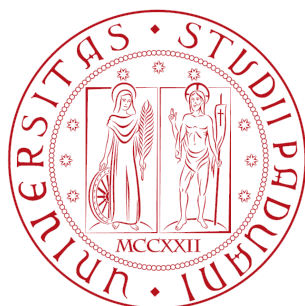


UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE
CORSO DI LAUREA MAGISTRALE
IN SCIENZE STATISTICHE



TEST DIREZIONALE PER MODELLI GRAFICI GAUSSIANI ANNIDATI

RELATORE: Prof. Nicola Sartori
Dipartimento di Scienze Statistiche

LAUREANDO: Novello Andrea
MATRICOLA N° 1035947

ANNO ACCADEMICO 2013/2014

Indice

Introduzione	5
1 I modelli grafici	7
1.1 Introduzione	7
1.2 Funzione di verosimiglianza	7
1.3 Grafi di indipendenza	9
1.4 I modelli grafici parametrici	10
1.4.1 Modelli grafici Gaussiani	11
1.4.2 La distribuzione marginale di S	12
1.5 Il test rapporto di verosimiglianza	13
2 Due test alternativi basati sulla verosimiglianza	15
2.1 Introduzione	15
2.2 Log-rapporto di verosimiglianza modificato	16
2.3 Test direzionali per famiglie esponenziali	17
2.4 Test Skovgaard per famiglie esponenziali	19
2.5 Approccio direzionale per modelli grafici gaussiani	20
3 Confronto di modelli grafici annidati	23
3.1 Introduzione	23
3.2 Il modello non saturo	23
3.3 Le matrici di informazione	25
3.4 Test direzionale per il confronto di modelli annidati	27
3.4.1 Test di Skovgaard per il confronto di modelli annidati	29
3.5 Analisi di due insiemi di dati reali	30

3.5.1	Introduzione	30
3.5.2	Analisi dei dati <i>marks</i>	30
3.5.3	Analisi dei dati <i>cow</i>	31
3.6	Studi di simulazione	34
3.6.1	Introduzione	34
3.6.2	Studi di simulazione per i dati <i>marks</i>	34
3.6.3	Studi di simulazione per i dati <i>cow</i>	37
	Conclusioni	53

Introduzione

Può capitare, durante lo studio di un fenomeno, di essere interessati a cogliere le relazioni che legano tra loro un certo numero di variabili casuali relative a fenomeni reali. Si pensi ad esempio ad uno studio di natura sociale che rileva per ogni comune italiano i tassi di disoccupazione, istruzione, volontariato e criminalità. Può essere interessante capire quante e quali siano le relazioni di indipendenza tra queste variabili. Inoltre, può essere opportuno verificare l'indipendenza tra due variabili al netto dell'informazione ricavata dalle restanti variabili in esame. A tal fine quindi, in questa tesi, definiremo un modello grafico attraverso la nozione di 'grafo di indipendenza' (condizionata). Inizialmente confronteremo due modelli grafici gaussiani annidati per verificare la plausibilità del modello annidato attraverso il test di log-rapporto di verosimiglianza.

Illustreremo poi due test alternativi più accurati del test di log-verosimiglianza, soprattutto nel caso in cui il numero di parametri del modello è elevato e la numerosità campionaria esigua. Il primo test è denominato test di log-verosimiglianza aggiustato di Skovgaard (1988) mentre il secondo è chiamato test direzionale. Definiremo quindi i due test, applicandoli al caso in cui il modello iniziale presenti già delle indipendenze prese per assunzione. Analizzeremo infine due insiemi di dati provando l'accuratezza dei test in esame tramite simulazione.

Nello specifico, il Capitolo 1 inizialmente sarà dedicato a richiamare i concetti basilari dell'inferenza di verosimiglianza. In seguito definiremo un modello grafico gaussiano attraverso la densità marginale di S . Presenteremo alla fine il test rapporto di verosimiglianza.

Nel Capitolo 2 verranno proposte due alternative al test di log-verosimiglianza:

il test di log-verosimiglianza aggiustato di Skovgaard e il test direzionale. Definiremo poi i due test nel caso dei modelli grafici gaussiani senza vincoli iniziali.

Nel Capitolo 3 si confronteranno i due test proposti con il test di log-rapporto di verosimiglianza. In particolare si considererà il caso in cui il modello di partenza presenti già dei vincoli posti per assunzione, ridefinendo il test direzionale per questo modello annidato. Successivamente verranno analizzati due insiemi di dati reali utilizzando il test di log-rapporto di verosimiglianza, il test direzionale e il test w^* di Skovgaard. Verranno quindi verificate le bontà dei tre test via simulazione.

Capitolo 1

I modelli grafici

1.1 Introduzione

In questo capitolo passeremo inizialmente in rassegna le nozioni base dell'inferenza statistica parametrica basata sulla verosimiglianza (Azzalini, 2001). Presenteremo poi i modelli grafici associandoli ai grafi di indipendenza. Ci interesserà trattare unicamente i modelli grafici gaussiani (Dempster, 1972) dato che saranno quelli che utilizzeremo nei capitoli successivi. Infine, illustreremo il test di log-rapporto di verosimiglianza che sarà confrontato con le proposte alternative nel capitolo successivo.

1.2 Funzione di verosimiglianza

Supponiamo di avere una matrice y $n \times p$ le cui righe sono realizzazioni indipendenti della stessa variabile casuale Y multivariata di ordine p . Supponiamo inoltre di non sapere quale sia la vera distribuzione F_0 di Y ma sappiamo che appartiene ad una famiglia parametrica \mathcal{F} particolare in cui i soli elementi ignoti sono i parametri da cui è indicizzata. Definiamo quindi

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

dove θ è il vettore dei parametri e Θ è lo spazio parametrico, cioè l'insieme dei valori che può assumere il vettore dei parametri. Per funzioni continue

possiamo anche scrivere

$$\mathcal{F} = \{f(y; \theta) : \theta \in \Theta \subseteq \mathbb{R}^p\}, \quad (1.1)$$

dove f è una qualsiasi funzione di densità. Questa formulazione riguarda la statistica parametrica in cui l'inferenza è sostanzialmente relativa agli ignoti parametri. Tornando al nostro modello F_0 abbiamo che esiste un particolare valore di θ : θ_0 , associato ad esso. Esistono diversi metodi per la stima del valore adeguato di θ_0 , nessuno di essi tuttavia ci assicura in generale di trovare il vero valore del parametro. Si cerca comunque di avvicinarsi il più possibile al valore corretto prendendo quello più plausibile in relazione ai dati osservati. Il metodo più utilizzato in ambito frequentista è il metodo della stima di massima verosimiglianza. Per un dato modello del tipo (1.1) una volta osservato il valore di y relativo alla funzione di densità $f(y, \theta)$, l'unico termine ignoto resta θ . Questa funzione ci dà la probabilità che a priori avevamo di osservare ciò che è stato effettivamente osservato.

Dovendo dare una preferenza tra due valori di θ sceglieremo quello che dà un maggior valore di $f(y, \theta)$, per il valore di y fissato. Riduciamo quindi il problema ad un semplice confronto tra $f(y, \theta')$ e $f(y, \theta'')$, del tipo $f(y, \theta')/f(y, \theta'')$. È evidente quindi che una costante moltiplicativa $c(y)$ non dipendente da θ si semplifica in un confronto di questo tipo.

Inoltre per un campione casuale semplice abbiamo che ciascuna delle singole osservazioni è indipendente e identicamente distribuita alle altre. La funzione di densità congiunta perciò è data dal semplice prodotto delle singole densità marginali associate a ciascuna osservazione.

Definiamo quindi la funzione di verosimiglianza $L(\theta)$ come

$$L(\theta) = L(\theta; y) = \prod_{i=1}^n c(y_i) f(y_i; \theta), \quad (1.2)$$

dove y_i sono le righe della matrice y definita in precedenza. È immediato capire che per trovare la stima del vero parametro θ_0 cercheremo il valore di θ che massimizza la (1.2). Ai fini inferenziali è comodo lavorare con la

trasformata logaritmica della (1.2)

$$l(\theta) = \sum_{i=1}^n \log f(y_i; \theta) \quad (1.3)$$

Dato che la funzione logaritmo è monotona crescente, massimizzare (1.2) è equivalente a massimizzare (1.3). Se (1.1) costituisce una famiglia esponenziale allora

$$l(\theta) = \varphi(\theta)^T u(y) - K(\theta), \quad (1.4)$$

dove $\varphi(\theta)$ è il parametro canonico, $u(y)$ è la statistica sufficiente minimale e $K(\theta)$ è funzione solo del parametro θ .

1.3 Grafi di indipendenza

Il concetto di grafo è utilizzato molto spesso in matematica e in informatica, oltre che in statistica. Un grafo è semplicemente una struttura formata da un insieme di nodi o vertici e di archi che collegano i nodi. Lauritzen (1996) spiega come un grafo possa essere descritto mediante appunto due insiemi, l'insieme dei vertici V e l'insieme degli archi E . Gli archi inoltre, possono essere diretti o non diretti. I primi sono rappresentati da una freccia e i secondi da un segmento che congiunge due nodi. Un esempio di grafo non diretto (senza archi diretti) è rappresentato in Figura 1.1.

Lauritzen nel suo lavoro spiega inoltre minuziosamente l'insieme delle proprietà strutturali di un grafo nelle sue varie forme. Possono essere utili le nozioni di *decomponibilità*, *riducibilità* e *collassabilità*. Non ci soffermeremo tuttavia sugli aspetti formali della teoria dei grafi ma focalizzeremo l'attenzione sul concetto generico di grafo di indipendenza.

Costruire un grafo di indipendenza (condizionata) partendo da un insieme di variabili casuali di numerosità p è in realtà molto semplice. Sia $Y = (Y_1, Y_2, \dots, Y_p)$ un vettore p -variato distribuito secondo una certa funzione di densità p -variata $f_Y(y)$ e sia G un grafo composto da un insieme V di vertici e un altro insieme E di archi. Ogni nodo appartenente a V rappresenta una delle p componenti di Y mentre l'assenza di un arco e in E che collega

due nodi v_1 e v_2 di V significa che le due variabili associate a v_1 e v_2 sono indipendenti condizionatamente alle variabili i cui nodi separano v_1 e v_2 .

Due variabili aleatorie R , Z sono condizionatamente indipendenti ad una terza variabile aleatoria W se la funzione di densità condizionata congiunta di R e Z rispetto a W può essere scritta come

$$f_{R,Z|W}(r, z) = f_{R|W}(r)f_{Z|W}(z).$$

Tramite un grafo è quindi possibile definire chiaramente la struttura di indipendenza (condizionata) che sorregge l'insieme delle p variabili di Y . Il numero massimo di elementi da cui può essere composto l'insieme E è $p(p-1)/2$.

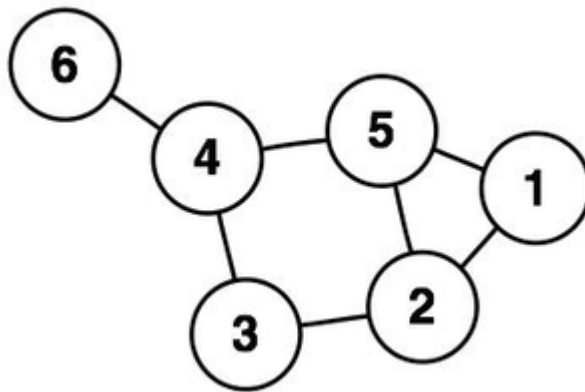


Figura 1.1: Esempio di grafo non diretto con 6 nodi e 7 archi

1.4 I modelli grafici parametrici

Un modello grafico in statistica è una famiglia di funzioni di densità di probabilità che incorpora un insieme specifico di vincoli di indipendenza condizionata visualizzabili in un grafo di indipendenza e quando si parla di “modellamento grafico” (graphical modelling) si intende scegliere un opportuno modello grafico per i dati. Possiamo classificare quindi la tecnica del *graphical modelling* come una delle tecniche statistiche parametriche. Il primo

passo da compiere quindi è associare all'insieme V dei vertici del grafo una funzione di densità o di ripartizione.

1.4.1 Modelli grafici gaussiani

Quelli che analizzeremo di seguito sono in realtà gli unici modelli grafici che interessano direttamente questa tesi. I modelli grafici gaussiani sono semplicemente dei modelli in cui le variabili casuali relative all'insieme dei vertici V hanno distribuzione normale multivariata, $Y \sim N_p(\mu, \Omega^{-1})$. In questo ambito la procedura per la verifica delle indipendenze viene anche chiamata procedura di selezione della covarianza (covariance selection procedure); il nome risulterà molto chiaro in seguito. L'assunzione più semplice per un insieme di variabili casuali continue è infatti che si distribuiscano congiuntamente come una normale multivariata.

Riconsideriamo quindi la matrice y e un grafo di indipendenza G . Le restrizioni associate al grafo G per il modello grafico gaussiano relativo alle righe di y si traducono nella presenza di uno o più zeri nella matrice di varianza inversa, indicata con Ω , in corrispondenza degli elementi ω_{ij} di Ω riferiti alle variabili che non presentano un arco in G che le collega (Wermuth, 1976). È di sostanziale interesse quindi procedere col fare inferenza su $\Omega = \Sigma^{-1}$. La densità di Y è quindi data da

$$f(y, \mu, \Omega) = \left[\frac{|\Omega|}{\sqrt{2\pi^p}} \right]^n \exp\left\{-\text{tr}\left[\frac{1}{2}(y - 1_n\mu^T)\Omega(y - 1_n\mu^T)^T\right]\right\},$$

dove μ rappresenta il vettore p -dimensionale delle medie mentre Ω è una matrice simmetrica e definita positiva, denominata matrice di concentrazione. Le stime di massima verosimiglianza per μ e Ω sono $\hat{\mu} = y^T 1_n/n$ e

$$\hat{\Omega} = \left\{ \frac{y^T y}{n} - \frac{y^T 1_n 1_n^T y}{n^2} \right\}^{-1}. \quad (1.5)$$

Il parametro μ tuttavia non è di sostanziale interesse per la nostra trattazione pertanto dovrebbe essere eliminato in qualche modo dal modello. Esistono diverse strade per eliminare μ ; potremmo fissarlo alla sua stima di massima

verosimiglianza e calcolare la verosimiglianza profilo. Tuttavia risulta piu' vantaggioso ricorrere alla distribuzione marginale di S .

1.4.2 La distribuzione marginale di S

Se nella stima di massima verosimiglianza sostituiamo le righe della matrice y con le variabili vasuali che le hanno generate otteniamo lo stimatore di massima verosimiglianza. Nello specifico, in accordo alla in (1.5) otteniamo che $S = \hat{\Omega}^{-1}n/(n-1)$ con $\hat{\Omega}^{-1}$ distribuito come

$$\hat{\Omega}^{-1} \sim W_p(n-1; \Omega^{-1}/n),$$

dove W_p è una variabile casuale Wishart di ordine p . Abbiamo quindi che la densità marginale di $X = \hat{\Omega}^{-1}$ è la seguente

$$f(x; \Omega^{-1}) = c^{-1} \left| \frac{\Omega^{-1}}{n} \right|^{-\frac{n-1}{2}} |x|^{\frac{n-1-p-1}{2}} \exp \left[-\frac{n}{2} \text{tr}(\Omega x) \right],$$

con $c = 2^{\frac{(n-1)p}{2}} (2\pi)^{\frac{p(p-1)}{4}} \prod_{i=1}^p \Gamma(\frac{n-i}{2})$. Eliminando le costanti moltiplicative otteniamo la funzione di verosimiglianza marginale

$$L(\Omega; x) = |\Omega|^{\frac{n-1}{2}} \exp \left[-\frac{n}{2} \text{tr}(\Omega x) \right],$$

con log-verosimiglianza

$$l(\Omega; S) = \frac{n-1}{2} \log |\Omega| - \frac{n-1}{2} \text{tr}(\Omega S). \quad (1.6)$$

Abbiamo quindi che S è la stima di massima verosimiglianza per Ω^{-1} relativa a (1.6).

1.5 Il test rapporto di verosimiglianza

Siamo interessati a valutare la plausibilità di una ipotesi riguardo al parametro θ . essa è definita

$$\begin{cases} H_0 : \theta_0 \in \Theta_0 \\ H_1 : \theta_1 \in \Theta_1 \end{cases},$$

dove H_0 definisce l'ipotesi nulla e H_1 l'ipotesi alternativa. L'ipotesi nulla è quella di cui stiamo verificando la plausibilità. Osserviamo che Θ_0 e Θ_1 costituiscono una partizione dello spazio parametrico Θ .

Definiamo inoltre test statistico come una funzione $T(y)$ dallo spazio campionario \mathcal{Y} alla partizione $\{\Theta_0, \Theta_1\}$. Il test statistico costituisce una partizione dello spazio campionario: da una parte avremo gli elementi di \mathcal{Y} tali che $T(y) = \Theta_0$, detta regione di accettazione, dall'altra i restanti elementi di \mathcal{Y} , ossia la regione di rifiuto. Si tratta in genere di trovare un valore critico T^* tale che se la funzione test è maggiore o minore di T^* , in relazione a come abbiamo definito il test, l'ipotesi nulla verrà rifiutata.

Il test statistico più utilizzato nella pratica in relazione alla funzione di verosimiglianza è il test rapporto di verosimiglianza. Definiamo la statistica test rapporto di verosimiglianza come

$$\lambda = \lambda(Y) = \frac{\sup_{\theta \in \Theta_0} L(\theta; Y)}{\sup_{\theta \in \Theta} L(\theta; Y)} = \frac{L(\hat{\theta}_0; Y)}{L(\hat{\theta}; Y)}.$$

È tuttavia più comune ai fini pratici considerare la trasformazione monotona

$$w(Y) = -2 \log \lambda(Y) = 2(l(\hat{\theta}) - l(\hat{\theta}_0)). \quad (1.7)$$

Si può mostrare che per $n \rightarrow \infty$ la variabile casuale $w(Y)$ riferita alla statistica test si distribuisce come un χ_d^2 dove d è il numero effettivo di parametri vincolati in θ_0 .

Nel caso di un modello grafico gaussiano con funzione di log-verosimiglianza

definita in (1.6) otteniamo che il test $w(Y)$ corrispondente è il seguente

$$w(Y) = -(n-1) \log |\tilde{\Omega} \hat{\Omega}^{-1}|, \quad (1.8)$$

dove $\tilde{\Omega}$ è lo stimatore di massima verosimiglianza vincolata di Ω sotto H_0 .

Il livello di significatività osservato, anche noto come p -value, nel caso del test rapporto di verosimiglianza si calcola come

$$\alpha_{oss} = \sup_{\theta \in \Theta_0} P\{w(Y) \geq w(y); \theta\}. \quad (1.9)$$

Avremo che per valori piccoli di α_{oss} rifiuteremo l'ipotesi nulla. Generalmente si rifiuta il test per $\alpha_{oss} < 0.05$. Il valore di soglia 0.05 tuttavia non è un vincolo restrittivo ma spesso una pura scelta arbitraria. Siccome nel nostro caso $w(Y)$ è distribuito asintoticamente secondo un χ_d^2 otteniamo che α_{oss} è approssimabile come complementamento della funzione di ripartizione di $w(Y)$ nel suo valore osservato $w(y)$.

Capitolo 2

Due test alternativi basati sulla verosimiglianza

2.1 Introduzione

Generalmente il test rapporto di verosimiglianza è il test più usato nella pratica per confrontare due modelli annidati in un contesto parametrico. Se la numerosità campionaria è molto elevata è possibile utilizzare l'approssimazione chi-quadrato per il test di ipotesi. Essa è tuttavia soggetta ad una bassa accuratezza, come per esempio in modelli grafici con molte variabili e poche osservazioni. È quindi utile poter disporre di approssimazioni più accurate del test in modo tale da avere p -value più attendibili. Di seguito introdurremo un approccio direzionale per il calcolo del p -value in un test di ipotesi per un generico parametro di interesse. I primi test direzionali sono stati proposti da Fraser e Massam (1985) e da Skovgaard (1988). L'approccio presentato nel seguito si riferisce a Davison *e altri* (2014) e consiste nel misurare il grado di allontanamento dall'ipotesi condizionatamente alla sua direzione. In questa tesi ci limiteremo a focalizzare l'attenzione ai modelli grafici gaussiani, che corrispondono ad una famiglia esponenziale in cui il parametro di interesse è funzione lineare del parametro canonico.

2.2 Log-rapporto di verosimiglianza modificato

Supponiamo ancora di avere un modello parametrico $f(y; \theta)$ dove y è una matrice $n \times p$ la cui i -esima riga rappresenta l' i -esima osservazione campionaria per le p variabili in esame e $\theta \in \mathcal{R}^p$ con $\hat{\theta} = \hat{\theta}(y)$ che massimizza la log-verosimiglianza $l(\theta; y) = \log f(y; \theta)$ (che semplifichiamo con $l(\theta)$). Indichiamo con y^0 il valore osservato della matrice y relativo ai dati la cui stima di massima verosimiglianza associata è $\hat{\theta}^0 = \hat{\theta}(y^0)$. Con $\psi(\theta)$ indichiamo invece il parametro di interesse d -dimensionale e verificheremo quindi l'ipotesi $H_\psi : \psi(\theta) = \psi$. Nel nostro caso avremo che $\theta = (\psi, \lambda)$ dove ψ è il parametro di interesse e λ parametro di disturbo. Indichiamo quindi con $\hat{\theta}_\psi$ lo stimatore di massima verosimiglianza vincolata di θ sotto H_ψ le cui componenti sono: $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$. Inizialmente potremmo approssimare $\hat{\theta}$ con una normale di media θ e matrice di varianza-covarianza stimata attraverso $j(\hat{\theta})^{-1}$ dove $j(\theta) = -\frac{\partial^2 l(\theta)}{\partial(\theta)\partial(\theta)^T}$ è la funzione di informazione di Fisher osservata. Riscriviamo quindi (1.7) in

$$w(\psi) = 2(l(\hat{\theta}) - l(\hat{\theta}_\psi))$$

Osserviamo però come il miglioramento da $O(n^{-1/2})$ a $O(n^{-1})$ nel passaggio da stimatore di massima verosimiglianza della distribuzione a test di log-rapporto di verosimiglianza è in qualche modo un artificio. Se $d = 1$ infatti il termine $O(n^{-1/2})$ dell'errore in ciascuna coda della distribuzione si annulla ma l'inferenza unilaterale non migliora. Skovgaard (2001) sostiene che la buona accuratezza dell'approssimazione r^* per l'inferenza di un parametro scalare di interesse sia dovuta sia all'errore relativo nell'approssimazione del test che alle proprietà nella regione di grandi deviazioni e propone un test per il caso vettoriale che si prospetta di mantenere l'accuratezza anche nelle code della distribuzione. La statistica proposta è una modificazione di $w(\psi)$, definita come

$$w^*(\psi) = w(\psi) \left(1 - \frac{\log \gamma(\psi)}{w(\psi)} \right)^2, \quad (2.1)$$

dove $\gamma(\psi)$ è un fattore di correzione. La statistica test è ancora distribuita approssimativamente come un χ_d^2 . Anche questa statistica offre una misura per

l'allontanamento dall'ipotesi nulla dove tutte le possibili direzioni rispetto ad essa sono mediate nel calcolo del p -value.

Al contrario quello che presenteremo è un test che misuri la distanza dall'ipotesi nulla incorporando l'informazione nei dati condizionandoci alla direzione rilevante da H_ψ .

2.3 Test direzionali per famiglie esponenziali

Assumiamo che il modello sia una famiglia esponenziale con parametro canonico $\varphi = \varphi(\theta)$ e con statistica sufficiente $u = u(y)$, quindi

$$f(y; \theta) = \exp[\varphi(\theta)^T u(y) - K\{\varphi(\theta)\}]h(y),$$

con funzione di log-verosimiglianza $l(\theta, y)$ definita da (1.4). Per convenienza centreremo inoltre la statistica sufficiente u nel suo punto osservato $u^0 = u(y^0)$; scriveremo quindi

$$l(\theta; s) = \varphi(\theta)^T s + l^0(\theta) = \varphi(\theta)^T (u - u^0) + l(\theta; u^0),$$

dove quindi $s = u - u^0$ e $l(\theta; s = 0) = l(\theta; u = u^0)$. Assumiamo inoltre che i parametri di interesse e di disturbo siano lineari in φ ; quindi $\varphi = \theta = (\psi, \lambda)$ e riscriveremo

$$l(\varphi; s) = \psi^T s_1 + \lambda^T s_2 + l^0(\psi, \lambda), \quad (2.2)$$

dove ψ e s_1 sono d -dimensionali e conseguentemente λ è q -dimensionale con $q = p - d$.

In questa struttura la distribuzione condizionata di s_1 dato s_2 dipende unicamente da ψ e ha a sua volta la forma di una famiglia esponenziale

$$f(s_1|s_2; \psi) = \exp(\psi^T s_1 - \tilde{K}(\psi))\tilde{h}_2(s_1). \quad (2.3)$$

Tuttavia sia \tilde{K} e \tilde{h}_2 raramente possono essere calcolati esplicitamente. Condizionarsi al valore osservato di s_2 è come fissare la stima di massima verosi-

miglianza sotto H_ψ che è $\hat{\varphi}_\psi = (\psi, \hat{\lambda}_\psi)$ e quando lo si calcola nella statistica sufficiente osservata u^0 si ottiene $\hat{\varphi}_\psi^0$.

L'approssimazione alla densità condizionata di $f(s_1|s_2; \psi)$ può essere espressa come

$$f(s; \psi)ds = c \exp[l(\hat{\varphi}_\psi^0; s) - l\{\hat{\varphi}(s); s\}] |j_{\varphi\varphi}(\hat{\varphi}(s); s)|^{-\frac{1}{2}} ds, s \in \mathcal{L}^0, \quad (2.4)$$

dove c è una costante di normalizzazione e \mathcal{L}^0 è lo spazio d -dimensionale dove giace s_1 una volta fissato $s_2 = 0$, oppure equivalentemente $\hat{\lambda}_\psi = \hat{\lambda}_\psi^0$. Inoltre la stima $\hat{\varphi}(s)$ è la soluzione in φ della equazione score: $s = -l_\varphi^0(\varphi) = -\frac{\partial l^0(\varphi)}{\partial \varphi}$ mentre la matrice di informazione osservata per φ si ottiene con $j_{\varphi\varphi}(\varphi; s) = -\frac{\partial^2 l(\varphi; s)}{\partial \varphi \partial \varphi^T}$.

La densità condizionata (2.4) può essere utilizzata per derivare un test di rapporto di verosimiglianza per il parametro ψ . Per ottenere p -value più accurati tuttavia noi ci condizioniamo alla direzione che si forma unendo s^0 e s_ψ nella retta che definiamo con \mathcal{L}^* . Notiamo che \mathcal{L}^* giace nel sottospazio \mathcal{L}^0 . Calcoliamo quindi un test unidimensionale misurando la grandezza di s ; tanto più grande sarà la distanza che separa il valore s^0 associato al campione e il valore s_ψ associato all'ipotesi nulla, tanto più saremo spinti a rifiutare l'ipotesi nulla. Abbiamo che s_ψ è quel valore di s per il quale $\varphi = \hat{\varphi}_\psi^0$ è la stima di massima verosimiglianza; dalla (2.2) otteniamo

$$s_\psi = -l_\varphi^0(\hat{\varphi}_\psi^0) = \begin{bmatrix} l_\psi^0(\hat{\varphi}_\psi^0) \\ 0 \end{bmatrix}, \quad (2.5)$$

dove il valore di s_ψ dipende da y^0 . Parametizziamo quindi \mathcal{L}^* con t , ottenendo

$$s(t) = s_\psi + t(s_0 - s_\psi) = (1 - t)s_\psi.$$

Le stime di massima verosimiglianza date da $\hat{\varphi}(s)$ variano quindi con $s(t)$. Con l'aumentare di t tracciano una curva nello spazio parametrico che passa attraverso la stima di massima verosimiglianza vincolata ($t = 0$) e attraverso la stima di massima verosimiglianza ($t = 1$). Riscriviamo quindi la (2.4)

riparametrizzata in t

$$h(t; \psi) = c \exp[l(\hat{\varphi}_\psi^0; s(t)) - l\{\hat{\varphi}(s(t)); s(t)\}] |j_{\varphi\varphi}[\hat{\varphi}\{s(t)\}; s(t)]|^{-\frac{1}{2}}. \quad (2.6)$$

Non è strettamente necessaria una riparametrizzazione del parametro di disturbo e utilizziamo $\hat{\varphi}_\psi = \operatorname{argsup}_{\psi(\varphi)=\psi} l^0(\varphi)$ per definire lo stimatore di massima verosimiglianza vincolata. Per misurare la divergenza dall'ipotesi nulla utilizzeremo quindi la distribuzione condizionata di $\|s\|$ dato il vettore $a = \frac{s}{\|s\|}$ calcolata attraverso un cambio di variabile da s a $(\|s\|, a)$. Lo Jacobiano della trasformazione è proporzionale a t^{d-1} dato che t rappresenta la distanza radiale e a il vettore dei $d - 1$ angoli. Otteniamo quindi

$$t^{d-1} h(t; \psi) \propto t^{d-1} \exp[l(\hat{\varphi}_\psi^0; s(t)) - l(\hat{\varphi}; s(t))] |j_{\varphi\varphi}(\hat{\varphi})|^{-\frac{1}{2}}.$$

Possiamo ora calcolare il p -value attraverso $h(t; \psi)$ semplicemente come la probabilità che t sia maggiore di 1 sotto H_ψ , condizionatamente al fatto di essere nella semiretta $s(t)$, $t \geq 0$. Otteniamo

$$p(\psi) = \frac{\int_1^{t_{max}} t^{d-1} h(t; \psi) dt}{\int_0^{t_{max}} t^{d-1} h(t; \psi) dt}, \quad (2.7)$$

dove $t = 0$ e $t = 1$ corrispondono rispettivamente a $s = s_\psi$ e $s = 0$. Il valore t_{max} è il massimo valore di t possibile per il quale la stima di massima verosimiglianza esiste se calcolata in $s(t)$. L'accuratezza teorica del test fa riferimento ad un errore relativo di $O(n^{-3/2})$ per valori continui in zone di deviazione locale. La funzione (2.7) è calcolata tramite un rapporto di due integrali scalari ottenuti tramite approssimazione numerica.

2.4 Test Skovgaard per famiglie esponenziali

Abbiamo lasciato in sospeso in precedenza il fattore $\gamma(\psi)$ che con la nostra nomenclatura per le variabili appartenenti a famiglie esponenziale si

ottiene tramite

$$\gamma(\psi) = \frac{\{(s - s_\psi)^T j_{\varphi\varphi}^{-1}(\hat{\varphi}_\psi)(s - s_\psi)\}^{d/2}}{w^{\frac{d}{2}-1}(\hat{\varphi} - \hat{\varphi}_\psi)^T(s - s_\psi)} \left[\frac{|j_{\varphi\varphi}(\hat{\varphi}_\psi)|}{|j_{\varphi\varphi}(\hat{\varphi})|} \right]^{\frac{1}{2}}. \quad (2.8)$$

Il calcolo del p-value si ottiene con $s = 0$ in corrispondenza a $y = y^0$. L'espressione (2.8) aggiusta la statistica w e consente alla statistica w^* di correggere il test w ovunque all'interno dello spazio parametrico.

2.5 Approccio direzionale per modelli grafici gaussiani

Torniamo quindi al caso di interesse, avremo ancora una volta che le osservazioni provengono da una normale $N_p(\mu, \Omega^{-1})$, dove μ e la matrice Ω sono sconosciuti e con quest'ultima definita positiva. Supponiamo quindi che y sia una matrice $n \times p$ in cui la i -esima riga è il vettore y_i^T . La funzione di log-verosimiglianza è stata definita in (1.6). Questo modello è saturo e la stima di massima verosimiglianza esiste se e solo se $\hat{\Omega}^{-1}$ è definita positiva. Se siamo intenzionati a verificare la possibilità che alcuni elementi al di fuori della diagonale siano 0 allora verificheremo la seguente ipotesi: $H_\psi : \psi = 0$. Sotto H_ψ la stima di massima verosimiglianza vincolata di θ è $\tilde{\theta} = \tilde{\Omega}$. In R è possibile ottenere la matrice $\tilde{\Omega}$ numericamente attraverso il pacchetto `gRim` con la funzione `cmod`. La statistica di log-rapporto di verosimiglianza è definita da (1.8). Il parametro canonico della famiglia esponenziale in esame è $\varphi = \theta = \Omega$ quindi

$$l(\varphi; Y) = l(\theta; Y).$$

Il valore atteso sotto H_ψ di s è s_ψ definito nella (2.5) e si ottiene da

$$s_\psi = -l_\Omega(\tilde{\Omega}) = \frac{n-1}{2}(\hat{\Omega}^{-1} - \tilde{\Omega}^{-1}).$$

Tramite $s(t) = (1-t)s_\psi$ otteniamo che

$$l(\varphi; t) = \frac{n-1}{2} \log |\Omega| - \frac{n-1}{2} \text{tr}\{\Omega[t\hat{\Omega}^{-1} + (1-t)\tilde{\Omega}^{-1}]\}. \quad (2.9)$$

Sostituendo Ω con $\hat{\Omega}(t)$ in (2.9) otteniamo:

$$l(\hat{\Omega}(t); t) = \frac{n-1}{2} \log |\hat{\Omega}(t)| - \frac{p(n-1)}{2}.$$

Analogamente sostituendo Ω con $\tilde{\Omega}$ otteniamo

$$l(\tilde{\Omega}; t) = \frac{n-1}{2} \log |\tilde{\Omega}| - \frac{p(n-1)}{2}$$

Possiamo quindi calcolare $g(t)$

$$g(t) \propto |\hat{\Omega}(t)|^{-\frac{n-1}{2}} |\tilde{\Omega}|^{\frac{n-1}{2}}.$$

La massimizzazione della verosimiglianza in $s(t)$ di questo modello porta a $\hat{\Omega}^{-1}(t) = \hat{\Sigma}(t)$ con $\hat{\Omega}(t) = t\hat{\Omega}^{-1} + (1-t)\tilde{\Omega}^{-1}$. L'ultimo valore di s lungo la linea $s(t)$ è $s(t_{max})$ dove t_{max} è calcolato numericamente aumentando gradualmente il valore di t fintanto che la matrice $\hat{\Omega}(t)$ risulta essere definita positiva. Definiamo quindi le componenti per il calcolo del p -value. Sapendo che, nel nostro caso

$$|j_{\varphi\varphi}[\hat{\varphi}\{s(t)\}; s(t)]|^{-\frac{1}{2}} \propto |\hat{\Omega}(t)|^{\frac{p+1}{2}},$$

otteniamo quindi che

$$h(t; \psi) \propto |\hat{\Omega}(t)|^{-\frac{(n-p-2)}{2}} \propto |t\hat{\Omega}^{-1} + (1-t)(\tilde{\Omega}^{-1})|^{\frac{n-p-2}{2}}.$$

Per quanto concerne il fattore γ nella (2.1), abbiamo

$$\gamma(\psi) = \frac{\frac{1}{2}[tr(\hat{\Omega}^{-1}\tilde{\Omega}\hat{\Omega}^{-1}\tilde{\Omega}) - p]^{\frac{d}{2}} |\Omega^{-1}\tilde{\Omega}|^{-\frac{(p+2)}{2}}}{\frac{1}{2}(tr(\hat{\Omega}^{-1}\tilde{\Omega}) - p)(-\log |\hat{\Omega}^{-1}\tilde{\Omega}|)^{\frac{d}{2}-1}}.$$

Capitolo 3

Confronto di modelli grafici annidati

3.1 Introduzione

In questo capitolo considereremo un modello grafico di partenza in cui alcuni elementi della matrice Ω sono nulli per assunzione. Calcoleremo quindi la matrice di informazione per il parametro di questo modello e adatteremo le formule del test direzionale. Useremo poi il test per confrontare modelli annidati in due insiemi di dati in cui il numero di variabili e numerosità campionaria differiscono. Infine verificheremo le prestazioni dei modelli via simulazione.

3.2 Il modello non saturo

Supponiamo che ci siano già delle indipendenze condizionate tra le variabili in esame. In altre parole assumiamo che la matrice Ω contenga già inizialmente degli elementi nulli al suo interno. Ciò a cui potremmo quindi essere interessati può essere riassunto dai due punti seguenti:

1. riscrivere la funzione di verosimiglianza associata al nuovo modello in esame

2. rieffettuare analogamente a quanto visto in precedenza il test di ipotesi volto a verificare la nullità di ulteriori elementi della matrice Ω partendo da una matrice con già degli elementi nulli.

Si può mostrare che il test log-rapporto di verosimiglianza per questo caso resta invariato con $\hat{\Omega}$ in (1.8) che diventa la stima di Ω basata sulle assunzioni del modello. Una strada per affrontare il problema ci è suggerita da Roverato e Whittaker (1996). Ciò che si può fare in questo caso è inizialmente vettorizzare la matrice Ω . La matrice è quadrata $p \times p$ quindi una vettorizzazione completa produrrebbe un vettore con p^2 elementi. A noi tuttavia interessano solamente gli elementi della matrice triangolare superiore (o equivalentemente inferiore) corrispondente ad Ω . Infatti, dato che la matrice è simmetrica non ci interessa avere parametri duplicati. Otteniamo quindi al massimo un totale di $p(p+1)/2$ parametri. La funzione di verosimiglianza relativa alla singola osservazione è

$$l(\omega; y) = \frac{1}{2} \log |\Omega| - \frac{1}{2} \text{tr}(\Omega yy^T) = \frac{1}{2} \log |\Omega| - \frac{1}{2} \omega^T J s,$$

dove ω ed s sono rispettivamente i vettori ottenuti dai $p(p+1)/2$ elementi distinti di Ω e yy^T , mentre J è la matrice diagonale formata dal valore 1 in corrispondenza degli elementi ω_{ii} e dal valore 2 in corrispondenza agli elementi di ω della forma ω_{ij} con $i \neq j$. Il vantaggio della vettorizzazione del parametro Ω è il fatto che ω e s possono essere ordinati in qualsiasi modo, purchè l'ordinamento che si utilizza sia lo stesso per entrambi i vettori. Utilizziamo quindi il seguente ordinamento

$$k = \{(i, j) : \omega_{ij} \neq 0, i \leq j\}.$$

$$h = \{(i, j) : \omega_{ij} = 0, i < j\}.$$

Possiamo ordinare arbitrariamente k e h ottenendo

$$k = \{k_1, k_2, \dots, k_v\}.$$

$$h = \{h_1, h_2, \dots, h_v\}.$$

Riordiniamo quindi gli elementi corrispettivamente a k ed h ottenendo

$$\omega = \begin{pmatrix} \omega_k \\ \omega_h \end{pmatrix} \quad s = \begin{pmatrix} s_k \\ s_h \end{pmatrix} \quad J_{kk} = \begin{bmatrix} J_{kk} & 0 \\ 0 & J_{hh} \end{bmatrix}.$$

Basandoci sul fatto che $\omega_h = 0$, la funzione di log-verosimiglianza può quindi essere espressa come

$$l(\omega_k; y) = \frac{1}{2} \log |\Omega| - \frac{1}{2} \omega_k^T J_{kk} s_k.$$

La statistica *score* ottenuta derivando $l(\omega_k; y)$ rispetto a ω_k ha la forma seguente

$$U_k = \frac{1}{2} J_{kk} (\sigma_k - s_k),$$

dove σ_k è ottenuto da Σ nello stesso modo in cui ω_k è ottenuto da Ω .

Nel caso però in cui partiamo da (1.6) per una matrice y $n \times p$ le funzioni di log-verosimiglianza e score necessitano di essere ricalcolate in modo opportuno. Infatti, in questo caso, la funzione di log-verosimiglianza si definisce come

$$l(\omega_k; y) = \frac{(n-1)}{2} \log |\Omega| - \frac{(n-1)}{2} \omega_k^T J_{kk} s_k \quad (3.1)$$

e la *score* come

$$U_k = \frac{(n-1)}{2} J_{kk} (\sigma_k - s_k), \quad (3.2)$$

dove ora il vettore s è ottenuto dalla matrice S .

3.3 Le matrici di informazione

Isserliss (1918) ha mostrato che

$$\text{var}(s) = Iss(\Sigma),$$

dove $I_{ss}(\Sigma)$ è la matrice con elemento generico (u, v) , dove u determina (i, j) e v determina (r, s) , in accordo con gli insiemi k e h di pagina 24

$$\text{cov}(s_{ij}, s_{rs}) = \sigma_{ir}\sigma_{js} + \sigma_{is}\sigma_{jr}.$$

La stima di massima verosimiglianza di σ quando h è l'insieme vuoto (cioè il modello saturo) è s mentre nel modello vincolato, dove cioè h non è l'insieme vuoto, la stima di massima verosimiglianza di σ_k è s_k . Se partizioniamo $I_{ss}(\Sigma)$ in accordo con l'insieme k ed h otteniamo che

$$I_{ss}(\Sigma) = \begin{bmatrix} I_{ss}(\Sigma)_{kk} & I_{ss}(\Sigma)_{kh} \\ I_{ss}(\Sigma)_{hk} & I_{ss}(\Sigma)_{hh} \end{bmatrix}$$

e di conseguenza

$$\text{var}(\hat{\sigma}_k) = I_{ss}(\Sigma)_{kk}.$$

Per un campione casuale semplice di dimensione n otteniamo che asintoticamente

$$\hat{\sigma}_k \sim N\left(\sigma_k, \frac{1}{n-1} I_{ss}(\Sigma)_{kk}\right).$$

Per quanto concerne invece il parametro ω_k , sappiamo che in una famiglia esponenziale la matrice di informazione attesa e osservata coincidono e si può calcolare in due modi:

1. attraverso la varianza della funzione score:

$$I_{\omega_k} = \text{var}(U_k) \tag{3.3}$$

2. oppure derivando la funzione score:

$$I_{\omega_k} = -\frac{\partial U_k}{\partial \omega_k^T}$$

Cox e Wermuth (1990) mostrano che per il modello saturo asintoticamente si ha

$$\hat{\omega} \sim N\left(\omega, \frac{1}{n-1} I_{SS}(\Omega)\right),$$

perciò applicando la regola per l'inversione di una matrice partizionata, l'inversa della matrice di informazione per ω_k risulta essere

$$I_{\omega_k}^{-1} = \frac{1}{n-1} [I_{SS}(\Omega)_{kk} - I_{SS}(\Omega)_{kh} [I_{SS}(\Omega)_{hh}]^{-1} I_{SS}(\Omega)_{hk}]. \quad (3.4)$$

Dalle equazioni (3.2) e (3.3) otteniamo che

$$I_{\omega_k}^{-1} = \left(-\frac{1}{2(n-1)} J_{kk} \frac{\partial \sigma_k}{\partial \omega_k^T} \right)^{-1},$$

che coincide con

$$I_{\omega_k}^{-1} = -2 \left(\frac{\partial \omega_k}{\partial \sigma_k^T} \right) J_{kk}^{-1} \quad (3.5)$$

e

$$\frac{\partial \omega_k}{\partial \sigma_k^T} = -\frac{1}{2(n-1)} (I_{SS}(\Omega)_{kk} - I_{SS}(\Omega)_{kh} [I_{SS}(\Omega)_{hh}]^{-1} I_{SS}(\Omega)_{hk}) J_{kk}$$

che una volta sostituita in (3.5) dà (3.4), come ci aspettavamo.

3.4 Test direzionale per il confronto di modelli annidati

In questo paragrafo otterremo l'espressione del test direzionale per il confronto di due modelli grafici normali annidati, nel caso in cui il modello maggiore non sia il modello saturo.

Otteniamo innanzitutto che data la score (3.2), s_ψ risulta essere

$$s_\psi = -l_{\omega_k}(\tilde{\omega}_k) = \frac{(n-1)}{2} J_{kk} (s_k - \tilde{\sigma}_k),$$

dove $\tilde{\sigma}_k$ è la stima di σ_k sotto H_ψ mentre

$$s(t) = (1 - t)s_\psi.$$

A questo punto data la funzione di log-verosimiglianza di ω_k definita in (3.1), otteniamo che la funzione in t risulta essere

$$l(\omega_k, t) = \frac{(n-1)}{2} \log |\Omega| - \frac{(n-1)}{2} \omega_k J_{kk} (\tilde{\sigma}_k + t(s_k - \tilde{\sigma}_k)), \quad (3.6)$$

dalla quale ricaviamo che

$$J_{kk} \hat{\sigma}_k(t) = J_{kk} (\tilde{\sigma}_k + t(s_k - \tilde{\sigma}_k))$$

e dato che J_{kk} è una matrice invertibile, otteniamo che

$$\hat{\sigma}_k(t) = (\tilde{\sigma}_k + t(s_k - \tilde{\sigma}_k)).$$

Sapendo che $\hat{\omega}_k(t)$ è ottenibile da

$$\hat{\Omega}_k(t) = \left[\hat{\Omega}^{-1} + t(\hat{\Omega}^{-1} - \tilde{\Omega}^{-1}) \right]^{-1},$$

in analogia a quanto fatto con ω_k per la matrice Ω , sostituiamo $\hat{\omega}_k(t)$ nella (3.6) ottenendo

$$l(\hat{\omega}_k(t), t) = \frac{(n-1)}{2} \log |\hat{\Omega}_k(t)| - \frac{(n-1)}{2} \hat{\omega}_k(t) J_{kk} \hat{\sigma}_k(t)$$

e analogamente sostituendo $\tilde{\omega}_k$ alla stessa otteniamo

$$l(\tilde{\omega}_k, t) = \frac{(n-1)}{2} \log |\tilde{\Omega}| - \frac{(n-1)}{2} \tilde{\omega}_k J_{kk} \hat{\sigma}_k(t).$$

Possiamo quindi calcolare $g(t)$ come

$$g(t) = \left[\frac{|\tilde{\Omega}|}{|\hat{\Omega}_k(t)|} \right]^{\frac{(n-1)}{2}} + \exp \left[\frac{(n-1)}{2} (\hat{\omega}_k(t) - \tilde{\omega}_k) J_{kk} \hat{\sigma}_k(t) \right],$$

inoltre dato che $|\tilde{\Omega}|$ non dipende da t otteniamo

$$g(t) \propto |\hat{\Omega}_k(t)|^{-\frac{(n-1)}{2}} + \exp \left[\frac{(n-1)}{2} (\hat{\omega}_k(t) - \tilde{\omega}_k) J_{kk} \hat{\sigma}_k(t) \right].$$

Si verifica che anche $\exp\{(\hat{\omega}_k(t) - \tilde{\omega}_k) J_{kk} \hat{\sigma}_k(t)\}$ è una funzione costante rispetto a t . Perciò otteniamo che

$$g(t) \propto |\hat{\Omega}_k(t)|^{-\frac{(n-1)}{2}}.$$

Per quanto riguarda la funzione di informazione sappiamo dalla (3.4) che

$$|j_{\omega_k \omega_k}(\omega_k)|^{-\frac{1}{2}} = \left(\frac{1}{n-1} |I_{SS}(\Omega)_{kk} - I_{SS}(\Omega)_{kh} [I_{SS}(\Omega)_{hh}]^{-1} I_{SS}(\Omega)_{hk}| \right)^{\frac{1}{2}}. \quad (3.7)$$

Pertanto infine otteniamo che

$$h(\omega_k, t) = |\hat{\Omega}_k(t)|^{-\frac{(n-1)}{2}} \times \left(\frac{1}{n-1} |I_{SS}(\Omega)_{kk} - I_{SS}(\Omega)_{kh} [I_{SS}(\Omega)_{hh}]^{-1} I_{SS}(\Omega)_{hk}| \right)^{\frac{1}{2}}.$$

Per il calcolo del p -value definito in (2.7) tuttavia è necessario ricorrere a tecniche numeriche. Utilizzeremo perciò la funzione di **R integrate**.

3.4.1 Test di Skovgaard per il confronto di modelli annidati

Si può mostrare che, in questo caso, il fattore di correzione $\gamma(\psi)$ si ottiene da

$$\gamma(\psi) = \frac{\left\{ \frac{1}{4} [J_{kk}(\tilde{\sigma}_k - \hat{\sigma}_k)]^T [(j_{\omega_k \omega_k}(\tilde{\omega}_k))^{-1}] [J_{kk}(\tilde{\sigma}_k - \hat{\sigma}_k)] \right\}^{d/2}}{[-\log |\hat{\Omega}^{-1} \tilde{\Omega}|]^{d/2-1} \frac{1}{2} (\hat{\omega}_k - \tilde{\omega}_k)^T J_{kk}(\tilde{\sigma}_k - \hat{\sigma}_k)} \left\{ \frac{|j_{\omega_k \omega_k}(\tilde{\omega}_k)|}{|j_{\omega_k \omega_k}(\hat{\omega}_k)|} \right\}^{1/2},$$

dove $j_{\omega_k \omega_k}(\omega_k)$ è definita in (3.7).

3.5 Analisi di due insiemi di dati reali

3.5.1 Introduzione

Quelle che svilupperemo di seguito saranno le analisi di due insiemi di dati, i dati *marks* e i dati *cow*. Il primo insieme di dati è stato scelto perchè contiene poche variabili con un numero sufficientemente elevato di unità statistiche mentre il secondo insieme di dati contiene un numero più elevato di variabili con numerosità campionaria più bassa. Successivamente verificheremo l'accuratezza dell'inferenza basata sui test di log-rapporto di verosimiglianza, di Skovgaard e direzionale nei confronti di questi due insiemi di dati attraverso studi di simulazione.

3.5.2 Analisi dei dati *marks*

I dati *marks* sono presenti nel pacchetto R *ggm*. Sono relativi ad 88 osservazioni riguardo a voti conseguiti dagli studenti in 5 materie scolastiche: *algebra*, *mechanics*, *vector*, *analysis*, *statistics*. Il numero di variabili è quindi relativamente basso rispetto al numero di osservazioni. Ci aspettiamo quindi che il test di log-rapporto di verosimiglianza e direzionale diano risultati piuttosto simili.

Partiamo inizialmente da un modello annidato rispetto a quello saturo. Assumiamo cioè che ci sia indipendenza condizionata tra alcune delle variabili in esame. Il modello annidato da cui partiamo inizialmente è rappresentato attraverso il grafo di Figura 3.1. In altre parole stiamo assumendo che il grafo associato al modello di partenza non abbia più archi rispetto a quanto rappresentato. Ci chiediamo se sia plausibile che le variabili *mechanics* e *algebra* siano congiuntamente indipendenti dalla variabile *vectors* che resta perciò completamente isolata nel grafo di indipendenza. Sotto l'ipotesi nulla H_ψ avremo quindi che il modello è associato al grafo di Figura 3.2.

Il test log-rapporto di verosimiglianza definito in (1.8) ha un p -value che corrisponde 1.07×10^{-11} ; il test aggiustato di Skovgaard invece risulta avere un p -value uguale a 1.92×10^{-11} ; mentre per quanto concerne il test direzionale abbiamo che il p -value risulta essere: 1.91×10^{-11} . Il test log-rapporto

di verosimiglianza quindi sembra portare ad un p -value due volte inferiore rispetto agli altri test. Tuttavia siamo di fronte a p -value talmente bassi da portare fortemente a rifiutare l'ipotesi nulla per ciascun test.

Proviamo ora a ridurre la dimensione del grafo sotto l'ipotesi nulla. Ci aspettiamo che con l'aumento della dimensione del parametro ψ il test di log-verosimiglianza sia meno affidabile degli altri test. Ci troviamo quindi a valutare l'ipotesi che le variabili *mechanics* e *algebra* siano condizionatamente indipendenti congiuntamente alle ipotesi del caso precedente. Il modelli sotto l'ipotesi alternativa e nulla sono rappresentati tramite i grafi di Figura 3.3. I p -value per i tre test sono: TRV 1.58×10^{-17} , directional test 8.27×10^{-17} , Skovgaard test 9.09×10^{-17} . Notiamo ora che i p -value sono diminuiti drasticamente rispetto al caso precedente evidenziando ancora che che gli archi che stiamo provando ad eliminare non possono essere eliminati dal modello. Notiamo però che ora il test di log-verosimiglianza da un p -value circa sei volte inferiore rispetto agli altri due.

Cambiamo ora il modello iniziale introducendo un grafo con una forma diversa, proviamo quindi ad eliminare quattro archi dal grafo relativo al modello sotto H_1 ; abbiamo perciò che in questo caso $d = \dim(\psi) = 4$. Il modello sotto H_1 e sotto H_ψ sono rappresentati dai grafi di Figura 3.4. I p -value per i tre test in questo caso sono: TRV 4.43×10^{-17} , directional 2.02×10^{-16} , Skovgaard w^* 2.15×10^{-16} . Sembra ancora che il test di log-verosimiglianza dia un p -value di molto inferiore rispetto agli altri due. Anche in questo caso comunque siamo fortemente spinti a rifiutare l'ipotesi nulla con ciascun test.

3.5.3 Analisi dei dati *cow*

I dati *cow* sono relativi ad esperimenti compiuti in un insieme di mucche per il controllo dell'azione di un parassita che agisce nel loro intestino. Durante il periodo del pascolo, da primavera ad autunno, i bovini possono ingerire delle larve che si sono sviluppate da delle uova precedentemente depositate nel foraggio del quale si cibano. Una volta infettato, l'animale viene privato delle proprie sostanze nutritive e le proprie difese immunitarie

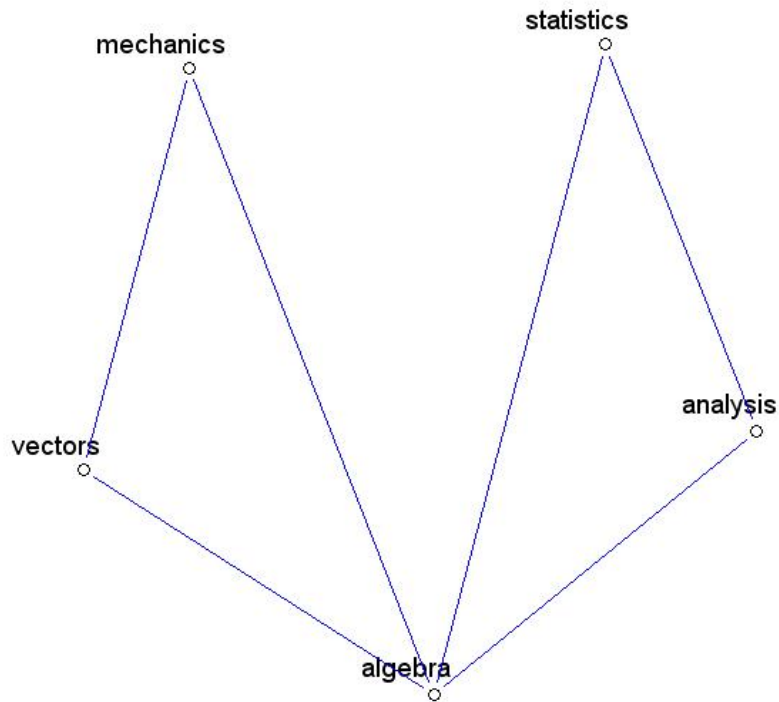


Figura 3.1: Grafo relativo al modello assunto di partenza per i dati *marks* con $\dim(\psi) = 2$, siamo quindi sotto H_1

si abbassano, lo rendono piu' incline alle malattie di altra natura e influiscono negativamente nella sua crescita. Il monitoraggio degli effetti dei trattamenti richiede che le osservazioni siano effettuate in tempistiche diverse durante il periodo del pascolo. In questi esperimenti i 60 animali sono stati assegnati casualmente ai due trattamenti, 30 per gruppo. Sono stati monitorati quindi i pesi degli animali in 11 momenti diversi, ognuno a distanza di 14 giorni dal precedente ad eccezione dell'undicesimo che è avvenuto 7 giorni dopo il decimo.

Vogliamo verificare che ci sia o meno una dipendenza markoviana di primo

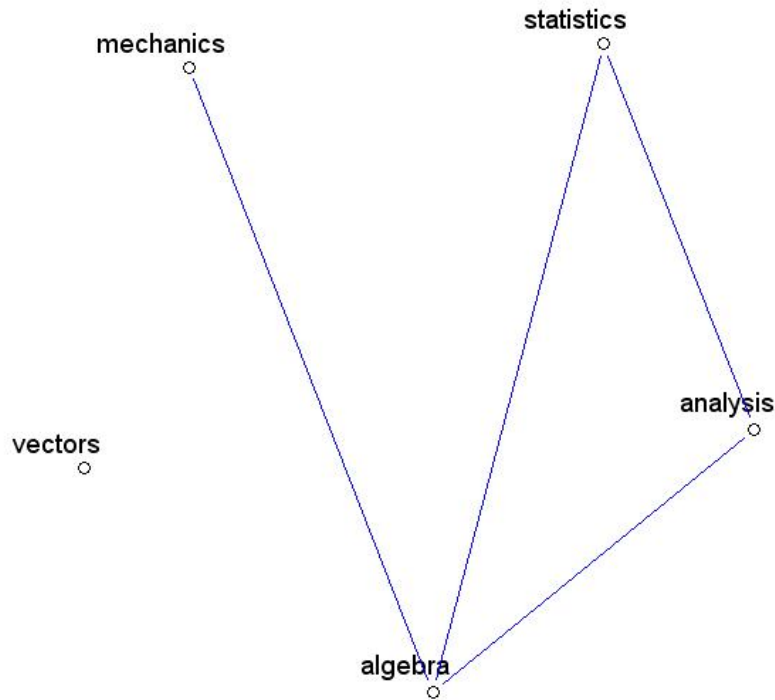


Figura 3.2: Grafo relativo al modello da verificare per i dati *marks* con $\dim(\psi) = 2$, siamo quindi sotto H_ψ

ordine tra i tempi dell'esperimento sapendo che la dipendenza è assunta come markoviana di ordine 2. Il grafo assunto e il grafo ipotetico sono riassunti in (3.5). I test di log-rapporto di verosimiglianza, direzionale e Skovgaard portano ad avere i seguenti p-value rispettivamente: 0.01, 0.014, 0.014. Il test di log-verosimiglianza ha un p -value inferiore rispetto agli altri due ma tutti ci spingono rifiutare l'ipotesi nulla. Non è quindi plausibile che la dipendenza sia markoviana di primo ordine.

3.6 Studi di simulazione

3.6.1 Introduzione

Sappiamo che sotto l'ipotesi nulla i p -value relativi ai test di nullità per i parametri di interesse dovrebbero essere distribuiti in modo uniforme nell'intervallo $(0, 1)$. Ciò significa che prendendo per esempio come soglia di rifiuto il 5% avremo che il 5% dei p -value staranno al di sotto della soglia nonostante sia vera H_ψ . Questo errore viene detto errore di primo tipo.

Ciò che accade in realtà però è che se il test è poco affidabile, come nel caso del test log-rapporto di verosimiglianza con n piccolo, dove cioè non è buona l'approssimazione asintotica, ci troveremo ad avere p -value che non si distribuiscono effettivamente in modo uniforme.

Nelle simulazioni noi partiremo stimando il modello sotto l'ipotesi nulla. Successivamente per ciascuna simulazione genereremo (tramite la funzione di R `rmvnorm` del pacchetto `mvtnorm`), un numero fissato m di osservazioni dal modello sotto H_ψ . Adatteremo quindi quel nuovo insieme di dati simulato ai due modelli in esame e calcoleremo i p -value utilizzando i test di log-rapporto di verosimiglianza, il w^* di Skovggard e il test direzionale; confronteremo poi i risultati. I modelli che useremo per le simulazioni saranno quelli delle analisi della sezione precedente. Lo scopo sarà valutare l'accuratezza dei test utilizzati nelle analisi. Mostriamo che il test direzionale migliora notevolmente l'approssimazione del test rapporto di verosimiglianza per i casi in cui n è troppo basso rispetto ad una dimensione del parametro di interesse che invece è relativamente grande. L'accuratezza sarà giudicata relativamente all'uniformità dei p -value, sia graficamente che attraverso opportune tabelle. Le stime dei modelli sono state calcolate utilizzando il pacchetto R `gRim`.

3.6.2 Studi di simulazione per i dati *marks*

Ritorniamo al primo modello in analisi per i dati *marks*. Abbiamo che il modello sotto H_1 è descritto dal grafo di Figura 3.1 mentre il modello sotto H_ψ è descritto dal grafo di Figura 3.2. Proviamo inizialmente a simulare dall'ipotesi nulla un campione di numerosità $m = 25$. Ci aspettiamo che,

data la bassa numerosità, i p -value del test di log-rapporto di verosimiglianza si discostano, seppur lievemente, dalla distribuzione ideale. I risultati della simulazione sono riportati in Figura 3.6 e in Tabella 3.1. Graficamente è facile osservare come il test di log-rapporto di verosimiglianza produca p -value piu' bassi rispetto alla distribuzione ideale. La Tabella mostra che questo avviene in particolare per le soglie piu' basse. Il test di Skovgaard e direzionale invece sono uniformemente piu' accurati. Aumentando la numerosità campionaria ($m = 50$) osserviamo (Figura 3.7 e Tabella 3.1) come le differenze tra i tre test siano già pressocchè nulle. Aumentando ulteriormente la numerosità campionaria osserviamo (Figura 3.8 e Tabella 3.1) come la situazione sia rimasta quasi invariata. Abbiamo perciò che l'approssimazione asintotica del test di log-verosimiglianza è davvero molto buona già con $m = 50$.

Nel secondo caso abbiamo che i modelli sotto H_1 e H_ψ sono descritti tramite i grafi di Figura 3.3. Simuliamo anche in questo caso utilizzando inizialmente $m = 25$. I risultati sono riportati in Figura 3.9 e in Tabella 3.2. Il peggioramento dell'approssimazione asintotica del test di log-verosimiglianza è lievemente peggiorata con l'aumento della dimensione del parametro ψ . Tuttavia è ancora sufficiente aumentare di poco la numerosità campionaria per far sì che le differenze tra i test siano quasi nulle. In Figura 3.10 e Figura 3.11 sono riportati i grafici per le simulazioni con $m = 50$ e $m = 100$.

Nel terzo caso abbiamo che i modelli sotto H_1 e H_ψ sono descritti tramite i grafi di Figura 3.4. Qui abbiamo che la dimensione del parametro di interesse è 4 e teoricamente il TRV dovrebbe essere peggiorato ulteriormente. Prendiamo ancora inizialmente $m = 25$. I risultati sono riportati in Figura 3.12 e in Tabella 3.3. Il peggioramento dell'approssimazione del test di log-verosimiglianza è ora osservabile piu' distintamente. È necessaria una numerosità campionaria molto elevata per migliorare l'accuratezza del TRV, mentre gli altri test offrono p -value attendibili per ciascuna delle scelte di m . In Figura 3.13 e Figura 3.14 sono riportati i grafici per le simulazioni con $m = 50$ e $m = 100$.

nominal	1,0	2,5	5,0	10,0	25,0	50,0	75,0	90,0	95,0	97,5	99,0
lrt	1,5	3,4	6,1	11,8	28,0	52,8	76,8	90,6	95,5	97,8	99,1
directional	0,9	2,5	4,9	9,6	24,9	50,0	75,2	89,8	95,0	97,6	99,0
Skovgaard W*	1,0	2,5	4,9	9,7	25,0	50,1	75,3	89,8	95,0	97,6	99,0
lrt	1,1	2,7	5,4	10,2	25,2	50,4	75,7	90,6	95,1	97,5	99,1
directional	1,1	2,6	5,3	9,8	24,7	50,0	75,5	90,5	95,1	97,5	99,0
Skovgaard W*	1,1	2,6	5,3	9,8	24,7	50,0	75,5	90,5	95,1	97,5	99,0
sim-error	0,1	0,2	0,2	0,3	0,4	0,5	0,4	0,3	0,2	0,2	0,1

Tabella 3.1: Risultati delle simulazioni per $m = 25, 150$ con grafi di Figura 3.1 e Figura 3.2 con R=10000 replicazioni

nominal (%)	1,0	2,5	5,0	10,0	25,0	50,0	75,0	90,0	95,0	97,5	99,0
lrt	1,4	3,4	6,7	12,3	27,8	52,8	76,9	91,2	95,4	97,7	98,9
directional	0,9	2,5	5,1	10,0	24,3	49,5	74,7	90,0	95,0	97,5	98,8
Skovgaard W*	0,9	2,5	5,1	10,0	24,4	49,3	74,3	89,6	94,9	97,2	98,7
lrt	1,1	2,8	5,6	11,1	26,4	51,5	76,1	90,5	95,2	97,7	99,0
directional	0,9	2,3	4,8	10,2	24,7	49,8	75,2	90,1	95,0	97,5	98,9
Skovgaard W*	1,0	2,3	4,9	10,1	24,7	49,8	74,9	90,0	95,0	97,3	98,9
lrt	1,1	2,7	5,2	10,4	26,0	51,3	75,5	90,3	95,4	97,8	99,0
directional	0,9	2,5	4,9	10,0	25,0	50,4	75,1	90,1	95,3	97,7	99,0
Skovgaard W*	0,9	2,5	4,9	9,9	25,1	50,1	75,0	90,1	95,3	97,7	99,0
sim-error	0,1	0,2	0,2	0,3	0,4	0,5	0,4	0,3	0,2	0,2	0,1

Tabella 3.2: Risultati delle simulazioni con i dati *marks* per $m = 25, 50, 100$ con grafi di Figura 3.3 con R=10000 replicazioni

nominal (%)	1,0	2,5	5,0	10,0	25,0	50,0	75,0	90,0	95,0	97,5	99,0
lrt	1,6	3,7	6,9	12,9	29,8	55,0	77,7	91,1	95,7	97,8	99,0
directional	1,0	2,7	5,4	10,0	26,1	50,7	75,4	89,9	95,1	97,6	98,9
Skovgaard W*	1,0	2,7	5,3	10,0	25,8	50,4	74,8	89,7	94,7	97,3	98,9
lrt	1,3	3,0	5,9	11,4	27,3	52,1	76,4	90,2	95,4	97,6	99,0
directional	1,1	2,6	4,9	10,2	25,3	50,2	75,2	89,8	95,0	97,5	99,0
Skovgaard W*	1,1	2,6	5,0	10,1	25,2	50,0	75,0	89,7	94,8	97,4	98,8
lrt	1,3	2,9	5,4	11,0	26,8	51,7	76,3	90,5	95,1	97,9	99,1
directional	1,2	2,6	5,0	10,4	26,0	50,8	75,8	90,3	95,0	97,9	99,1
Skovgaard W*	1,2	2,7	5,1	10,3	26,0	50,7	75,6	90,3	94,9	97,9	99,1
sim-error	0,1	0,2	0,2	0,3	0,4	0,5	0,4	0,3	0,2	0,2	0,1

Tabella 3.3: Risultati delle simulazioni con i dati *marks* per i grafi di Figura 3.4 con $m = 25, 50, 100$ con R=10000 replicazioni

3.6.3 Studi di simulazione per i dati *cow*

Prendiamo ancora in esame i modelli descritti dai grafi di Figura 3.5. Con una dimensione del parametro di interesse così grande ($\dim(\psi) = 9$) il test di log-rapporto di verosimiglianza è molto inaffidabile e produce p -value decisamente troppo bassi, per basse numerosità campionarie. Il test direzionale invece, così come il test w^* di Skovgaard, sembra davvero produrre p -value estremamente accurati nonostante la bassa numerosità campionaria (Figura 3.15 e Tabella 3.4). Vediamo dalle Figure 3.16 e 3.17 che nemmeno una numerosità più elevata basta per far sì che il test log-rapporto di verosimiglianza sia attendibile come il test direzionale.

nominal (%)	1,0	2,5	5,0	10,0	25,0	50,0	75,0	90,0	95,0	97,5	99,0
lrt	2,1	4,8	8,7	15,4	33,1	58,4	80,4	92,6	96,3	98,1	99,1
directional	1,0	2,5	5,0	10,0	25,0	50,1	74,7	89,8	95,0	97,5	98,9
Skovgaard W^*	1,0	2,6	5,1	10,1	24,9	50,0	74,3	89,6	94,8	97,4	98,8
lrt	1,6	3,4	6,7	12,4	28,8	54,6	78,2	91,3	95,6	97,8	99,3
directional	1,1	2,6	5,0	9,9	24,9	50,5	75,3	90,1	94,9	97,5	99,1
Skovgaard W^*	1,1	2,6	5,0	10,0	24,9	50,4	75,1	90,0	94,8	97,5	99,0
lrt	1,2	3,0	5,9	11,2	26,9	52,6	76,8	90,8	95,5	97,9	99,2
directional	1,1	2,6	5,2	10,1	25,1	50,7	75,6	90,2	95,2	97,7	99,1
Skovgaard W^*	1,1	2,6	5,2	10,1	25,1	50,6	75,5	90,1	95,1	97,7	99,1
sim-error	0,1	0,2	0,2	0,3	0,4	0,5	0,4	0,3	0,2	0,2	0,1

Tabella 3.4: Risultati delle simulazioni con i dati *cow* per $m = 25, 50, 100$ con grafi di Figura 3.5 con $R=10000$ replicazioni

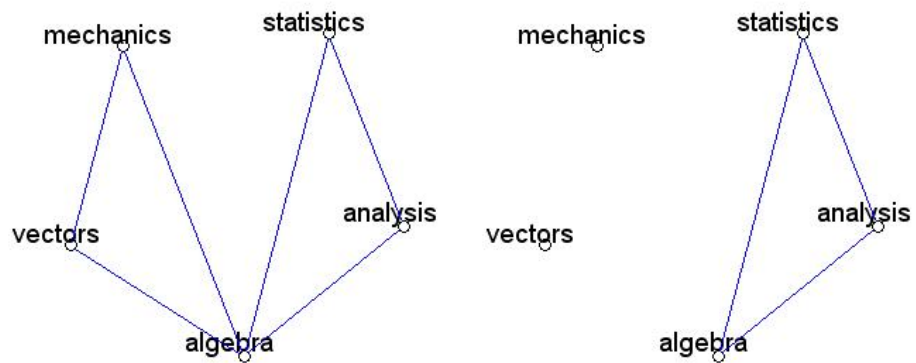


Figura 3.3: Grafi per dati *marks* con $\dim(\psi) = 3$, a sinistra abbiamo quello relativo al modello assunto (sotto H_1), a destra il modello da verificare (sotto H_ψ)

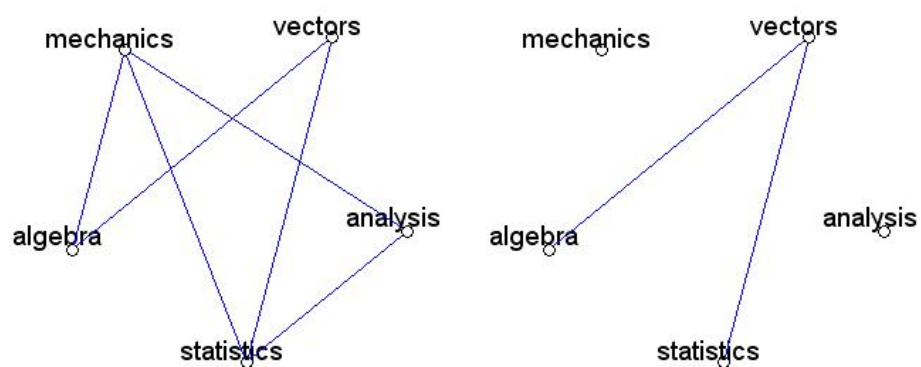


Figura 3.4: Grafi per dati *marks* con $\dim(\psi) = 4$, a sinistra abbiamo quello relativo al modello assunto (sotto H_1), a destra il modello da verificare (sotto H_ψ)

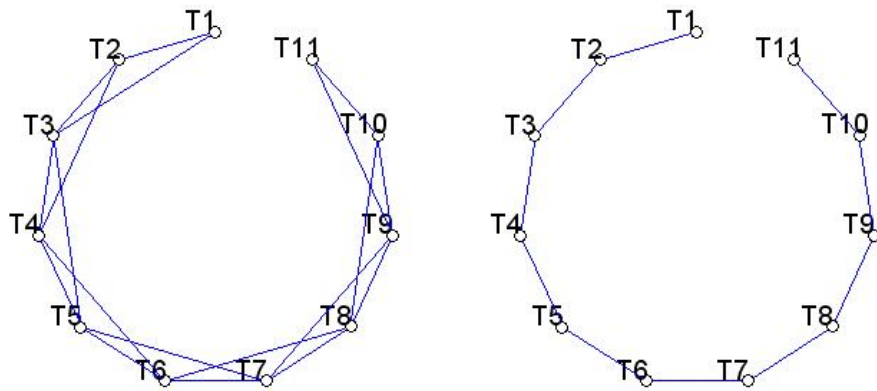


Figura 3.5: Grafi per dati *cow* con $\dim(\psi = 9)$, a sinistra abbiamo quello relativo al modello assunto (sotto H_1), a destra il modello da verificare (sotto H_ψ)

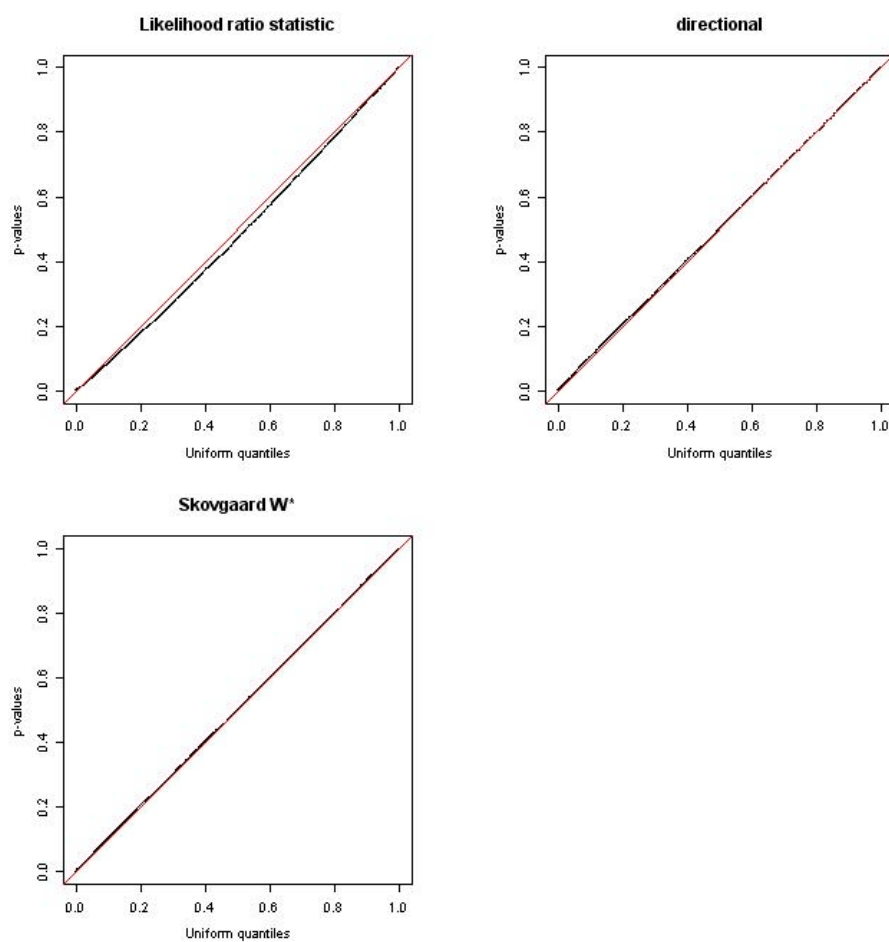


Figura 3.6: Risultati delle simulazioni con i dati *marks* per $m=25$ relativi ai grafi di Figura 3.1 e Figura 3.2 con $R=10000$ repliche

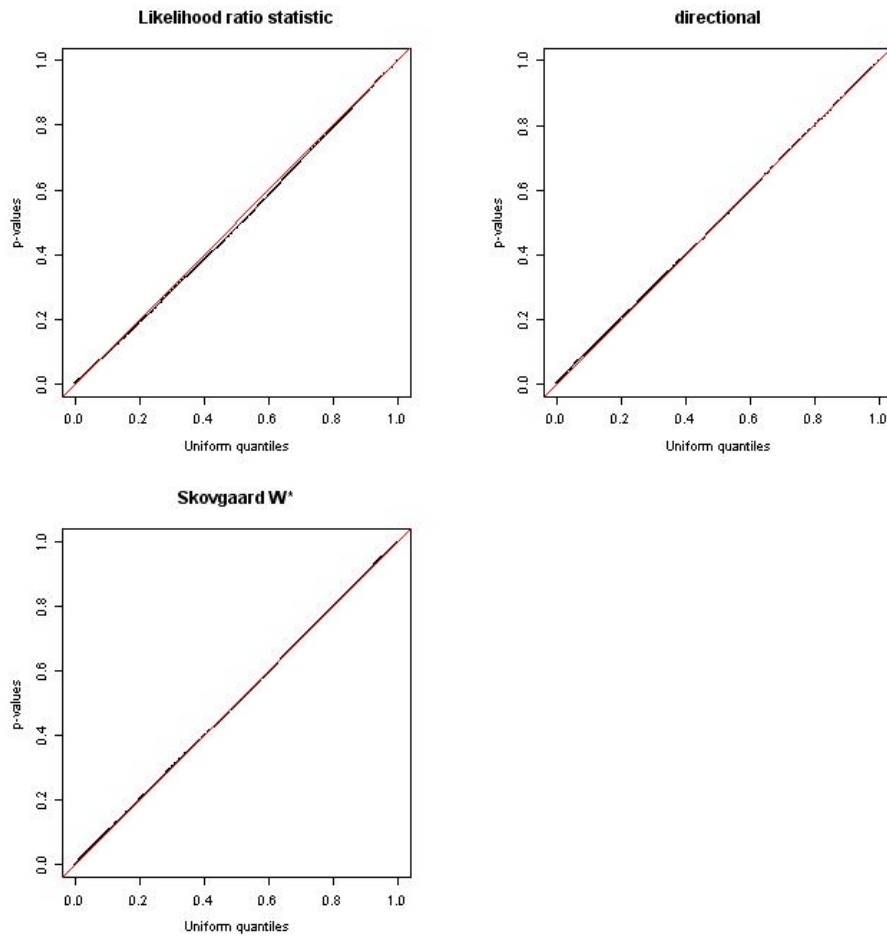


Figura 3.7: Risultati delle simulazioni con i dati *marks* per $m=50$ relativi ai grafi di Figura 3.1 e Figura 3.2 con $R=10000$ replicazioni

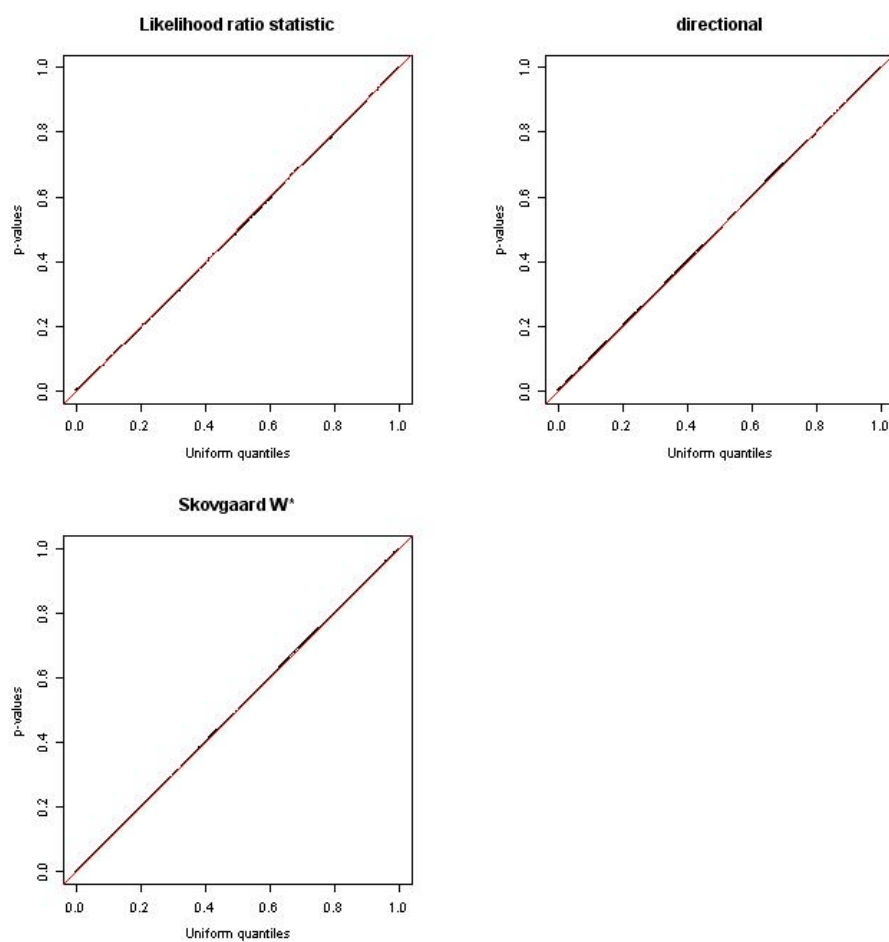


Figura 3.8: Risultati delle simulazioni con i dati *marks* per $m=150$ relative ai grafi di Figura 3.1 e Figura 3.2 con $R=10000$ replicazioni

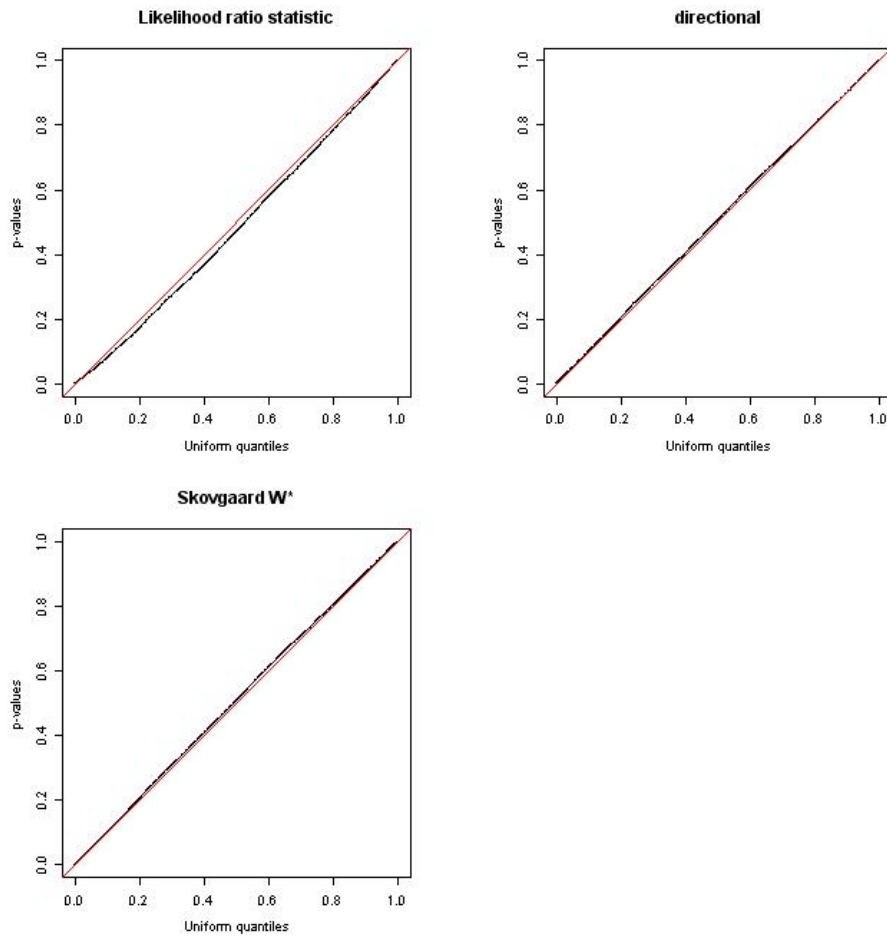


Figura 3.9: Risultati delle simulazioni con i dati *marks* per $m=25$ relativi ai grafi di Figura 3.3 con $R=10000$ replicazioni

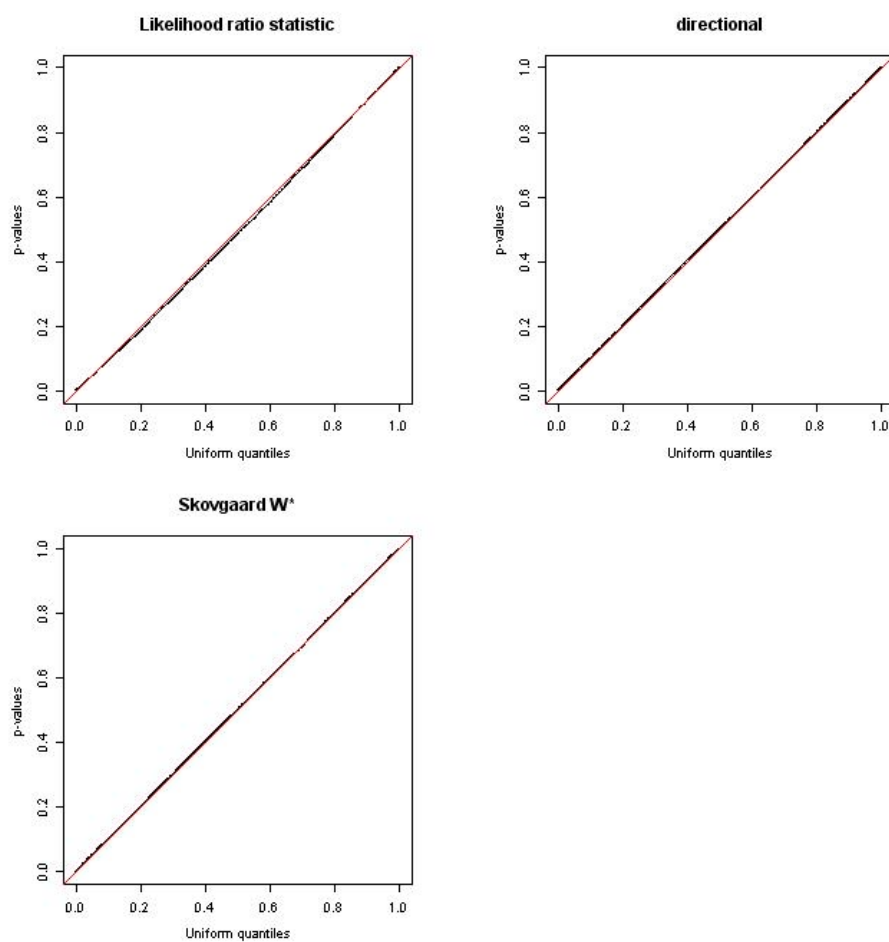


Figura 3.10: Risultati delle simulazioni con i dati *marks* per $m=50$ relativi ai grafi di Figura 3.3 con $R=10000$ replicazioni

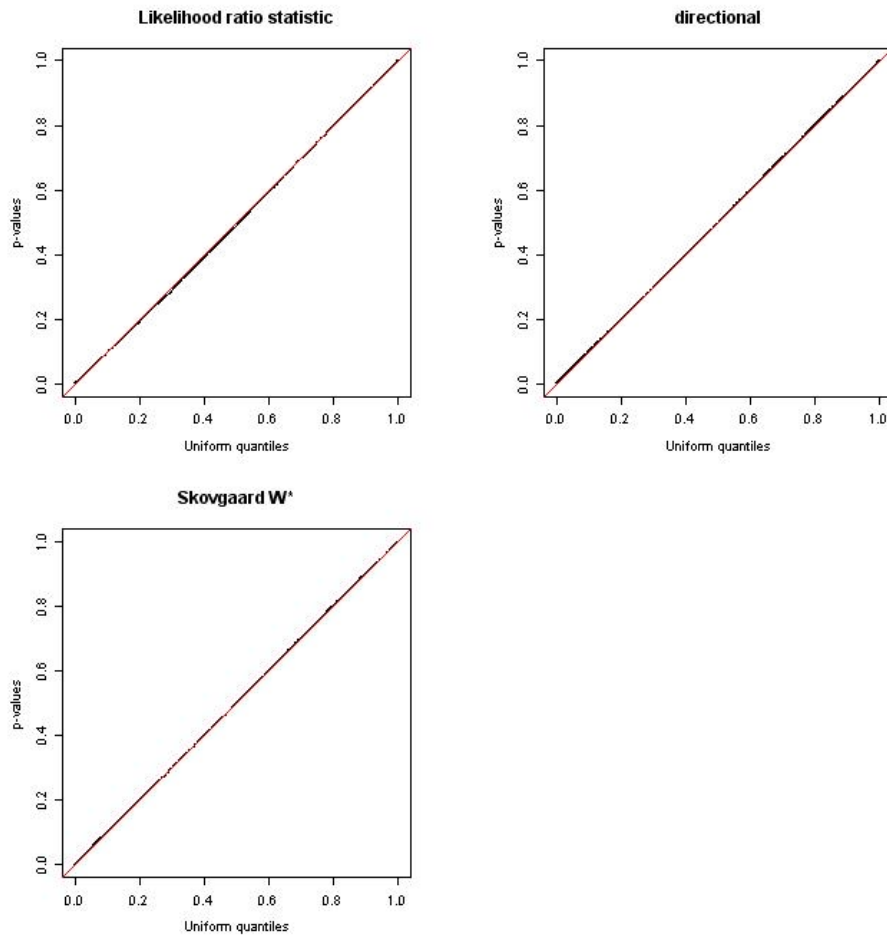


Figura 3.11: Risultati delle simulazioni con i dati *marks* per $m=100$ relativi ai grafi di Figura 3.3 con $R=10000$ replicazioni

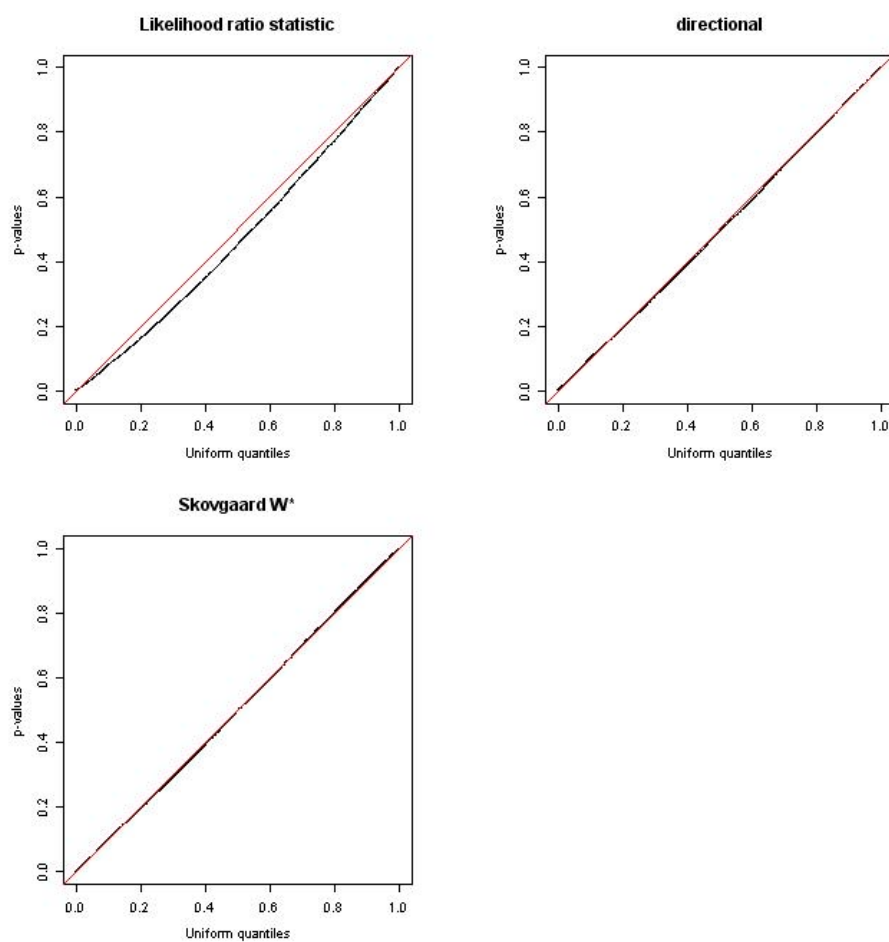


Figura 3.12: Risultati delle simulazioni con i dati *marks* per $m = 25$ relativi ai grafi di Figura 3.4 con $R=10000$ replicazioni

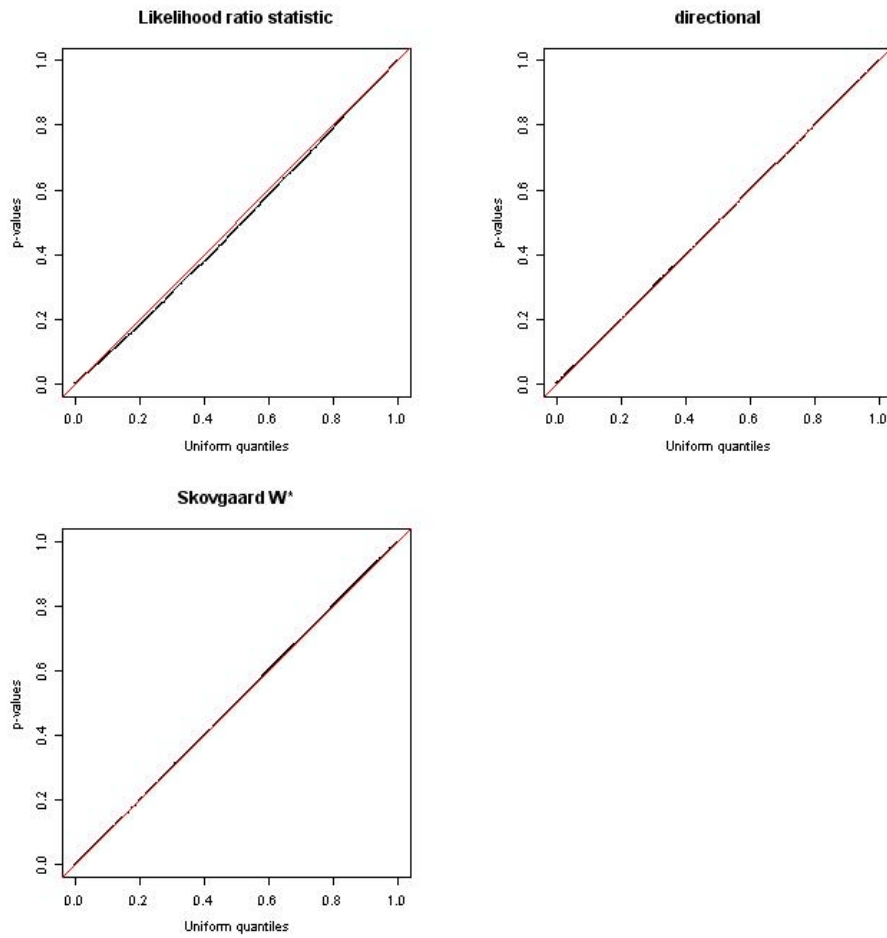


Figura 3.13: Risultati delle simulazioni con i dati *marks* per $m = 50$ relativi ai grafi di Figura 3.4 con $R=10000$ replicazioni

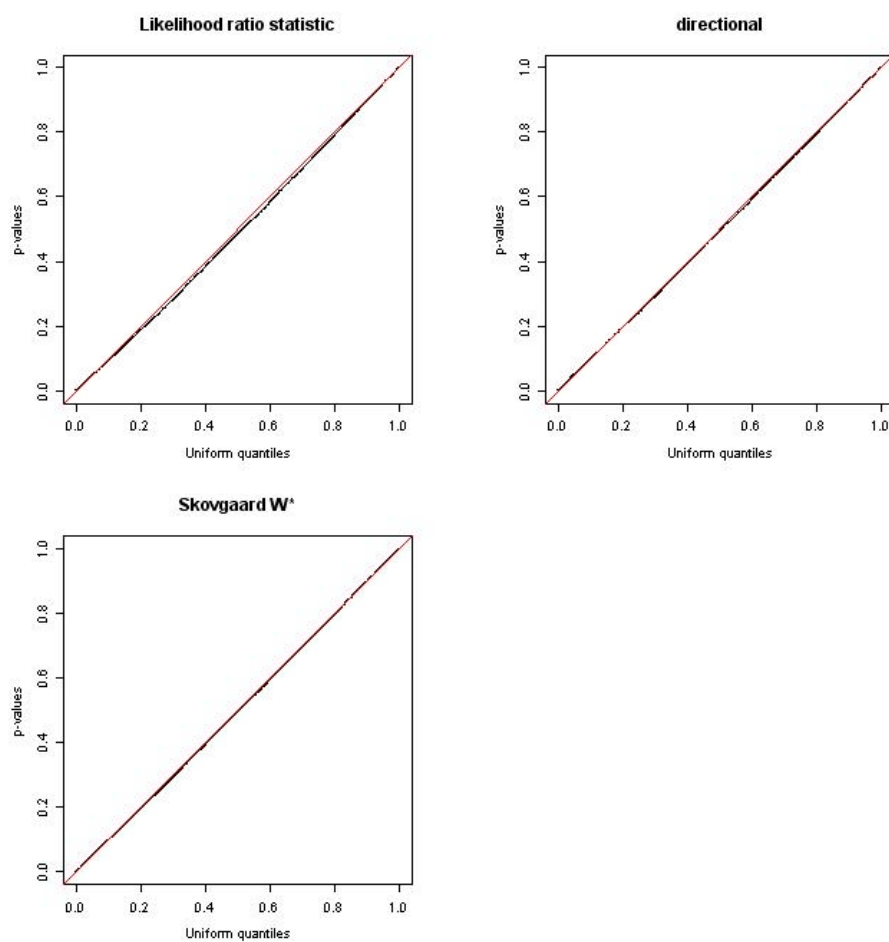


Figura 3.14: Risultati delle simulazioni con i dati *marks* per $m=100$ relativi ai grafi di Figura 3.4 con $R=10000$ replicazioni

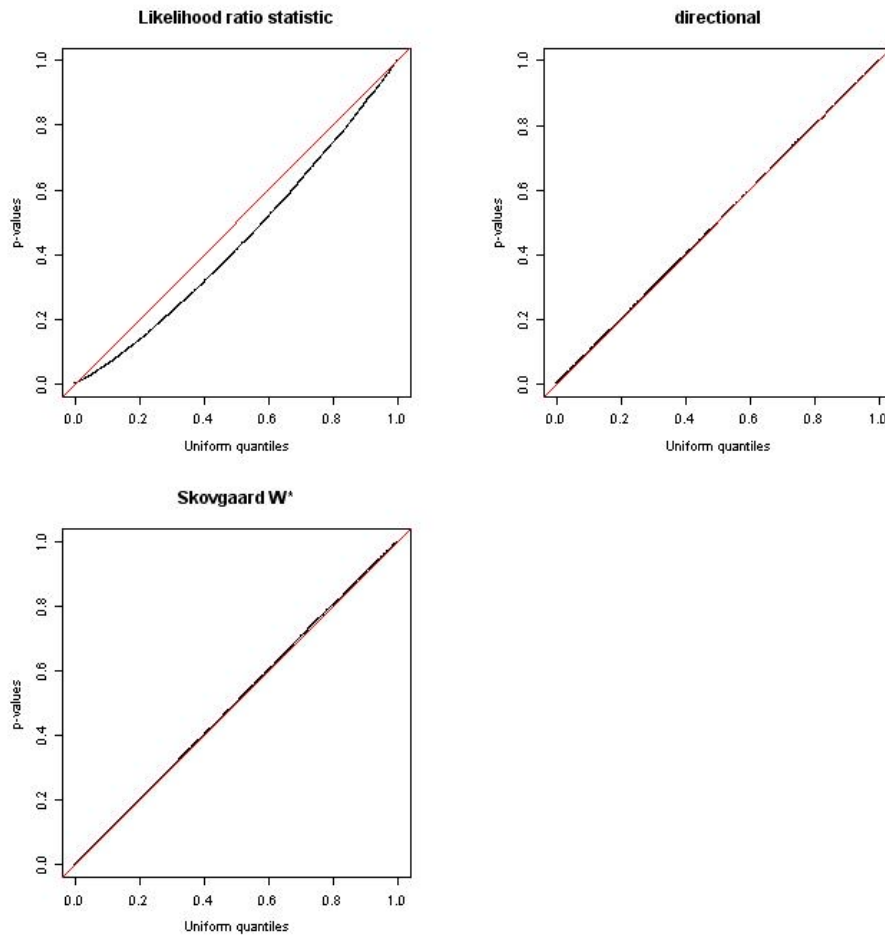


Figura 3.15: Risultati delle simulazioni per $m=25$ relative ai grafi di Figura 3.5 per i dati *cow* con $R=10000$ repliche

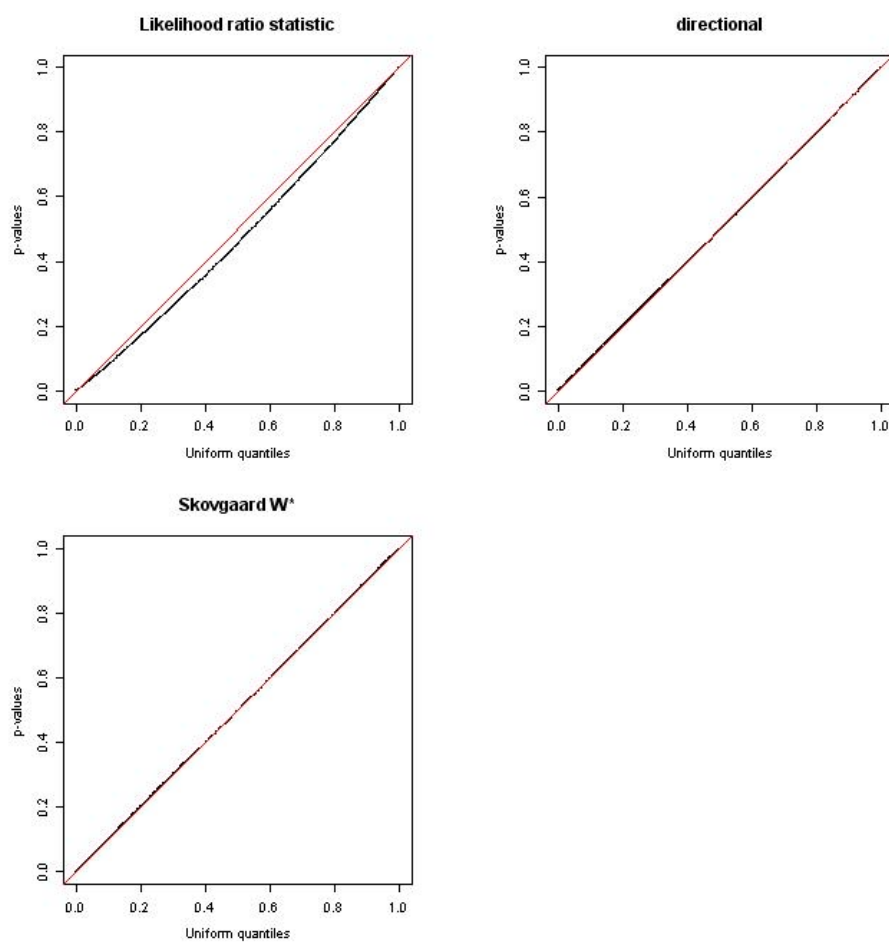


Figura 3.16: Risultati delle simulazioni per $m=50$ relative ai grafi Figura 3.5 per i dati *cow* con $R=10000$ replicazioni

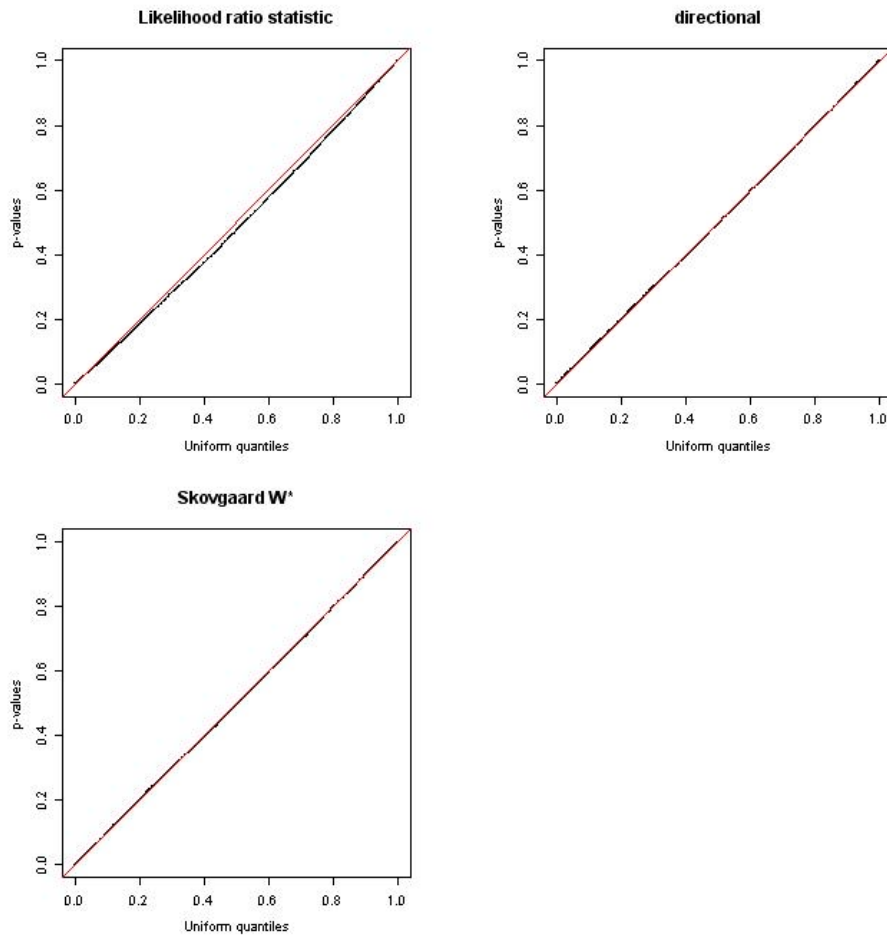


Figura 3.17: Risultati delle simulazioni per $m=100$ relative ai grafi Figura 3.5 per i dati *cow* con $R=10000$ repliche

Conclusioni

Nel corso di questa tesi abbiamo utilizzato diversi test basati sulla verosimiglianza per fare inferenza sulla matrice di concentrazione di un modello gaussiano. In tutti i casi abbiamo utilizzato la verosimiglianza marginale ottenuta dalla distribuzione marginale di S . L'utilità dei test di Skovgaard (2001) e direzionale è dovuta alla bassa accuratezza del test rapporto di verosimiglianza in corrispondenza di un gran numero di parametri vincolati sotto l'ipotesi nulla.

Abbiamo scelto di adottare il test direzionale per la verifica di ipotesi che un modello saturo abbia in realtà un insieme di indipendenze condizionate rappresentabili tramite un grafo di indipendenza. Successivamente il modello è stato riadattato per il caso in cui inizialmente siano già presenti degli elementi nulli nella matrice di concentrazione Ω .

Nel Capitolo 3 abbiamo analizzato due insiemi di dati per la verifica di ipotesi connessa ad un grafo di indipendenza annidato ad un primo grafo di partenza in cui abbiamo già delle indipendenze condizionate assunte. Più precisamente sono state compiute tre verifiche di ipotesi per il primo dataset e una singola verifica di ipotesi per il secondo insieme di dati. Per ciascuna ipotesi poi si è verificata la bontà dei test tramite simulazione variando di volta in volta la numerosità campionaria dei dataset simulati.

Infine, si è mostrato come il test Skovgaard e in particolare il test direzionale siano uniformemente più accurati sia per basse che per alte numerosità campionarie, in particolare quando il numero di parametri vincolati è alto. Ulteriori verifiche possono essere fatte con insiemi di dati con un gran numero di variabili e basse repliche campionarie. Un classico esempio è offerto dai dati di genomica, caratterizzati da un elevato numero di variabili rispetto

ad esigue numerosità campionarie. Ulteriori approfondimenti sui temi trattati sono consultabili su Davison *e altri* (2014), Wermuth (1976), Lauritzen (1996), Darroch *e altri* (1980), Wood (1993), Barndorff-Nielsen e Cox (1979), Butler (2007), Brazzale (2007), Brazzale *e altri* (2007) e Fraser e Reid (2006).

Bibliografia

- Azzalini A. (2001). *Inferenza statistica Una Presentazione Basata sul Concetto di Verosimiglianza*. Springer, Padova.
- Barndorff-Nielsen O. E.; Cox D. R. (1979). Edgeworth and saddlepoint approximation with statistical applications (with discussion). *J.R. Statist. Soc. B*, **41**, 279–312.
- Brazzale A. R.; Davinson A. C.; Reid N. (2007). *Applied asymptotics*. Cambridge: Cambridge University Press.
- Butler R. (2007). *Saddlepoint approximation with applications*. Cambridge: Cambridge University Press.
- Cox D.; Wermuth N. (1990). An approximation to maximum likelihood estimates in reduced models. *Biometrika*, **77**, 747–761.
- Darroch J. N.; Lauritzen S.; Speed T. P. (1980). Markov fields and log-linear interaction models for contingency table. *Annals of Statistics*.
- Davison A.; Fraser. D. A. S.; Reid N.; Sartori N. (2014). Accurate directional inference for vector parameters in linear exponential families. *Journal of the American Statistical Association*, **109**, 302–314.
- Dempster A. (1972). Covariance selection. *Biometrics*, **28**, 157–175.
- Fraser D. A. S.; Massam I. (1985). Conical tests: Observed levels of significance and confidence regions. *Statist. Hefte*, **26**, 1–17.

- Fraser D. A. S.; Reid N. (2006). Assessing a vector parameter. *Student*, **5**, 247–256.
- Isserliss L. (1918). On a formula for the product-moment correlation of any order of a normal frequency distribution in any number of variables. *Biometrika*, **12**, 134–139.
- Lauritzen S. L. (1996). *Graphical Models*. Oxford: Oxford University Press.
- Ross S. M. (2006). *Calcolo delle probabilità*. APOGEO.
- Roverato A.; Whittaker J. (1996). Standard errors for the parameters of graphical gaussian models. *Statistics and Computing*, **6**, 297–302.
- Skovgaard I. M. (1988). Saddlepoint expansions for directional test probabilities. *J. Roy. Statist. Soc.*, **50**, 3–32.
- Skovgaard I. M. (2001). Likelihood asymptotics. *Scand. J. Statist.*, **28**, 3–32.
- Wermuth (1976). Analogies between multiplicative models in contingency tables and covariance selections. *Biometrics*, **32**, 95–108.
- Wood (1993). Saddlepoint approximation to the cdf of some statistics with non-normal limit distributions. *J. Am. Statist. Assoc.*, **88**, 680–686.