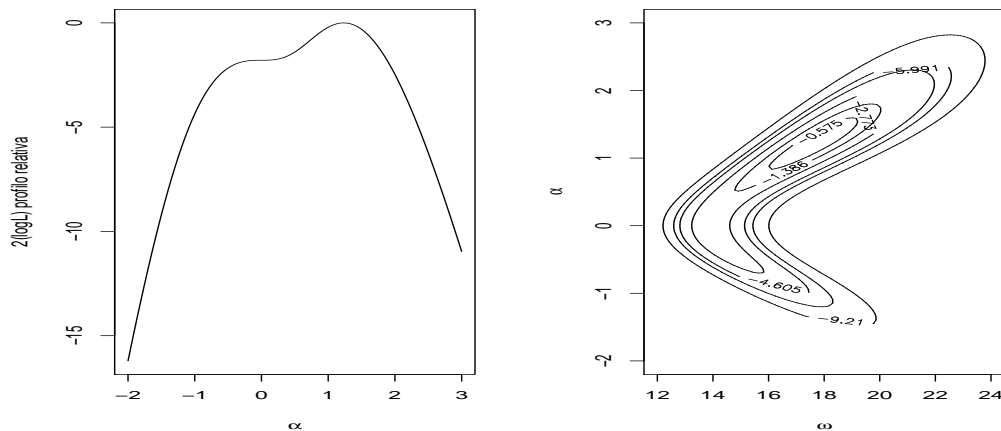


Figura 1.2: Dati AIS sul peso: $2\log$ verosimiglianza profilo relativa per α , a sinistra e per (ω, α) sulla destra

1.6 Aspetti statistici

1.6.1 Verosimiglianza e matrice d'informazione

La funzione di logverosimiglianza per il modello (1.3) relativa a una singola osservazione y è:

$$\ell(\xi, \omega, \alpha; y) = -\frac{1}{2} \log(\omega^2) - \frac{1}{2} \left(\frac{y - \xi}{\omega} \right)^2 + \log \left(2\Phi \left(\alpha \frac{y - \xi}{\omega} \right) \right) \quad (1.12)$$

la cui massimizzazione viene ottenuta per via numerica.

Il comportamento della funzione di verosimiglianza, in un intorno di $\alpha = 0$, presenta alcune caratteristiche poco piacevoli, che rendono problematica la stima dei parametri. Si vedano a riguardo i grafici in fig. 1.2. Questi si riferiscono all'usuale dataset adottato in questo filone di letteratura: i dati dell'AIS (*Australian Institute of Sport*). Questo dataset contiene le misurazioni di 13 variabili effettuate su 202 atleti australiani. L'esempio proposto utilizza solamente la variabile relativa al peso corporeo (in kg).

I grafici mostrano la logverosimiglianza profilo relativa per α (grafico di sinistra) e per (ω, α) (grafico a curve di livello di destra). Entrambi non presentano l'usuale andamento quadratico tipico delle funzioni di logverosimiglianza, mostrando invece un evidente punto di stazionarietà

in $\alpha = 0$. Questo andamento risulta tipico del modello e non specifico del *dataset* come evidenziato in Azzalini (1985) e successivamente in Arnold et al. (1993) e in Azzalini & Capitanio (1999).

In Azzalini (1985) troviamo l'espressione della matrice di informazione attesa di Fisher per il modello $Y = \xi + \omega Z$ definito in (1.3) data dall'espressione:

$$\mathbf{I} = \begin{pmatrix} (1 + \alpha^2 a_0) / \omega^2 & \left(b\delta \left(\frac{1+2\alpha^2}{1+\alpha^2} \right) + \alpha^2 a_1 \right) / \omega^2 & \left(\frac{b}{(1+\alpha^2)^{3/2}} - \alpha a_1 \right) / \omega \\ \left(b\delta \left(\frac{1+2\alpha^2}{1+\alpha^2} \right) + \alpha^2 a_1 \right) / \omega^2 & (2 + \alpha^2 a_2) / \omega^2 & -\alpha a_2 / \omega^2 \\ (2 + \alpha^2 a_2) / \omega^2 & -\alpha a_2 / \omega^2 & a_2 \end{pmatrix}$$

dove le quantità a_k con $k = 0, 1, 2$ sono delle funzioni del parametro di forma α da calcolare numericamente e definite come:

$$a_k = a_k(\alpha) = \mathbb{E} \left[Z^k \left(\frac{\phi(\alpha Z)}{\Phi(\alpha Z)} \right)^2 \right]. \quad (1.13)$$

Un aspetto problematico della matrice di informazione attesa per questo modello statistico è che diventa singolare quando $\alpha \rightarrow 0$.

I risultati sulla distribuzione asintotica delle stime di massima verosimiglianza affermano che, in condizioni di regolarità del problema di stima, dato un campione di numerosità n :

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I(\theta)^{-1})$$

dove θ in questo caso è il vettore dei parametri $\theta = (\xi, \omega^2, \alpha)$ e dove $I(\theta)$ è la matrice d'informazione attesa per i parametri appena descritta. Nel modello in esame, poichè si è evidenziata la possibilità che la matrice sia non invertibile, non è possibile utilizzare questi risultati.

In Rotnizsky et al. (2000) si sono studiati casi in cui la matrice d'informazione è singolare ed è stata sviluppata una teoria asintotica non-standard, mostrando come la distribuzione delle stime possa essere bimodale. In particolare in Chiogna (2005) si è studiato nel dettaglio la distribuzione asintotica delle stime di massima verosimiglianza per la normale asimmetrica

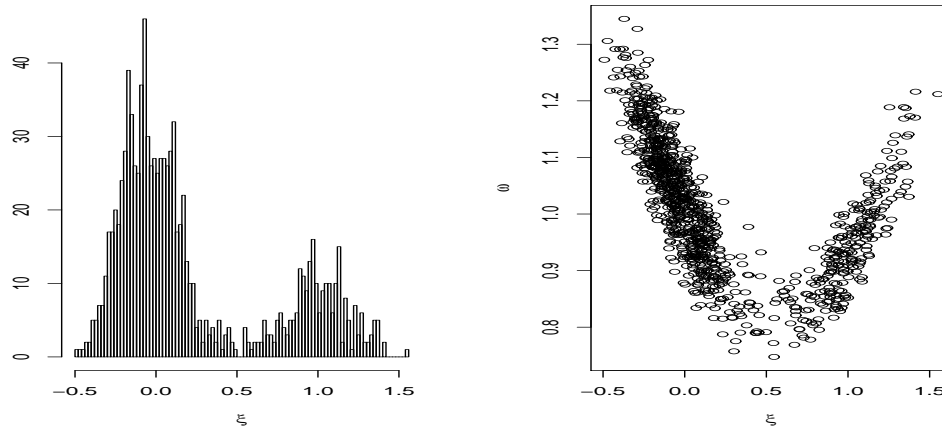


Figura 1.3: Distribuzione empirica delle stime di massima verosimiglianza da campioni simulati; istogramma per $\hat{\mu}$ sulla sinistra e grafico di dispersione per $(\hat{\mu}, \hat{\sigma})$ sulla destra

dimostrando tra l'altro, che in situazioni come quella descritta il tasso di convergenza delle stime è più lento dell'usuale $O_p(n^{-1/2})$.

Sebbene la singolarità della matrice d'informazione attesa si limiti esclusivamente al punto $\alpha = 0$ per $n \rightarrow \infty$, nella pratica, anche per campioni grandi questa caratteristica interessa anche un intorno più o meno ampio di $\alpha = 0$.

Per rendere l'idea di questo si vedano i risultati di uno studio di simulazione in Arellano-Valle & Azzalini (2008). Dopo aver generato 5000 campioni di numerosità $n = 200$ da una $SN(0, 1, 1)$, si sono calcolate le stime di massima verosimiglianza per $(\hat{\xi}, \hat{\omega}^2, \hat{\alpha})$. La distribuzione empirica delle stime di massima verosimiglianza risulta evidentemente bimodale. Nella figura 1.3 è riportata la distribuzione empirica delle stime di massima verosimiglianza, in cui la bimodalità risulta più che evidente.

1.6.2 Parametrizzazione centrata

Per ovviare ai problemi legati all'inferenza sui parametri (ξ, ω^2, α) appena menzionati, Azzalini (1985) propone una riparametrizzazione del modello. Partendo dall'identità

$$Y = \xi + \omega Z = \mu + \sigma Z_0 \quad (1.14)$$

dove Z ha una distribuzione $SN(\alpha)$ del tipo (1.2), e posto:

$$\mu_z = \mathbb{E}(Z) = b\delta, \quad \sigma_z = \mathbb{V}ar(z) = 1 - (b\delta)^2$$

definiamo $Z_0 = \frac{1}{\sigma_z}(Z - \mu_z)$, come la standardizzazione di Z .

La parametrizzazione alternativa è data da $(\mu, \sigma^2, \gamma_1)$ le cui espressioni esplicite in termini dei parametri originali sono:

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

dove γ_1 è l'indice di asimmetria, già ottenuto dal calcolo dei momenti (cfr. par. 1.5). Questa parametrizzazione è nota in letteratura con il nome di *parametrizzazione centrata* (o CP utilizzando il relativo acronimo dalla lingua inglese) in quanto viene introdotta a partire dalla variabile centrata Z_0 , mentre la parametrizzazione fin qui utilizzata è chiamata *parametrizzazione diretta* (o DP).

L'utilizzo della parametrizzazione centrata offre sicuramente numerosi benefici. Da un lato semantico, i nuovi parametri hanno un significato chiaro, più intuitivo e familiare. Come nel modello normale, infatti, μ e σ^2 rappresentano esattamente la media e la varianza della distribuzione, mentre il parametro γ_1 , rappresentando l'indice di asimmetria della distribuzione, risulta più logicamente collegato a questa rispetto al parametro α .

Da un lato più pratico e inferenziale, la parametrizzazione centrata elimina le caratteristiche che rendevano difficoltose le operazioni d'inferenza sulla normale asimmetrica. Si veda a riguardo come in fig.1.4 cambi la verosimiglianza per i nuovi parametri γ_1 e σ^2 rispetto a quella mostrata in fig.1.2, e come essa presenti un andamento più regolare, senza alcun punto di stazionarietà.

Anche il problema della singolarità della matrice viene risolto tramite la parametrizzazione centrata, rendendo quindi possibile applicare gli usuali metodi dell'inferenza asintotica.

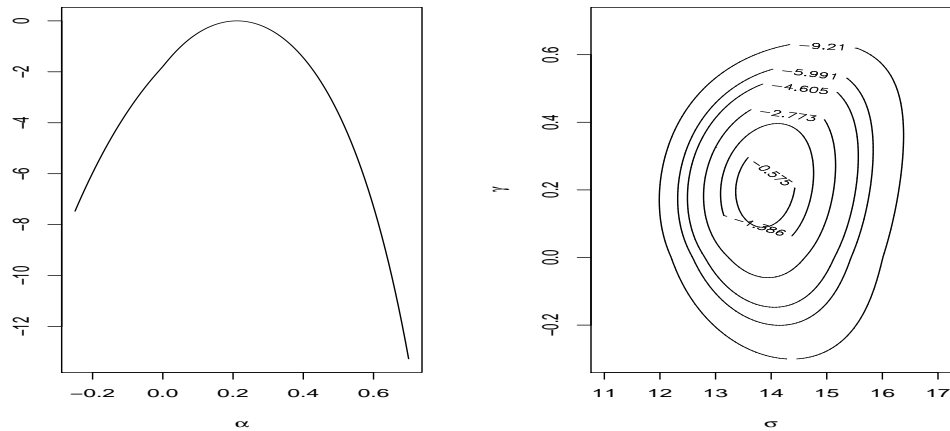


Figura 1.4: Dati AIS sul peso: 2logverosimiglianza profilo relativa per γ_1 e per (σ^2, γ_1)

Si veda a riguardo come cambiano i risultati della distribuzione empirica delle stime di massima verosimiglianza con la nuova parametrizzazione, effettuate sugli stessi campioni della fig. 1.3, riportati in fig. 1.5.

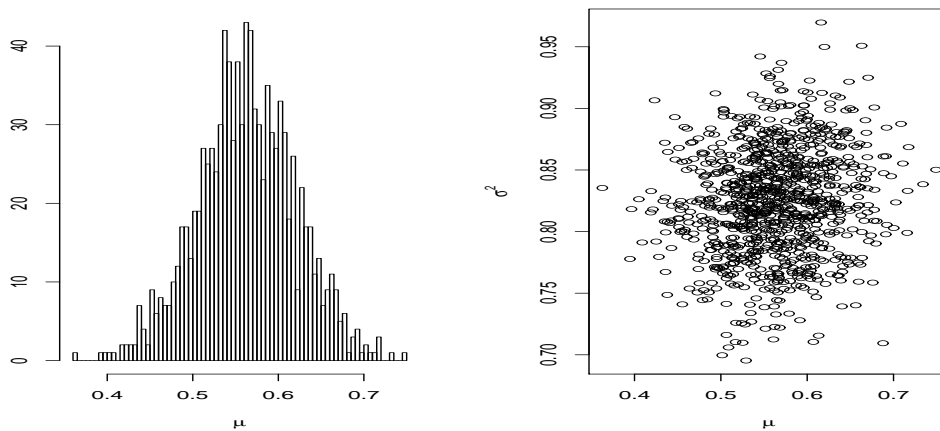


Figura 1.5: Distribuzione empirica delle stime di massima verosimiglianza da campioni simulati; istogramma per $\hat{\mu}$ sulla sinistra e grafico di dispersione per $(\hat{\mu}, \hat{\sigma}^2)$ sulla destra

Capitolo 2

La distribuzione normale asimmetrica estesa

In questo capitolo si vuole presentare la distribuzione di probabilità *normale asimmetrica estesa* (abbreviata con l'acronimo dalla lingua inglese ESN, per *Extended Skew-Normal*), tracciandone le caratteristiche principali, e sottolineando le differenze con la *normale asimmetrica*.

La normale asimmetrica estesa compare già nel primo lavoro di Azzalini (1985) e viene successivamente studiata in Arnold et al. (1993), Azzalini & Capitanio (1999), Arnold & Beaver (2000) e da Capitanio et al. (2003).

In letteratura è stato mostrato che così come la normale asimmetrica è una famiglia di distribuzioni che generalizza le distribuzioni normali, così essa è generalizzata dalla normale asimmetrica estesa. Questa distribuzione, oltre ai tre parametri della normale asimmetrica, prevede un quarto parametro e, quando esso assume il valore 0, ci si riconduce alla normale asimmetrica.

Dopo aver riportato alcuni possibili fenomeni che danno luogo a questo tipo di distribuzione, è riportata la funzione generatrice dei momenti e i primi quattro momenti sono confrontati con quelli della normale asimmetrica.

Verranno inoltre confrontate le due distribuzioni per quanto riguarda l'asimmetria, la curtosi

e la pesantezza delle code. Sarà poi descritto il comportamento della curva al variare dei parametri e le relative distribuzioni limite quando essi raggiungono i punti limite dei rispettivi domini.

2.1 Generazione

2.1.1 Campioni selezionati

Si riprende quanto riportato al paragrafo 1.3.1 ammettendo la possibilità di troncamento latente in un qualsiasi punto del dominio di u . Per completezza si riportano tutti i passaggi.

Si supponga che $\tilde{f}(v, z)$ sia la funzione di densità di una normale bivariata a media zero e varianza unitaria con correlazione pari a δ .

Immaginiamo di osservare un campione selezionato, ossia solamente gli individui che superano una certa soglia (diciamo $-\tau$, per motivi di notazione che si chiariranno in seguito) per quanto riguarda v . La densità congiunta diventa quindi

$$f(v, z) = \begin{cases} \frac{\tilde{f}(v, z)}{1 - \Phi(-\tau)} & v \geq -\tau \\ 0 & \text{altrimenti.} \end{cases} \quad (2.1)$$

Il denominatore $1 - \Phi(-\tau)$ ha la funzione di costante di normalizzazione.

Immaginiamo che sia di interesse lo studio della sola marginale z . Marginalizziamo dunque rispetto a v :

$$\begin{aligned} \int_{-\tau}^{+\infty} f(v, z) dz &= \frac{1}{1 - \Phi(-\tau)} \int_{-\tau}^{+\infty} \frac{1}{2\pi\sqrt{1 - \delta^2}} \exp \left\{ -\frac{1}{2} \left(\frac{v^2 - 2\delta vz + z^2}{1 - \delta^2} \right) \right\} dv \\ &= \frac{1}{\Phi(\tau)} \frac{1}{\sqrt{1 - \delta^2}} \int_{-\tau}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{v^2 - 2\delta vz + \delta^2 z^2}{1 - \delta^2} \right) \right\} \times \\ &\quad \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{z^2 - \delta^2 z^2}{1 - \delta^2} \right) \right\} dv \\ &= \frac{1}{\Phi(\tau)} \frac{1}{\sqrt{1 - \delta^2}} \phi(z) \int_{-\tau}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{v - \delta z}{\sqrt{1 - \delta^2}} \right)^2 \right\} dv. \end{aligned}$$

Effettuando la sostituzione $u = \frac{v-\delta z}{\sqrt{1-\delta^2}}$ e ponendo $\alpha = \frac{\delta}{\sqrt{1-\delta^2}}$, $\alpha_0(\tau) = \tau\sqrt{1+\alpha^2}$ otteniamo:

$$f_{ESN}(z; \alpha, \tau) = \frac{1}{\Phi(\tau)} \phi(z) \Phi(\alpha_0(\tau) + \alpha z). \quad (2.2)$$

Definiamo la (2.2) come la densità di una distribuzione *normale asimmetrica estesa*, con parametro di forma α e parametro di troncamento τ e scriveremo $Z \sim ESN(\alpha, \tau)$. Come per la normale asimmetrica ammettiamo variazioni di posizione e scala. Siano $\xi \in \mathbb{R}$ e $\omega \in \mathbb{R}^+$; allora definita $Y = \xi + \omega Z$ la densità di Y sarà:

$$f_{ESN}(z; \xi, \omega^2, \alpha, \tau) = \frac{1}{\omega \Phi(\tau)} \phi\left(\frac{y-\xi}{\omega}\right) \Phi\left(\alpha_0(\tau) + \alpha \frac{y-\xi}{\omega}\right) \quad (2.3)$$

e scriveremo che $Y \sim ESN(\xi, \omega^2, \alpha, \tau)$. Si noti che quando $\alpha = 0$ ci si riconduce alla distribuzione normale, per qualsiasi valore del parametro τ e che, quando $\tau = 0$, ci si riconduce alla distribuzione (1.2).

2.1.2 Distribuzione condizionata di una SN_2

Sia Y una variabile casuale normale asimmetrica bidimensionale del tipo (1.5), ossia $Y \sim SN_2(\tilde{\xi}, \tilde{\Omega}, \tilde{\alpha})$, con $\tilde{\xi} = (\xi_1, \xi_2)$, $\tilde{\alpha} = (\alpha_1, \alpha_2)$ e

$$\tilde{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}.$$

Se osserviamo una delle marginali, diciamo $Y_1 = y_1$, ricaviamo la distribuzione condizionata di $Y_2|Y_1 = y_1$. Detti:

$$\xi = \xi_2 + \omega_{21}\omega_{11}^{-1}(y_1 - \xi_1),$$

$$\omega = \omega_{22} - \omega_{21}\omega_{11}^{-1}\omega_{12},$$

$$\alpha = \frac{\alpha_1 + \omega_{12}/\sqrt{\omega_{11}\omega_{22}}\alpha_2}{(1 + \alpha_2^T \tilde{\omega}_{22.1} \alpha_2)^{1/2}},$$

$$\tau = \alpha\sqrt{\omega_{11}}^{-1}(y_1 - \xi_1)$$

dove $\bar{\omega} = \frac{\omega}{\omega_{22}}$, per la densità condizionata risulta che:

$$(Y_2|Y_1 = y_1) \sim ESN(\xi, \omega, \alpha, \tau),$$

come mostrato in Azzalini & Dalla Valle (1996) e successivamente in Azzalini & Capitanio (1999).

2.2 Momenti

La funzione generatrice dei momenti è facilmente ricavabile sfruttando i lemmi 1.3 e 1.4. Dopo alcuni sviluppi algebrici si ottiene l'espressione:

$$M_Z(t) = \frac{1}{\Phi(\tau)} \exp\left\{\frac{t^2}{2}\right\} \Phi(\delta t + \tau) \quad (2.4)$$

per la funzione generatrice dei momenti e

$$K_Z(t) = -\log(\Phi(\tau)) + \frac{t^2}{2} + \zeta_0(\delta t + \tau) \quad (2.5)$$

per la funzione generatrice dei cumulanti, nella quale compare la funzione $\zeta_0(x)$. Essa e le rispettive derivate sono definite come:

$$\begin{aligned} \zeta_0(x) &= \log(\Phi(x)), \\ \zeta_m(x) &= \frac{d^m}{dx^m} \zeta_0(x) \quad (m = 1, 2, \dots). \end{aligned}$$

Derivando la 2.5 e calcolandone il valore in $t = 0$ otteniamo i momenti della 2.2, definiti come

$$\mathbb{E}[Z] = \zeta_1(\tau)\delta, \quad (2.6)$$

$$\text{Var}[Z] = 1 + \delta^2\zeta_2(\tau), \quad (2.7)$$

$$\gamma_1(Z) = \frac{\delta^3\zeta_3(\tau)}{(1 + \delta^2\zeta_2(\tau))^{3/2}}, \quad (2.8)$$

$$\gamma_2(Z) = \frac{\delta^4 \zeta_4(\tau)}{(1 + \delta^2 \zeta_2(\tau))^2}, \quad (2.9)$$

dove, come per il par.1.5, γ_1 , γ_2 sono il terzo e il quarto cumulante standardizzato, noti anche rispettivamente come coefficiente di asimmetria e di curtosi; cfr. Capitanio et al. (2003) pag. 131 e 142. I momenti coincidono a quelli del par.1.5, nel caso in cui il parametro $\tau = 0$.

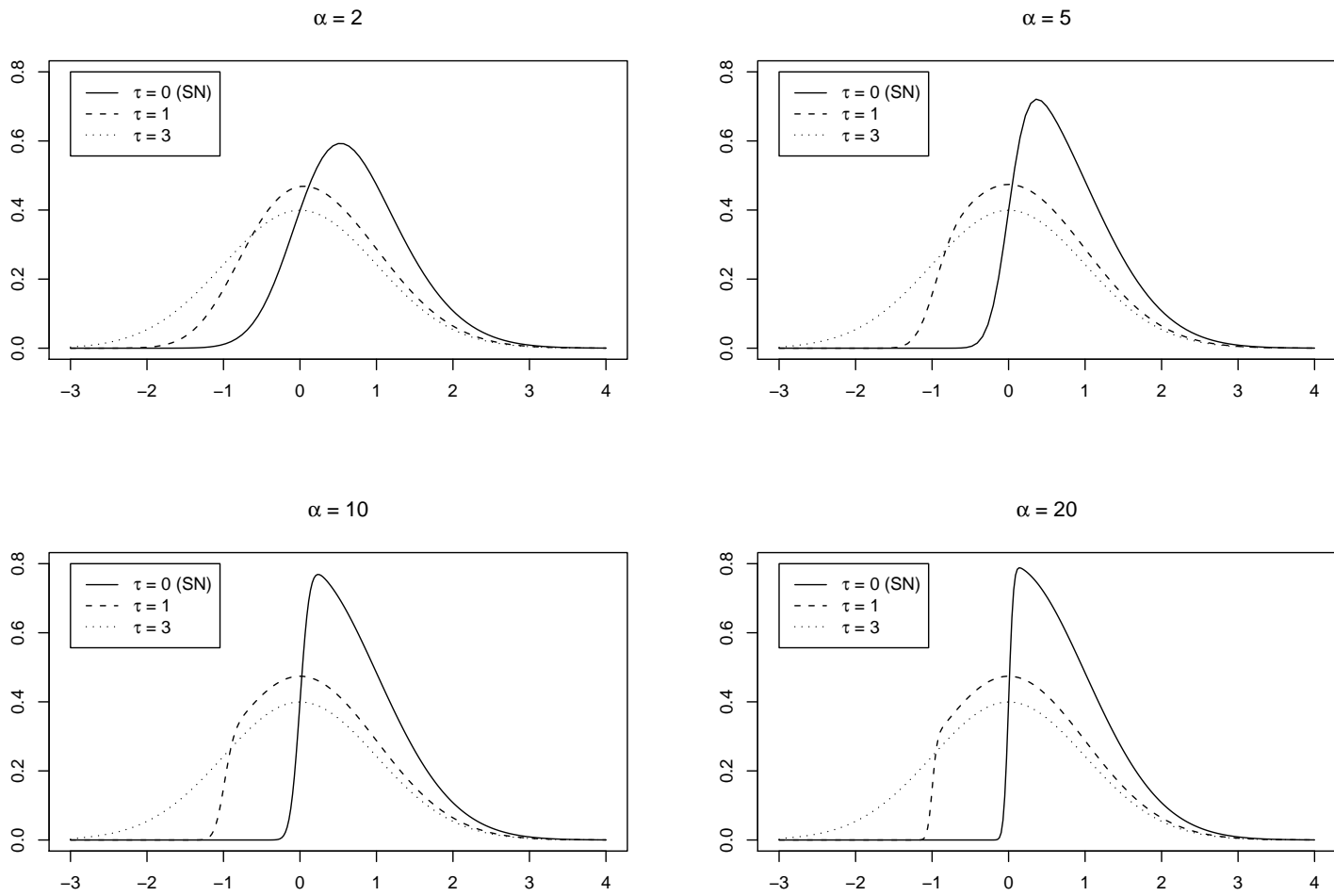
2.3 Comportamento della densità

Nelle figure 2.1, 2.2, 2.3, si riportano alcuni grafici relativi alla densità (2.2), per mostrare in che modo l'introduzione del parametro τ influisca sulla sua forma.

A differenza del parametro di forma α , che per valori opposti rispetto allo 0 porta a cambiamenti speculari sulla forma della densità, il parametro τ comporta differenti variazioni sulla forma della curva, a seconda che esso vari nel semiasse negativo o in quello positivo.

Si pensi infatti al significato del parametro τ per come lo si è presentato al paragrafo 2.1. Esso rappresenta il condizionamento al troncamento di una variabile latente correlata con la variabile d'interesse. Se questo condizionamento tende a $-\infty$ (e quindi per $\tau \rightarrow +\infty$), esso tende ad essere trascurabile. Si veda a riguardo la fig. 2.1 in cui sono riportati i grafici relativi a valori positivi del parametro τ . Come si può vedere, per valori crescenti del parametro la densità tende a perdere la sua caratteristica di asimmetria, assumendo una forma *normale*, anche per valori del parametro di asimmetria elevati.

Se al contrario ci si condiziona ad osservare la variabile d'interesse solo per quegli individui che superano un certo valore *alto* della variabile latente (e quindi per $\tau \rightarrow -\infty$), il condizionamento è molto forte, e stiamo osservando la coda della marginale. In fig. 2.2 si vede che per valori del parametro negativi *piccoli* in valore assoluto la curva mantiene una certa asimmetria già presente per il caso $\tau = 0$ e inoltre si nota una traslazione crescente della moda. Guardando i grafici in figura 2.3 è evidente come per valori prossimi allo 0 del parametro α , il decrescere di τ fa perdere totalmente l'asimmetria alla curva continuando a portare una traslazione crescente della moda. Anche per curve sensibilmente asimmetriche come quelle riportate nei quadranti

Figura 2.1: Densità per alcune scelte di $\tau \geq 0$

bassi della figura 2.3, il decrescere di τ fa diminuire l'asimmetria, facendo sembrare le curve delle normali con una determinata media dipendente dal valore di τ e varianza.

Questo comportamento della densità in funzione del parametro τ risulta nella pratica problematico. Un campione proveniente da una normale asimmetrica estesa con un valore di τ negativo, ma elevato in valore assoluto, e un campione proveniente da una normale con una determinata media e varianza saranno molto simili.

Inoltre per valori di τ intorno allo 0 la forma della densità cambia ben poco rispetto a quello di una normale asimmetrica, potendo di fatto approssimare il valore di τ a 0, come già evidenziato in Azzalini & Capitanio (1999).

Tutte queste considerazioni, che preannunciano la problematica dell'identificabilità del modello, saranno trattate nel cap. 3.

2.3.1 Asimmetria

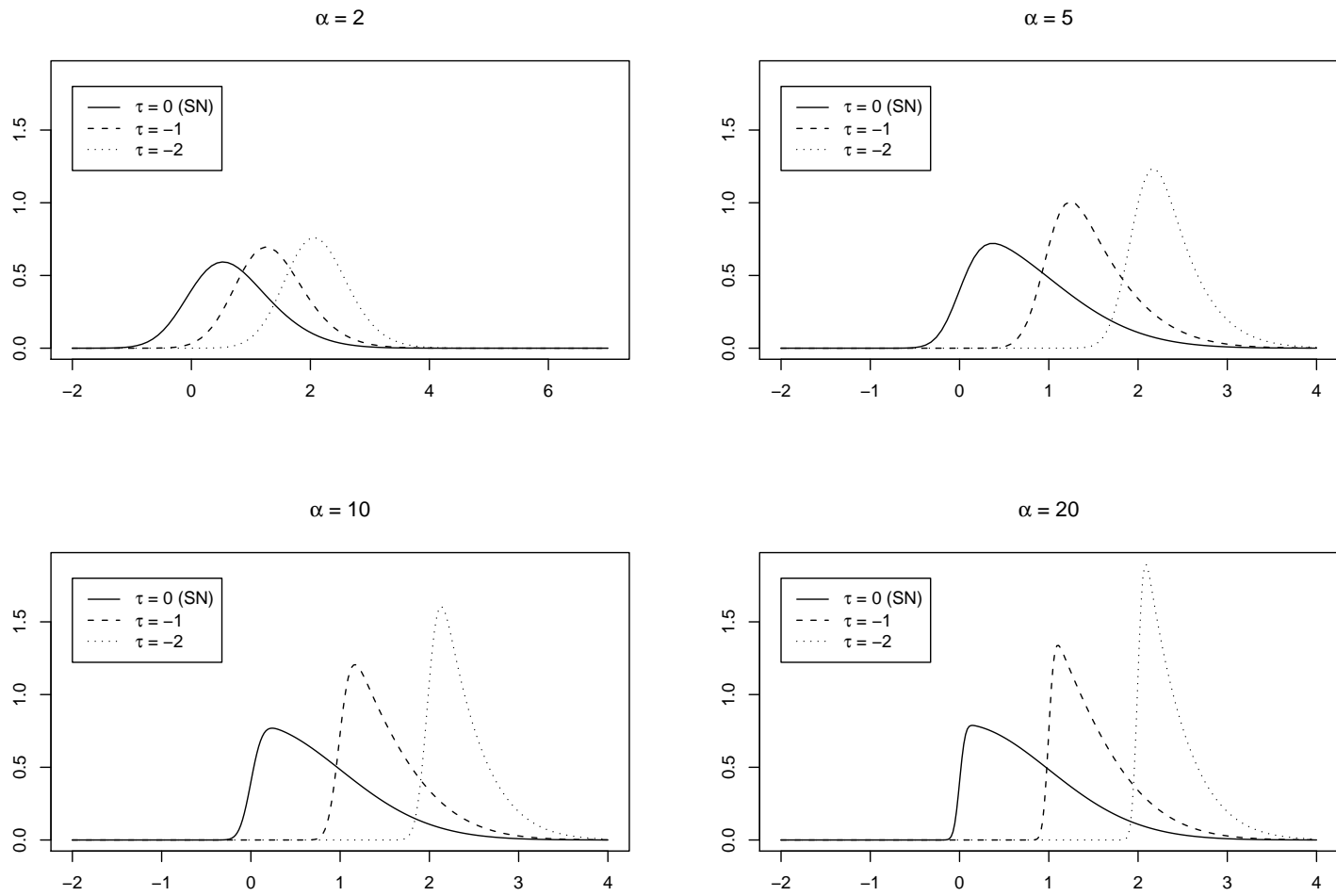
Studiamo nel dettaglio l'asimmetria di questo modello. La normale asimmetrica ha un indice di asimmetria che è costretto a rientrare in un campo di variazione compreso tra il ± 0.955 circa.

La normale asimmetrica estesa, a sua volta, permette un certo grado di asimmetria che risulta essere una funzione, non solo del parametro α , ma anche del parametro τ . Entrambi questi parametri entrano infatti nell'espressione dell'indice γ_1 presentata al par. 2.2. Il modello presenta comunque un determinato campo di variazione dell'indice di asimmetria, che si ha per $\alpha \rightarrow \pm\infty$ e $\tau \rightarrow -\infty$. Esso è compreso approssimativamente nell'intervallo $(-1.995, 1.995)$. Per vedere come il parametro τ influisce sull'indice di asimmetria si veda la fig. 2.4

Come già evidenziato precedentemente, per valori positivi di τ l'indice di asimmetria tende velocemente a zero, portando di fatto a una normale asimmetrica estesa, *non asimmetrica*.

A seconda del valore di α l'asimmetria cresce per valori prossimi allo 0 di τ , per poi decrescere fino ad annullarsi con una velocità minore rispetto a quella evidenziata per τ crescente. Tuttavia per $\alpha \rightarrow \pm\infty$ l'asimmetria raggiunge e continua ad assumere il suo massimo anche per τ decrescente.

Nelle figure 2.5 e 2.6 è sempre riportato il valore dell'indice di asimmetria γ_1 , in funzione questa volta di α per alcune scelte del parametro τ . Risulta anche qui evidente che, anche per

Figura 2.2: Densità per alcune scelte di $\tau \leq 0$

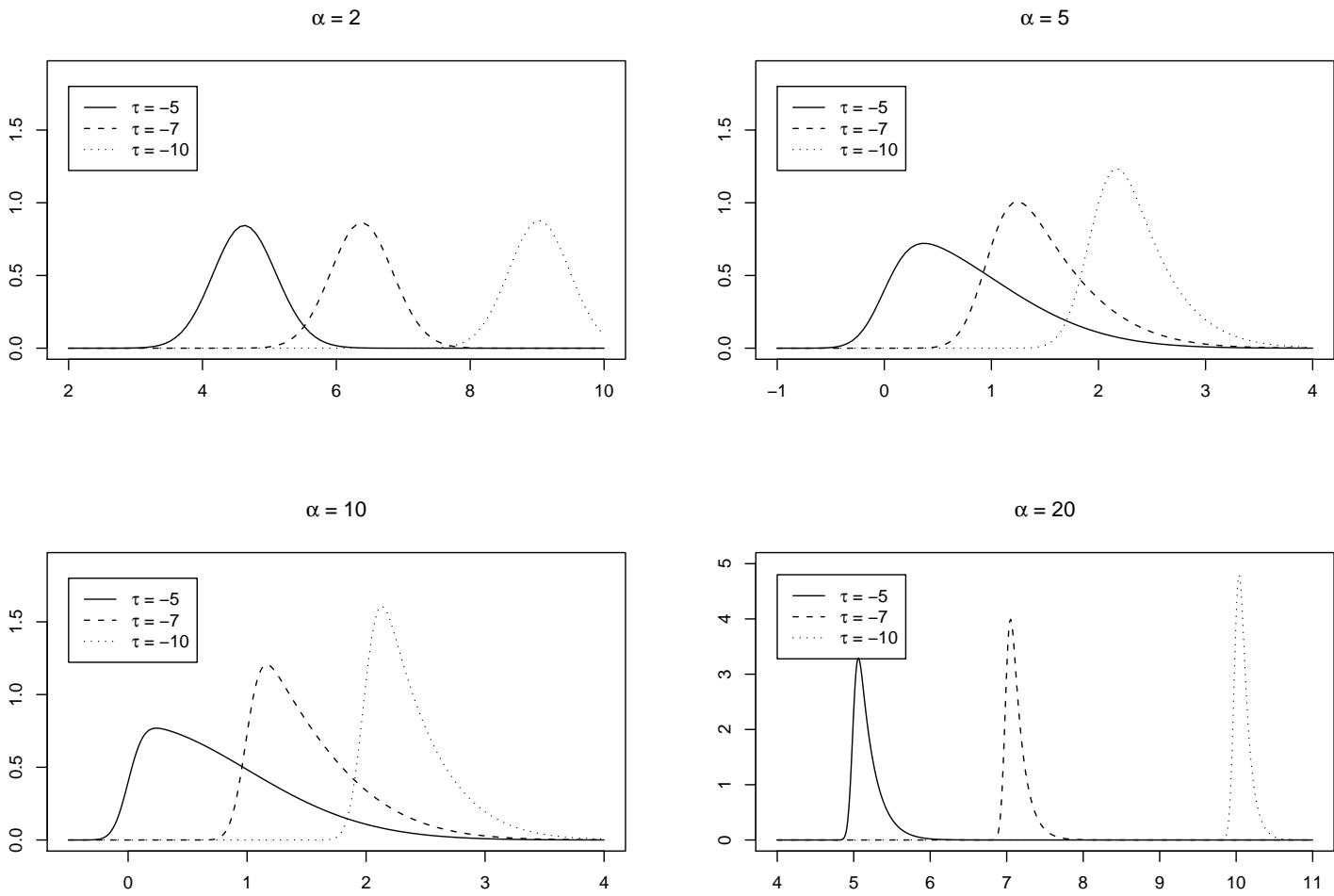


Figura 2.3: Densità per alcune scelte di $\tau < 0$. (Il quarto grafico riporta una scala per l'asse delle ordinate differente)

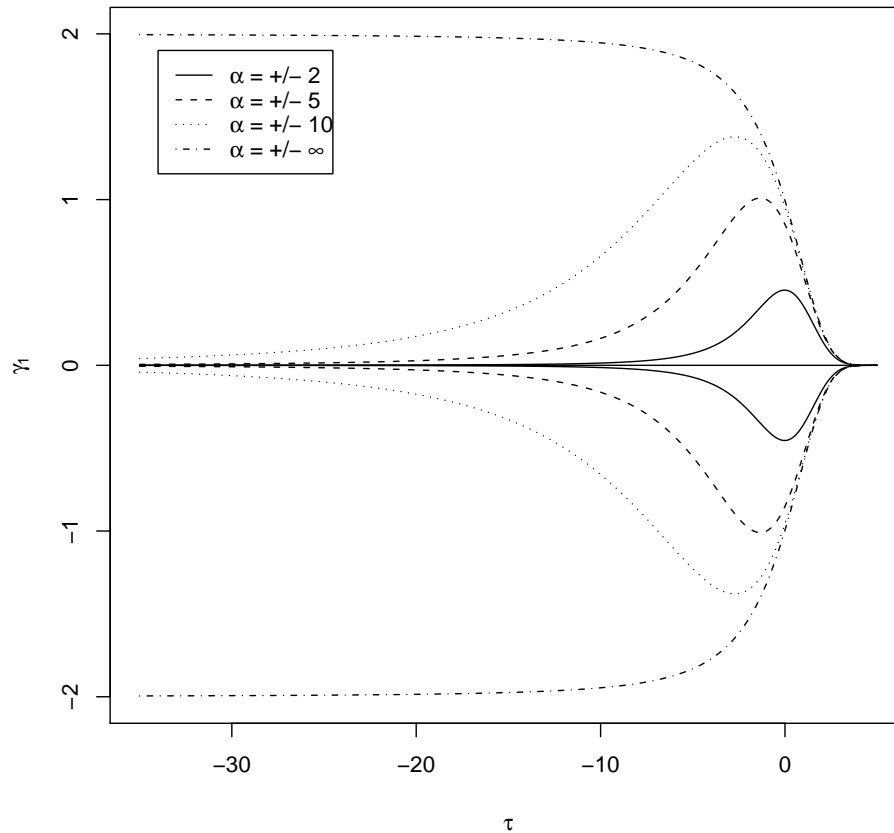


Figura 2.4: Indice di asimmetria γ_1 in funzione di τ per alcune scelte di α

valori grandi di α , quando τ è positivo l'asimmetria è confinata a variare in una prossimità tale dello 0 da renderla di fatto trascurabile (si veda in particolare la linea punteggiata della fig.2.5 che rappresenta il campo di variazione di γ_1 quando $\tau = 3$).

Per valori negativi e piccoli in valore assoluto di τ , invece, vediamo come siano plausibili valori dell'indice di asimmetria anche esterni all'intervallo $(-0.955, 0.955)$ della normale asimmetrica.

Si noti inoltre la differenza tra la linea tratteggiata e quella punteggiata, nel secondo grafico della figura 2.6. La linea tratteggiata indica che per $\tau = -10$ per valori piccoli di α si ha una asimmetria minore, del caso specifico $\tau = 0$, ma maggiore se α cresce. Nel caso in cui

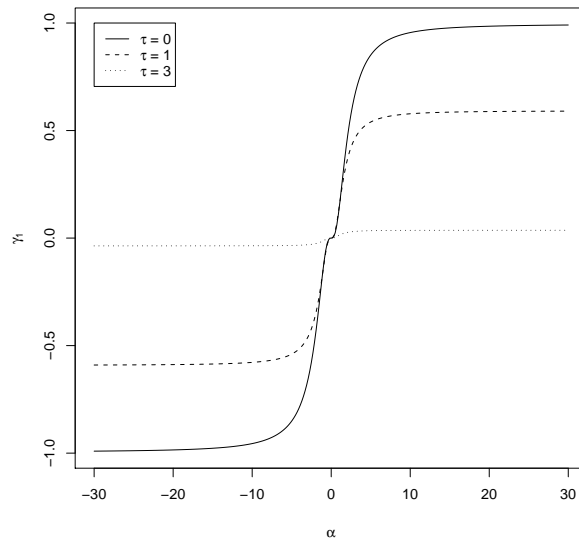


Figura 2.5: Indice di asimmetria γ_1 in funzione di α per alcune scelte di $\tau \geq 0$

$\tau = -30$ (linea punteggiata), vediamo che l'asimmetria è sempre minore del caso specifico della normale asimmetrica.

2.3.2 Curtosi

Così come per l'indice di asimmetria, anche l'indice di curtosi, nella normale asimmetrica estesa, presenta un campo di variazione più ampio rispetto a quello della normale asimmetrica.

Si vedano i grafici in fig. 2.7 e fig. 2.8 rappresentanti il valore assunto da γ_2 in funzione di α e τ . Sia nei grafici della figura 2.7 che nella fig. 2.8 si nota che al variare di τ l'indice può assumere anche valori non ammessi nel caso particolare di $\tau = 0$.

L'indice di curtosi ha un comportamento per $\tau \rightarrow \pm\infty$ simile all'indice di asimmetria appena descritto ossia tende ad annullarsi.

2.3.3 Comportamenti limite

Studiamo analiticamente cosa succede alla densità nel caso in cui il valore del parametro $\tau \rightarrow \pm\infty$.

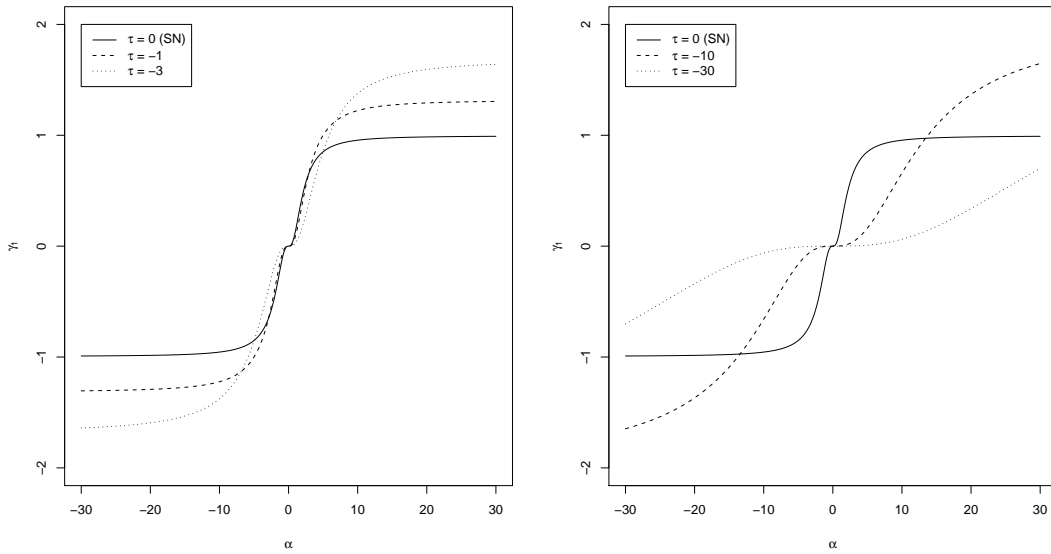


Figura 2.6: Indice di asimmetria γ_1 in funzione di α per alcune scelte di $\tau \leq 0$

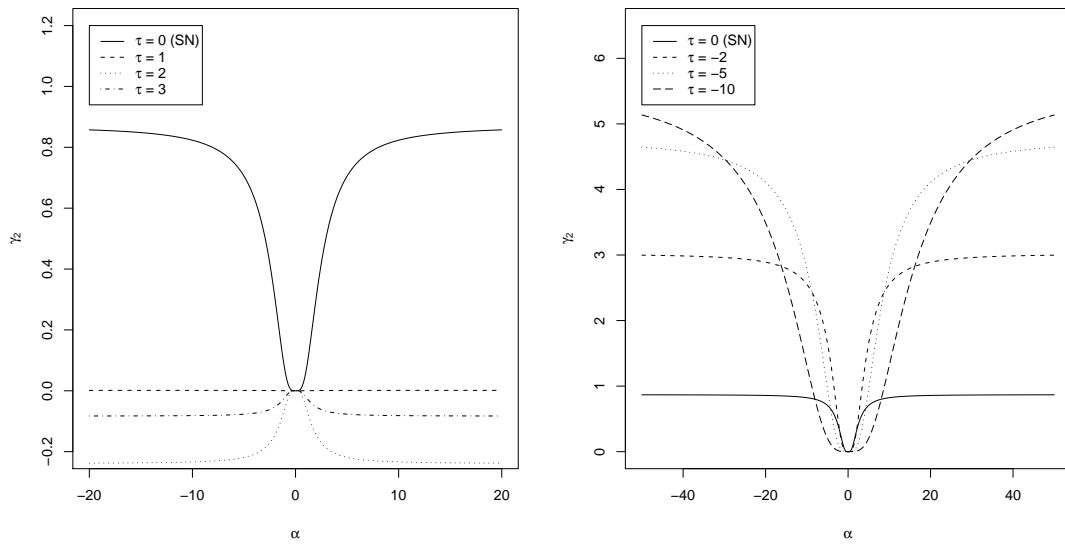


Figura 2.7: Indice di curtosi γ_2 in funzione di α : a sinistra per $\tau \geq 0$ e a destra per $\tau \leq 0$

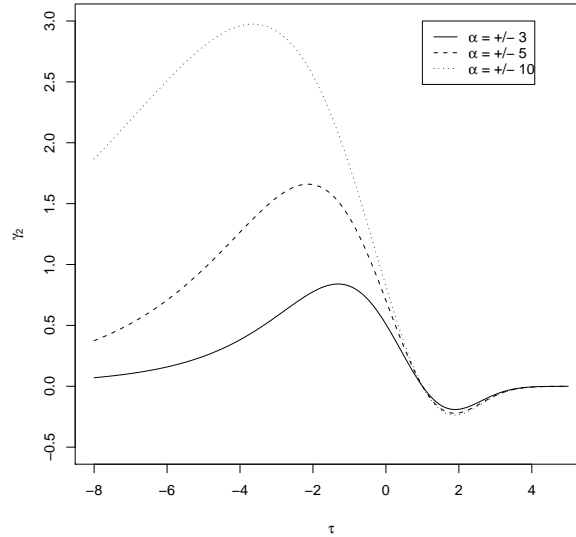


Figura 2.8: Indice di curtosi γ_2 in funzione di τ

Per $\tau \rightarrow -\infty$ abbiamo che, per il teorema di de l'Hospital:

$$\begin{aligned}
 \lim_{\tau \rightarrow -\infty} \frac{\phi(z)\Phi(\alpha_0(\tau) + \alpha z)}{\Phi(\tau)} &= \lim_{\tau \rightarrow -\infty} \frac{\phi(z)\phi(\alpha_0(\tau) + \alpha z)\sqrt{\alpha^2 + 1}}{\phi(\tau)} \\
 &= \lim_{\tau \rightarrow -\infty} \frac{\frac{1}{2\pi} \exp\left\{-\frac{z^2}{2}\right\} \sqrt{\alpha^2 + 1} \exp\left\{-\frac{1}{2}(\alpha_0(\tau) + \alpha z)^2\right\}}{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\tau^2}{2}\right\}} \\
 &= \lim_{\tau \rightarrow +\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1 - \delta^2}} \exp\left\{-\frac{1}{2} \left(\frac{z + \delta\tau}{\sqrt{1 - \delta^2}}\right)^2\right\} \\
 &= \lim_{\tau \rightarrow +\infty} \frac{1}{\sqrt{1 - \delta^2}} \phi\left(\frac{z + \delta\tau}{\sqrt{1 - \delta^2}}\right),
 \end{aligned}$$

dove si è posto $\delta = \frac{\alpha}{\sqrt{\alpha^2 + 1}}$. La distribuzione tende ad essere una *normale* di media $-\delta\tau$ e varianza $1 - \delta^2$, annullando come già osservato l'asimmetria. Anche il calcolo dei momenti per $\tau \rightarrow -\infty$ mostra quanto appena affermato, infatti:

$$\lim_{\tau \rightarrow -\infty} \mathbb{E}[Z] = \lim_{\tau \rightarrow -\infty} \zeta_1(\tau)\delta = +\infty,$$

$$\begin{aligned}\lim_{\tau \rightarrow -\infty} \text{Var}[Z] &= 1 + \delta^2 \lim_{\tau \rightarrow -\infty} \zeta_2(\tau) \\ &= 1 - \delta^2, \\ \lim_{\tau \rightarrow -\infty} \gamma_1[Z] &= \lim_{\tau \rightarrow -\infty} \delta^3 \zeta_3(\tau) = 0.\end{aligned}$$

Pensando alla generazione della distribuzione, così come la si è presentata al paragrafo 2.1, il tendere a $-\infty$ di τ è implicato da un condizionamento molto forte all'osservare gli individui che superano un valore sulla coda della variabile latente. A causa di ciò la stessa variabile d'interesse sarà osservata sulla coda destra. Questo fatto giustifica lo spostamento della moda evidenziato empiricamente dai grafici alle pag.24 e 25, e la distribuzione limite con media tendente a $+\infty$ appena dimostrata.

Per $\tau \rightarrow +\infty$, invece, la densità tende a quella di una normale di media ξ e varianza ω^2 , come era intuibile dai grafici a pagina 22. Infatti è evidente che:

$$\lim_{\tau \rightarrow +\infty} \frac{1}{\Phi(\tau)} \phi\left(\frac{x - \xi}{\omega}\right) \Phi\left(\alpha_0(\tau) + \alpha\left(\frac{x - \xi}{\omega}\right)\right) = \phi\left(\frac{x - \xi}{\omega}\right).$$

In questo caso il condizionamento effettuato per valori molto bassi della marginale latente tende ad essere di fatto trascurabile.

2.3.4 Code a confronto

Si confrontano ora le code della normale asimmetrica estesa con quelle della normale asimmetrica, calcolando il limite del rapporto delle densità della prima sulla densità della seconda per $x \rightarrow \pm\infty$. Il calcolo è stato fatto considerando $\alpha \geq 0$; i risultati sarebbero qualitativamente analoghi se considerassimo $\alpha \leq 0$ ma, in questo caso, dovremo scambiare il risultato valido per $\tau \rightarrow +\infty$ con quello per $\tau \rightarrow -\infty$.

Per quanto riguarda la coda destra facilmente otteniamo:

$$\begin{aligned}\lim_{x \rightarrow +\infty} \frac{\frac{1}{\Phi(\tau)} \phi(x) \Phi(\alpha_0(\tau) + \alpha x)}{2\phi(x) \Phi(\alpha x)} &= \frac{1}{2\Phi(\tau)} \lim_{x \rightarrow +\infty} \frac{\Phi(\alpha_0(\tau) + \alpha x)}{\Phi(\alpha x)} \\ &= \frac{1}{2\Phi(\tau)},\end{aligned}$$

ossia che la normale asimmetrica estesa risulta avere la coda destra più o meno pesante di quella della normale asimmetrica in funzione del parametro τ . In particolare per $\tau < 0$ è più pesante la coda della normale asimmetrica estesa mentre, per $\tau > 0$, è vero il contrario.

Per quanto riguarda la coda sinistra, invece risulta che:

$$\lim_{x \rightarrow -\infty} \frac{\frac{1}{\Phi(\tau)} \phi(x) \Phi(\alpha_0(\tau) + \alpha x)}{2\phi(x) \Phi(\alpha x)} = \frac{1}{2\Phi(\tau)} \lim_{x \rightarrow -\infty} \frac{\Phi(\alpha_0(\tau) + \alpha x)}{\Phi(\alpha x)}$$

Utilizzando il teorema di de l'Hospital,

$$\begin{aligned} \lim_{x \rightarrow -\infty} \frac{\frac{1}{\Phi(\tau)} \phi(x) \Phi(\alpha_0(\tau) + \alpha x)}{2\phi(x) \Phi(\alpha x)} &= \frac{1}{2\Phi(\tau)} \lim_{x \rightarrow -\infty} \frac{\phi(\alpha_0(\tau) + \alpha x)}{\phi(\alpha x)} \\ &= \frac{1}{2\Phi(\tau)} \lim_{x \rightarrow +\infty} \exp \left\{ \alpha_0(\tau) \alpha x + \frac{1}{2} \alpha_0(\tau)^2 \right\} \\ &= +\infty. \end{aligned}$$

Il rapporto tra le densità di una normale asimmetrica estesa e di una normale asimmetrica per $x \rightarrow -\infty$ diverge all'infinito con la velocità di $\exp(\alpha_0(\tau)\alpha x)$. La coda sinistra della normale asimmetrica estesa risulta quindi più pesante.

2.4 Problemi di identificabilità del modello

Si è già accennato alla possibilità di avere un modello difficilmente identificabile. Prima di analizzare come in pratica questo renda non semplici le operazioni di inferenza, presentate nel prossimo capitolo, si vuole far percepire concretamente questa caratteristica del modello con l'ausilio di un grafico. In figura 2.9 sono state tracciate le densità di una normale asimmetrica estesa, di una normale asimmetrica e di una normale. Si noti come le tre curve siano molto simili.

Per la curva relativa alla *ESN* il valore di $\tau = -3$ porta ad una curva ben poco asimmetrica nonostante $\alpha = 3$. Nel caso della normale asimmetrica il valore del parametro α è basso e pari solo a 1. Di fatto l'esigua asimmetria può essere trascurata, e anche la densità di una normale, con valori opportunamente scelti per media e varianza, mostra un grafico molto

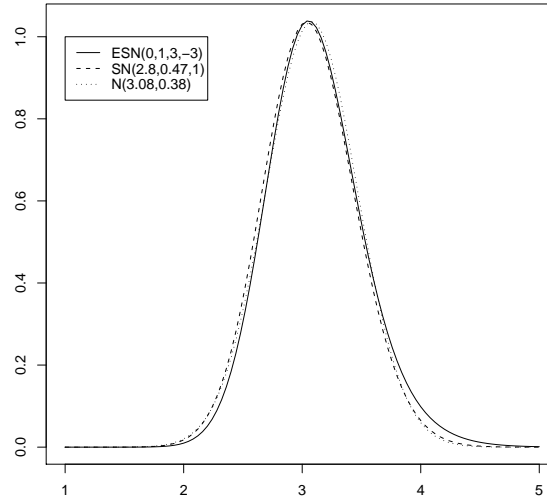


Figura 2.9: Densità di una $ESN(0, 1.3, -3)$ (linea continua), di una $SN(2.8, 0.47, 1)$ (linea tratteggiata) e di una normale $N(3.08, 0.38)$ (linea punteggiata)

simile ai precedenti. Togliendo la legenda dal grafico, infatti, sarebbe molto difficile riuscire a individuare quale curva segua una distribuzione normale, quale SN e quale ESN .

2.5 Caso multivariato

Si riporta la densità della distribuzione normale asimmetrica estesa d -dimensionale che verrà utilizzata nel caso particolare $d = 2$ nel capitolo 4 per la costruzione di regioni di tolleranza.

Definiamo normale asimmetrica estesa d -dimensionale con vettore di posizione ξ , matrice di scala Ω , vettore di asimmetria α e parametro τ , una una funzione da \mathbb{R}^d in \mathbb{R} con densità di probabilità:

$$f(y; \xi, \Omega, \alpha, \tau) = \phi_d(y - \xi; \Omega) \Phi(\alpha_0 + \alpha^T \omega^{-1}(y - \xi)) / \Phi(\tau) \quad (2.10)$$

dove $\alpha_0 = \tau(1 + \alpha^T \Omega^{-1} \alpha)^{1/2}$, $\phi_d(y - \xi; \Omega)$ è la densità di una normale d -dimensionale con vettore delle medie pari a ξ e matrice di varianza Ω .

Capitolo 3

Aspetti statistici

In questo capitolo verranno presentati alcuni aspetti di carattere statistico-inferenziale della distribuzione normale asimmetrica estesa.

Nel paragrafo 3.1 viene presentata la funzione di verosimiglianza della distribuzione e ne viene ricavato il vettore punteggio. Nel paragrafo 3.2 viene ricavata la matrice di informazione attesa di Fisher, risultato non noto in letteratura. Essa sarà confrontata con quella della normale asimmetrica riportata al cap. 2 e, in particolare, ne verrà dimostrata la singolarità in $\alpha = 0$.

Nel paragrafo 3.3 verrà studiata la funzione di verosimiglianza al fine di ottenere le stime di massima verosimiglianza per il modello. Verrà dimostrato che, come nel caso della normale asimmetrica, la funzione di logverosimiglianza profilo per α presenta un punto di stazionarietà in $\alpha = 0$ e si mostrerà inoltre che la logverosimiglianza profilo per τ è scarsamente informativa in quanto a inferenza sul parametro. Verranno poi discussi alcuni metodi di massimizzazione numerica della verosimiglianza, mostrando come in alcuni casi si presentino situazioni di quasi *non identificabilità* del modello.

Nel paragrafo 3.4, un esperimento di simulazione simile a quello di Arellano-Valle & Azzalini (2008) studierà il comportamento della distribuzione delle stime di massima verosimiglianza.

In fine, nel paragrafo 3.5, viene discussa una parametrizzazione alternativa che ripercorre concettualmente la parametrizzazione centrata della normale asimmetrica.

3.1 Funzione di verosimiglianza e vettore punteggio

Sia $X \sim ESN(\xi, \omega^2, \alpha, \tau)$. Data un'osservazione x di X , la funzione di logverosimiglianza risulta essere:

$$\ell(\xi, \omega^2, \alpha, \tau; x) = -\zeta_0(\tau) - \frac{1}{2} \log(\omega^2) - \frac{1}{2} \left(\frac{x - \xi}{\omega} \right)^2 + \zeta_0 \left(\alpha_0 + \alpha \frac{x - \xi}{\omega} \right). \quad (3.1)$$

Posto $z = \frac{x - \xi}{\omega}$, il calcolo delle derivate parziali porta al vettore punteggio:

$$\ell' = \begin{pmatrix} \frac{1}{\omega} [z - \alpha \zeta_1(\alpha_0 + \alpha z)] \\ \frac{1}{\omega} [z^2 - \alpha z \zeta_1(\alpha_0 + \alpha z) - 1] \\ \zeta_1(\alpha_0 + \alpha z) \left(\frac{\tau \alpha}{\sqrt{\alpha^2 + 1}} + z \right) \\ \zeta_1(\alpha_0 + \alpha z) \sqrt{\alpha^2 + 1} - \zeta_1(\tau) \end{pmatrix}.$$

Le equazioni di verosimiglianza $\frac{d}{d\theta} \ell(\theta) = 0$ non hanno una soluzione esplicita. Il calcolo delle stime di massima verosimiglianza dovrà quindi essere effettuato per via numerica (cfr. 3.3.2).

Esplicitiamo inoltre

$$\zeta_1(x) = \frac{\phi(x)}{\Phi(x)}, \quad (3.2)$$

$$\zeta_2(x) = -(\zeta_1(x)x + [\zeta_1(x)]^2), \quad (3.3)$$

derivata prima e seconda della funzione $\zeta_0(\cdot)$ introdotta a pag. 20 poiché entreranno numerose volte nelle prossime espressioni.

3.2 Matrice d'informazione

In questo paragrafo si riportano tutti i passaggi relativi al calcolo della matrice d'informazione attesa e alcune considerazioni di carattere qualitativo su di essa.

3.2.1 Calcolo della matrice d'informazione

Calcolando le derivate seconde parziali del vettore punteggio otteniamo le espressioni:

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \xi^2} &= \frac{1}{\omega^2} (\alpha^2 \zeta_2(\alpha_0 + \alpha z) - 1), \\
\frac{\partial^2 \ell}{\partial \xi \partial \omega} &= \frac{1}{\omega^2} (-2z + \alpha \zeta_1(\alpha_0 + \alpha z) + \alpha^2 z \zeta_2(\alpha_0 + \alpha z)), \\
\frac{\partial^2 \ell}{\partial \xi \partial \alpha} &= -\frac{1}{\omega} \left(\frac{\tau \alpha^2 + \alpha z \sqrt{1 + \alpha^2}}{\sqrt{\alpha^2 + 1}} \zeta_2(\alpha_0 + \alpha z) + \zeta_1(\alpha_0 + \alpha z) \right), \\
\frac{\partial^2 \ell}{\partial \xi \partial \tau} &= -\frac{\alpha \sqrt{\alpha^2 + 1}}{\omega} \zeta_2(\alpha_0 + \alpha z), \\
\frac{\partial^2 \ell}{\partial \omega^2} &= \frac{1}{\omega^2} (-3z^2 + 2\alpha z \zeta_1(\alpha_0 + \alpha z) + (\alpha z)^2 \zeta_2(\alpha_0 + \alpha z) + 1), \\
\frac{\partial^2 \ell}{\partial \omega \partial \alpha} &= -\frac{z}{\omega} \left(\left(\frac{\tau \alpha^2}{\sqrt{\alpha^2 + 1}} + \alpha z \right) \zeta_2(\alpha_0 + \alpha z) + \zeta_1(\alpha_0 + \alpha z) \right), \\
\frac{\partial^2 \ell}{\partial \omega \partial \tau} &= -\frac{\alpha z \sqrt{\alpha^2 + 1}}{\omega} \zeta_2(\alpha_0 + \alpha z), \\
\frac{\partial^2 \ell}{\partial \alpha^2} &= \left(\frac{\tau \alpha}{\sqrt{\alpha^2 + 1}} + z \right)^2 \zeta_2(\alpha_0 + \alpha z) + \frac{\tau}{(\alpha^2 + 1)^{3/2}} \zeta_1(\alpha_0 + \alpha z), \\
\frac{\partial^2 \ell}{\partial \alpha \partial \tau} &= (\alpha \tau + z \sqrt{\alpha^2 + 1}) \zeta_2(\alpha_0 + \alpha z) + \frac{\alpha}{\sqrt{\alpha^2 + 1}} \zeta_1(\alpha_0 + \alpha z), \\
\frac{\partial^2 \ell}{\partial \tau^2} &= (\alpha^2 + 1) \zeta_2(\alpha_0 + \alpha z) - \zeta_2(\tau).
\end{aligned}$$

Le espressioni ottenute per via algebrica sono state confrontate con le derivate numeriche calcolate tramite la funzione `hessian` del `package numDeriv` di R per varie combinazioni dei parametri e di x , riportando, per ogni combinazione, dei valori sostanzialmente uguali. Le differenze in valore assoluto tra i risultati ottenuti per via numerica e tramite le espressioni algebriche sono sempre di un ordine di grandezza inferiore alla precisione di calcolo impostata nella funzione `hessian` e, ragionevolmente, sono quindi da attribuirsi ad errore numerico.

Prima di passare al calcolo degli elementi della matrice d'informazione attesa si introduce il seguente lemma:

Lemma 3.1. *Sia Z una variabile casuale tale che $Z \sim ESN(0, 1, \alpha, \tau)$, $\zeta_1(\cdot)$ la funzione definita in (3.2) e $g(x) = \alpha\tau + \sqrt{\alpha^2 + 1}x$, allora:*

$$\mathbb{E}[f(Z)\zeta_1(\alpha_0 + \alpha Z)] = \frac{\zeta_1(\tau)}{\sqrt{\alpha^2 + 1}} \mathbb{E}[f(g^{-1}(U))]$$

dove $U \sim N(0, 1)$ e $f(\cdot)$ è una qualunque funzione da \mathbb{R} in \mathbb{R} tale per cui gli integrali coinvolti esistano.

Dimostrazione.

$$\begin{aligned} \mathbb{E}[f(Z)\zeta_1(\alpha_0 + \alpha Z)] &= \int_{\mathbb{R}} f(z) \frac{\phi(\alpha_0 + \alpha z)}{\Phi(\alpha_0 + \alpha z)} \times \frac{1}{\Phi(\tau)} \phi(z) \Phi(\alpha_0 + \alpha z) dz \\ &= \frac{1}{\Phi(\tau)} \int_{\mathbb{R}} f(z) \phi(z) \phi(\alpha_0 + \alpha z) dz \\ &= \frac{1}{\Phi(\tau)} \int_{\mathbb{R}} f(z) \frac{1}{2\pi} \exp \left\{ -\frac{1}{2} (z^2 + (\alpha_0 + \alpha z)^2) \right\} dz \\ &= \frac{1}{\Phi(\tau)} \int_{\mathbb{R}} f(z) \frac{1}{2\pi} \exp \left\{ -\frac{1}{2} (\alpha_0^2 + 2z\alpha\alpha_0 + (\alpha^2 + 1)z^2) \right\} dz \\ &= \frac{1}{\Phi(\tau)} \int_{\mathbb{R}} f(z) \frac{1}{2\pi} \exp \left\{ -\frac{1}{2} \left(\alpha\tau + \sqrt{\alpha^2 + 1}z \right)^2 \right\} \times \\ &\quad \exp \left\{ -\frac{\tau^2}{2} \right\} dz \\ &= \frac{\phi(\tau)}{\Phi(\tau)} \int_{\mathbb{R}} f(z) \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\alpha\tau + \sqrt{\alpha^2 + 1}z \right)^2 \right\} dz \end{aligned}$$

Effettuiamo una sostituzione, ponendo $u = g(z)$. Semplici calcoli portano a:

$$\begin{aligned} \mathbb{E}[f(Z)\zeta_1(\alpha_0 + \alpha Z)] &= \frac{\zeta_1(\tau)}{\sqrt{\alpha^2 + 1}} \int_{\mathbb{R}} f(g^{-1}(u)) \phi(u) du \\ &= \frac{\zeta_1(\tau)}{\sqrt{\alpha^2 + 1}} \mathbb{E}[f(g^{-1}(U))]. \end{aligned}$$

□

Si riportano quindi alcuni risultati utili nel calcolo dei valori attesi delle derivate parziali.

Utilizzando il Lemma 3.1 otteniamo:

$$\begin{aligned}
\mathbb{E}[\zeta_1(\alpha_0 + \alpha Z)] &= \frac{\zeta_1(\tau)}{\sqrt{\alpha^2 + 1}}, \\
\mathbb{E}[Z\zeta_1(\alpha_0 + \alpha Z)] &= -\alpha\tau \frac{\zeta_1(\tau)}{\alpha^2 + 1}, \\
\mathbb{E}[Z^2\zeta_1(\alpha_0 + \alpha Z)] &= (1 + \alpha^2\tau^2) \frac{\zeta_1(\tau)}{(\alpha^2 + 1)^{3/2}}, \\
\mathbb{E}[Z^3\zeta_1(\alpha_0 + \alpha Z)] &= -(3\alpha\tau + \alpha^3\tau^3) \frac{\zeta_1(\tau)}{(\alpha^2 + 1)^2}, \\
\mathbb{E}[(\alpha_0 + \alpha Z)\zeta_1(\alpha_0 + \alpha Z)] &= \frac{\tau\zeta_1(\tau)}{\alpha^2 + 1}, \\
\mathbb{E}[\zeta_2(\alpha_0 + \alpha Z)] &= -\frac{\tau\zeta_1(\tau)}{\alpha^2 + 1} - \mathbb{E}[\zeta_1(\alpha_0 + \alpha Z)^2],
\end{aligned}$$

dove l'ultima espressione si ottiene utilizzando il risultato (3.3).

Gli elementi della matrice d'informazione risultano essere:

$$\begin{aligned}
-\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \xi^2}\right] &= \frac{1}{\omega^2} (1 - \alpha^2 \mathbb{E}[\zeta_2(\alpha_0 + \alpha Z)]) \\
&= \frac{1}{\omega^2} (1 + \alpha^2 \mathbb{E}[\zeta_1(\alpha_0 + \alpha Z)(\alpha_0 + \alpha Z)] + \alpha^2 \mathbb{E}[\zeta_1(\alpha_0 + \alpha Z)^2]) \\
&= \frac{1}{\omega^2} (1 + \delta^2 \tau \zeta_1(\tau) + \alpha^2 a_0), \tag{3.4}
\end{aligned}$$

$$\begin{aligned}
-\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \xi \partial \omega}\right] &= \frac{1}{\omega^2} \left(2\mathbb{E}[Z] - \alpha \mathbb{E}[\zeta_1(\alpha_0 + \alpha Z)] - \alpha^2 \mathbb{E}[Z\zeta_2(\alpha_0 + \alpha Z)] \right) \\
&= \frac{1}{\omega^2} \left(2\delta\zeta_1(\tau) - \delta\zeta_1(\tau) + \alpha^2 \mathbb{E}[Z\zeta_1(\alpha_0 + \alpha Z)(\alpha_0 + \alpha Z)] + \right. \\
&\quad \left. \alpha^2 \mathbb{E}[Z\zeta_1(\alpha_0 + \alpha Z)^2] \right) \\
&= \frac{1}{\omega^2} \left(\delta\zeta_1(\tau) + \alpha^2 \alpha_0 \mathbb{E}[Z\zeta_1(\alpha_0 + \alpha Z)] + \alpha^3 \mathbb{E}[Z^2\zeta_1(\alpha_0 + \alpha Z)] + \right. \\
&\quad \left. \alpha^2 a_1 \right) \\
&= \frac{1}{\omega^2} \left(\delta\zeta_1(\tau) + \alpha^2 \tau \sqrt{\alpha^2 + 1} \left(\frac{-\alpha\tau\zeta_1(\tau)}{\alpha^2 + 1} \right) + \right.
\end{aligned}$$

$$\begin{aligned}
& \alpha^3 \left((1 + \alpha^2 \tau^2) \frac{\zeta_1(\tau)}{(\alpha^2 + 1)^{3/2}} \right) + \alpha^2 a_1 \\
= & \frac{1}{\omega^2} \left(\delta \zeta_1(\tau) \left(\frac{2\alpha^2 + 1}{\alpha^2 + 1} \right) - \frac{\alpha^3 \tau^2 \zeta_1(\tau)}{(1 + \alpha^2)^{3/2}} + \alpha^2 a_1 \right) \\
= & \frac{1}{\omega^2} \left[\zeta_1(\tau) \delta \left(\frac{1 + 2\alpha^2 - (\alpha^2 + 1)\delta^2 \tau^2}{\alpha^2 + 1} \right) + \alpha^2 a_1 \right], \tag{3.5}
\end{aligned}$$

$$\begin{aligned}
- \mathbb{E} \left[\frac{\partial^2 \ell}{\partial \xi \partial \alpha} \right] &= \frac{1}{\omega} \left(\mathbb{E}[\zeta_1(\alpha_0 + \alpha Z)] + \mathbb{E} \left[\frac{\tau \alpha^2 + \alpha Z \sqrt{\alpha^2 + 1}}{\sqrt{\alpha^2 + 1}} \zeta_2(\alpha_0 + \alpha Z) \right] \right) \\
&= \frac{1}{\omega} \left(\frac{\zeta_1(\tau)}{\sqrt{\alpha^2 + 1}} + \tau \alpha \delta \mathbb{E}[\zeta_2(\alpha_0 + \alpha Z)] + \alpha \mathbb{E}[Z \zeta_2(\alpha_0 + \alpha Z)] \right) \\
&= \frac{1}{\omega} \left(\frac{\zeta_1(\tau)}{\sqrt{\alpha^2 + 1}} - \tau \alpha \delta \left(\frac{\tau \zeta_1(\tau)}{\alpha^2 + 1} + a_0 \right) - \right. \\
&\quad \left. \alpha (\alpha_0 \mathbb{E}[Z \zeta_1(\alpha_0 + \alpha Z)] + \alpha \mathbb{E}[Z^2 \zeta_1(\alpha_0 + \alpha Z)]) - \alpha a_1 \right) \\
&= \frac{1}{\omega} \left(\frac{\zeta_1(\tau)}{\sqrt{\alpha^2 + 1}} - \tau \alpha \delta \left(\frac{\tau \zeta_1(\tau)}{\alpha^2 + 1} + a_0 \right) + \alpha \tau^2 \delta \zeta_1(\tau) - \right. \\
&\quad \left. \delta^2 \frac{1 + \alpha^2 \tau^2}{\sqrt{\alpha^2 + 1}} \zeta_1(\tau) - \alpha a_1 \right) \\
&= \frac{1}{\omega} \left(\frac{\zeta_1(\tau)}{\sqrt{\alpha^2 + 1}} - \tau^2 \delta^2 \frac{\zeta_1(\tau)}{\sqrt{\alpha^2 + 1}} - \tau \alpha \delta a_0 + \alpha \tau^2 \delta \zeta_1(\tau) - \right. \\
&\quad \left. \delta^2 \frac{1 + \alpha^2 \tau^2}{\sqrt{\alpha^2 + 1}} \zeta_1(\tau) - \alpha a_1 \right) \\
&= \frac{1}{\omega} \left(\frac{\alpha^2 + 1 - \tau^2 \alpha^2 + \alpha^2 \tau^2 (\alpha^2 + 1) - \alpha^2 - \alpha^4 \tau^2}{(1 + \alpha^2)^{3/2}} \zeta_1(\tau) - \right. \\
&\quad \left. \tau \alpha \delta a_0 - \alpha a_1 \right) \\
&= \frac{1}{\omega} \left(\frac{\zeta_1(\tau)}{(1 + \alpha^2)^{3/2}} - \tau \alpha \delta a_0 - \alpha a_1 \right), \tag{3.6}
\end{aligned}$$

$$\begin{aligned}
- \mathbb{E} \left[\frac{\partial^2 \ell}{\partial \xi \partial \tau} \right] &= \frac{\alpha \sqrt{\alpha^2 + 1}}{\omega} \mathbb{E}[\zeta_2(\alpha_0 + \alpha Z)] \\
&= - \frac{\alpha \sqrt{\alpha^2 + 1}}{\omega} \mathbb{E}[\zeta_1(\alpha_0 + \alpha Z)(\alpha_0 + \alpha Z)] + \mathbb{E}[\zeta_1(\alpha_0 + \alpha Z)^2] \\
&= - \frac{1}{\omega} \left(\alpha \sqrt{\alpha^2 + 1} \frac{\tau \zeta_1(\tau)}{\alpha^2 + 1} + \alpha \sqrt{\alpha^2 + 1} a_0 \right)
\end{aligned}$$

$$= -\frac{1}{\omega} \left(\delta\tau\zeta_1(\tau) + \alpha\sqrt{\alpha^2 + 1}a_0 \right), \quad (3.7)$$

$$\begin{aligned} -\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \omega^2} \right] &= \frac{1}{\omega^2} \left(3\mathbb{E}[Z^2] - 2\alpha\mathbb{E}[Z\zeta_1(\alpha_0 + \alpha Z)] + \alpha^2\alpha_0\mathbb{E}[Z^2\zeta_1(\alpha_0 + \alpha Z)] + \right. \\ &\quad \left. \alpha^3\mathbb{E}[Z^3\zeta_1(\alpha_0 + \alpha Z)] + \alpha^2\mathbb{E}[Z^2\zeta_1(\alpha_0 + \alpha Z)^2] - 1 \right) \\ &= \frac{1}{\omega^2} \left(3(1 + \delta^2(\zeta_2(\tau) + \zeta_1(\tau)^2)) + 2\delta^2\tau\zeta_1(\tau) + \frac{\alpha^2\alpha_0(1 + \alpha^2\tau^2)\zeta_1(\tau)}{(\alpha^2 + 1)^{3/2}} - \right. \\ &\quad \left. (3\alpha^4\tau + \alpha^6\tau^3)\frac{\zeta_1(\tau)}{(\alpha^2 + 1)^2} + \alpha^2a_2 - 1 \right) \\ &= \frac{1}{\omega^2} \left(2 + \alpha^2a_2 - 3\delta^2\zeta_1(\tau)\tau + 2\delta^2\tau\zeta_1(\tau) + \frac{\alpha^2\tau(1 + \alpha^2\tau^2)\zeta_1(\tau)}{(\alpha^2 + 1)} \right. \\ &\quad \left. \frac{(3\alpha^4\tau + \alpha^6\tau^3)\zeta_1(\tau)}{(\alpha^2 + 1)^2} \right) \\ &= \frac{1}{\omega^2} (2 + \alpha^2a_2 + \delta^4\tau(\tau^2 - 3)\zeta_1(\tau)), \quad (3.8) \end{aligned}$$

$$\begin{aligned} -\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \omega \partial \alpha} \right] &= \frac{1}{\omega} \left(\mathbb{E}[Z\zeta_1(\alpha_0 + \alpha Z)] + \mathbb{E} \left[\left(Z \frac{\tau\alpha^2}{\sqrt{\alpha^2 + 1}} + \alpha Z^2 \right) \times \right. \right. \\ &\quad \left. \left. \zeta_2(\alpha_0 + \alpha Z) \right] \right) \\ &= \frac{1}{\omega} \left(\mathbb{E}[Z\zeta_1(\alpha_0 + \alpha Z)] - \tau^2\alpha^2\mathbb{E}[Z\zeta_1(\alpha_0 + \alpha Z)] - \right. \\ &\quad \left. \tau\alpha^2\delta\mathbb{E}[Z^2\zeta_1(\alpha_0 + \alpha Z)] - \alpha\alpha_0\mathbb{E}[Z^2\zeta_1(\alpha_0 + \alpha Z)] - \right. \\ &\quad \left. \alpha^2\mathbb{E}[Z^3\zeta_1(\alpha_0 + \alpha Z)] - \tau\alpha\delta a_1 - \alpha a_2 \right) \\ &= \frac{1}{\omega} \left(-\alpha\tau\frac{\zeta_1(\tau)}{\alpha^2 + 1} + \tau^3\alpha^3\frac{\zeta_1(\tau)}{\alpha^2 + 1} - \tau\alpha^2\delta\frac{\zeta_1(\tau) + \alpha^2\tau^2\zeta_1(\tau)}{(\alpha^2 + 1)^{3/2}} - \right. \\ &\quad \left. \alpha\tau\frac{\zeta_1(\tau) + \alpha^2\tau^2\zeta_1(\tau)}{(\alpha^2 + 1)} + \alpha^2(3\alpha\tau + \alpha^3\tau^3)\frac{\zeta_1(\tau)}{(\alpha^2 + 1)^2} - \right. \\ &\quad \left. \tau\alpha\delta a_1 - \alpha a_2 \right) \\ &= \frac{1}{\omega} \left(\frac{-2\alpha\tau\zeta_1(\tau)}{(1 + \alpha^2)^2} - \tau\alpha\delta a_1 - \alpha a_2 \right), \quad (3.9) \end{aligned}$$

$$\begin{aligned}
-\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \omega \partial \tau} \right] &= \frac{\alpha \sqrt{\alpha^2 + 1}}{\omega} \mathbb{E} [Z \zeta_2(\alpha_0 + \alpha Z)] \\
&= \frac{\alpha \sqrt{\alpha^2 + 1}}{\omega} \left(-\alpha_0 \mathbb{E} [Z \zeta_1(\alpha_0 + \alpha Z)] - \alpha \mathbb{E} [Z^2 \zeta_1(\alpha_0 + \alpha Z)] - a_1 \right) \\
&= \frac{\alpha \sqrt{\alpha^2 + 1}}{\omega} \left(\alpha_0 \alpha \tau \frac{\zeta_1(\tau)}{\alpha^2 + 1} - \alpha(1 + \alpha^2 \tau^2) \frac{\zeta_1(\tau)}{(\alpha^2 + 1)^{3/2}} - a_1 \right) \\
&= \frac{\alpha \sqrt{\alpha^2 + 1}}{\omega} \left(\frac{\alpha \zeta_1(\tau)(\tau^2 - 1)}{(1 + \alpha^2)^{3/2}} - a_1 \right), \tag{3.10}
\end{aligned}$$

$$\begin{aligned}
-\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \alpha^2} \right] &= -\mathbb{E} \left[\left(\frac{\tau \alpha}{\sqrt{\alpha^2 + 1}} + Z \right)^2 \zeta_2(\alpha_0 + \alpha Z) \right] - \frac{\tau}{(\alpha^2 + 1)^{3/2}} \mathbb{E} [\zeta_1(\alpha_0 + \alpha Z)] \\
&= \mathbb{E} \left[\left(\tau^2 \delta^2 + 2\delta \tau Z + Z^2 \right) \left(\zeta_1(\alpha_0 + \alpha Z)(\alpha_0 + \alpha Z) + [\zeta_1(\alpha_0 + \alpha Z)]^2 \right) \right] - \\
&\quad \frac{\tau}{(\alpha^2 + 1)^{3/2}} \mathbb{E} [\zeta_1(\alpha_0 + \alpha Z)] \\
&= \left(\tau^3 \delta^2 \sqrt{\alpha^2 + 1} - \frac{\tau}{(\alpha^2 + 1)^{3/2}} \right) \mathbb{E} [\zeta_1(\alpha_0 + \alpha Z)] + \\
&\quad \left(\tau^2 \delta^2 \alpha + 2\delta \tau^2 \sqrt{\alpha^2 + 1} \right) \mathbb{E} [Z \zeta_1(\alpha_0 + \alpha Z)] + \\
&\quad \left(2\delta \tau \alpha + \tau \sqrt{\alpha^2 + 1} \right) \mathbb{E} [Z^2 \zeta_1(\alpha_0 + \alpha Z)] + \alpha \mathbb{E} [Z^3 \zeta_1(\alpha_0 + \alpha Z)] + \\
&\quad \tau^2 \delta^2 a_0 + 2\delta \tau a_1 + a_2 \\
&= \zeta_1(\tau) \left(\tau^3 \delta^2 - \frac{\tau}{(\alpha^2 + 1)^2} - \tau^3 \delta^4 - 2\delta^2 \tau^3 + \frac{2\tau \alpha^2}{(\alpha^2 + 1)^2} + \frac{\tau}{\alpha^2 + 1} + \right. \\
&\quad \left. \frac{2\tau^3 \alpha^4}{(\alpha^2 + 1)^2} + \tau^3 \delta^2 - \frac{\alpha^2(3\tau + \alpha^2 \tau^3)}{(\alpha^2 + 1)^2} \right) + \tau^2 \delta^2 a_0 + 2\delta \tau a_1 + a_2 \\
&= \zeta_1(\tau) \left(-\frac{\tau}{(\alpha^2 + 1)^2} + \frac{2\tau \alpha^2}{(\alpha^2 + 1)^2} + \frac{\tau}{\alpha^2 + 1} - \frac{3\tau \delta^2}{(\alpha^2 + 1)} \right) + \\
&\quad \tau^2 \delta^2 a_0 + 2\delta \tau a_1 + a_2 \\
&= (\tau \delta)^2 a_0 + 2\tau \delta a_1 + a_2, \tag{3.11}
\end{aligned}$$

$$\begin{aligned}
-\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \alpha \partial \tau} \right] &= -\mathbb{E} \left[(\alpha \tau + Z \sqrt{\alpha^2 + 1}) \zeta_2(\alpha_0 + \alpha Z) \right] - \delta \mathbb{E} [\zeta_1(\alpha_0 + \alpha Z)] \\
&= \mathbb{E} \left[(\alpha \tau + Z \sqrt{\alpha^2 + 1}) \left(\zeta_1(\alpha_0 + \alpha Z)(\alpha_0 + \alpha Z) + \right. \right.
\end{aligned}$$

$$\begin{aligned}
& \left. \zeta_1(\alpha_0 + \alpha Z)^2 \right] - \delta \mathbb{E} [\zeta_1(\alpha_0 + \alpha Z)] \\
= & \alpha \tau \alpha_0 \mathbb{E} [\zeta_1(\alpha_0 + \alpha Z)] + (2\alpha^2 \tau + \tau) \mathbb{E} [Z \zeta_1(\alpha_0 + \alpha Z)] + \\
& \alpha \sqrt{\alpha^2 + 1} \mathbb{E} [Z^2 \zeta_1(\alpha_0 + \alpha Z)] - \delta \mathbb{E} [\zeta_1(\alpha_0 + \alpha Z)] + \alpha \tau a_0 + \\
& \sqrt{\alpha^2 + 1} a_1 \\
= & \alpha \tau^2 \zeta_1(\tau) - (2\alpha^2 \tau + \tau) \alpha \tau \frac{\zeta_1(\tau)}{\alpha^2 + 1} + \alpha(1 + \alpha^2 \tau^2) \frac{\zeta_1(\tau)}{\alpha^2 + 1} - \\
& \alpha \frac{\zeta_1(\tau)}{\alpha^2 + 1} + \alpha \tau a_0 + \sqrt{\alpha^2 + 1} a_1 \\
= & \alpha \tau a_0 + \sqrt{\alpha^2 + 1} a_1, \tag{3.12}
\end{aligned}$$

$$\begin{aligned}
- \mathbb{E} \left[\frac{\partial^2 \ell}{\partial \tau^2} \right] &= -(\alpha^2 + 1) \mathbb{E} [\zeta_2(\alpha_0 + \alpha Z)] + \zeta_2(\tau) \\
&= (\alpha^2 + 1) \left(\frac{\tau \zeta_1(\tau)}{\alpha^2 + 1} + a_0 \right) - \tau \zeta_1(\tau) - \zeta_1(\tau)^2 \\
&= (\alpha^2 + 1) a_0 - \zeta_1(\tau)^2. \tag{3.13}
\end{aligned}$$

dove le quantità a_k con $k = 0, 1, 2$ sono ridefinite, per il modello *ESN*, in modo simile alla (1.13) come

$$a_k = \mathbb{E} \left[Z^k \left(\zeta_1(\tau \sqrt{\alpha^2 + 1} + \alpha Z) \right)^2 \right]. \tag{3.14}$$

Esse sono funzioni sia del parametro α che, a differenza delle analoghe quantità della (1.13), del quarto parametro τ . Il calcolo di queste quantità sarà effettuato tramite integrazione numerica analogamente al caso $\tau = 0$.

La matrice d'informazione attesa per una singola osservazione tratta dalla distribuzione (2.3) è quindi riportata a pag. 42.

Ogni elemento della matrice ottenuto per via algebrica è stato confrontato con l'analogo ottenuto tramite integrazione numerica delle derivate parziali. Si è utilizzata la *routine integrate* di R sia per il calcolo della matrice d'informazione numerica sia per gli elementi a_k della (3.14). Il confronto tra i risultati indica l'uguaglianza delle due procedure. Si riscontrano tuttavia alcuni problemi numerici per il calcolo della matrice d'informazione attesa per valori negativi estremi di τ , nell'ordine di $\tau < -15$. L'integrazione numerica degli elementi

$$\mathbf{I} = \begin{pmatrix} \frac{1}{\omega^2} (1 + \delta^2 \tau \zeta_1(\tau) + \alpha^2 a_0) & \frac{1}{\omega^2} \left(\zeta_1(\tau) \delta \left(\frac{1+2\alpha^2 - (\alpha^2+1)\delta^2 \tau^2}{\alpha^2+1} \right) + \alpha^2 a_1 \right) & \frac{1}{\omega} \left(\frac{\zeta_1(\tau)}{(\alpha^2+1)^{3/2}} - \tau \alpha \delta a_0 - \alpha a_1 \right) & -\frac{1}{\omega} (\delta \tau \zeta_1(\tau) + \alpha \sqrt{\alpha^2+1} a_0) \\ - & \frac{1}{\omega^2} (2 + \alpha^2 a_2 + \delta^4 \tau (\tau^2 - 3) \zeta_1(\tau)) & \frac{1}{\omega} \left(\frac{-2\alpha \tau \zeta_1(\tau)}{(\alpha^2+1)^2} - \tau \alpha \delta a_1 - \alpha a_2 \right) & \frac{\alpha \sqrt{\alpha^2+1}}{\omega} \left(\frac{\alpha \zeta_1(\tau) (\tau^2-1)}{(\alpha^2+1)^{3/2}} - a_1 \right) \\ - & - & (\tau \delta)^2 a_0 + 2\tau \delta a_1 + a_2 & \alpha \tau a_0 + \sqrt{\alpha^2+1} a_1 \\ - & - & - & (\alpha^2+1)a_0 - \zeta_1(\tau)^2 \end{pmatrix}$$

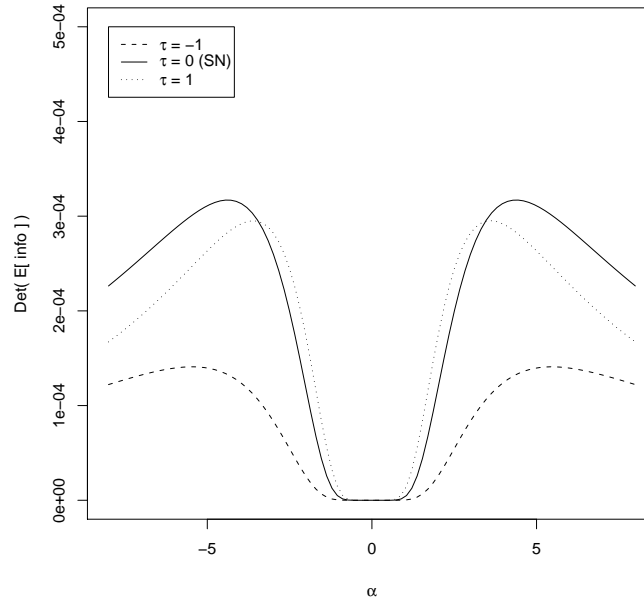


Figura 3.1: Determinante della matrice d'informazione attesa in funzione di α

della (3.14), infatti, è particolarmente instabile a causa della presenza della del fattore $\Phi(\tau)$ al denominatore della densità del modello che entra all'interno dell'integrale. In appendice è riportato il codice R utilizzato per il calcolo.

3.2.2 Caratteristiche della matrice d'informazione

È facile verificare l'uguaglianza degli elementi della matrice a pag. 42 con quelli della matrice a pag. 13 quando $\tau = 0$. Tuttavia le analogie con la matrice d'informazione per il modello normale asimmetrica non riguardano solamente questo caso particolare.

Una caratteristica comune, valida per ogni τ , è che la matrice d'informazione tende a essere singolare per $\alpha \rightarrow 0$. Si veda a riguardo il grafico in figura 3.1 che riporta un fascio di curve per cui ad ogni curva corrisponde una particolare scelta di τ ; in ascissa è riportato il valore del parametro α e in ordinata il valore del determinante della matrice d'informazione calcolato in funzione di α . I parametri ξ e ω^2 sono fissati rispettivamente a 0 e 1.

Come si può vedere, il determinante della matrice tende a 0 al tendere di $\alpha \rightarrow 0$. Questo

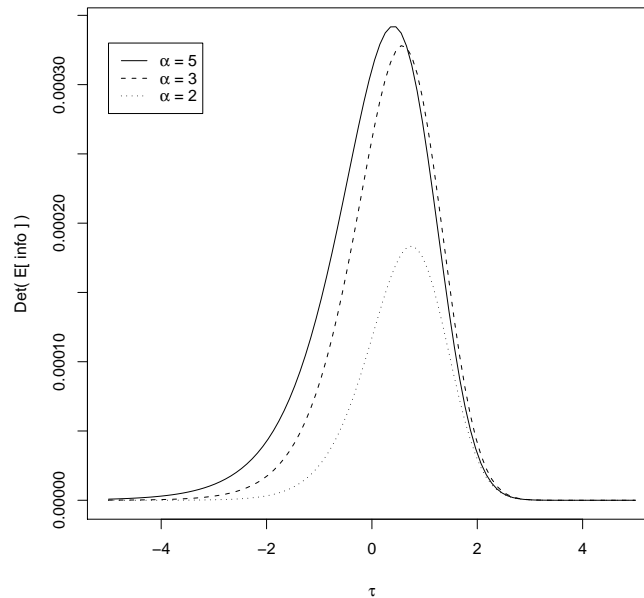


Figura 3.2: Determinante della matrice d'informazione attesa in funzione di τ

succede per qualsiasi scelta del parametro τ , il quale influisce solamente sulla velocità con cui il determinante si annulla.

Vediamo lo stesso tipo di grafico invertendo il ruolo di α e τ . Le osservazioni di pag. 32 sui casi limite quando $\tau \rightarrow \pm\infty$, infatti, suggeriscono lo studio della matrice d'informazione per questo comportamento limite del parametro τ . In fig. 3.2, possiamo infatti vedere come la matrice d'informazione abbia un determinante nullo quando $\tau \rightarrow \pm\infty$ e che esso decresca molto velocemente al punto di essere di fatto approssimabile a zero, anche per valori finiti a cavallo dello 0 e piccoli in valore assoluto.

Vediamo infine un grafico a curve di livello (fig. 3.3) in cui il determinante della matrice d'informazione è espresso in funzione di entrambi α e τ .

Sempre considerando le osservazioni di pagina 32, e i risultati algebrici del paragrafo 2.3.3, sappiamo che quando $\tau \rightarrow \pm\infty$ la densità tende ad essere approssimabile a quella di una normale, nello specifico $X \sim N(\xi, \omega^2)$ per il limite positivo e $(X + \delta\tau) \sim N(0, 1 - \delta^2)$ per il limite negativo. Anche la matrice d'informazione rispecchia questo comportamento.

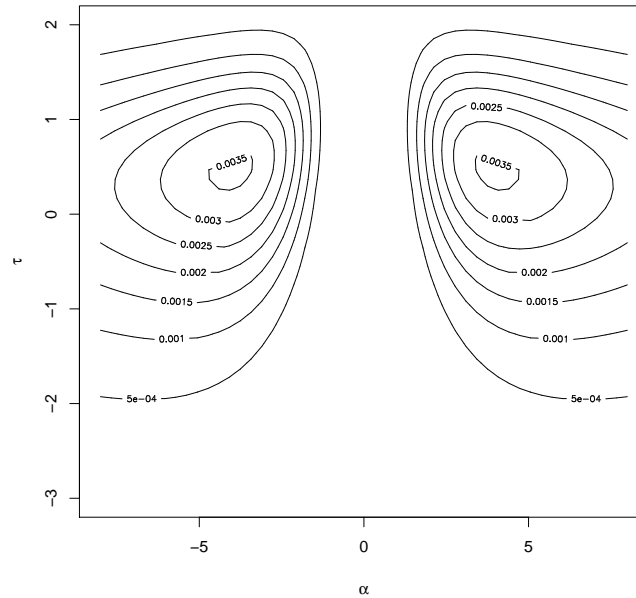


Figura 3.3: Determinante della matrice d'informazione attesa. Curve di livello in funzione di α e τ

Vediamo il primo caso per cui $\tau \rightarrow +\infty$. L'informazione di $\hat{\sigma}^2$, stimatore del parametro di scala σ^2 , nel modello normale $N(\mu, \sigma^2)$ è ricavabile calcolando il valore atteso delle derivate parziali (cambiato di segno) della funzione di logverosimiglianza; si veda ad esempio Azzalini (2001a) pag. 81–82. L'informazione per $\hat{\sigma}^2$ risulta:

$$I(\sigma^2) = 1/(2\sigma^4).$$

Analogamente possiamo ricavare l'informazione per σ , radice quadrata di σ^2 :

$$I(\sigma) = -\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \sigma^2} \right] = \frac{2}{\sigma^2},$$

che è analoga all'informazione per ω riportata nell'equazione (3.8) della matrice d'informazione attesa del modello *ESN*. Infatti il secondo e il terzo addendo all'interno delle parentesi della (3.8) tendono entrambi a 0 quando $\tau \rightarrow +\infty$. Analogamente succede per l'informazione rela-

tiva al parametro di scala ξ , pari a $1/\omega^2$ esattamente come per il parametro di posizione della normale. Gli elementi i_{12} e i_{21} tendono a 0.

Analizziamo il caso in cui $\tau \rightarrow -\infty$. Abbiamo visto a pagina 29 che la densità della normale asimmetrica estesa tende a quella di una $N(-\delta\tau, 1 - \delta^2)$. La matrice d'informazione di questo modello è una matrice diagonale con $\text{diag}\left(\frac{1}{1-\delta^2}, \frac{1}{2(1-\delta^2)^2}\right)$. Poiché $\lim_{x \rightarrow -\infty} \mathbb{E}[\zeta_2(x)] = -1$, facilmente si vede che l'elemento (3.4) della matrice tende esattamente a $\frac{1}{1-\delta^2}$. La dimostrazione che il secondo elemento diagonale della matrice tende a $\frac{1}{2(1-\delta^2)^2}$ è meno immediata, a causa della presenza dell'elemento a_2 che non può essere che calcolato algebricamente. Tuttavia, verifiche numeriche testimoniano che l'elemento (3.8) tende a $\frac{1}{2(1-\delta^2)^2}$ quando $\tau \rightarrow -\infty$.

Un'ultima caratteristica della matrice d'informazione attesa calcolata riguarda sempre l'elemento (3.8). L'addendo $\tau(\tau^2 - 3)\zeta_1(\tau) \sim -\tau^4$ per $\tau \rightarrow \infty$. Allora poiché tutto l'elemento resti positivo, l'addendo a_2 , a sua volta, deve crescere almeno tanto rapidamente quanto τ^4 . Anche questo fatto risulta difficoltoso da dimostrare algebricamente, ma sempre tramite verifiche numeriche risulta che il valore di a_2 è sempre superiore o uguale a quello di τ^4 .

3.3 Stima di massima verosimiglianza

3.3.1 Peculiarità della verosimiglianza

Una prima caratteristica della funzione di verosimiglianza, già riscontrata nel modello SN , è la presenza di un punto di stazionarietà per la verosimiglianza profilo per α in $\alpha = 0$.

Questa caratteristica rimane valida anche per la ESN . Così come nel modello SN quando $\alpha = 0$ la densità si riconduce a quella di una normale così succede anche per il modello ESN . Questo fatto e in generale le analogie tra i due modelli per quanto riguarda α suggeriscono la presenza di un punto di stazionarietà in $\alpha = 0$ di cui, di seguito, si riporta la dimostrazione analitica.

Si definisce punto di stazionarietà di una funzione $f(x)$ un punto x_0 tale che il gradiente $\nabla f(x_0) = 0$. Il gradiente della logverosimiglianza è stato riportato a pag 34. La funzione di logverosimiglianza profilo per α è la (3.1) con al posto di ξ, ω^2, τ i valori $\xi_{(\alpha=0)}^*, \omega_{(\alpha=0)}^{2*}, \tau_{(\alpha=0)}^*$, definiti come le stime di massima verosimiglianza quando $\alpha = 0$. Le equazioni di stima per i

parametri dato (z_1, \dots, z_n) , campione di numerosità n da una distribuzione di probabilità del tipo (2.3), sono quindi:

$$\frac{1}{\omega} \sum_{i=1}^n (z_i - \alpha \zeta_1(\alpha_0 + \alpha z_i)) = 0, \quad (3.15)$$

$$\frac{1}{\omega} \sum_{i=1}^n (z_i^2 - \alpha z_i \zeta_1(\alpha_0 + \alpha z_i) - 1) = 0, \quad (3.16)$$

$$\sum_{i=1}^n \left(\zeta_1(\alpha_0 + \alpha z_i) \left(\frac{\tau \alpha}{\sqrt{\alpha^2 + 1}} + z_i \right) \right) = 0, \quad (3.17)$$

$$\sqrt{\alpha^2 + 1} \sum_{i=1}^n \zeta_1(\alpha_0 + \alpha z_i) - n \zeta_1(\tau) = 0. \quad (3.18)$$

Quando $\alpha = 0$, l'equazione di stima (3.15) si semplifica portando a $\xi_{(\alpha=0)}^* = \bar{y}$, media campionaria, mentre la (3.16) porta a stimare $\omega_{(\alpha=0)}^{2*} = \frac{1}{n} \sum (y_i - \bar{y})^2$, varianza campionaria non corretta. Il calcolo di $\tau_{(\alpha=0)}^*$ non è effettuabile poiché la (3.18) risulta identicamente verificata per ogni τ . Il gradiente risulta quindi esattamente nullo per $\alpha = 0$, il che dimostra la presenza di un punto di stazionarietà in $\alpha = 0$.

In figura 3.4 sono riportati i grafici della logverosimiglianza profilo relativa per α di tre campioni di numerosità 500. Sulla sinistra è riportata la logverosimiglianza per subcampioni di numerosità 200 mentre sulla destra per l'intero campione. I tre campioni sono generati rispettivamente da una $ESN(0, 1, 1, -1)$, da una $ESN(0, 1, 1, 0)$ (e quindi di fatto una $SN(0, 1, 1)$) e da una $ESN(0, 1, 1, 1)$. In tutti i grafici risulta evidente il punto di stazionarietà per $\alpha = 0$ a prescindere dalla numerosità del campione. Inoltre, osservando la scala delle ordinate, notiamo, come c'era da aspettarsi, che la logverosimiglianza risulta maggiormente informativa utilizzando, in fase di stima, tutte e 500 le osservazioni.

Si analizza ora la verosimiglianza profilo per il quarto parametro τ . Come già sottolineato, valori maggiori di 3 e minori di -6 portano a densità che di asimmetrico hanno ben poco, e con differenze tra loro minime. Distribuzioni con parametro τ rispettivamente pari a 4 o a 20, presentano infatti una forma del tutto simile a quella di una normale standardizzata, mentre valori negativi molto grandi in valore assoluto portano, fissato il parametro di posizione ξ , a un

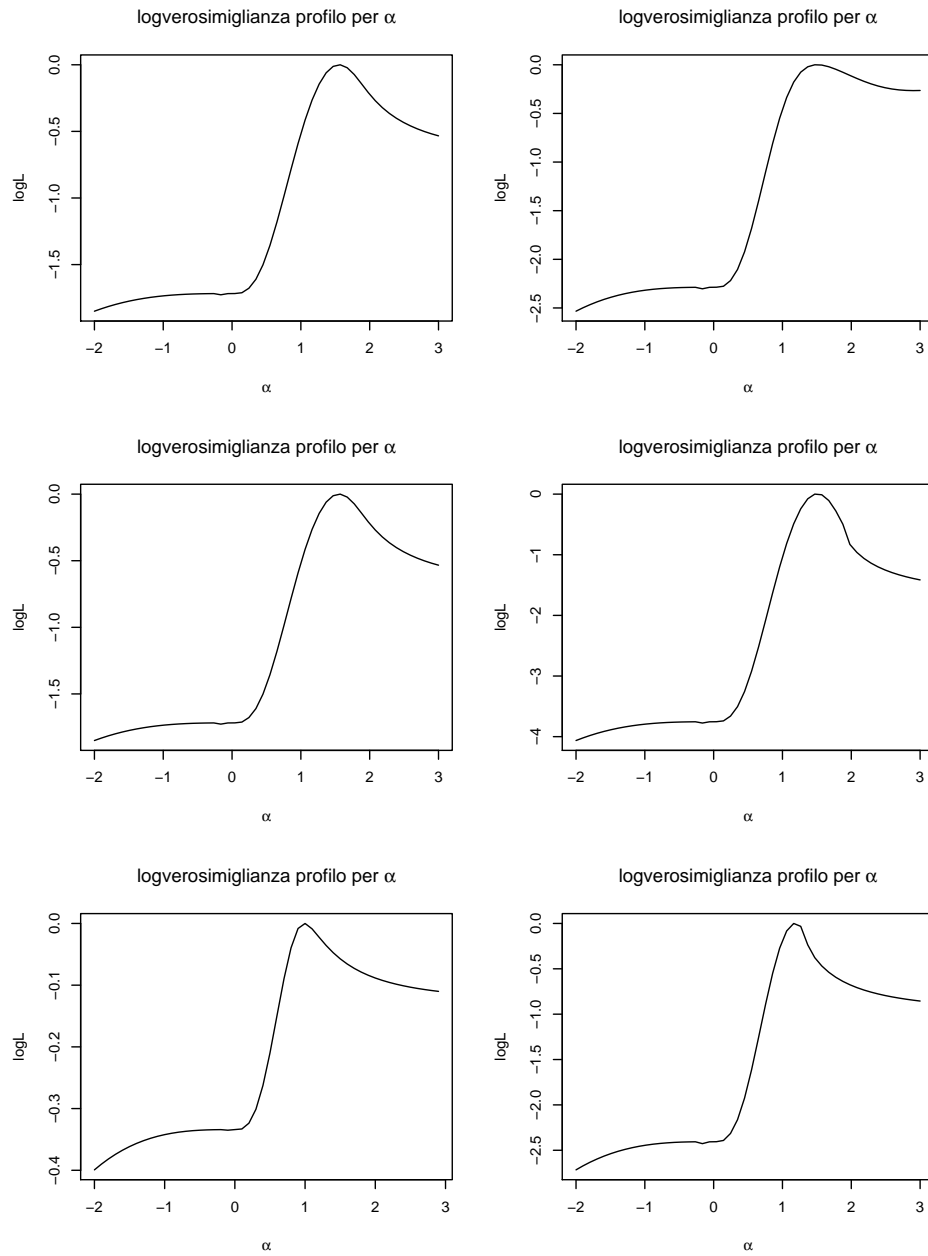


Figura 3.4: Logverosimiglianza profilo relativa per α in campioni di numerosità $n = 500$ simulati (dall'altro verso il basso) da $ESN(0, 1, 1, -1)$, $ESN(0, 1, 1, 0)$ e $ESN(0, 1, 1, 1)$. Nei grafici di sinistra la logverosimiglianza profilo è calcolata considerando solamente le prime 200 osservazioni di ogni campione

traslamento della curva. Valori negativi minori di -6 portano inoltre una instabilità numerica della funzione $\Phi(\tau)$. Per questi motivi la massimizzazione della verosimiglianza per i restanti parametri, fissato il parametro τ , è stata effettuata solamente per $-6 \leq \tau \leq 2$. Nelle figure 3.5 sono riportati alcuni dei grafici ottenibili per gli stessi campioni della figura 3.4.

A colpo d'occhio è immediatamente palpabile il fallimento della stima di massima verosimiglianza nell'avvicinarsi al vero valore del parametro. Tutti i grafici, nonostante provengano da distribuzioni differenti, presentano il massimo della curva per valori di τ positivi e minori di 1.

Inoltre non è raro ottenere grafici come quelli in figura 3.6 in cui la stima di massima verosimiglianza è molto distante dal vero valore del parametro, o addirittura tendente a $-\infty$. In particolare i grafici della figura 3.6 rappresentano la logverosimiglianza profilo relativa per τ calcolata per due particolari campioni di numerosità 500 da una $ESN(0, 1, 1, 0)$. Questo tipo di situazioni non è raro e si verifica abbastanza frequentemente. Si è riscontrata una stima tendente a $-\infty$ circa una volta su 10 mentre quella in prossimità di 2 per circa un terzo delle volte. In seguito (cfr par. 3.4) si vedranno in dettaglio queste caratteristiche tramite uno studio di simulazione.

Se guardiamo i valori in ordinata, tuttavia, questa presunta distorsione delle stime viene ridimensionata da un'altra caratteristica della verosimiglianza: la scarsa informatività. L'ascissa dei grafici della fig. 3.5 riporta infatti la logverosimiglianza relativa $\ell(\hat{\theta}) - \ell(\theta)$ che sappiamo, a meno della moltiplicazione per 2, distribuirsi asintoticamente come un χ_1^2 . Qualsiasi valore di questa distribuzione ragionevole per verificare l'ipotesi che $\tau = \tau_0$ è inferiore al valore minimo dei grafici, il che comporta che non possiamo rifiutare l'ipotesi che il parametro valga un qualsiasi valore dell'intervallo in esame. Tutte queste considerazioni rimangono valide, come è evidente, anche aumentando la dimensione del campione.

3.3.2 Calcolo delle stime

Come più volte sottolineato, le equazioni di verosimiglianza non hanno una soluzione esplicita. Il calcolo delle stime di massima verosimiglianza, deve quindi effettuato tramite massimizzazione numerica della funzione di logverosimiglianza.

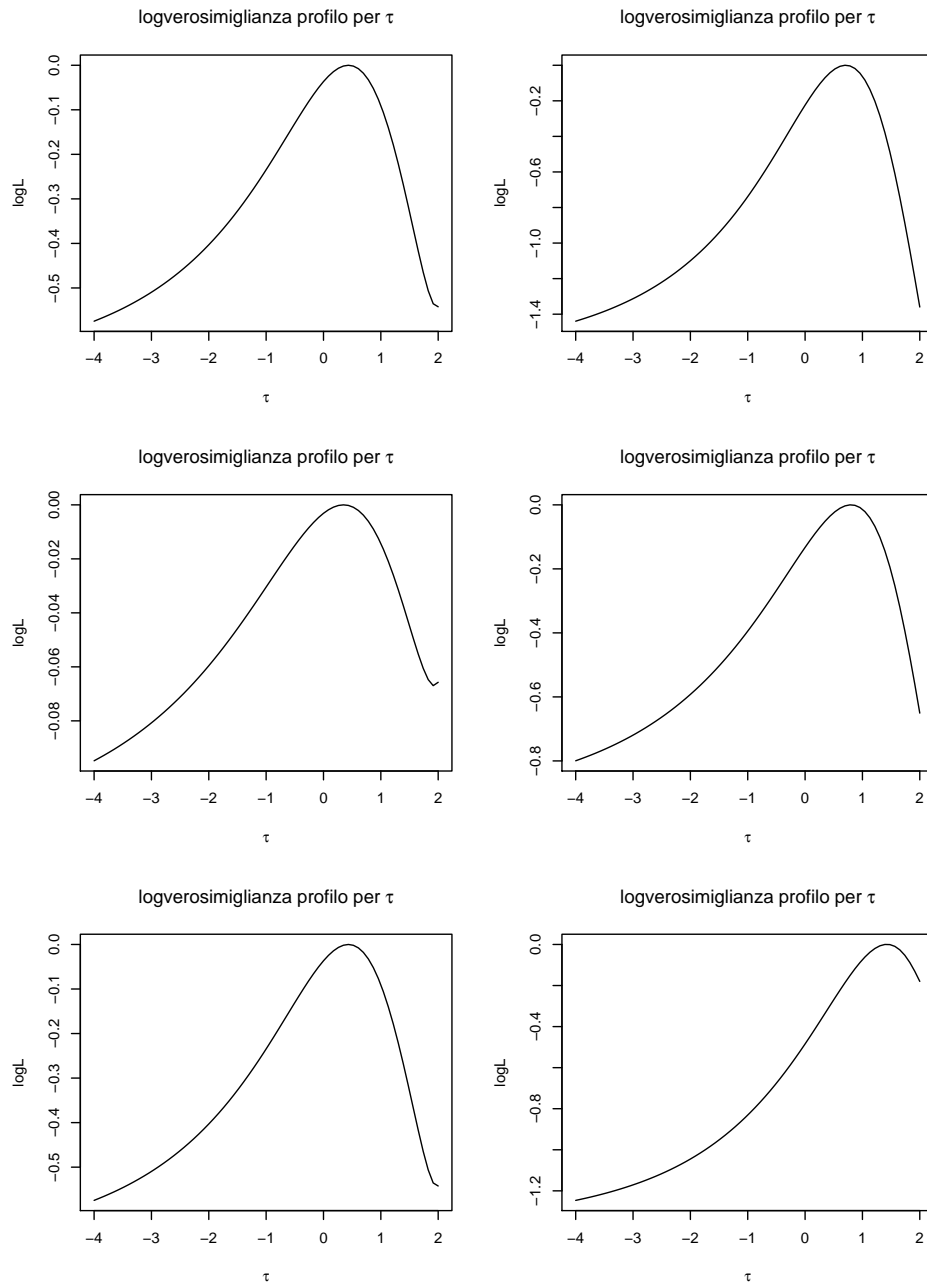


Figura 3.5: Logverosimiglianza profilo relativa per τ in campioni di numerosità $n = 500$ simulati (dall'alto verso il basso) da $ESN(0, 1, 1, -1)$, $ESN(0, 1, 1, 0)$ e $ESN(0, 1, 1, 1)$. Nei grafici di sinistra la logverosimiglianza profilo è calcolata considerando solamente le prime 200 osservazioni di ogni campione

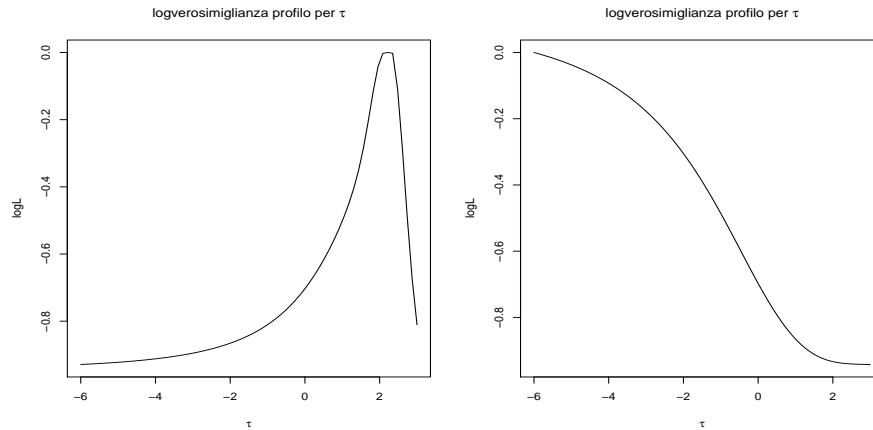


Figura 3.6: Logverosimiglianza profilo per τ in campioni di numerosità $n = 500$ simulati da una $ESN(0, 1, 1, 0)$. Entrambi i grafici mostrano il massimo della logverosimiglianza in punti dello spazio parametrico distanti dal vero valore del parametro τ

Per fare ciò sono state create alcune funzioni in R. Una massimizzazione è stata effettuata tramite il metodo *quasi-Newton* L-BFGS-B, già presente in R (si veda Byrd et al. (1995)). Il codice della funzione è riportato in appendice.

Sono state effettuate alcune prove, simulando campioni e applicando il metodo di stima per alcuni valori di inizializzazione dei parametri. Quello che si nota da questo piccolo studio è che l'algoritmo fatica ad arrivare al massimo globale e, a seconda del punto di inizializzazione dello stesso, converge a punti di minimo locale. Anche metodi di *simulated annealing*, (si è utilizzato il metodo SANN, di R), portano agli stessi risultati. Per rendere l'idea si riporta un esempio relativo alla simulazione di un campione di numerosità 200 da una $ESN(0, 1, 1, 0)$. I risultati sono riportati nella tabella 3.1.

Le righe dalla 2 alla 4 della tabella 3.1 mostrano valori molto simili delle stime e della funzione di logverosimiglianza, uguale fino alla quarta cifra decimale. Si noti come i valori delle stime per queste righe non si avvicinino ai veri valori dei parametri della distribuzione da cui il campione è stato generato, anche nonostante il fatto che, ad esempio, le stime nella seconda riga sono state ottenute inizializzando l'algoritmo esattamente sui veri valori della distribuzione.

Si osservino adesso le stime riportate nella prima riga e ottenute inizializzando l'algoritmo

con il metodo dei momenti. I primi tre momenti campionari sono stati trasformati nei parametri ξ , ω^2 e α tramite le equazioni inverse alle (2.6)–(2.8) (cfr. par. 3.5) e τ è stato fissato a 0. Il valore di verosimiglianza a cui converge l'algoritmo risulta maggiore degli altri, anche se i valori delle stime risultano essere ancora più distanti dai veri valori.

In Capitano et al. (2003), è stato mostrato che per questo modello la stima numerica dei parametri può avere difficoltà a convergere, anche per elevate numerosità campionarie. Per questi motivi è stata avanzata la proposta di utilizzare la verosimiglianza profilo per τ , la cui massimizzazione è stata studiata per i campioni simulati.

Per completezza è stata poi effettuata anche la massimizzazione della logverosimiglianza profilo per α .

La tabella 3.1 riporta tutti i risultati relativi a questi metodi di stima.

L'algoritmo di massimizzazione della logverosimiglianza profilo per τ converge ad un valore della funzione di verosimiglianza, di poco superiore a quello raggiunto dagli altri metodi. Tuttavia anch'esso è associato a dei valori delle stime molto diversi sia dai veri valori dei parametri sia dai valori ottenuti tramite le altre procedure. La causa di ciò può essere trovata nel fatto che, come mostrato in precedenza, la stima di massima verosimiglianza per τ può essere molto distante dal vero valore del parametro. Si noti infatti che il valore alto di $\hat{\tau}$ porta la verosimiglianza ad *aggiustare* la stima degli altri parametri. Nella riga 5 della tabella, l'alto valore di $\hat{\tau}$, che porterebbe, fissato α , a una perdita di asimmetria della distribuzione, è compensato dall'alto valore di $\hat{\alpha}$. Nell'ultima riga, invece, la traslazione della densità verso valori negativi dovuta al valore negativo di $\hat{\tau}$ congiuntamente al valore negativo di $\hat{\alpha}$ (il segno di $\hat{\alpha}$ determina il verso della traslazione) è compensata dal valore di $\hat{\xi}$.

In fig. 3.7 sono state tracciate le curve delle densità stimate, ed esse risultano sostanzialmente equivalenti e quasi sovrapponibili.

La scarsa informatività della verosimiglianza legata al parametro τ e l'estrema versatilità del modello, che include sia il caso *SN* che il caso normale per $\tau \rightarrow \pm\infty$, lo rendono di fatto poco utilizzabile. I problemi evidenziati nel corso di questo capitolo mostrano che le stime di massima verosimiglianza possono essere sensibilmente differenti dai veri valori dei parametri, portando tuttavia a densità pressoché equivalenti.

	Funz. obiet.	Inizializz.	$\hat{\xi}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\tau}$	$\log L$
1	$\ell(\theta)$	momenti	4.1963	1.3745	-1.6850	-2.8215	-48.4709
2	$\ell(\theta)$	(0,1,1,0)	0.5082	0.7738	0.0037	-0.4030	-48.7047
3	$\ell(\theta)$	(0.5,0.6,0,0)	0.5085	0.7738	0.0045	-0.0000	-48.7047
4	$\ell(\theta)$	(0,1,0,0)	0.5091	0.7738	0.0035	-0.0418	-48.7047
5	$\ell_{P(\tau)}$	momenti	0.5214	0.7860	-15.0000	2.4783	-47.5344
6	$\ell_{P(\alpha)}$	momenti	5.4127	1.5077	-1.8980	-3.4184	-48.4702

Tabella 3.1: Stime di massima verosimiglianza per vari metodi ed inizializzazioni. (Campione con $n = 200$ da $ESN(0, 1, 1, 0)$)

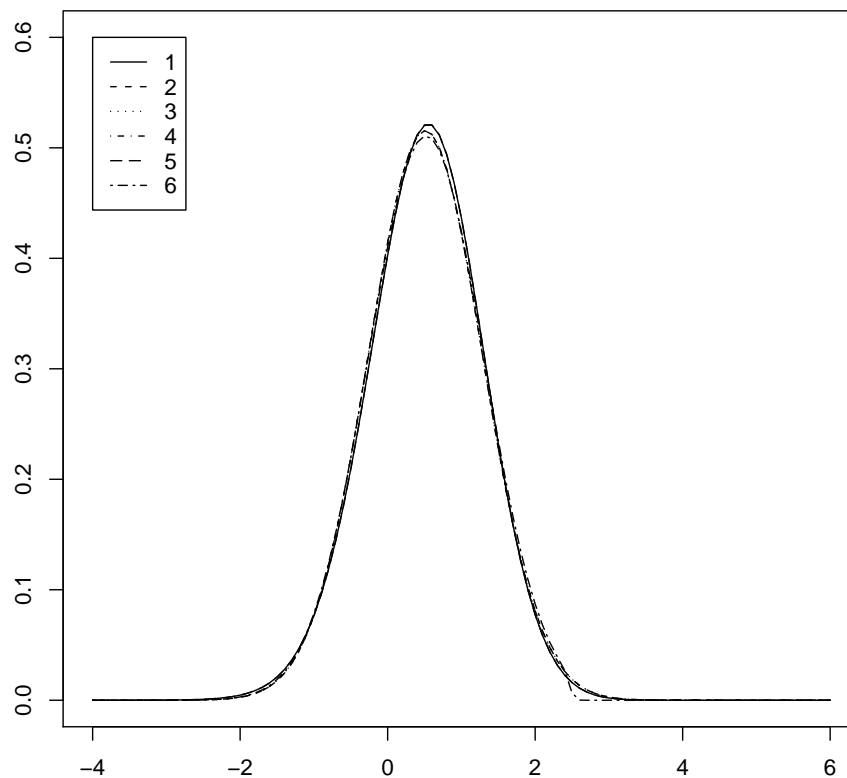


Figura 3.7: Funzioni di densità con i parametri della tabella 3.1

3.4 Studio numerico della distribuzione delle stime

Si è effettuato un piccolo studio di simulazione per cercare di studiare la distribuzione delle stime di massima verosimiglianza. Nei grafici a pagina 14 la distribuzione empirica delle stime per il modello normale asimmetrica presentava una evidente bimodalità causata dalla singolarità della matrice d'informazione attesa in $\alpha = 0$, caso studiato in Rotnizsky et al. (2000).

Sugli stessi campioni dei grafici di pagina 14 è stato adattato un modello *ESN* calcolando le stime sia tramite massimizzazione diretta della verosimiglianza, sia utilizzando la verosimiglianza profilo per τ . I risultati per le due procedure sono riportati nelle figure 3.8 e 3.9, rispettivamente per la massimizzazione contemporanea dei quattro parametri e per quella con τ fissato. Le figure rappresentano i grafici di dispersione e gli istogrammi delle stime di massima verosimiglianza. La presenza di una buona quantità di stime di α visibilmente lontane dal vero valore del parametro e che superano in valore assoluto 200 rende poco leggibile il grafico relativo alla massimizzazione diretta della logverosimiglianza a cavallo dello 0. Per questo sono stati tolti i ben 148 campioni relativi a questi valori delle stime di α e i restanti sono stati rappresentati, analogamente alla fig. 3.8, nella fig. 3.10.

Analizziamo in prima battuta i risultati relativi alla massimizzazione della logverosimiglianza profilo per τ . Gli istogrammi delle frequenze sulla diagonale della figura mostrano contemporaneamente per la distribuzione di $\hat{\xi}$ una decisa concentrazione intorno alla moda in prossimità dello 0, e delle code pesanti, con valori decisamente distanti dal vero valore del parametro. Si veda anche la fig. 3.11. La distribuzione di $\hat{\omega}^2$ è concentrata in valori prossimi all'unità, ma anch'essa con una coda destra visibilmente pesante. L'istogramma relativo alle stime di $\hat{\alpha}$ è molto particolare. Esso infatti presenta una bimodalità a cavallo dello 0 intorno a -1 e 2 circa e due decisi picchi esattamente in -15 e 5 . Su 1000 stime ben 136 sono esattamente -15 e 264 sono pari a 5 . L'istogramma è riproposto ingrandito nella fig. 3.11.

Infine la distribuzione empirica di $\hat{\tau}$ (vincolata alla scansione per valori compresi in $(-6, 4)$) mostra un'evidente moda in 2 , e un picco nell'estremo inferiore dell'intervallo, dovuto probabilmente a una logverosimiglianza profilo crescente per $\tau \rightarrow -\infty$ come quella riportata in

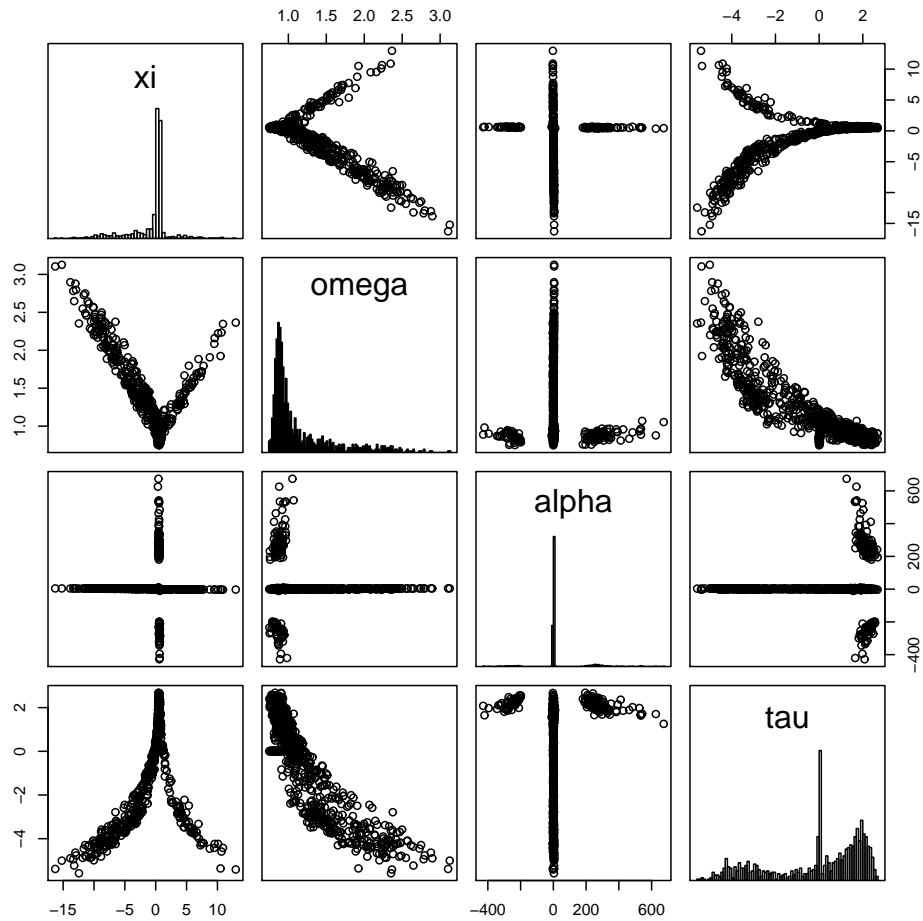


Figura 3.8: Diagrammi di dispersione e istogrammi delle distribuzioni empiriche delle stime di massima verosimiglianza da 1000 campioni di numerosità 200 da una $ESN(0, 1, 1, 0)$. La massimizzazione è stata effettuata contemporaneamente per tutti i parametri

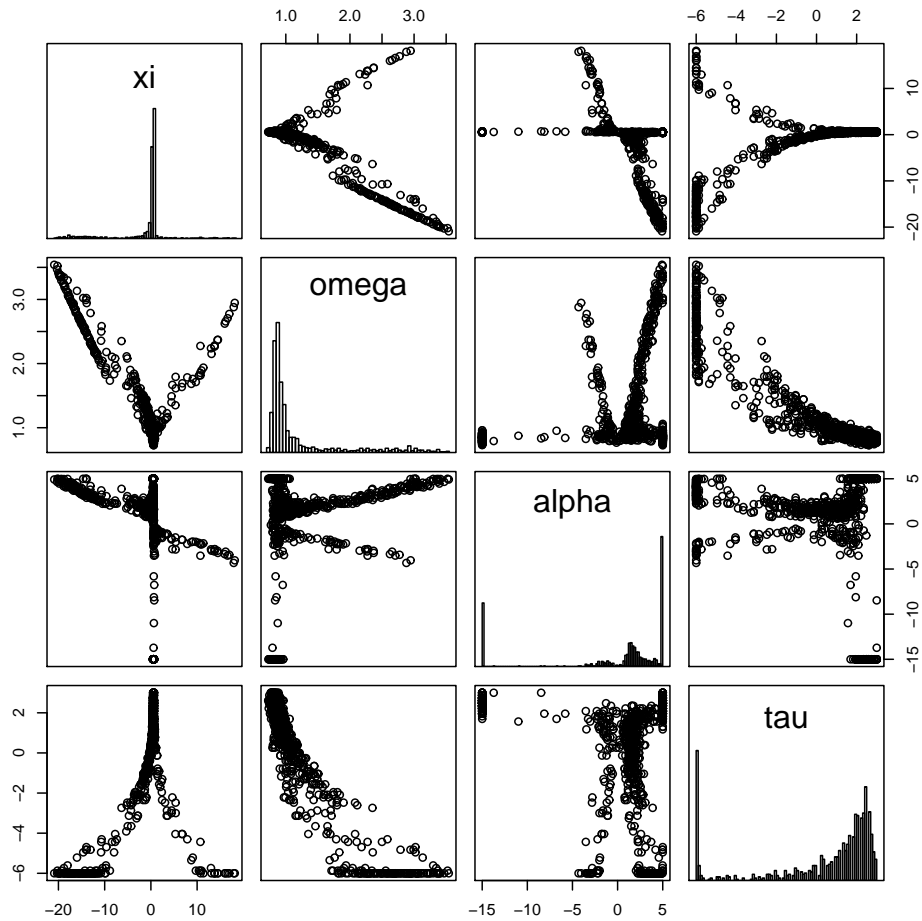


Figura 3.9: Diagrammi di dispersione e istogrammi delle distribuzioni empiriche delle stime di massima verosimiglianza da 1000 campioni di numerosità 200 da una $ESN(0, 1, 1, 0)$. La massimizzazione è stata effettuata sulla logverosimiglianza profilo per τ

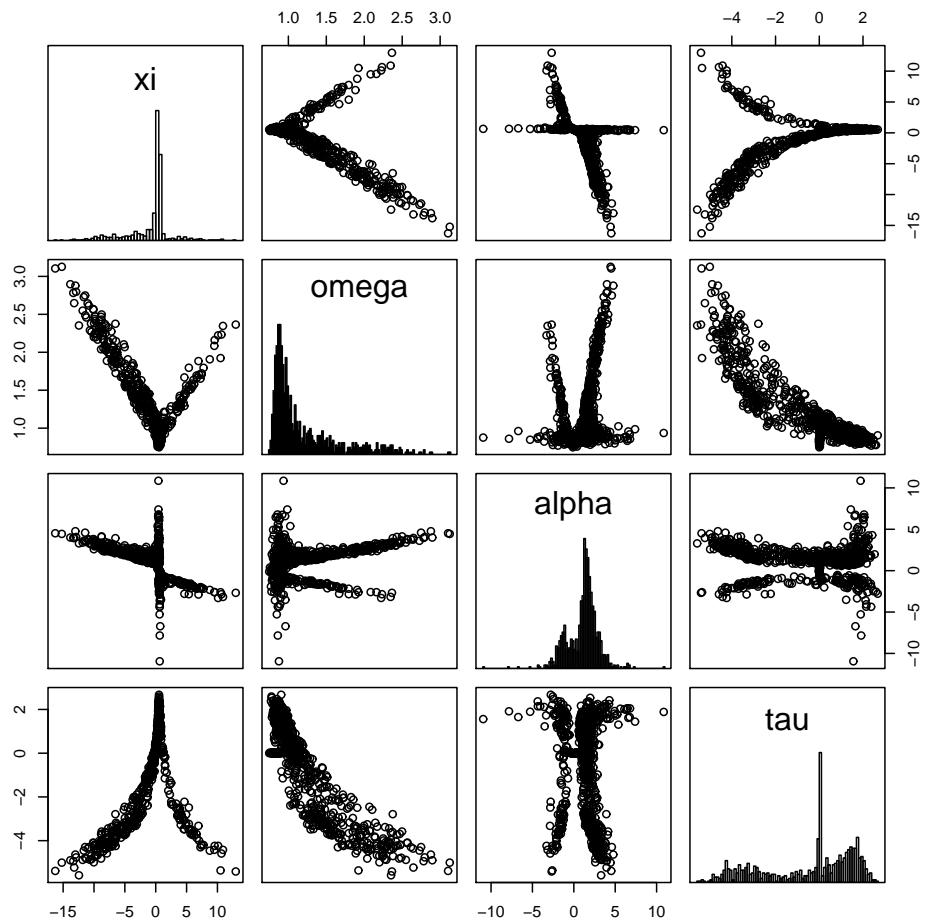


Figura 3.10: Diagrammi di dispersione e istogrammi delle distribuzioni empiriche delle stime di massima verosimiglianza come da fig. 3.8. Si sono eliminati i 152 campioni con $|\hat{\alpha}| > 150$

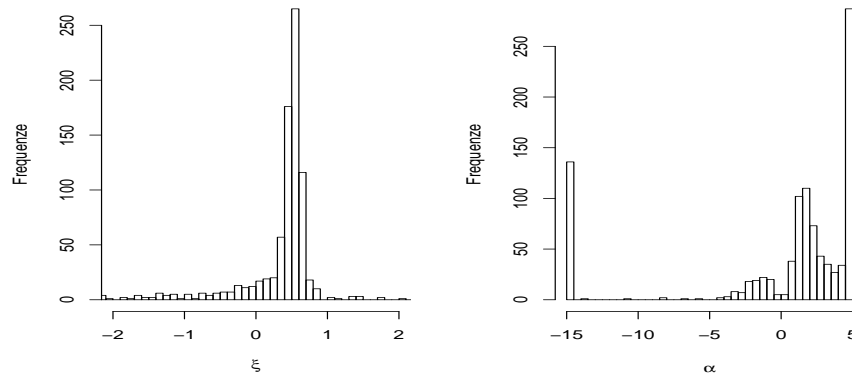


Figura 3.11: Distribuzione empirica della stime di massima verosimiglianza. Dettaglio della fig. 3.8. L'istogramma a sinistra riporta le frequenze di ξ nell'intervallo $(-2, 2)$. L'istogramma a destra riporta le frequenze di α

fig. 3.6.

I grafici di dispersione sottolineano quanto detto in precedenza. Si riscontra infatti correlazione tra il decrescere di τ e lo spostamento del parametro di posizione, nella direzione data dal segno del parametro α . Le nuvole di punti dei grafici di dispersione per ξ e ω^2 sono simili a quelle della fig. 1.3 di pag. 14 ad eccezione della presenza di più punti tra le due nuvole.

Analizziamo ora i risultati relativi alla massimizzazione diretta di tutti i quattro parametri della logverosimiglianza. I grafici sulla diagonale relativi ai parametri $\hat{\xi}$ e $\hat{\omega}$ sono qualitativamente simili a quelli appena analizzati. L'istogramma per α invece presenta come già detto 148 osservazioni superiori in valore assoluto a 200. La presenza di forte concentrazione di valori distorti si vedeva già nella figura 3.9, ma solamente nei valori 5 e -15 .

Dalla distribuzione delle stime di massima verosimiglianza possiamo evidenziare principalmente tre caratteristiche:

- le stime di massima verosimiglianza risultano spesso distanti dai veri valori dei parametri, il che porta a delle distribuzioni empiriche con code pesanti o peggio con mode molto distanti dal valore centrale della distribuzione;
- la verosimiglianza tende ad *aggiustare* le stime degli altri parametri quando uno di essi

risulta distante nello spazio parametrico dal vero valore del parametro. La correlazione di τ con il parametro di scala ξ o con il parametro di asimmetria α è marcatamente evidente;

- la bimodalità evidenziata per il modello SN non è evidente marginalmente mentre lo è per le distribuzioni congiunte.

Le procedure descritte sono state ripetute sui medesimi campioni aumentati di 300 unità. Qualitativamente, i risultati relativi ai nuovi campioni di numerosità 500 sono analoghi. I grafici per i nuovi campioni, equivalenti a quelli delle pagg. 55–57, sono riportati per completezza in appendice (cfr. pag. 92), senza voler appesantire questo capitolo.

3.5 Una parametrizzazione alternativa

In Azzalini (1985) viene proposta una riparametrizzazione (parametrizzazione centrata o CP) per il modello SN , al fine di risolvere il problema della singolarità della matrice d'informazione, causa dell'inusuale distribuzione delle stime di massima verosimiglianza. Nel paragrafo 1.6.2 si è già discusso dei benefici che questa parametrizzazione porta. Visto che il modello ESN ripropone le problematiche del modello SN quali la singolarità della matrice d'informazione in $\alpha = 0$, il punto di stazionarietà della logverosimiglianza profilo per α nel punto 0, la non usuale distribuzione delle stime di massima verosimiglianza, nonché la scarsa interpretabilità dei parametri ξ , ω^2 e α , è naturale chiedersi se una riparametrizzazione del modello, che ricalchi quella della SN , possa portare gli stessi benefici.

È stata quindi studiata una forma di parametrizzazione centrata. Se da un lato la scelta dei primi tre, dei quattro parametri, è stata fatta seguendo la strada ben nota in letteratura, per il quarto parametro si sono studiate alcune possibilità.

Poiché i primi tre parametri sono in corrispondenza con i primi tre momenti centrati, con il loro significato di media, varianza e asimmetria, il primo pensiero è stato di introdurre come quarto parametro il quarto momento centrato noto anche come indice di curtosi. Nonostante la famiglia ESN permetta variazioni dell'indice di curtosi in un intervallo più ampio di quello

del modello SN , lo studio della densità in funzione di τ non ha effetti univoci sulla curva della densità e, in particolare, non sembra regolarne direttamente la curtosi. Inoltre il passaggio tra una parametrizzazione alla nuova avrebbe creato non pochi problemi, in quanto τ compare sempre come argomento della funzione $\zeta_0(\cdot)$ e delle sue derivate rendendone quindi difficoltosa l'esplicitazione.

La proposta è sicuramente più pratica e consiste nel non riparametrizzare il quarto parametro. Volendo essere più specifici, poniamo

$$Y = \xi + \omega Z = \mu + \sigma Z_0 \quad (3.19)$$

dove Z ha una distribuzione $ESN(0, 1, \alpha, \tau)$ e siano

$$\mu_z(\tau) = \mathbb{E}(Z) = \zeta_1(\tau)\delta, \quad \sigma_z^2(\tau) = \mathbb{V}ar(Z) = 1 + \delta^2\zeta_2(\tau)$$

il valore atteso e la varianza della variabile Z , entrambi funzioni di τ . Definiamo Z_0 , come fatto per la normale asimmetrica, come la standardizzazione di Z , ossia:

$$Z_0 = \frac{1}{\sigma_z(\tau)} (Z - \mu_z(\tau)).$$

La *parametrizzazione centrata* proposta per il modello ESN è data dal vettore $\theta_{CP}(\xi, \omega^2, \alpha, \tau) = (\mu, \sigma^2, \gamma_1, \tau)$ le cui espressioni esplicite in funzione dei vecchi parametri sono:

$$\begin{aligned} \mu &= \mathbb{E}(Y) = \xi + \omega\mu_z = \xi + \omega\delta\zeta_1(\tau), \\ \sigma^2 &= \mathbb{V}ar(Y) = \omega^2\sigma_z^2 = \omega^2(1 + \delta^2\zeta_2(\tau)), \\ \gamma_1 &= \frac{\mathbb{E}\{(Y - \mathbb{E}(Y))^3\}}{\mathbb{V}ar(Y)^{3/2}} = \frac{\delta^3\zeta_3(\tau)}{(1 + \delta^2\zeta_2(\tau))^{3/2}}, \\ \tau &= \tau. \end{aligned}$$

La proprietà di equivarianza rispetto a riparametrazioni della stima di massima verosimiglianza è verificata se la riparametrizzazione è una funzione biunivoca dallo spazio Θ_{DP} allo spazio Θ_{CP} . Prima di analizzare i benefici che la parametrizzazione proposta porta, si dimostra

quindi questa caratteristica. Per fare ciò mostriamo che esiste in tutto il dominio di θ_{CP} la funzione inversa $\theta_{DP}(\mu, \sigma^2, \gamma_1, \tau) = (\xi, \omega^2, \alpha, \tau)$. Le equazioni per il passaggio dallo spazio Θ_{CP} allo spazio Θ_{DP} sono calcolate come

$$\mu = \xi + \omega\delta\zeta_1(\tau) \Rightarrow \xi = \mu - \omega\delta\zeta_1(\tau) \quad (3.20)$$

$$\sigma^2 = \omega^2(1 + \delta^2\zeta_2(\tau)) \Rightarrow \omega^2 = \sigma^2/(1 + \delta^2\zeta_2(\tau)) \quad (3.21)$$

$$\gamma_1 = \frac{\delta^3\zeta_3(\tau)}{(1 + \delta^2\zeta_2(\tau))^{3/2}} \Rightarrow (\zeta_3(\tau)^{2/3} - \gamma_1^{2/3}\zeta_2(\tau))\delta^2 = \gamma_1^{2/3}$$

$$\delta = (\gamma_1) \frac{\gamma_1^{1/3}}{\sqrt{\zeta_3(\tau)^{2/3} - \gamma_1^{2/3}\zeta_2(\tau)}} \quad (3.22)$$

$$\tau = \tau \quad (3.23)$$

dove, in particolare, la (3.22) esplicita δ in funzione di γ_1 e δ è la trasformazione (chiaramente biunivoca) di α per cui $\alpha = \frac{\delta}{\sqrt{1-\delta^2}}$, e la (3.23), fa corrispondere τ a sé stesso.

La matrice d'informazione per la nuova parametrizzazione si ottiene sfruttando l'usuale risultato:

$$I_{\Theta_{CP}} = \Delta^T I_{\Theta_{DP}} \Delta$$

dove $I_{\Theta_{DP}}$ è la matrice ricavata a pag 42 e Δ è la matrice delle derivate dei parametri DP rispetto ai nuovi:

$$\Delta = \begin{pmatrix} 1 & \Delta_{12} & \Delta_{13} & \Delta_{14} \\ 0 & \Delta_{22} & \Delta_{23} & \Delta_{24} \\ 0 & 0 & \Delta_{33} & \Delta_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

dove

$$\Delta_{12} = -\gamma_1^{1/3} \frac{\sqrt{\zeta_1(\tau)}}{\zeta_3(\tau)^{2/3}},$$

$$\Delta_{13} = -\frac{1}{3} \frac{\sigma^2}{\gamma_1^{2/3}} \frac{\sqrt{\zeta_1(\tau)}}{\zeta_3(\tau)^{1/3}},$$

$$\Delta_{14} = -\frac{1}{2} \sigma \gamma_1^{1/3} \frac{\zeta_2(\tau)\zeta_3(\tau)^{2/3} - \frac{2}{3}\zeta_3(\tau)^{-1/3}\zeta_4(\tau)\zeta_1(\tau)}{\zeta_1(\tau)^{1/2}\zeta_3(\tau)},$$

$$\begin{aligned}
\Delta_{22} &= \frac{\zeta_3(\tau)^{2/3} - \gamma_1^{2/3} \zeta_2(\tau)}{\zeta_3(\tau)^{2/3}}, \\
\Delta_{23} &= -\frac{2}{3} \frac{\sigma^2 \zeta_2(\tau)}{\gamma_1^{1/3} \zeta_3(\tau)^{2/3}}, \\
\Delta_{24} &= -\sigma^2 \gamma_1^{2/3} \frac{\zeta_3(\tau)^{5/3} - (2/3) \zeta_2(\tau) \zeta_3(\tau)^{-1/3} \zeta_4(\tau)}{\zeta_3(\tau)^{4/3}}, \\
\Delta_{33} &= \frac{1}{3} \gamma_1^{-2/3} \frac{1}{\sqrt{\zeta_3(\tau)^{2/3} - \gamma_1^{2/3} (1 + \zeta_2(\tau))}} - \frac{1}{2} \gamma_1^{1/3} \frac{(-2/3)(1 + \zeta_2(\tau)) \gamma_1^{-1/3}}{\left(\zeta_3(\tau)^{2/3} - \gamma_1^{2/3} (1 + \zeta_2(\tau))\right)^{3/2}}, \\
\Delta_{34} &= -\frac{1}{2} \gamma_1^{1/3} \frac{(2/3) \zeta_3(\tau)^{-1/3} \zeta_4(\tau) - \gamma_1^{2/3} \zeta_3(\tau)^{3/2}}{\zeta_3(\tau)^{2/3} - \gamma_1^{2/3} (1 + \zeta_2(\tau))}.
\end{aligned}$$

Il passaggio alla nuova parametrizzazione ha, come nel modello SN , il vantaggio di eliminare la singolarità della matrice d'informazione attesa per $\gamma_1 = 0$. Tuttavia, non riparametrizzando il parametro τ , la matrice continua ad essere singolare al divergere di questo.

Dal lato del calcolo delle stime di massima verosimiglianza, la massimizzazione tramite la nuova parametrizzazione per via numerica è poco fattibile. Per la normale asimmetrica, infatti, il parametro γ_1 può variare solamente nell'intervallo $(-0.995, 0.995)$, in funzione solamente di α tramite $\delta = \alpha / \sqrt{\alpha^2 + 1}$, che a sua volta deve essere minore in valore assoluto di 1. Il passaggio da una parametrizzazione all'altra, fatta eccezione per i casi di γ_1 che superano il dominio e che sono sempre aggiustati, riporta valori di δ sempre ammissibili, e quindi è possibile il passaggio ad α . Nella normale asimmetrica estesa, invece, le cose sono più complicate in quanto il valore di γ_1 varia in funzione sia di α che di τ . La massimizzazione numerica della verosimiglianza per tutti i nuovi parametri porta non raramente l'algoritmo a valutare la funzione di logverosimiglianza per valori di γ_1 e τ incompatibili tra loro, dando inevitabili problemi numerici (quali il calcolo di radici quadrate di numeri negativi).

Utilizzando la nuova parametrizzazione sulle stime ottenute dalla massimizzazione della funzione con la parametrizzazione diretta, abbiamo sicuramente il vantaggio di avere dei parametri con un significato più chiaro. Applicando inoltre la nuova parametrizzazione alle stime ottenute al par. 3.4, la distribuzione di queste risulta decisamente più regolare, togliendo la bimodalità e le code pesanti. Nelle figure 3.12 e 3.13 sono riportati i grafici analoghi a quelli di pag. 55–57

adottando questa volta la parametrizzazione centrata.

Volendo ricapitolare i benefici della nuova parametrizzazione:

- essa rimuove la singolarità della matrice in $\alpha = 0$;
- il significato dei primi tre parametri è chiaro e rappresentativo delle caratteristiche della densità;
- la distribuzione delle stime di massima verosimiglianza migliora notevolmente.

Tuttavia essa presenta dei limiti quali:

- il quarto parametro, non venendo riparametrizzato, continua a portare con sé i problemi di singolarità della matrice d'informazione attesa quanto $\tau \rightarrow \pm\infty$;
- il significato del quarto parametro rimane vincolato a quello di troncamento latente con cui si è presentata la distribuzione al paragrafo 2.1;
- la difficoltà numerica di ottimizzare la logverosimiglianza dovuta a vincoli tra i parametri.

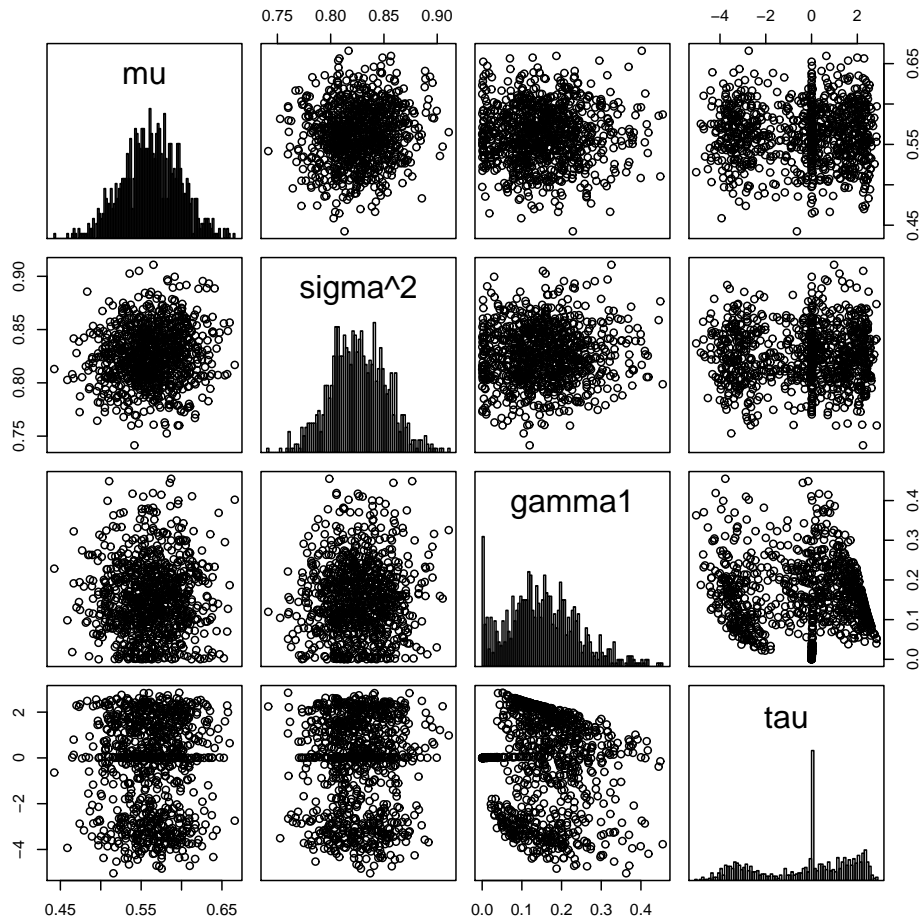


Figura 3.12: Diagrammi di dispersione e istogrammi delle distribuzioni empiriche delle stime di massima verosimiglianza per la parametrizzazione centrata da 1000 campioni di numerosità 200 da una $ESN(0, 1, 1, 0)$. (cfr. fig. 3.8)

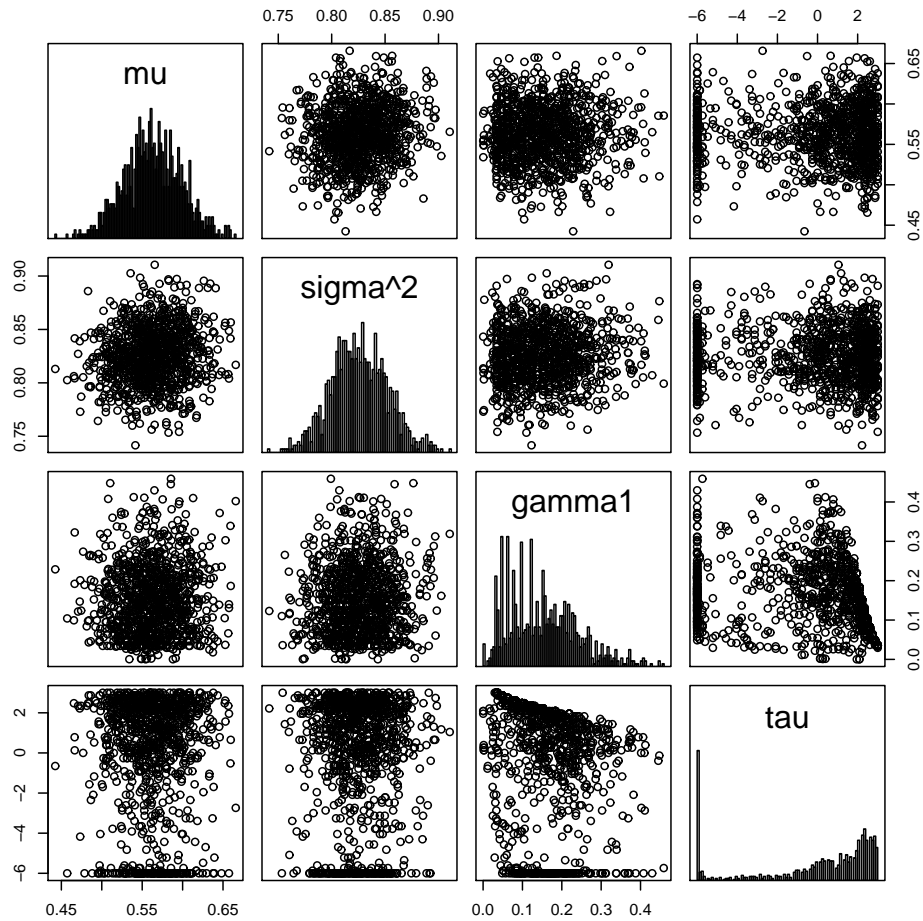


Figura 3.13: Diagrammi di dispersione e istogrammi delle distribuzioni empiriche delle stime di massima verosimiglianza per la parametrizzazione centrata da 1000 campioni di numerosità 200 da una $ESN(0, 1, 1, 0)$. (cfr. fig. 3.9)

Capitolo 4

Regioni di tolleranza per la *ESN* doppia

La costruzione di regioni con assegnata probabilità p è un genere di problema con risvolti sia dal lato teorico che pratico, come ad esempio la costruzione delle regioni di tolleranza. Seguendo le tracce di Azzalini (2001b) e sfruttando le proprietà discusse nei precedenti capitoli quando il parametro $\tau \rightarrow \pm\infty$, si discuterà un approccio per la costruzione di regioni di tolleranza per una distribuzione normale asimmetrica estesa doppia.

4.1 Per la normale e la normale asimmetrica

Si riportano, in questo paragrafo, i risultati dell'articolo di Azzalini (2001b). Sia Z una variabile normale d -dimensionale con matrice di correlazione Ω . È noto il risultato per cui, se c_p è il p -esimo quantile di una distribuzione χ_d^2 , allora

$$R_N = \{x : x^T \Omega^{-1} x \leq c_p\} \quad (4.1)$$

è la regione con assegnata probabilità p di volume minimo.

In forza della proprietà 4 riportata a pagina 7, la (4.1) continua a essere una regione di esatta probabilità anche per la normale asimmetrica, ma non è più quella a volume minimo.

Azzalini (2001b) propone una regione del tipo:

$$R_{SN} = \{x : f(x) \geq f_0\} \quad (4.2)$$

dove f è una funzione di densità del tipo (1.2) e f_0 un opportuno valore della funzione dipendente da p , Ω e α .

Poiché la forma quadratica $Z^T \Omega^{-1} Z$ nel caso normale può essere ottenuta partendo dalla funzione di densità come:

$$2 \log \phi_d(z; \Omega) - d \log(2\pi) - \log |\Omega|,$$

se sostituiamo il logaritmo della densità della normale con quello della densità di una normale asimmetrica, possiamo pensare a

$$R_{SN} = \{x : 2 \log f(x) \geq -c_p - d \log(2\pi) - \log |\Omega|\} \quad (4.3)$$

come una regione candidata allo scopo.

Si riporta l'esperimento di simulazione presente in Azzalini (2001b) per cui si sono esaminati i risultati empirici della regola (4.3) per il caso $d = 2$ per una serie di valori di probabilità p , nello specifico:

$$p = (0.99, 0.975, 0.95, 0.90, 0.80, 0.70, 0.50, 0.30, 0.20, 0.10, 0.05, 0.025, 0.01).$$

Il grafico in figura 4.1 riporta i punti, rappresentati con dei tondi, aventi come ascissa il valore dei percentili teorici calcolati per i valori di probabilità p assegnata e, come ordinata, $c_{\tilde{p}}$, i percentili calcolati sulle probabilità empiriche di copertura \tilde{p} ottenute dell'applicazione della regola (4.3) a un campione di numerosità 10^6 da una *SN* con $\alpha = (2, 6)$ e $\omega = -0.5$.

Quello che si nota è che i punti giacciono su una retta parallela alla bisettrice. In altre parole,

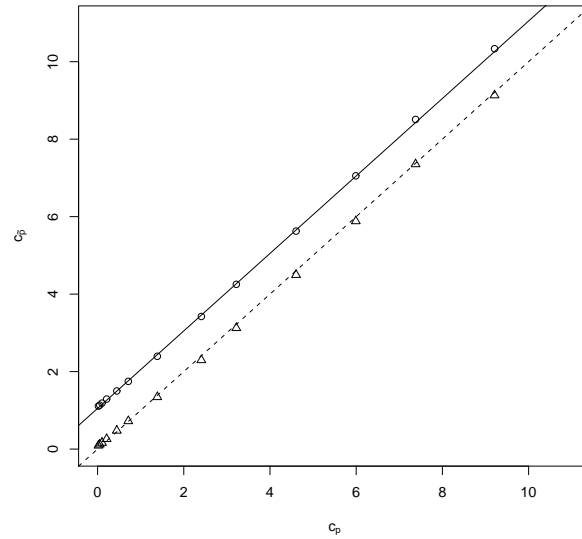


Figura 4.1: Quantili relativi ai valori di probabilità osservati contro i valori fissati, dati i parametri $\alpha_1 = 2$, $\alpha_2 = 6$ e $\omega = -0.5$

sembra esserci una relazione del tipo

$$c_{\bar{p}} = h + c_p \quad (4.4)$$

e quindi sembra opportuno inserire nella regola (4.3) un termine di correzione h .

Questo termine di correzione risulta essere funzione del grado di asimmetria della curva, ossia funzione di α . Nella figura 4.2 è riportato il valore di h in funzione di $\alpha^* = \sqrt{\alpha^T \Omega \alpha}$ per alcune scelte di α e del termine di correlazione ω . La figura 4.2 suggerisce che h sia una funzione monotona crescente di α^* mentre la figura 4.3 mostra una relazione di proporzionalità tra α^* e il termine $(\log(e^{h/2} - 1))^{-1}$ con rapporto -0.6373 . Queste considerazioni si traducono nell'espressione

$$\hat{h} = 2 \log(1 + \exp\{-1.544/\alpha^*\}). \quad (4.5)$$

come funzione interpolante per h .

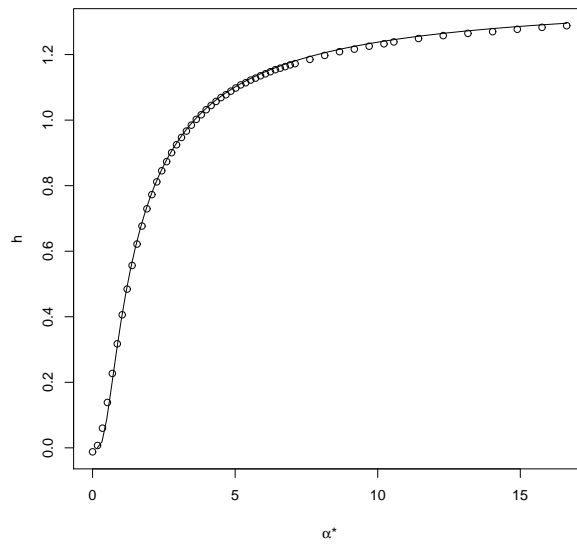


Figura 4.2: Valori osservati di h in funzione di α^* e funzione interpolante

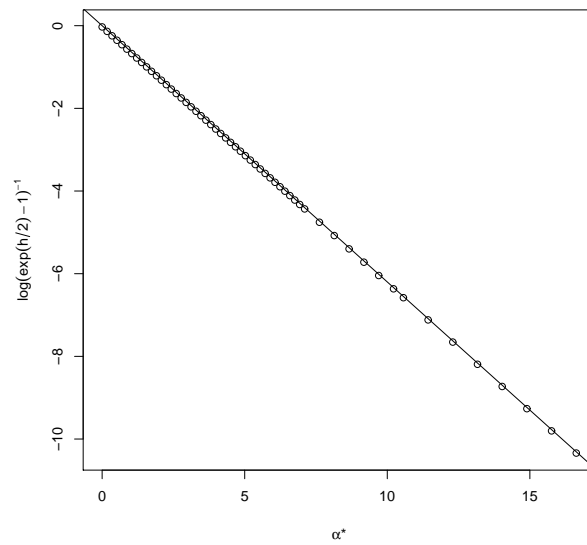


Figura 4.3: Valori osservati di $1/\log(\exp(h/2) - 1)$ in funzione di α^*

p	0.990	0.975	0.950	0.900	0.800	0.700	0.500	0.300	0.200	0.100	0.050	0.025	0.010
\tilde{p}	0.994	0.986	0.971	0.940	0.881	0.819	0.698	0.583	0.528	0.476	0.447	0.433	0.426
\hat{p}	0.990	0.975	0.947	0.894	0.790	0.683	0.489	0.303	0.212	0.122	0.079	0.058	0.045

Tabella 4.1: Valori di probabilità fissati e osservati con $\alpha_1 = 2$, $\alpha_2 = 6$ e $\omega = -0.5$

Sistemando la (4.3) con l'aggiunta di h , ossia

$$R_{SN} = \{x : 2 \log f(x) \geq -c_p - d \log(2\pi) - \log |\Omega| + 2 \log(1 + \exp\{-1.544/\alpha^*\})\}, \quad (4.6)$$

otteniamo una regione di probabilità assegnata per la normale asimmetrica che, applicata al medesimo campione utilizzato in precedenza, mostra dei risultati più soddisfacenti. I punti segnati con dei triangoli della figura 4.1 sono ottenuti dall'applicazione della regola (4.6).

Nella tabella 4.1 sono riportati i valori delle probabilità assegnate p e delle probabilità empiriche di copertura ottenute dalla (4.3) (\tilde{p}) e dalla (4.6) (\hat{p}).

4.2 Per la normale asimmetrica estesa

La costruzione di regioni di assegnata probabilità per la normale asimmetrica estesa ha un'evidente e a prima vista insormontabile difficoltà: se $Z \sim ESN_2(\xi, \Omega, \alpha, \tau)$, la relativa forma quadratica $Z^T \Omega^{-1} Z$ non si distribuisce come un χ_2^2 se $\tau \neq 0$.

4.2.1 Casi limite: $\tau \rightarrow \pm\infty$

Il fatto che la forma quadratica relativa alla normale asimmetrica non si distribuisca come un χ_2^2 limita la possibilità di identificare delle regioni di assegnata probabilità. Tuttavia sappiamo che, in almeno tre casi, questa caratteristica torna a valere, nello specifico quando $\alpha = 0$, quando $\tau = 0$ e quando $\tau \rightarrow \pm\infty$ ¹. Poiché variazioni di posizione e scala lasciano invariato il supporto teorico, fissiamo da qui in seguito, senza perdita di generalità, $\xi = (0, 0)$ e Ω matrice di covarianza con 1 sulla diagonale principale e ω all'infuori di essa.

Al paragrafo 2.3.3 si è dimostrato, nel caso scalare, che quando $\tau \rightarrow \pm\infty$ la distribuzione tende ad essere una distribuzione normale. Di fatto le considerazioni del paragrafo 2.3 evi-

¹si vedrà che nel caso $\tau \rightarrow +\infty$ bisognerà considerare la forma quadratica in $(Z + \delta\tau)$

denziano che questo comportamento limite sia già riscontrabile per valori finiti del parametro grandi in valore assoluto.

Sfruttando questa caratteristica si pensa perciò di approssimare la densità della normale asimmetrica estesa con una densità di una normale. Nello specifico, se

$$Z \sim ESN_2(0, \Omega, \alpha, \tau),$$

vorremmo approssimare la distribuzione di Z con quella di \tilde{Z} dove

$$\tilde{Z} \sim N_2(\mu(\alpha, \tau), \Sigma(\alpha, \tau)). \quad (4.7)$$

Il vettore delle medie $\mu(\tau, \alpha)$, così come la matrice di varianza $\Sigma(\alpha, \tau)$, avrà delle espressioni in funzione di α e τ .

Seguendo degli sviluppi simili a quelli del paragrafo 2.3.3, otteniamo le distribuzioni limite per $\tau \rightarrow \pm\infty$ nel caso bivariato. Per $\tau \rightarrow +\infty$ il calcolo risulta molto semplice e si mostra che:

$$Z \dot{\sim} N_2(0, \Omega) \text{ per } \tau \rightarrow +\infty. \quad (4.8)$$

Quando $\tau \rightarrow -\infty$, invece, la distribuzione limite ha un'espressione tutt'altro che immediata da ricavare. Seguendo uno sviluppo simile a quello di pag. 29, applicando cioè una volta il teorema di de l'Hospital, otteniamo:

$$\begin{aligned} \lim_{\tau \rightarrow -\infty} \phi_2(z; \Omega) \Phi(\tau(1 + \alpha^T \Omega \alpha)^{1/2} + \alpha^T z) / \Phi(\tau) = \\ (1 + \alpha^T \Omega \alpha)^{1/2} \lim_{\tau \rightarrow -\infty} \phi_2(z; \Omega) \phi(\tau(1 + \alpha^T \Omega \alpha)^{1/2} + \alpha^T z) / \phi(\tau). \end{aligned}$$

Al contrario di quanto fatto nel caso scalare, portare all'interno di un'unica funzione esponenziale tutte le quantità non permette di riconoscere nessun tipo di espressione esplicitamente riconducibile a una normale. Tuttavia dopo lunghi e tediosi calcoli, di cui si riporano in

appendice i passaggi principali (cfr. pag. 95), si riesce ad ottenere che:

$$Z + \delta\tau \overset{\sim}{\sim} N_2(0, \tilde{\Omega}) \text{ per } \tau \rightarrow -\infty. \quad (4.9)$$

La matrice di varianza e covarianza della (4.9) è data dall'espressione:

$$\tilde{\Omega} = \begin{pmatrix} 1 - \delta_1^2 & \delta_1\delta_2 - \omega \\ \delta_1\delta_2 - \omega & 1 - \delta_2^2 \end{pmatrix}, \quad (4.10)$$

dove $\delta = (1 + \alpha^T \Omega \alpha)^{-1/2} \alpha^T \Omega$, e $\delta = (\delta_1, \delta_2)^T$.

Utilizzando le approssimazioni (4.8) e (4.9), possiamo costruire delle regioni di assegnata probabilità per la normale asimmetrica estesa utilizzando la regola:

$$R_{ESN} = \{x : 2 \log f_i(x) \geq -c_p - d \log(2\pi) - \log |\Omega_i|\} \quad (4.11)$$

dove a seconda che τ sia positivo o negativo, $f_i(\cdot)$ sarà la distribuzione della (4.8) o della (4.9) e analogamente per la matrice di varianza Ω_i .

È stato effettuato un esperimento simile a quello di Azzalini (2001b) simulando campioni di numerosità 10^6 da alcune distribuzioni normali asimmetriche, su cui è stata applicata la regola (4.11) per alcuni valori fissati di probabilità p rispettivamente utilizzando l'approssimazione (4.9) per i primi due campioni e la (4.8) per i secondi. In fig. 4.4 e nella tabella 4.2 si riportano i grafici e le tabelle analoghi a quelli della fig. 4.1 e della tab. 4.1. Essi si riferiscono a particolari scelte del parametro τ , ma non sono tuttavia casi particolari e sono stati scelti, per l'appunto, come quattro casi tipici.

p	0.990	0.975	0.950	0.900	0.800	0.700	0.500	0.300	0.200	0.100	0.050	0.025	0.010
\tilde{p} (a)	0.989	0.973	0.946	0.893	0.790	0.687	0.489	0.292	0.193	0.097	0.049	0.024	0.010
\tilde{p} (b)	0.959	0.928	0.886	0.818	0.702	0.598	0.411	0.240	0.158	0.078	0.039	0.020	0.008
\tilde{p} (c)	0.993	0.982	0.962	0.922	0.836	0.746	0.549	0.339	0.228	0.114	0.058	0.030	0.012
\tilde{p} (d)	0.990	0.975	0.950	0.901	0.801	0.701	0.501	0.299	0.199	0.099	0.050	0.025	0.010

Tabella 4.2: Valori di probabilità osservati e fissati, dati i parametri $\alpha_1 = 2$, $\alpha_2 = 6$ e $\omega = -0.5$. Ogni riga si riferisce ai medesimi campione dei grafici della fig. 4.4. Valori di τ : $\tau = -10$ per la (a), $\tau = -2$ per la (b), $\tau = 1$ per la (c), $\tau = 4$ per la (d)

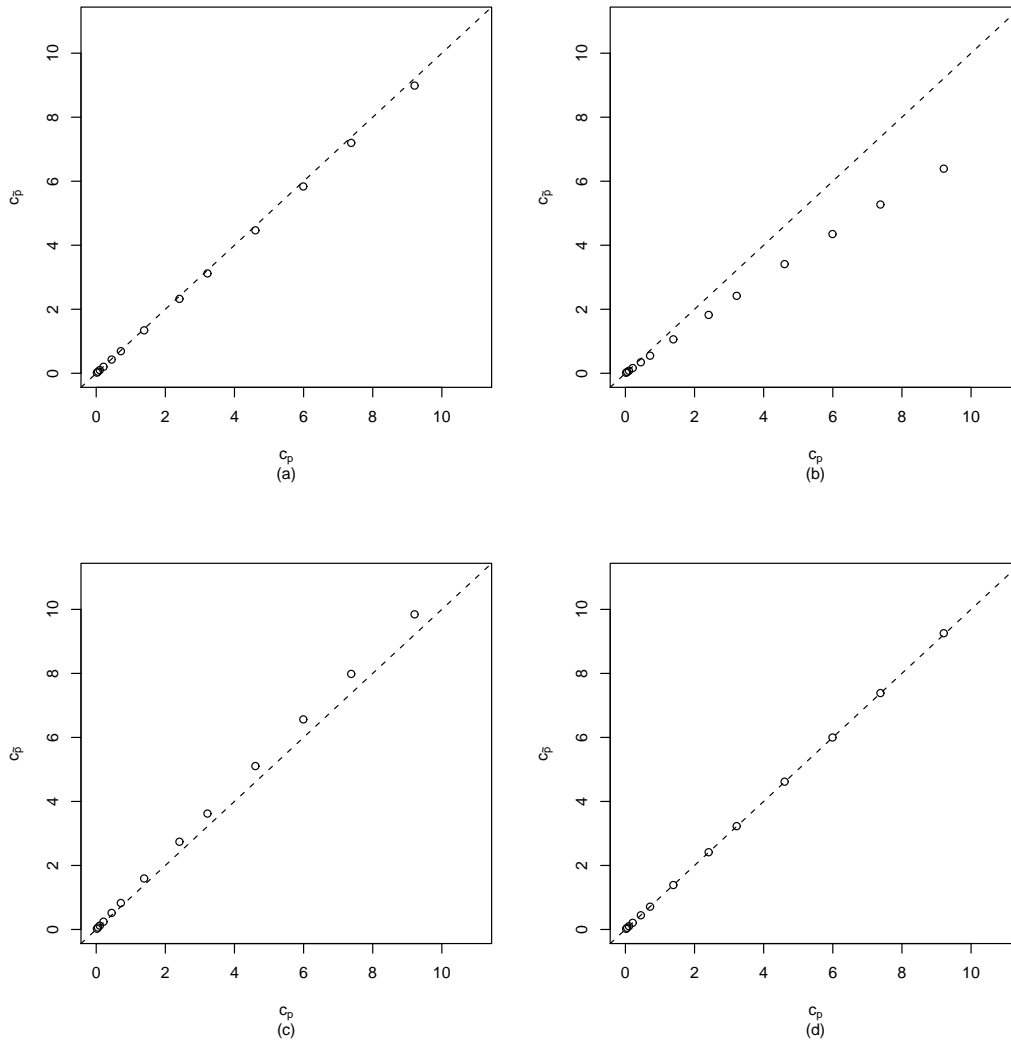


Figura 4.4: Quantili relativi ai valori di probabilità osservati contro i valori fissati, dati i parametri $\alpha_1 = 2$, $\alpha_2 = 6$ e $\omega = -0.5$. Per ogni grafico è stato fatto variare il valore del parametro τ : $\tau = -10$ per la (a), $\tau = -2$ per la (b), $\tau = 1$ per la (c), $\tau = 4$ per la (d)

Come c'era da aspettarsi la regola (4.11) funziona molto bene quando la distribuzione normale asimmetrica degenera a una normale, mentre per valori di τ intorno allo 0 mostra dei risultati decisamente non soddisfacenti. Si veda come nel grafico (a) e nel grafico (d) della figura 4.4, relativi a valori *alti* in valore assoluto del parametro τ , i punti sono quasi perfettamente allineati sulla bisettrice, mentre per i restanti grafici ciò non si verifica. Dai grafici è evidente che le regole portano ad individuare regioni di probabilità differente da quella data per valori di τ intorno allo 0. È proprio in questi casi, infatti, che la proprietà per cui $Z^T \Omega^{-1} Z \sim \chi_2^2$ non è valida.

4.2.2 Casi per valori finiti di τ

Nonostante la mancanza di una distribuzione di riferimento per la forma quadratica della normale asimmetrica estesa per valori finiti di τ diversi da 0, si spera che utilizzando un approccio simile a quello di Azzalini (2001b), sintetizzato nella regola (4.6), si riesca comunque a costruire delle regioni di assegnata probabilità, magari *aggiustando* qualche coefficiente o *aggiungendo* qualche costante.

Applicando a 1250 campioni (facendo variare sia α che τ) la regola (4.3), otteniamo ancora un comportamento simile a quello evidenziato nella normale asimmetrica, riassunto nella figura 4.5.

Come per la normale asimmetrica osserviamo una relazione del tipo

$$c_{\bar{p}} = h + c_p$$

dove questa volta il termine di correzione h sarà una funzione sia di α^* che di τ . In figura 4.6 si riporta il valore di h in funzione di α^* per i 1250 campioni. I diversi simboli del grafico rappresentano diversi valori di τ . In aggiunta è stata tracciata la curva relativa alla funzione interpolante (4.5).

A prima vista sembra che una funzione che descrive h in funzione di entrambi α e τ non sia

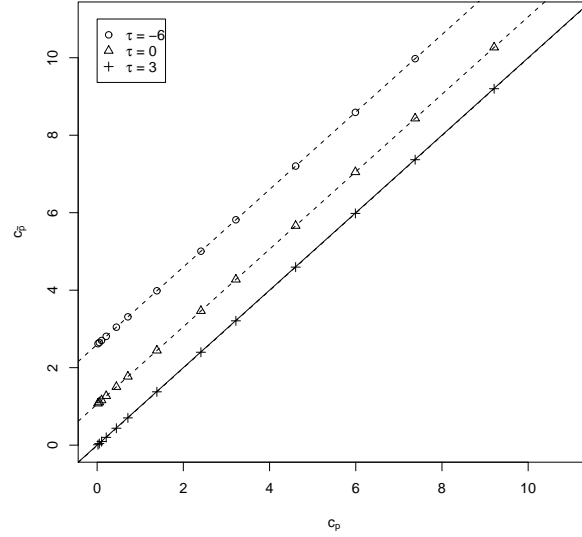


Figura 4.5: Quantili relativi ai valori di probabilità osservati contro i valori fissati applicando la regola (4.3), per alcuni valori di τ

altro che la (4.5) moltiplicata per un coefficiente funzione di τ . In altre parole:

$$\hat{h} = k(\tau)2 \log(1 + \exp\{-1.544/\alpha^*\}). \quad (4.12)$$

Per approssimare $k(\tau)$, è stato calcolato il rapporto tra $h(\tau_i)$ e $h(0)$ per ogni valore di α^* della figura 4.6. Per ogni τ è stato poi calcolato il valore di k come:

$$k(\tau_i) = \frac{1}{l} \sum_{j=1}^l \frac{h(\tau_i, \alpha_j^*)}{h(0, \alpha_j^*)}, \quad (4.13)$$

dove l è il numero di valori di α^* , in questo caso $l = 25$.

Resta da trovare una forma per una funzione di τ per $k(\tau)$. Per fare ciò si sono riportati su un grafico (cfr. fig. 4.7) i valori di k calcolati dalla (4.13) in funzione di τ .

L'andamento di k non sembra particolarmente facile da interpretare tramite una funzione di τ . A riguardo si sono provati vari tipi di funzione: polinomiali, trigonometriche, esponenziali, senza riuscire a trovare una funzione con una interpolazione particolarmente accurata.

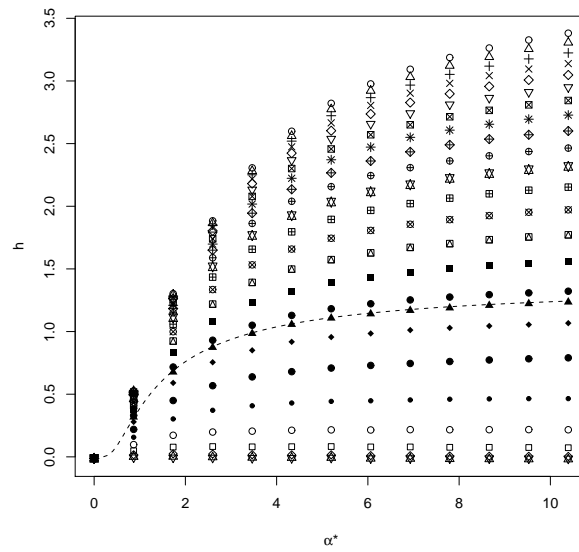


Figura 4.6: Valori osservati di h in funzione di α^* . La linea tratteggiata rappresenta la funzione interpolante (4.5)

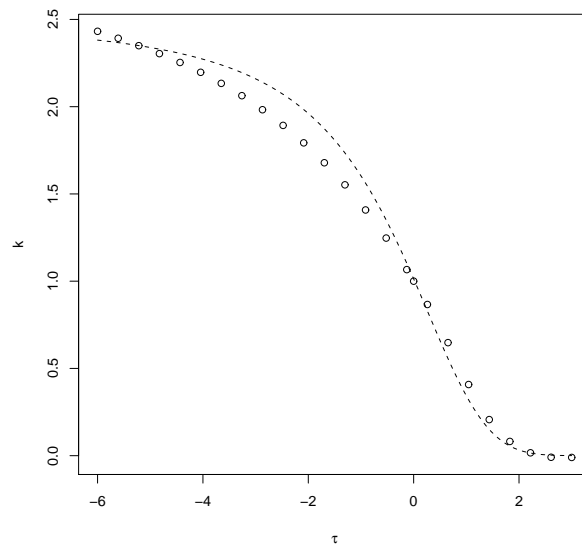


Figura 4.7: Valori osservati di k in funzione di τ . La linea tratteggiata rappresenta la funzione interpolante (4.14)

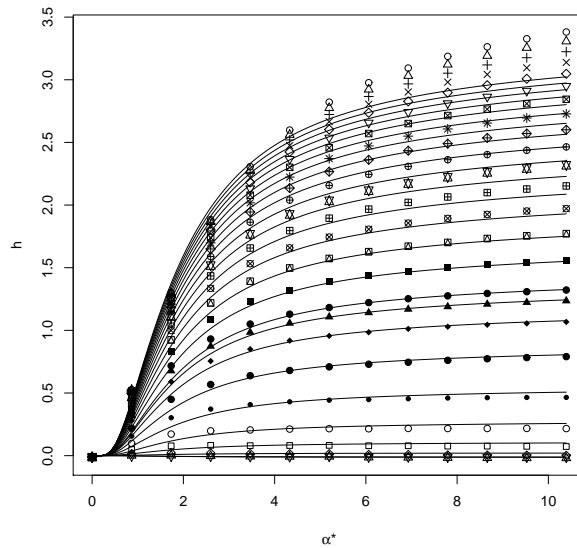


Figura 4.8: Valori osservati di h in funzione di α^* . Le linee continue rappresentano le funzioni interpolanti (4.12)

L'espressione migliore (la cui curva è stata tracciata nella figura 4.7) è del tipo:

$$\hat{k} = \frac{\pi^2}{4} \zeta_2(\tau)^2 \quad (4.14)$$

in modo che calcolata in $\tau = 0$ sia $\hat{k} = 1$.

Utilizzando la (4.14) possiamo tracciare in un grafico analogo a quello in fig. 4.6 le curve per ogni τ . Si veda a riguardo la fig. 4.8. L'interpolazione sembra buona per le curve piu basse, corrispondenti nello specifico a valori positivi del parametro τ , meno per valori negativi.

La regola (4.14) funziona molto bene per valori positivi di τ non solo praticamente ma per taluni aspetti anche concettualmente. Infatti h tende ad essere nullo al tendere di $\tau \rightarrow +\infty$ andando nuovamente a sottolineare il comportamento limite della densità esposto e sfruttato al sottoparagrafo precedente.

Per valori negativi del parametro τ , invece, le curve relative alla (4.12) sovrastimano h per valori piccoli di α^* mentre lo sottostimano per valori grandi di α^* .

Se osserviamo i singoli addendi della (4.13) in funzione di α^* , notiamo che sintetizzarli con

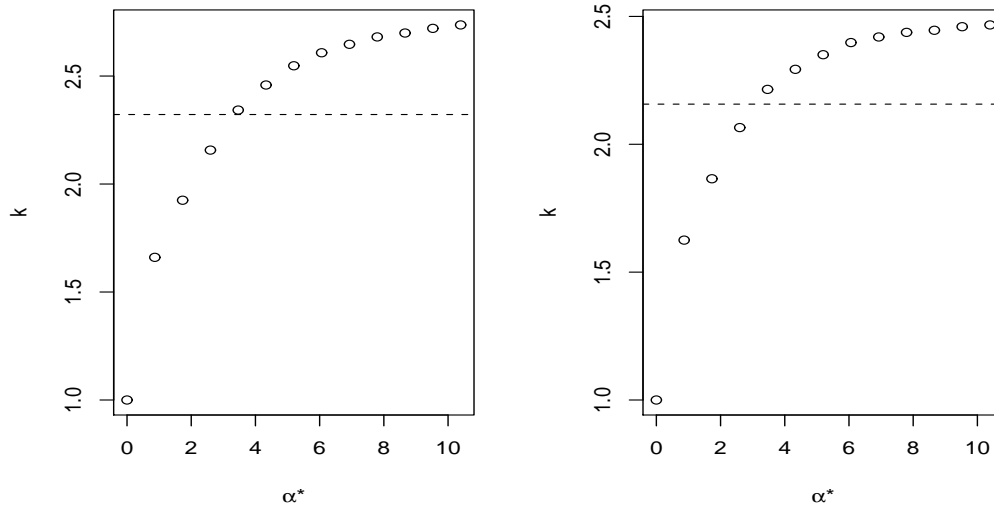


Figura 4.9: Valori osservati di k in funzione di α^* per $\tau = -6$ (a sinistra) e $\tau = -4.43$ (a destra). Le linee tratteggiate rappresentano la media aritmetica dei k

la media aritmetica non sembra opportuno. Si veda a riguardo la fig. 4.9 in cui la relazione tra k e α^* è chiaramente monotona crescente.

Nonostante ciò si è preferito non complicare ulteriormente la funzione che dovrebbe stimare il coefficiente k da inserire nella (4.12) lasciando quindi la (4.14) invariata. Nella figura 4.10 si osservano i risultati per i medesimi campioni della fig. 4.4 e così nella tabella 4.3.

p	0.990	0.975	0.950	0.900	0.800	0.700	0.500	0.300	0.200	0.100	0.050	0.025	0.010
\tilde{p} (a)	0.989	0.973	0.946	0.893	0.790	0.687	0.489	0.292	0.193	0.097	0.049	0.024	0.010
\tilde{p} (b)	0.959	0.928	0.886	0.818	0.702	0.598	0.411	0.240	0.158	0.078	0.039	0.020	0.008
\tilde{p} (c)	0.993	0.982	0.962	0.922	0.836	0.746	0.549	0.339	0.228	0.114	0.058	0.030	0.012
\tilde{p} (d)	0.990	0.975	0.950	0.901	0.801	0.701	0.501	0.299	0.199	0.099	0.050	0.025	0.010

Tabella 4.3: Valori di probabilità osservati e fissati, dati i parametri $\alpha_1 = 2$, $\alpha_2 = 6$ e $\omega = -0.5$. Ogni riga si riferisce ai medesimi campioni dei grafici della fig. 4.4. Valori di τ : $\tau = -10$ per la (a), $\tau = -2$ per la (b), $\tau = 1$ per la (c), $\tau = 4$ per la (d). Il calcolo delle probabilità empiriche di copertura è stato effettuato con la (4.12)

Questo secondo approccio funziona decisamente meglio rispetto al precedente per valori finiti di τ , senza arrivare tuttavia a risultati particolarmente accurati. Inoltre sembra funzionare egregiamente anche per quei valori di τ per cui è garantita l'approssimazione a una normale in

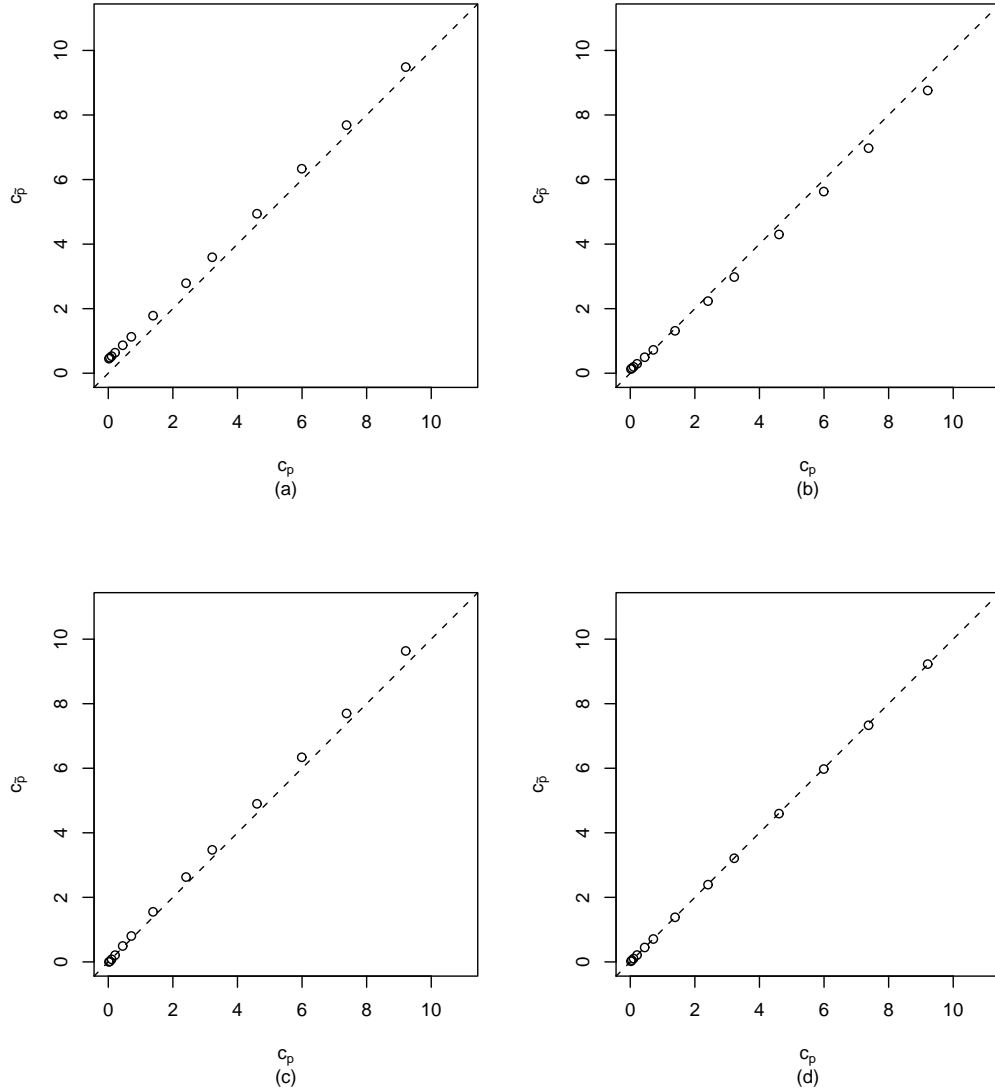


Figura 4.10: Quantili relativi ai valori di probabilità osservati contro i valori fissati, dati i parametri $\alpha_1 = 2$, $\alpha_2 = 6$ e $\omega = -0.5$. Per ogni grafico è stato fatto variare il valore del parametro τ : $\tau = -10$ per la (a), $\tau = -2$ per la (b), $\tau = 1$ per la (c), $\tau = 4$ per la (d). Il calcolo delle probabilità empiriche di copertura è stato effettuato con la (4.12)

particolare per i τ positivi e grandi.

Volendo trarre delle conclusioni sulla base degli studi e delle proprietà osservate e delle numerose prove empiriche effettuate, possiamo affermare che la costruzione di regioni con assegnata probabilità funziona molto bene in quei casi in cui la proprietà secondo cui $Z^T \Omega^{-1} Z \sim \chi_2^2$ è verificata. In questi casi si sono utilizzate delle regole basate sull'approssimazione della normale asimmetrica estesa alla distribuzione normale per $\tau \rightarrow \pm\infty$. Si è poi visto che per valori di τ a cavallo dello 0 ma diversi da 0 (valori che, come più volte sottolineato nei precedenti capitoli, a fini inferenziali sono più che auspicabili) la costruzione di regole per individuare regioni con assegnata probabilità è meno semplice. Tuttavia utilizzando una regola che tiene conto sia di α che di τ , si è riusciti ad ottenere dei risultati comunque soddisfacenti.

Conclusioni

La distribuzione normale asimmetrica estesa è, per costruzione, un modello più flessibile della normale asimmetrica. Considerando il meccanismo di generazione tramite troncamento latente, essa generalizza il caso della normale asimmetrica permettendo quindi maggior flessibilità e capacità di adattamento a fenomeni empirici quali l'osservazione di campioni selezionati. Le sue caratteristiche probabilistiche, inoltre, permettono variazioni dell'indice di asimmetria e di curtosi maggiori rispetto alla normale asimmetrica.

Il quarto parametro, si è visto, non porta a cambiamenti univoci sulla forma della densità al suo variare nel semiasse negativo o in quello positivo, come invece fanno gli altri parametri. Si è dimostrato inoltre che per $\tau \rightarrow \pm\infty$, la distribuzione degenera a una normale; in un caso ($\tau \rightarrow +\infty$) standardizzata e centrata in 0, nel secondo ($\tau \rightarrow -\infty$) con una media proporzionale a $-\tau$ e una varianza funzione del parametro di asimmetria α .

Se si lavora in campo inferenziale, questa distribuzione presenta non poche problematiche. La matrice d'informazione attesa risulta singolare non solo per $\alpha = 0$ ma anche quando $\tau \rightarrow \pm\infty$. Inoltre il determinante di questa è prossimo allo zero anche per valori finiti di τ .

La logverosimiglianza risulta difficile da massimizzare; una superficie irregolare e la presenza di punti di massimo relativo fanno sì che diversi algoritmi numerici, per diversi valori di inizializzazione, diano delle stime dei parametri distanti tra loro e spesso distanti dal vero valore di ogni parametro.

La logverosimiglianza profilo per τ risulta molto poco informativa sul valore del parametro, e questo porta spesso gli algoritmi numerici a non stimarlo correttamente. Questo è compensato

dal fatto che la verosimiglianza *aggiusta* le stime degli altri parametri portando nella pratica a densità equivalenti, per valori diversi dei parametri. Queste caratteristiche evidenziano un problema di *quasi* non identificabilità del modello.

Per quanto riguarda la costruzione di regioni di assegnata probabilità, il fatto che nella normale asimmetrica estesa d -dimensionale la relativa forma quadratica non goda della proprietà di distribuirsi come un χ_d^2 (per τ finito e $\tau \neq 0$), rende difficoltosa la costruzione di queste regioni. Se da un lato si è dimostrata un'efficace regola di costruzione di regioni di assegnata probabilità p nei casi per cui $\tau \rightarrow \pm\infty$, dall'altro si è presentata una regola che funziona egregiamente anche per valori finiti del parametro.

Appendice A

Codice R

A.1 Calcolo della matrice d'informazione attesa

Si riporta il codice R implementato per il calcolo della matrice d'informazione attesa.

```
desn<-function (x, location = 0, scale = 1, shape = 0,
trunc=0, dp = NULL, log = FALSE)
{
  if (!is.null(dp)) {
    if (!missing(shape))
      stop("You cannot set both component parameters and dp")
    location <- dp[1]
    scale <- dp[2]
    shape <- dp[3]
    trunc <- dp[4]
  }
  z <- (x - location)/scale
  if (log)
    y <- -zeta(0,trunc) - log(scale) - z^2/2 + zeta(0,
      trunc * ( sqrt( shape^2 + 1)) + shape * z)
  else y <- (1/pnorm(trunc)) * dnorm(z) * pnorm(trunc * (sqrt( shape^2 + 1)) +
    shape * z)/scale
  replace(y, scale <= 0, NaN)
}
```

```

esn.Einfo<-function(dp=NULL)
{
  xi<-dp[1]
  omega<-dp[2]
  alpha<-dp[3]
  tau<-dp[4]

  #calcolo numericamente le quantità ak

a0<-integrate( function(x) desn(x, 0, 1, alpha, tau) *
  zeta(1, tau*sqrt(1+alpha^2)+alpha*x)^2, -Inf, Inf,
  rel.tol = .Machine$double.eps^.5 )$value
a1<-integrate( function(x) desn(x, 0, 1, alpha, tau) * x *
  zeta(1, tau*sqrt(1+alpha^2)+alpha*x)^2, -Inf, Inf,
  rel.tol = .Machine$double.eps^.5 )$value
a2<-integrate( function(x) desn(x, 0, 1, alpha, tau) * x^2 *
  zeta(1, tau*sqrt(1+alpha^2)+alpha*x)^2, -Inf, Inf,
  rel.tol = .Machine$double.eps^.5 )$value

  delta<-alpha/sqrt(1+alpha^2)
  E.z<-ze(1, tau)*delta

i11<-(1/omega^2) * (1 + (alpha^2)*a0 + (delta^2)*tau*ze(1, tau) )
i12<-(1/omega^2) * (E.z*((1 + 2*alpha^2)/(1 + alpha^2)) -
((alpha^3)*(tau^2)*ze(1, tau))/((1+alpha^2)*sqrt(1+alpha^2))
+ (alpha^2)*a1)
i13<-(1/omega) * (ze(1, tau)/(1+alpha^2)^(3/2) - tau*alpha*delta*a0
- alpha*a1)
i14<-((-alpha * sqrt(alpha^2+1)) / omega) * a0 - delta*tau*ze(1, tau)/omega
i22<-(1/omega^2) * (2 + (alpha^2)*a2 + delta^4*ze(1,tau)*tau*(tau^2-3) )
i23<- (1/omega) * ( -2*alpha*tau*ze(1, tau)/((1+alpha^2)^2) -
tau*alpha*delta*a1 - alpha*a2)
i24<-((alpha*sqrt(1+alpha^2))/omega) * (-a1 + (alpha*ze(1, tau)*(tau^2-1))/
(1+alpha^2)^(3/2) )
i33<-((tau*delta)^2)*a0 + 2*tau*delta*a1 + a2
i34<- alpha*tau*a0 + sqrt(1 + alpha^2)*a1
i44<- (1+alpha^2)*a0 - (ze(1, tau))^2

i<-c(i11,i12,i13,i14,i12,i22,i23,i24,i13,i23,i33,i34,i14,i24,i34,i44)
I_dp<-matrix(i,byrow=T,ncol=4)

```



```

aux<-list(a.int = c(a0,a1,a2))
list(info.dp = I_dp, aux = aux)
}

```

A.2 Calcolo delle stime di massima verosimiglianza

Massimizzazione diretta di tutti i parametri

```

esn.mle<-function(y, dp = NULL, meth="L-BFGS-B"){
  if(is.null(dp)){
    mu<-mean(y)
    sig<-var(y)
    gammal<-sum((y - mu)^3)/(length(y) * sd(y)^3)
    tau<-0
    dp <- as.vector(c(cp.to.dp(c(mu,sig,gammal)),tau))
  }

  opt <- optim(dp, fn = esn.dev.dp, method = meth, # gr = esn.dev.gh,
    lower = c(-Inf, 10 * .Machine$double.eps, -Inf, -Inf),
    upper = c(Inf, Inf, Inf, Inf), control = list(iter.max = 100, abs.tol = 1e-05),
    y = y, trace = FALSE, hessian = FALSE)
  par<-c(opt$par)
  list( param = par, logL = -0.5 * opt$value)
}

```

Massimizzazione logverosimiglianza profilo per τ

```

esn.mle.prof.t<-function(y, iter = 70, tau = seq(-6,3,length=70), plot=TRUE)
{
  mu<-mean(y)
  sig<-var(y)
  gammal<-sum( (y - mu)^3) / (length(y) * sd(y)^3)
  out<-NULL
  dp<-as.vector(cp.to.dp(c(mu,sig,gammal)))
  for(i in 1:iter)
  {
    opt <- optim(dp, fn = esn.dev.prof.t, method = "L-BFGS-B",
      lower = c(-Inf, 10 * .Machine$double.eps, -15),
      upper = c(Inf, Inf, 5), control = list(iter.max = 100, abs.tol = 1e-08),
      y = y, tau=tau[i], trace = FALSE, hessian = FALSE)
  }
}

```

```

    if(opt$convergence==0)    out<-rbind(out,c(opt$par,tau[i],opt$value))

  }
out<-data.frame(out)
names(out)<- c("xi","omega","alpha","tau","logL")
if (plot)
{
  windows()
  plot(out$tau,-0.5*(out$logL-out[out$logL==min(out$logL),5]),type='l',
  xlab=expression(tau), ylab="logL",
  main=expression("logverosimiglianza profilo per " * tau))
  abline(h=-qchisq(.95,1)/2,lty=2)
}

MLE<-out[out$logL==min(out$logL),]
invisible(list(param = out[out$logL==min(out$logL),1:4] ,
  logL = -0.5 * out[out$logL==min(out$logL),5] ,
  tau=out$tau, proflogL=-0.5*(out$logL-out[out$logL==min(out$logL),5])) )
}

```

Massimizzazione logverosimiglianza profilo per α

```

esn.mle.prof.a<-function(y, iter = 50, alpha = seq(-4,4,length=iter), plot=TRUE){

  mu<-mean(y)
  sig<-var(y)
  tau<-0
  par<-c(cp.to.dp(c(mu,sig,0)),tau=tau)[-3]
  out<-NULL
  if(cp){gamma1<- seq(-0.9,0.9,length=iter)
    alpha<- gamma1.to.lambda(gamma1)
  }
  for(i in 1:50)
  {
    opt<-optim(par, fn = esn.dev.prof.a, method = "L-BFGS-B", gr = esn.dev.gh.prof.a,
    lower = c(-Inf, 10 * .Machine$double.eps, -10),
    upper = c(Inf, Inf, Inf), control = list(iter.max = 100, abs.tol = 1e-05),
    y = y, alpha=alpha[i], trace = FALSE, hessian = FALSE)
    out<-rbind(out,c(opt$par[1:2],alpha[i],opt$par[3],opt$value))
  }
  out<-data.frame(out)
  names(out)<- c("xi","omega","alpha","tau","logL")
}

```

```

if(plot) plot(out$alpha,-0.5*(out$logL-out[out$logL==min(out$logL),5]),type='l',
  xlab=expression(alpha), ylab="logL",
  main=expression("log-verosimiglianza profilo per " * alpha))
MLE<-out[out$logL==min(out$logL),]
list(param = MLE[1:4], logL = -0.5 * MLE[5])
}

```

Massimizzazione logverosimiglianza profilo per α, τ

```

esn.mle.prof.at<-function(y, iter = 100, alpha = seq(-10,10,length=iter),
  tau = seq(-10,7,length=iter), plot=TRUE){
  out<-matrix(NA,ncol=iter,nrow=iter)
  mu<-mean(y)
  sig<-var(y)
  par<-c(cp.to.dp(c(mu,sig,0)))[-3]
  for(i in 1:iter)
  {
    for(j in 1:iter)
    {
      dp<-c(par,tau[i],alpha[j])
      opt<-optim(dp, fn = esn.dev.prof.at, method = "L-BFGS-B",
        lower = c(-Inf, 10 * .Machine$double.eps),
        upper = c(Inf, Inf), control = list(iter.max = 100, abs.tol = 1e-05),
        y = y, alpha=alpha[j], tau=tau[i], trace = FALSE, hessian = FALSE)
      out[i,j]<- -0.5 * opt$value
    }
  }
  out==max(out)
  if(plot){
    contour(tau,alpha,out, xlab=expression(tau), ylab=expression(alpha),
      main=expression("log-verosimiglianza profilo per " * alpha * " e " * tau))
  }
}

```


Appendice B

Ulteriori grafici e risultati

B.1 Distribuzione delle stime

Si riportano i grafici relativi ai commenti di pag. 59.

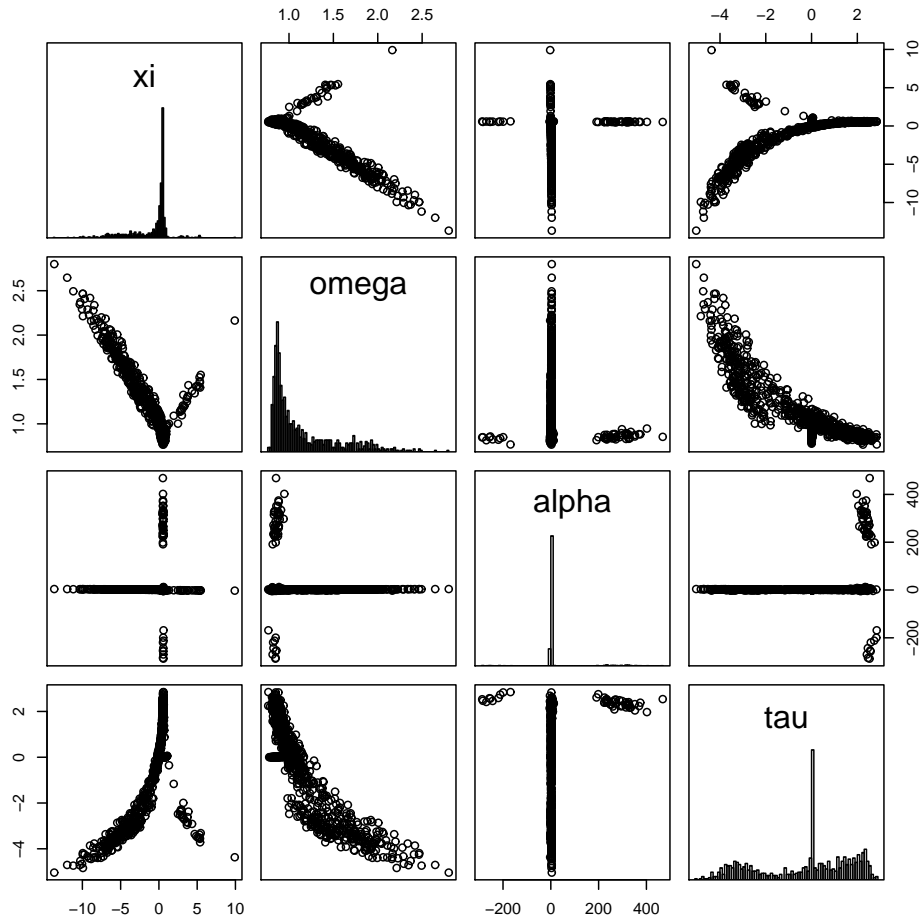


Figura B.1: Diagrammi di dispersione e istogrammi delle distribuzioni empiriche delle stime di massima verosimiglianza da 1000 campioni di numerosità 500 da una $ESN(0, 1, 1, 0)$. Le prime 200 osservazioni di ogni campione sono le stesse su cui si sono calcolati i grafici in fig. 3.8. La massimizzazione è stata effettuata contemporaneamente per tutti i parametri

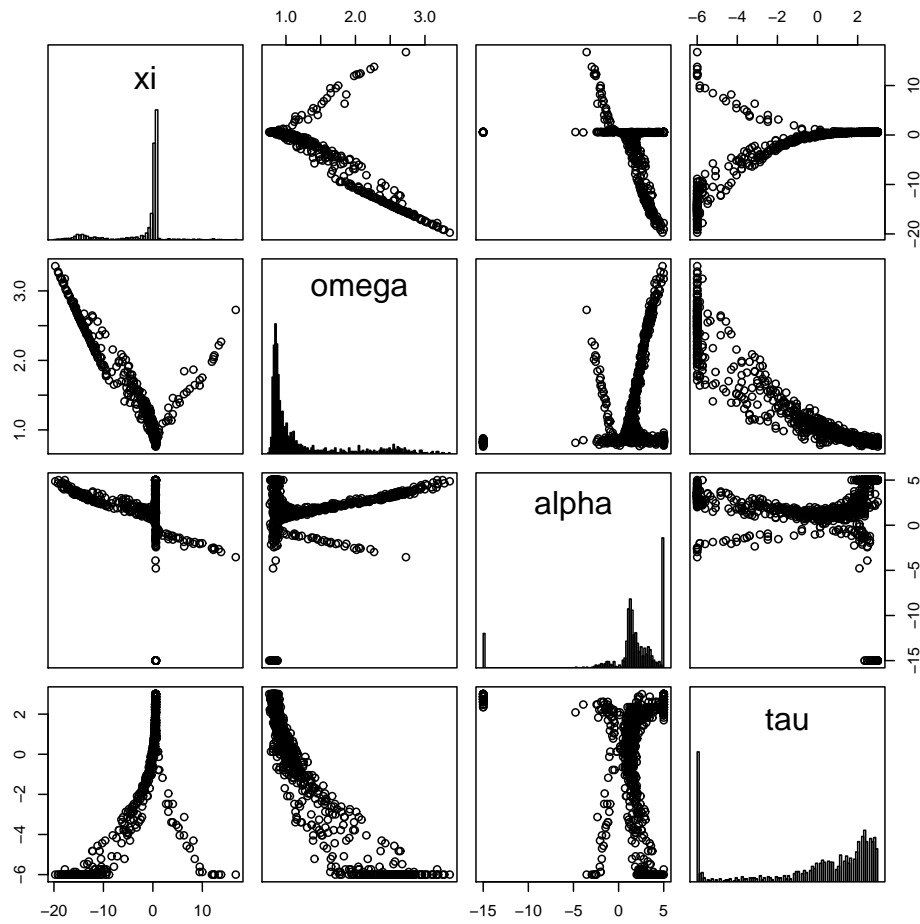


Figura B.2: Diagrammi di dispersione e istogrammi delle distribuzioni empiriche delle stime di massima verosimiglianza da 1000 campioni di numerosità 500 da una $ESN(0, 1, 1, 0)$. Le prime 200 osservazioni di ogni campione sono le stesse su cui si sono calcolati i grafici in fig. 3.9. La massimizzazione è stata effettuata sulla logverosimiglianza profilo per τ

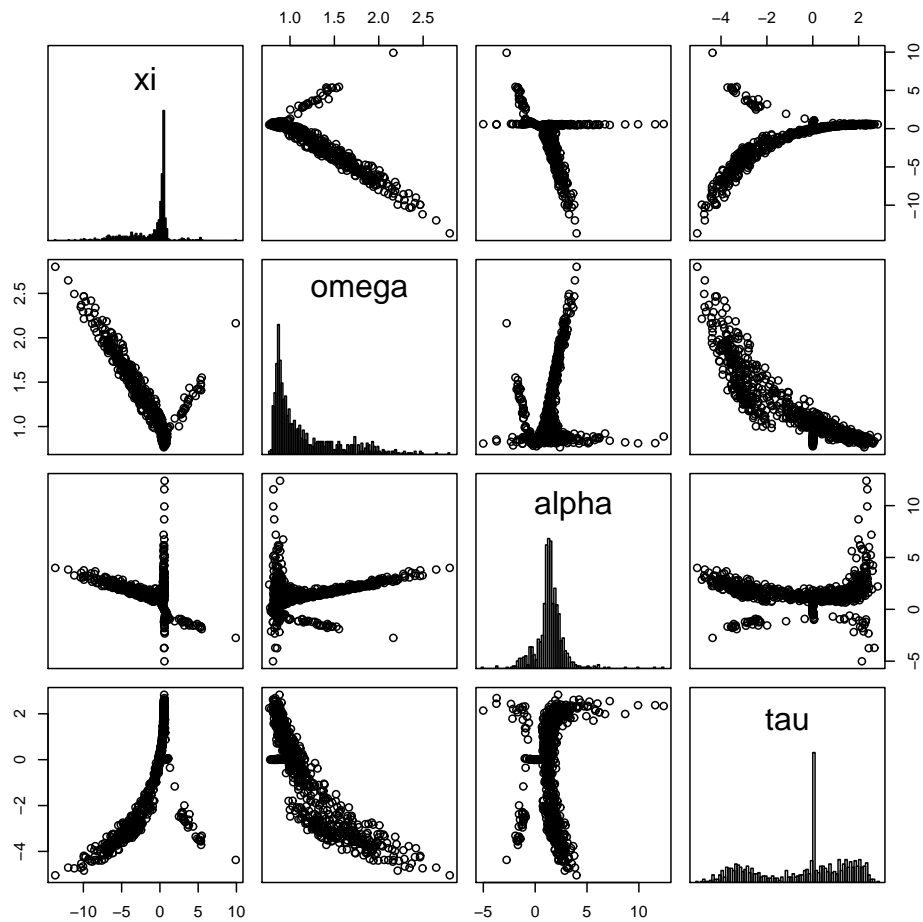


Figura B.3: Distribuzione empirica delle stime di massima verosimiglianza come da fig.B.1. Si sono eliminati i 152 campioni con $|\hat{\alpha}| > 150$

B.2 Distribuzioni limite ESN_2

Si riportano i passaggi principali relativi al calcolo dei limiti per $\tau \rightarrow \pm\infty$ per la densità della normale asimmetrica estesa doppia (cfr. cap. 4, pag. 73).

Detta $f_2(z; \Omega, \alpha, \tau)$ la densità della normale asimmetrica, per $\tau \rightarrow +\infty$, molto semplicemente si ha che:

$$\begin{aligned} \lim_{\tau \rightarrow +\infty} f_2(z; \Omega, \alpha, \tau) &= \lim_{\tau \rightarrow +\infty} \phi_2(z; \Omega) \Phi(\tau(1 + \alpha^T \Omega \alpha)^{1/2} + \alpha^T z) / \Phi(\tau) \\ &= \phi_2(z; \Omega) \end{aligned}$$

Per $\tau \rightarrow -\infty$, i calcoli sono decisamente più complicati. In virtù del teorema di de l'Hospital possiamo porre:

$$\begin{aligned} \lim_{\tau \rightarrow -\infty} f_2(z; \Omega, \alpha, \tau) &= \lim_{\tau \rightarrow -\infty} \phi_2(z; \Omega) \Phi(\tau(1 + \alpha^T \Omega \alpha)^{1/2} + \alpha^T z) / \Phi(\tau) \\ &= (1 + \alpha^T \Omega \alpha)^{1/2} \lim_{\tau \rightarrow -\infty} \phi_2(z; \Omega) \phi(\tau(1 + \alpha^T \Omega \alpha)^{1/2} + \alpha^T z) / \phi(\tau). \end{aligned}$$

Utilizzando l'uguaglianza

$$\delta = \frac{1}{(1 + \alpha^T \Omega \alpha)^{1/2}} \Omega \alpha, \quad (\text{B.1})$$

possiamo porre

$$\begin{aligned} \lim_{\tau \rightarrow -\infty} f_2(z; \Omega, \alpha, \tau) &= \frac{1}{(1 - \delta^T \Omega^{-1} \delta)^{1/2}} \lim_{\tau \rightarrow -\infty} \phi_2(z; \Omega) \phi(\tau(1 + \alpha^T \Omega \alpha)^{1/2} + \alpha^T z) / \phi(\tau) \\ &= \frac{1}{(1 - \delta^T \Omega^{-1} \delta)^{1/2}} \lim_{\tau \rightarrow -\infty} \frac{1}{2\pi|\Omega|} \exp \left\{ -\frac{1}{2} z^T \Omega^{-1} z \right\} \times \\ &\quad \exp \left\{ -\frac{1}{2} \left(\tau(1 + \alpha^T \Omega \alpha)^{1/2} + \alpha^T z \right)^2 \right\} \exp \left\{ \frac{\tau^2}{2} \right\} \\ &= \frac{1}{(1 - \delta^T \Omega^{-1} \delta)^{1/2}} \frac{1}{2\pi|\Omega|} \lim_{\tau \rightarrow -\infty} \exp \left\{ -\frac{1}{2} A \right\}, \quad (\text{B.2}) \end{aligned}$$

dove

$$A = z^T \Omega^{-1} z + \left(\tau(1 + \alpha^T \Omega \alpha)^{1/2} + \alpha^T z \right)^2 - \tau^2$$

$$\begin{aligned}
&= z^T \Omega^{-1} z + \tau^2 \alpha^T \Omega \alpha + (\alpha^T z)^2 + 2\tau \alpha^T z (1 + \alpha^T \Omega \alpha)^{1/2} \\
&= z^T \Omega^{-1} z + \tau^2 \frac{\delta^T \Omega^{-1} \delta}{1 - \delta^T \Omega^{-1} \delta} + \frac{(\delta^T \Omega^{-1} z)^2}{1 - \delta^T \Omega^{-1} \delta} + 2\tau \frac{\delta^T \Omega^{-1} z}{1 - \delta^T \Omega^{-1} \delta} \\
&= \frac{1}{1 - \delta^T \Omega^{-1} \delta} \left(z^T \Omega^{-1} z + (z^T \Omega^{-1} z \delta^T \Omega^{-1} \delta) + \tau^2 \delta^T \Omega^{-1} \delta + \right. \\
&\quad \left. (\delta^T \Omega^{-1} z)^2 + 2\tau \delta^T \Omega^{-1} z \right),
\end{aligned}$$

tenuto conto della (B.1). A questo punto non è semplice individuare esplicitamente all'interno dell'esponenziale una qualche forma quadratica tale da poter ricondursi a una normale doppia. Per fare ciò è infatti necessario esprimere esplicitamente ogni prodotto matriciale in termini scalari.

Posto $\delta = (\delta_1, \delta_2)^T$, $z = (z_1, z_2)^T$, esplicitando ogni addendo all'interno della (B.3) si ottiene:

$$z^T \Omega^{-1} z = \frac{1}{1 - \omega^2} (z_1^2 - 2\omega z_1 z_2 + z_2^2), \quad (\text{B.3})$$

$$\begin{aligned}
(z^T \Omega^{-1} z \delta^T \Omega^{-1} \delta) &= \frac{1}{(1 - \omega^2)^2} \left(-z_1^2 \delta_1^2 + 2\omega \delta_1 \delta_2 z_1^2 - \delta_2^2 z_1^2 + 2\omega z_1 z_2 \delta_1^2 + \right. \\
&\quad \left. 2\omega z_1 z_2 \delta_2^2 - 4\omega^2 \delta_1 \delta_2 z_1 z_2 - z_2^2 \delta_1^2 + 2\omega \delta_1 \delta_2 z_2^2 - \right. \\
&\quad \left. \delta_2^2 z_2^2 \right), \quad (\text{B.4})
\end{aligned}$$

$$\tau^2 \delta^T \Omega^{-1} \delta = \frac{\tau^2}{1 - \omega^2} (\delta_1^2 - 2\omega \delta_1 \delta_2 + \delta_2^2), \quad (\text{B.5})$$

$$\begin{aligned}
(\delta^T \Omega^{-1} z)^2 &= \frac{1}{(1 - \omega^2)^2} \left(z_1^2 \delta_1^2 + \omega^2 \delta_2^2 z_1^2 \omega^2 \delta_1^2 z_2^2 - 2\omega \delta_1 \delta_2 z_1^2 - 2\omega \delta_1^2 z_1 z_2 + \right. \\
&\quad \left. - 2\omega \delta_2^2 z_1 z_2 + 2\delta_1 \delta_2 z_1 z_2 + 2\omega^2 \delta_1 \delta_2 z_1 z_2 - 2\omega \delta_2^2 z_1 z_2 \right), \quad (\text{B.6})
\end{aligned}$$

$$2\tau \delta^T \Omega^{-1} z = \frac{2\tau}{1 - \omega^2} (z_1 \delta_1 - \omega \delta_2 z_1 - \omega \delta_1 z_2 + \delta_2 z_2). \quad (\text{B.7})$$

Raccogliendo tutti i termini dalla (B.3) alla (B.7), dopo alcuni passaggi algebrici, possiamo finalmente ricondursi a una forma quadratica in z . Sfruttando l'indicazione delle espressioni a pag.29, poniamo come vettore delle medie della distribuzione limite $(-\delta_1 \tau, -\delta_2 \tau)^T$. Allora, sviluppando la generica forma quadratica di una normale doppia di media $\mu = (\mu_1, \mu_2)^T$ e

matrice di varianza \mathbf{S} e uguagliando i coefficienti delle potenze delle variabili e delle medie marginali otteniamo delle espressioni per i termini della matrice \mathbf{S}^{-1} che portano all'espressione:

$$\tilde{\Omega} = \begin{pmatrix} 1 - \delta_1^2 & \delta_1 \delta_2 - \omega \\ \delta_1 \delta_2 - \omega & 1 - \delta_2^2 \end{pmatrix}. \quad (\text{B.8})$$

Affinchè la densità $f(z; \alpha, \tau)$ quando $\tau \rightarrow -\infty$ possa essere scritta come densità di una normale $N(-\delta\tau, \tilde{\Omega})$, il determinante di $\tilde{\Omega}$ deve comparire anche elevato alla $-1/2$ all'esterno dell'esponenziale. Riprendendo la (B.2) e sviluppando la quantità alla destra del limite a meno di $\frac{1}{2\pi}$ si ottiene:

$$\begin{aligned} \frac{1}{(1 - \delta^T \Omega^{-1} \delta)^{1/2} |\Omega|} &= \frac{\left((1 - \omega^2)^{-1} (1 - \omega^2 - \delta_1^2 + 2\delta_1 \delta_2 \omega - \delta_2^2) \right)^{1/2}}{\left((1 - \omega^2)^{-1} (1 - \omega^2 - \delta_1^2 + 2\delta_1 \delta_2 \omega - \delta_2^2) \right) (1 - \omega^2)^{1/2}} \\ &= \left(1 - \omega^2 - \delta_1^2 + 2\delta_1 \delta_2 \omega - \delta_2^2 \right)^{-1/2}, \end{aligned}$$

che è esattamente il determinante della matrice $\tilde{\Omega}$ elevato alla $-1/2$.

Alla luce di ciò si può concludere che:

$$\lim_{\tau \rightarrow +\infty} f_2(z; \Omega, \alpha, \tau) = \phi_2(z + \delta\tau; \tilde{\Omega}).$$

Bibliografia

- Aigner, D. J. (1964). The sum of values from a normal and a truncated normal distribution. *Technometrics*, 4, 469–471.
- Andel, J., Netuka, I., & Zvára, K. (1984). On threshold autoregressive processes. *Kybernetika*, 20, 89–106.
- Arellano-Valle, R. & Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *J. Multivariate Analysis*, 99, 1362–1382.
- Arnold, B. & Beaver, R. (2000). Hidden truncation model. *Sankhya, series A*, 62, 22–35.
- Arnold, B., Cox, D., Bottai, M., & Robins, J. (1993). The non truncated marginal of a truncated bivariate normal distribution. *Psychometrika*, 58, 471–488.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, 12, 171–178.
- Azzalini, A. (1986). Further result on a class of distributions which includes the normal ones. *Statistica*, XLVI, 171–178.
- Azzalini, A. (2001a). *Inferenza statistica. Una presentazione basata sul concetto di verosimiglianza*. Springer-Verlag Italia.
- Azzalini, A. (2001b). A note on regions of given probability of the skew-normal distribution. *Metron*, 59(3-4), 27–34.

- Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scand. J. Statist.*, 32(2), 159–188.
- Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *J. R. Stat. Soc. series B*, 61, 579–602.
- Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83, 715–726.
- Birnbaum, Z. W. (1950). Effect of linear truncation on a multinormal population. *Ann. Math. Statist.*, 21, 272–279.
- Byrd, R. H., Lu, P., & Nocedal, J. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. on Scientific and Statistical Computing*, 16, 1190–1208.
- Capitanio, A., Azzalini, A., & Stanghellini, E. (2003). Graphical models for skew-normal variates. *Scand. J. Statist.*, 30, 129–144.
- Chiogna, M. (1998). Some results on the scalar skew-normal distribution. *J. Ital. Statist. Soc.*, 7, 1–13.
- Chiogna, M. (2005). A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Stat. Meth. & Appl.*, 14, 331–341.
- de Helguero, F. (1909). Sulla rappresentazione analitica delle statistiche abnormali. *Atti del IV congresso internazionale dei matematici*, III, 288–299.
- Edgeworth, F. Y. (1886). The law of error and the elimination of chance. *Philosophical Magazine*, 21, 308–324.
- Genton, M. G. (2004). *Skew-elliptical distributions and their applications. A journey beyond normality*. Chapman & Hall/CRC.
- Pace, L. & Salvan, A. (2001). *Introduzione alla statistica. Inferenza, verosimiglianza, modelli*, volume 2. CEDAM.

- Pearson, K. (1893). Asymmetrical frequency curves. *Nature*, 48, 615–616.
- Pearson, K. (1895). On skew probability curves. *Nature*, 52, 317.
- Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, 27(7), 859–870.
- Rotnizsky, A., Cox, D., Bottai, M., & Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, 6, 243–284.
- Soriani, N. (2007). La distribuzione t -asimmetrica: analisi discriminante e regioni di tolleranza. Tesi di laurea, Facoltà di Scienze Statistiche, Università di Padova.
- Weinstein, M. A. (1964). The sum of values from a normal and a truncated normal distribution. *Technometrics*, 6, 104–105.

Ringraziamenti

In quest'ultima pagina voglio ringraziare chi, a suo modo, mi ha aiutato in questi cinque anni di Università.

In primo luogo ringrazio il prof. Azzalini. Grazie per avermi fatto apprezzare molti particolari della Statistica sia durante i corsi di quest'anno sia durante questo lavoro di tesi, grazie per avermi fatto lavorare intensamente e avermi permesso di imparare molto. Un grazie in particolare per avermi portato a Bertinoro, per avermi dato la possibilità di ascoltare e conoscere alcuni validissimi statistici. Ringrazio inoltre il prof. Scarpa per le animate chiacchierate sulla Statistica e sui suoi possibili approcci, per consigli sui corsi e le lezioni mai noiose. Ringrazio Bruno per tutto il resto. Un grazie va anche alla prof.ssa Capitanio che pazientemente mi ha ascoltato su questioni *skew*, al prof. di Fonzo per la disponibilità e per i consigli sulla scelta dell'argomento di tesi, al prof Masarotto per la disponibilità e sopra tutto per avermi convinto, ancora liceale, ad iscrivermi a questa facoltà.

Ringrazio calorosamente le *ragazze del primo piano*, in ordine di ufficio: Francesca, Delfina, Paola e Fiorenza per la pazienza dimostrata e la simpatia con cui mi hanno sempre accolto nelle mie vicende *burocratiche*.

Un grazie a chi in questi anni ha studiato e discusso di Statistica con me. Gli amici di SGI: Steno, Cecchin, Giacomo, Beppe, Mirko e Clara ma anche chi non si vede più come il vecchio Flebo, la Dany e Frizz. Gli amici di Statistica Laboratorio: Pippo e Nicola, la Checca, la Marghe, Paolo, Michele, Massimiliano. In più: Nicola, Federica, Ilaria, Andrea, Flavio, Barbara, Elisa e Anna. Un grazie ai colleghi tutor, Davide e Margherita, per i sabati mattina a fare integrali.

Un grazie agli amici di sempre per tutto e di più: Giò, la sua tromba e la sua enorme bontà. Fabio, il suo essere in ritardo e il suo modo di fare l'uomo di *business*. Simone, la sua vaghezza e la sua innata simpatia. Roberto, la sua capacità di scomparire e la sua continua ricerca del senso delle cose.

Un caro saluto e un abbraccio vanno a Rita che sicuramente sarebbe stata contenta con me per questo traguardo/partenza che è la laurea.

Ringrazio la mia famiglia per avermi portato fin qui. Mamma e papà per i consigli durante questi cinque anni e l'interesse e la preoccupazione dimostrata per gli ultimi sviluppi della tesi e di quello che verrà. Geppo perchè uno così è troppo simpatico. La nonna Terry per le fresche spremute dopo gli esami.

Grazie a Valeria per aver sopportato i miei sproloqui su integrali che non venivano, algoritmi che non convergevano e derivate sbagliate. Grazie per il lavoro di rilettura e individuazione degli errori ma sopra tutto grazie per tutto il supporto morale, la presenza e la sicurezza che ogni giorno non mi ha fatto mancare.

Gli ultimi tre grazie, i più importanti, vanno a Dio, al Rock e alla birra, per il dono della vita, per l'attitudine a viverla al meglio e per le pause riflessive che ogni tanto bisogna prendersi.

*It's a long way to the top
if you wanna rock 'n' roll*