

UNIVERSITÀ DEGLI STUDI DI PADOVA

CORSO DI LAUREA MAGISTRALE IN SCIENZE STATISTICHE



Indici non parametrici per famiglie
parametriche

con particolare riferimento alla t asimmetrica

Tesi di Laurea Magistrale

Relatore:

Chiar.mo Prof. Adelchi Azzalini
Dipartimento di Scienze Statistiche

Laureando:

Emanuele Giorgi

ANNO ACCADEMICO 2011-2012

*A tutti coloro che
mi hanno amato
e mi amano.*

Indice

1	Introduzione	3
1.1	Distribuzioni con code pesanti	4
1.2	Obiettivo della trattazione	7
2	L'indice di posizione	11
2.1	Le funzioni <i>depth</i> : definizione e proprietà	11
2.2	Selezione dell'indice di posizione	13
2.2.1	Concetti di simmetria multidimensionale	14
2.2.2	Mediane multidimensionali	15
2.3	La mediana di Tukey (o dei semispazi)	18
2.3.1	Concetti di base	18
2.3.2	Aspetti campionari dello stimatore non parametrico della funzione HS	21
2.3.3	Funzione HS e distribuzione ST	25
2.4	La varianza dello stimatore quantile	39
3	Misure matriciali di concordanza	41
3.1	Indici di concordanza	41
3.2	Matrici di concordanza	53
4	L'indice di asimmetria	57
4.1	Concetti di asimmetria unidimensionale	57
4.2	L'indice di Yule-Bowley	63
5	L'indice di curtosi	67
5.1	Concetti di curtosi e misure basate sui quantili	67
5.1.1	Caso unidimensionale	67
5.1.2	Estensione al caso multidimensionale	71

5.2	L'indice ϕ_F nel caso della funzione <i>depth</i> HS	74
5.2.1	Un'approssimazione di C_p per $d = 2$	74
5.2.2	Una famiglia di ordinamenti di curtosi	77
5.2.3	Il caso della SC	80
6	Considerazioni finali	87
6.1	Una sintesi	87
6.2	Applicazione: dati sui rendimenti finanziari	91
6.3	Software: le principali routine	94
A	Codice R	101
	Bibliografia	123

Capitolo 1

Introduzione

In statistica la distribuzione di probabilità di una variabile casuale X viene generalmente sintetizzata attraverso quattro fondamentali caratteristiche: la posizione, la variabilità, l'asimmetria e la curtosi. Questi attributi distributivi sono usualmente misurati attraverso indici basati (o in qualche modo associati) sui momenti della distribuzione. Abbiamo così la media, la varianza ed i momenti standardizzati del terzo e del quarto ordine. Se X è una variabile casuale d -dimensionale è necessario sintetizzare, oltre alle già menzionate caratteristiche distributive, anche la struttura di dipendenza per almeno ciascuna coppia delle componenti di X .

Nel caso multidimensionale gli indici più noti ed ampiamente impiegati per misurare la posizione e la struttura di dipendenza sono il vettore delle medie $E[X] = \mu$, ossia il vettore dei momenti del primo ordine di ciascuna delle componenti di X , e la matrice di covarianza $E[(X - \mu)(X - \mu)^T] = \Sigma$, in cui vengono coinvolti i momenti del primo e del secondo ordine di ciascuna componente del vettore casuale X . Nel caso dell'asimmetria e della curtosi multidimensionali, diversi indici sono stati proposti nel tentativo di generalizzare i momenti standardizzati unidimensionali del terzo e quarto ordine. Risultati importanti a tal proposito sono forniti in Mardia (1970, 1974), Srivastava (1984) e Malkovitch & Afifi (1973).

Diversi campi della scienza moderna, in virtù proprio della semplicità e dell'elevata trattabilità analitica del modello gaussiano, si basano pesantemente sull'assunzione che le variabili casuali sotto indagine seguano una distribuzione normale. Tuttavia le soluzioni a cui si giunge dipendono spesso, a volte in modo critico, dall'assunzione distributiva. L'impiego della distribuzione gaussiana diventa allora opinabile, in particolar modo quando i dati

mostrano “code” pesanti ed asimmetria.

A partire dalla seconda metà del secolo scorso l’attenzione si è spostata sempre di più verso distribuzioni di probabilità che esibiscono code più pesanti della distribuzione gaussiana. Tali distribuzioni costituiscono uno strumento fondamentale nella modellazione di eventi rari, in finanza, in idrologia, nella modellazione di dati navigazionali, in climatologia ed in numerosi altri campi della scienza.

Tuttavia la parametrizzazione di distribuzioni di probabilità con code pesanti è spesso costituita da un maggior numero di parametri rispetto al modello gaussiano, pertanto un maggior numero di momenti di ordine maggiore viene coinvolto al fine di sintetizzare la distribuzione attraverso la sua posizione, variabilità, asimmetria e curtosi. Il problema che viene dunque sollevato è l’impossibilità di descrivere i diversi attributi distributivi mediante i momenti, poiché questi nel caso di distribuzioni con code pesanti non sempre esistono.

1.1 Distribuzioni con code pesanti

Intuitivamente una distribuzione a code pesanti è una distribuzione per la quale le probabilità di eventi estremi tendono a zero molto lentamente. Tuttavia non si è a tutt’oggi giunti ad una definizione universalmente condivisa di distribuzione con code pesanti. Va detto inoltre che, nel tentativo di categorizzare distribuzioni a code pesanti, l’intento è quello di caratterizzare la forma della distribuzione, non la scala; pertanto una distribuzione con varianza maggiore rispetto ad un’altra non ha necessariamente code più pesanti.

Seppur non c’è unanimità nel stabilire quale sia la soglia tra una distribuzione con code pesanti ed una con code leggere, la maggior parte degli statistici è d’accordo nel considerare la distribuzione gaussiana come una a code leggere e la distribuzione t di Student come una a code pesanti se i gradi di libertà sono bassi. Mandelbrot (1963), ad esempio, considera effettivamente distribuzioni con code pesanti se e solo se la varianza risulta essere infinita. Mandelbrot (1963) e Fama (1963) discutono anche la frequente occorrenza di varianze infinite nell’analisi dei dati economici e l’inefficienza del metodo dei minimi quadrati. Gumbel (1962) considera distribuzioni a code leggere tutte quelle per le quali il massimo delle statistiche d’ordine ha una

distribuzione limite con funzione di ripartizione

$$F(x) = \exp[-\exp(-x)], \quad x \in \mathbb{R}$$

mentre se il massimo ha distribuzione limite con funzione di ripartizione

$$F(x) = \exp[-x^{-k}], \quad x > 0, \quad k > 0$$

allora la distribuzione è considerata avere code pesanti. Bryson (1974) considera distribuzioni unidimensionali con code pesanti tutte quelle distribuzioni per le quali la quantità $E[X - x | X > x]$ è crescente per valori grandi di x o non esiste. Mantel (1976) invece propone un criterio basato sul rapporto tra funzione di sopravvivenza $1 - F(x)$ e la funzione di densità $f(x)$: se tale rapporto diverge per $x \rightarrow \infty$ allora la distribuzione è considerata avere code pesanti. Come si può immaginare, tutti questi criteri non sono tra loro equivalenti. Ad esempio la distribuzione log-normale è a code pesanti secondo i criteri di Bryson e Mantel, ma a code leggere secondo il criterio di Gumbel. A prescindere da quale sia il criterio adottato per definire una distribuzione a code pesanti, una delle implicazioni più importanti è che nella maggior parte dei casi distributivi in questione, non tutti, o addirittura alcuno, dei momenti esistono.

La distribuzione con code pesanti più nota storicamente è la t di Student. Se $X \sim N(0, 1)$ allora la variabile casuale

$$Y = \frac{X}{\sqrt{\frac{\chi_\nu^2}{\nu}}} \sim t_\nu, \quad (1.1)$$

dove χ_ν^2 è un chi-quadro con ν gradi di libertà indipendente da X , è detta avere distribuzione t di Student con ν gradi di libertà. Tale distribuzione è apparsa per la prima volta nella letteratura statistica in Student (1908) pubblicato da William Gosset. Il parametro dei gradi di libertà ν è associato alla forma della distribuzione e ne determina la relativa pesantezza delle code. Se infatti $\nu \rightarrow \infty$ otteniamo la distribuzione normale; per ν finito la distribuzione è caratterizzata da code più pesanti rispetto alla distribuzione normale ed è proprio tale aspetto che ha reso la distribuzione t di Student di notevole interesse. Va detto inoltre che per $\nu > 25$ all'incirca, la distribuzione t di Student è ben approssimata dalla distribuzione normale, pertanto possiamo effettivamente considerare la distribuzione t di Student a code pesanti se ν è inferiore ad una decina.

Tecniche di stima robuste sono state proposte nei modelli di regressione lineare assumendo, ad esempio, una distribuzione t di Student per il termine d'errore, a differenza del classico approccio basato sull'assunzione distributiva normale (Zellner, 1976). Tuttavia se i gradi di libertà della distribuzione t Student sono pari a ν , allora i momenti di ordine maggiore o uguale di ν non esistono. Nel caso in cui $\nu = 1$, ossia quando abbiamo una distribuzione Cauchy, non siamo in grado di definire alcuno dei momenti distributivi. Inoltre valori piccoli di ν ostacolano l'inferenza sul modello statistico t di Student. In un modello di regressione lineare t di Student, la log-verosimiglianza non possiede alcun massimo globale per i coefficienti di regressione in corrispondenza di valori relativamente piccoli di ν . Una revisione critica di un approccio sia bayesiano che frequentista al modello di regressione lineare con errori indipendenti distribuiti secondo una t di Student con gradi di libertà ν ignoti è presentata in Fernández & Steel (1999).

La distribuzione di probabilità *slash* è un altro esempio di distribuzione a code pesanti ed è ottenuta nel seguente modo. Se $X \sim N(0, 1)$, allora

$$Y = \frac{X}{U^{1/q}}, \quad (1.2)$$

dove U è una distribuzione uniforme in $(0, 1)$ indipendente da X e $q > 0$, si ha che Y è distribuita come una *slash* con q gradi di libertà (Kafadar, 1988). I parametri di posizione e di scala possono essere definiti considerando $X \sim N(\mu, \sigma^2)$. L'utilizzo primario di questa distribuzione è negli studi di simulazione, dove rappresenta un caso estremo con il quale confrontare la prestazione di diverse procedure statistiche. In uno studio di simulazione la prestazione di una procedura statistica viene valutata simulando da diverse distribuzioni di probabilità. La distribuzione *slash* è infatti vista come l'estremo opposto della distribuzione normale ed una procedura statistica che ha un buon comportamento in entrambi questi due estremi avrà allora una buona prestazione su una ampia gamma di distribuzioni caratterizzate da una diversa pesantezza delle code. I gradi di libertà q regolano inoltre la curtosi della distribuzione: se $q = 1$ si ha una distribuzione *slash* canonica e per $q \rightarrow \infty$ otteniamo la distribuzione normale. Inoltre i momenti distributivi di ordine maggiore od uguale a q non esistono. Ulteriori caratteristiche della distribuzione *slash* sono analizzate in Rogers & Tukey (1972). Una possibile generalizzazione della distribuzione *slash* al caso asimmetrico è fornita da Wang & Genton (2006).

Numerose altre distribuzioni a code pesanti possono essere ottenute da note distribuzioni come la Gamma inversa, la Weibull e la Pareto fissando il parametro di forma per valori minori o uguali di 1. In tutti questi casi, nessuno dei momenti distributivi esiste.

Ulteriori approfondimenti sulle distribuzioni a code pesanti sono forniti da Bryson (1974).

1.2 Obiettivo della trattazione

L'obiettivo che ci proponiamo è l'individuazione di un insieme di indici multidimensionali, denotato dal vettore $\psi \in \Psi \subseteq \mathbb{R}^p$, che, generati e studiati in un contesto non parametrico, siano in grado di sintetizzare le diverse caratteristiche di una famiglia parametrica di distribuzioni \mathcal{F} senza assumere l'esistenza dei momenti distributivi.

Prima di procedere oltre diamo la definizione di matrice di dispersione: $S(X)$ è una matrice di dispersione per il vettore casuale d -dimensionale X se $S(X)$ è una matrice $d \times d$ simmetrica semi-definita positiva ed equivariante, ossia per ogni matrice non singolare A di dimensioni $d \times d$ e per ogni vettore $b \in \mathbb{R}^d$ si ha

$$S(AX + b) = AS(X)A^T.$$

Denotato con $\theta \in \Theta \subseteq \mathbb{R}^p$ il parametro di indicizzazione della famiglia di distribuzioni \mathcal{F} , rivolgeremo in particolare la nostra attenzione alle famiglie parametriche per le quali le componenti di θ possono essere rappresentate in una struttura del tipo

$$\theta^T = (\xi^T, ([\Omega_{ij}]_{1 \leq i \leq j \leq d})^T, \alpha^T, \nu), \quad (1.3)$$

dove $\xi, \alpha \in \mathbb{R}^d$, Ω è una matrice di dispersione di dimensioni $d \times d$, indicando con $[\Omega_{ij}]_{1 \leq i \leq j \leq d}$ il vettore degli elementi sulla diagonale principale di Ω e gli elementi sopra di questa, ed infine $\nu \in \mathbb{R}$. Ciascun elemento è direttamente o indirettamente associato alla posizione, alla struttura di dipendenza e variabilità, all'asimmetria ed alla curtosi della distribuzione, rispettivamente. Spesso l'associazione del parametro θ con tali caratteristiche distributive avviene rispettivamente mediante i momenti del primo, secondo, terzo e quarto ordine. Tuttavia tale associazione viene meno nel caso delle distribuzioni prive dei momenti distributivi e l'interpretazione delle diverse componenti di θ diventa incerta. In seguito il nostro interesse verrà in particolar modo rivolto a tali distribuzioni.

Il vettore di indici ψ avrà una struttura analoga a quella di θ , ossia

$$\psi^T = (\eta^T, ([T_{ij}]_{1 \leq i \leq j \leq d})^T, \gamma^T, \phi) \quad (1.4)$$

dove $\eta, \gamma \in \mathbb{R}^d$, T matrice di dimensioni $d \times d$ simmetrica e definita positiva ma non necessariamente equivariante, indicando con $[T_{ij}]_{1 \leq i \leq j \leq d}$ il vettore degli elementi sulla diagonale principale di T e gli elementi sopra di questa, ed infine $\phi \in \mathbb{R}$. Ogni elemento è questa volta una diretta misura di ciascuna delle quattro caratteristiche distributive, la cui esistenza è indipendente dall'esistenza dei momenti distributivi. Abbiamo inoltre che $p = (d^2 + 5d + 2)/2$.

La famiglia ellittica è un esempio di famiglia di famiglie parametriche del tipo espressa nella (1.3), fatta eccezione per la mancanza del parametro di asimmetria in quanto simmetrica per definizione.

Definizione 1.2.1. *Sia X il vettore casuale d -dimensionale con distribuzione ellittica, ossia $X \sim E_d(\xi, \Omega, g)$. Allora la funzione di densità è della forma*

$$f(x) = c|\Sigma|^{-1/2}g((x - \xi)^T\Omega^{-1}(x - \xi)), \quad x \in \mathbb{R}^d \quad (1.5)$$

dove c è la costante di normalizzazione e $g(\cdot)$ una qualche funzione da \mathbb{R}^d a \mathbb{R}^+ ed integrabile.

Il concetto di simmetria ellittica è infatti uno dei concetti di simmetria multidimensionale caratterizzante tale tipo di distribuzioni (vedi Definizione 2.2.2).

La distribuzione t di Student multidimensionale è un esempio di distribuzione ellittica e la sua funzione di densità è

$$t_d(x; \nu) = \frac{\Gamma((\nu + d)/2)}{|\Omega|^{1/2}(\pi\nu)^{d/2}\Gamma(\nu/2)}(1 + Q_x/\nu)^{-(\nu+d)/2}, \quad (1.6)$$

dove

$$Q_x = (x - \xi)\Omega^{-1}(x - \xi).$$

In seguito indicheremo che $X \sim T_d(\xi, \Omega, \nu)$. La rappresentazione stocastica della t di Student d -dimensionale è del tutto simile al caso unidimensionale dato dalla (1.1). Più esattamente abbiamo che se

$$Y = \xi + \frac{X}{\sqrt{\frac{X^2}{\nu}}},$$

dove $X \sim N_d(0, \Omega)$ indipendente da χ_ν^2 , allora Y ha distribuzione t di Student d -dimensionale con parametro di posizione ξ , matrice di dispersione Ω e ν gradi di libertà. Diverse generalizzazioni della t Student unidimensionale al caso multidimensionale sono discusse nel Capitolo 5 di Kotz & Nadarajah (2004). Noi ci limiteremo a considerare la t di Student con funzione di densità data dalla (1.6), in quanto è la più diretta e naturale generalizzazione della t di Student al caso multidimensionale. Una difficoltà posta dalla (1.6) è data dal fatto che nel caso in cui Ω è una matrice diagonale, le variabili casuali marginali (distribuite come t di Student unidimensionali) risultano essere incorrelate ma non indipendenti. Infatti non è possibile fattorizzare la (1.6) nel prodotto di funzioni di densità marginali di variabili casuali t di Student indipendenti tra loro.

Oltre alla maggior pesantezza delle code, un'ulteriore dipartenza dall'assunzione distributiva di normalità è data dall'asimmetria. La famiglia di distribuzioni "skew- t " (ST) (Azzalini & Capitanio, 2003) combina in sé caratteristiche sia di asimmetria che di maggior pesantezza delle code rispetto al modello gaussiano, rappresentando così uno strumento flessibile per la modellazione dei dati (Azzalini & Genton, 2008). In seguito rivolgeremo una particolare attenzione a questa distribuzione di probabilità ed indicheremo con $X \sim ST_d(\xi, \Omega, \alpha, \nu)$ la variabile casuale d -dimensionale la cui funzione di densità è

$$f(x) = 2t_d(x; \nu)T_1 \left(\alpha^T \omega^{-1}(x - \xi) \left(\frac{\nu + d}{Q_x + \nu} \right)^{1/2}; \nu + d \right), x \in \mathbb{R}^d \quad (1.7)$$

dove

$$\omega = \text{diag}([\Omega_{ii}]_{i \leq d})^{1/2}$$

e $T_1(\cdot; \nu + d)$ è la funzione di ripartizione di una t di Student unidimensionale con $\nu + d$ gradi di libertà. Come si può notare la ST non è altro che una generalizzazione del modello t di Student, dove il parametro aggiuntivo α regola il grado di asimmetria della distribuzione.

Nel corso della trattazione verrà a ricoprire un ruolo di primo piano la cosiddetta forma canonica della ST. La forma canonica della ST si ha quando la matrice di dispersione Ω coincide con la matrice identità I_d ed il parametro di forma, che regola l'asimmetria distributiva, ha al massimo una componente non nulla (Azzalini & Capitanio, 1999). Il successivo risultato mette in relazione la forma canonica con una qualsiasi altra distribuzione della famiglia ST.

Lemma 1.2.2. *Sia $Y \sim \text{ST}_d(0, \Omega, \alpha, \nu)$ allora esiste una trasformazione lineare $Y^* = AY$ tale che $Y^* \sim \text{ST}_d(0, I_d, \alpha^*, \nu)$ dove al massimo una componente di α^* è non nulla.*

Dimostrazione. In virtù del fatto che le forme lineari della ST multidimensionale hanno espressioni del tutto analoghe a quelle delle distribuzione multidimensionale normale asimmetrica, la dimostrazione è analoga a quella fornita per la normale asimmetrica nella Proposizione 4 di Azzalini & Capitanio (1999).

Per la scomposizione di Cholesky abbiamo che $\Omega = BB^T$, con B matrice triangolare inferiore. Se α_B è il parametro responsabile dell'asimmetria distributiva per la variabile $B^{-1}Y$, allora la matrice A sarà del tipo $P^T B^{-1}$, con P matrice ortonormale avente una colonna nella direzione di α_B . \square

Inoltre se $X \sim \text{ST}_d(\xi, \Omega, \alpha, \nu)$ allora la rispettiva forma canonica avrà l'unica componente non nulla per il vettore α^* pari a

$$(\alpha^T \omega^{-1} \Omega \omega^{-1} \alpha)^{1/2},$$

analogamente a quanto avviene per la normale asimmetrica (Azzalini & Capitanio, 1999).

Nel tentativo di individuare indici multidimensionali non parametrici per famiglie parametriche, si cercherà in particolare di sfruttare il concetto di quantile. Più esattamente, si farà ricorso al concetto di funzione *depth* (Zuo & Serfling, 2000a), di cui si parlerà da subito nel Capitolo 2, al fine di stabilire una logica di ordinamento nello spazio multidimensionale che ci permetta di misurare l'estremità di un dato punto $x \in \mathbb{R}^d$ rispetto ad una distribuzione di probabilità F e di stabilire un concetto di mediana multidimensionale. Ci focalizzeremo in particolare sulla funzione *depth* di Tukey (1975) da cui ha origine una particolare mediana multidimensionale con delle proprietà che la rendono preferibile, secondo i criteri da noi utilizzati, rispetto ad altre possibili mediane multidimensionali. Nel Capitolo 3 si cercherà invece di individuare una misura matriciale di associazione per una data distribuzione multidimensionale al fine di disporre di un indice che funga da surrogato della matrice di covarianza. Nel Capitolo 4 indagheremo indici unidimensionali di asimmetria basati sui quantili e nel Capitolo 5 le funzioni *depth* torneranno a svolgere un ruolo primario nella definizione di un indice multidimensionale di curtosi, proposto in Wang & Serfling (2005).

Capitolo 2

L'indice di posizione

In questo capitolo affronteremo il problema di scelta di un indice di posizione la cui definizione non assuma l'esistenza dei momenti della distribuzione di probabilità. Definiremo nel §2.1 metodi non parametrici basati sulle funzioni *depth* e tra queste ci concentreremo in particolar modo sulla funzione *depth* proposta da Tukey (1975), che, come si vedrà nel §2.2, permette di definire una mediana multidimensionale con proprietà che la rendono preferibile rispetto ad altre mediane multidimensionali proposte in letteratura. Nell'ultima parte contestualizzeremo la discussione alla distribuzione ST, mettendo in risalto aspetti legati alla simmetria radiale ed alcune proprietà della funzione *depth* di Tukey (HS).

Indicheremo inoltre con $\|x\| = \sqrt{\sum_i x_i^2}$ e con $\|x\|_1 = \sum_i |x_i|$ rispettivamente la norma euclidea e la norma 1 del vettore $x \in \mathbb{R}^d$.

2.1 Le funzioni *depth*: definizione e proprietà

La necessità di definire una funzione che misuri quanto un determinato punto $x \in \mathbb{R}^d$ risulta essere estremo per una data distribuzione F o rispetto ad un insieme di punti d -dimensionali $\{x_1, \dots, x_n\}$ (specialmente nel caso in cui $d > 2$ dove la visualizzazione grafica fallisce) ha dato sviluppo a metodi non parametrici in grado di definire una logica di ordinamento nello spazio multidimensionale. Le funzioni *depth* possono essere viste come una generalizzazione del metodo dei ranghi nel caso unidimensionale sebbene esistono forti differenze con questo. Come chiariremo più avanti, nel caso multidimensionale la logica di ordinamento indotta dalla funzione *depth* è “dal centro

verso l'esterno", contrapposta alla logica lineare "dal più piccolo al più grande" nel caso unidimensionale (Liu, Prelius & Singh, 1999). Come in Zuo & Serfling (2000a), limiteremo la nostra attenzione alle funzioni *depth* non negative e limitate e definiamo quattro proprietà desiderabili. Indicheremo con \mathcal{F} una generica famiglia di distribuzioni e con F_X la distribuzione in \mathcal{F} del vettore casuale d -dimensionale X .

Definizione 2.1.1. *Sia $D(.,.) : \mathbb{R}^d \times \mathcal{F} \rightarrow \mathbb{R}^+$ una funzione limitata e che soddisfa le seguenti proprietà.*

1. *Invarianza: sia F_{AX+b} la distribuzione della variabile casuale $AX + b$, allora l'uguaglianza $D(Ax + b; F_{AX+b}) = D(x; F_X)$ risulta vera per ogni vettore casuale $x \in \mathbb{R}^d$, per ogni matrice A invertibile di dimensioni $d \times d$ e per ogni vettore $b \in \mathbb{R}^b$.*
2. *Massimalità: se la distribuzione F è simmetrica rispetto ad un punto θ , secondo una qualche definizione di simmetria, allora $D(\theta; F) = \sup_{x \in \mathbb{R}^d} D(x; F)$ è valida per ogni $F \in \mathcal{F}$ avente come centro di simmetria θ .*
3. *Monotonia rispetto al centro: per ogni $F \in \mathcal{F}$ avente come punto di massima depth θ , si ha che $D(x; F) \leq D(\alpha x + (1 - \alpha)\theta; F)$ per ogni $\alpha \in [0, 1]$.*
4. *Limite nullo all'infinito: $D(x; F) \rightarrow 0$, per $\|x\| \rightarrow \infty$ per ogni $F \in \mathcal{F}$.*

Allora $D(., F)$ è una funzione statistica *depth*.

La versione campionaria di $D(x; F)$ può essere definita sostituendo F con la distribuzione empirica \hat{F} . In seguito ci riferiremo alle funzioni *depth* nell'accezione fornita dalla Definizione 2.1.1 omettendo l'aggettivo "statistica". In base a queste quattro proprietà ed insieme a considerazioni relative al punto di rottura, Zuo & Serfling (2000a) affermano che la mediana di Tukey, che come vedremo non è altro che il punto di massimo globale di una particolare funzione *depth*, risulta essere più attraente rispetto ad altre mediane.

Per poter procedere oltre abbiamo bisogno delle seguenti definizioni.

Definizione 2.1.2. *L'insieme $\{x \in \mathbb{R}^d : D(x; F) = t\}$ è detto curva di livello di funzione *depth* pari a t .*

Definizione 2.1.3. *L'insieme $\{x \in \mathbb{R}^d : D(x; F) \geq t\}$ è la regione racchiusa dalla curva di livello t , e viene denotata con $R(t)$.*

Definizione 2.1.4. *Definita la misura di probabilità $P(\cdot; F)$ su \mathbb{R}^d , l'insieme*

$$C_p = \bigcap_t \{R(t) : P(R(t); F) \geq p\}$$

viene definito come la p -esima regione centrale.

Secondo quest'ultima definizione C_p non è altro che la più piccola regione racchiusa dalle curve di livello che delimita una massa di probabilità p . Se F è assolutamente continua e la sua funzione di densità è non nulla ovunque allora

$$C_p = R(t_p)$$

dove si ha che $R(t_p)$ soddisfa la condizione che $P(R(t_p); F) = p$.

La frontiera di C_p denotata con $q(p)$, verrà riferita come p -esima curva di livello indotta dalla funzione *depth* D . Inoltre se la funzione di densità f è non nulla ovunque allora $q(p)$ è la curva di livello $\{x \in \mathbb{R}^d : D(x; F) = t_p\}$. Nel caso particolare in cui $X \sim E_d(\mu, \Sigma, g)$ si ha inoltre che

$$D(x; F) = h((x - \mu)^T \Sigma^{-1} (x - \mu))$$

per una qualche funzione h non crescente (risultato fornito nel Teorema 3.3 di Zuo & Serfling (2000b)). Abbiamo pertanto che

$$R(t_p) = \{x \in \mathbb{R}^d : (x - \mu)^T \Sigma^{-1} (x - \mu) \leq r_p^2\}$$

per un qualche r_p . Nel caso delle distribuzioni ellittiche $q(p)$ corrisponde alla curva di livello che sottende una regione con massa di probabilità p ed ha pertanto forma ellittica.

2.2 Selezione dell'indice di posizione

Indici di posizione e simmetria multidimensionale sono due concetti legati dal fatto che in una distribuzione simmetrica, secondo una qualche definizione di simmetria, è particolarmente auspicabile che un qualsiasi indice di posizione sia in accordo con il rispettivo centro di simmetria. I più noti concetti di simmetria multidimensionale, in ordine crescente di generalità, sono la simmetria *sferica*, la simmetria *ellittica*, la simmetria *centrale* e la simmetria *radiale*.

2.2.1 Concetti di simmetria multidimensionale

Definizioni di simmetria possono essere fornite in termini di proprietà strutturali della funzione di ripartizione o della funzione caratteristica o della funzione di densità. Nel caso multidimensionale, uno dei modi più convenienti per definire un particolare tipo di simmetria è quello di esprimerla in termini di invarianza della distribuzione di un vettore casuale “centrato” $X - \mu$ per un’opportuna famiglia di trasformazioni. Indicheremo in seguito con il simbolo $\stackrel{d}{=}$ l’uguaglianza in distribuzione tra due variabili casuali.

Definizione 2.2.1. *Un vettore casuale $X \in \mathbb{R}^d$ ha distribuzione con simmetria sferica rispetto a $\mu \in \mathbb{R}^d$ se*

$$X - \mu \stackrel{d}{=} A(X - \mu)$$

per ogni matrice ortogonale A di dimensioni $d \times d$.

Nel caso di variabili casuali assolutamente continue con distribuzione con simmetria sferica, la funzione di densità può essere espressa nella forma $g((x - \mu)^T(x - \mu))$, $x \in \mathbb{R}^d$, per una qualche funzione g da \mathbb{R}^d a \mathbb{R}^+ ed integrabile. Tra le distribuzioni simmetriche sfericamente la più nota distribuzione è la normale multidimensionale con matrice di covarianza pari a $\sigma^2 I_d$.

Un’equivalente definizione alla Definizione 1.2.1 della famiglia di distribuzioni ellittiche e che stabilisce il concetto di simmetria ellittica è la seguente.

Definizione 2.2.2. *Sia Y una variabile casuale k -dimensionale con simmetria sferica rispetto all’origine. Allora per ogni matrice A di dimensioni $k \times d$ tale che $A^T A = \Sigma$ e $\text{rango}(\Sigma) = k \leq d$ e per ogni $\mu \in \mathbb{R}^d$ si ha che la variabile casuale*

$$X = A^T Y + \mu$$

ha distribuzione con simmetria ellittica.

Se X è assolutamente continua, la funzione di densità è della forma data dalla (1.5), per una qualche funzione g da \mathbb{R}^d a \mathbb{R}^+ integrabile. Se A è inoltre una matrice $d \times d$ non singolare, allora la funzione di densità di X può essere scritta nella forma $|A|^{-1} g_0(A^{-1}(x - \mu))$, con g_0 funzione di densità di una distribuzione sfericamente simmetrica rispetto l’origine. Le curve di livello per distribuzioni simmetriche in modo ellittico hanno forma ellissoidale. Nell’insieme delle famiglie di distribuzioni ellittiche possono essere

pertanto individuate valide alternative al modello normale nel caso in cui si vuole rilassare l'ipotesi di normalità e mantenere allo stesso tempo specifiche strutture distributive. Tuttavia un più ampio rilassamento dell'ipotesi di normalità rispetto alla simmetria ellittica è dato dalla simmetria centrale.

Definizione 2.2.3. *Un vettore casuale $X \in \mathbb{R}^d$ ha distribuzione simmetrica centralmente rispetto a $\mu \in \mathbb{R}^d$ se*

$$X - \mu \stackrel{d}{=} \mu - X.$$

La funzione di densità di distribuzioni simmetriche centralmente, se esiste, soddisfa la seguente identità: $f(x - \mu) = f(-x - \mu)$. La simmetria centrale rappresenta l'estensione non parametrica più diretta della simmetria unidimensionale. Tuttavia una più generale definizione di simmetria multidimensionale è data dalla simmetria radiale.

Definizione 2.2.4. *Un vettore casuale $X \in \mathbb{R}^d$ ha distribuzione simmetrica radialmente rispetto a $\mu \in \mathbb{R}^d$ se*

$$\frac{X - \mu}{\|X - \mu\|} \stackrel{d}{=} \frac{\mu - X}{\|X - \mu\|}.$$

Due proprietà del centro di simmetria radiale μ , rilevanti nel successivo sviluppo di questo capitolo, sono:

- P1 a meno che la distribuzione del vettore casuale X sia concentrata su una retta, il centro di simmetria radiale μ è unico (Lemma 2.3 in Zuo & Serfling, 1999);
- P2 ogni iperpiano passante per μ divide \mathbb{R}^d in due semipiani aperti con uguale probabilità, pari ad $1/2$ se X è continuo (Corollario 2.1 in Zuo & Serfling, 1999).

La simmetria radiale rappresenta il concetto di simmetria più generale tra quelli appena introdotti. Approfondimenti ed equivalenti caratterizzazioni dei concetti di simmetria introdotti sono presentati in Serfling (2006).

2.2.2 Mediane multidimensionali

Definiamo ora una generica misura di posizione per una classe di distribuzioni di probabilità \mathcal{F} in \mathbb{R}^d mediante il funzionale statistico $\psi : \mathcal{F} \rightarrow \mathbb{R}^d$. Un

criterio con il quale possiamo discriminare tra le varie proposte di mediana multidimensionale è il seguente:

$C1$ per ogni $F \in \mathcal{F}$ simmetrica attorno ad un unico punto μ , secondo una qualche definizione di simmetria fornita in §2.2.1, si ha che $\psi(F) = \mu$.

Un ulteriore criterio che richiediamo è quello relativo alla proprietà di equivarianza:

$C2$ sia $Y = AX + b$ allora $\psi(Y) = A\psi(X) + b$, per ogni matrice $d \times d$ non singolare e vettore $b \in \mathbb{R}^d$.

Insieme a $C1$ e $C2$, la robustezza della misura di posizione risulta essere una proprietà fondamentale nel contesto in cui stiamo operando, poiché nel caso di distribuzione prive dei momenti distributivi la probabilità di dover trattare valori anomali è relativamente alta. Questi stessi criteri sono stati considerati come prioritari anche in Zuo & Serfling (1999).

La robustezza di uno stimatore verrà inoltre quantificata per mezzo del punto di rottura. Il punto di rottura di uno stimatore T , indicato con ϵ^* , è la più piccola frazione di contaminazione del campione sotto la quale avviene “la rottura” ossia

$$\epsilon^* = \min \left\{ \frac{m}{m+n} : \sup_{Y_m} |T(X_n \cup Y_m) - T(X_n)| = \infty \right\}$$

dove X_n è un campione di dimensione n ed Y_m rappresenta il campione contaminato di dimensione m .

Elenchiamo ora diverse mediane multidimensionali in grado di competere tra loro in base ai criteri $C1$, $C2$ e di robustezza.

1. La mediana per componenti: per una distribuzione continua F in \mathbb{R}^d la mediana per componenti, ossia il vettore delle mediane delle distribuzioni marginali, può essere vista come il punto η in \mathbb{R}^d che minimizza la norma 1, ossia

$$\eta = \arg \min_{x \in \mathbb{R}^d} E \|X - x\|_1. \quad (2.1)$$

2. La mediana spaziale, (Grower, 1974; Brown, 1983): per una distribuzione continua F in \mathbb{R}^d , non concentrata su una retta, la mediana spaziale

è il punto η_s in \mathbb{R}^d che minimizza la norma euclidea, ossia

$$\eta_s = \arg \min_{x \in \mathbb{R}^d} E\|X - x\|. \quad (2.2)$$

Tale mediana può essere vista anche come il punto in \mathbb{R}^d che massimizza la seguente funzione *depth* calcolata in $x \in \mathbb{R}^d$

$$(1 + E\|X - x\|)^{-1}.$$

3. La mediana di Tukey: per una distribuzione continua F la mediana di Tukey M_F può essere vista come il punto x in \mathbb{R}^d che massimizza la seguente funzione *depth* calcolata in $x \in \mathbb{R}^d$

$$\min_{\|u\|=1} P(u^T X < u^T x).$$

Maggiori dettagli verranno forniti in §2.3.1.

4. La *simplicial Liu depth median* (Liu, 1988, 1990): per una distribuzione continua F in \mathbb{R}^d la *simplicial Liu depth median* è il punto β in \mathbb{R}^d che massimizza la probabilità di appartenere ad un simpleso casuale in \mathbb{R}^d , ossia

$$\beta = \arg \max_{x \in \mathbb{R}^d} P_F(x \in S[X_1, \dots, X_{d+1}]) \quad (2.3)$$

dove X_1, \dots, X_{d+1} è un campione casuale semplice dalla distribuzione F e $S[x_1, \dots, x_{d+1}]$ denota un simpleso d -dimensionale con vertici x_1, \dots, x_{d+1} .

Un'altra mediana multidimensionale molto studiata in letteratura è la mediana di Oja (1983), che rispetta sia $C1$ e $C2$ ma ha punto di rottura pari a 0. Essendo il punto di rottura, come già osservato, la proporzione di osservazioni anomale che uno stimatore è in grado di tollerare prima di fornire stime arbitrariamente grandi, si ha che la mediana di Oja non dà alcuna garanzia in termini di robustezza, pertanto la escludiamo a priori.

Nel caso in cui la distribuzione di X sia radialmente simmetrica si ha che tutte le mediane sopracitate coincidono e pertanto soddisfano il criterio $C1$. Il secondo criterio $C2$ risulta invece essere soddisfatto solamente dalla mediana di Tukey e dalla *simplicial Liu depth median*. Per quanto riguarda la robustezza riportiamo brevemente quali sono i punti di rottura per le versioni campionarie delle varie mediane multidimensionali.

1. La mediana per componenti: il punto di rottura è $1/2$.
2. La mediana spaziale: il punto di rottura è $1/2$.
3. La mediana di Tukey: il punto di rottura è $1/3$.
4. La *simplicial Liu depth median*: il punto di rottura è $1/(d + 2)$.

L'unica mediana che risulta rispettare entrambi i criteri $C1$ e $C2$, ed ha un punto di rottura relativamente alto è la mediana di Tukey che analizzeremo in dettaglio nel resto di questo capitolo. Una rassegna delle varie mediane multidimensionali, con discussione delle loro principali proprietà, è presentata in Small (1990).

2.3 La mediana di Tukey (o dei semispazi)

2.3.1 Concetti di base

Focalizzeremo ora l'attenzione su un particolare tipo di funzione *depth* che, in base alle quattro proprietà della Definizione 2.1.1, risulta avere un comportamento preferibile rispetto ad altre possibili funzioni *depth*. Tale funzione *depth* è stata introdotta in Tukey (1975), nel quale c'è il tentativo generalizzare il concetto di mediana campionaria unidimensionale allo spazio bidimensionale. La generalizzazione al caso multidimensionale è fornita da Donoho & Gasko (1992). Va detto che molti anni prima in Hotelling (1929), nell'ambito della teoria dei giochi, questa mediana era stata già introdotta in insiemi in cui è definita una misura di distanza tra punti.

In Hotelling (1929) il problema è motivato da considerazioni sulla localizzazione del sito di massimo beneficio finanziario. Consideriamo una spiaggia e due gelatai che vogliono posizionare i loro chioschi al fine di vendere più gelati possibili. Assumiamo che la spiaggia sia una retta lungo la quale i clienti si distribuiscono secondo una variabile casuale C con distribuzione F (vedi Figura 2.1). Assumiamo inoltre che i clienti siano tutti ugualmente interessati a comprare i gelati e che nella scelta del chiosco da cui comprare il gelato, la distanza sia l'unico criterio discriminante, ossia i clienti sceglieranno il chiosco a loro più vicino. Se i due chioschi sono ugualmente vicini allora uno di questi verrà scelto con probabilità $1/2$. Infine poniamo il vincolo che la posizione dei due chioschi non può coincidere. La domanda che sorge è la seguente: come posizioneranno i gelatai i loro chioschi se il gelataio

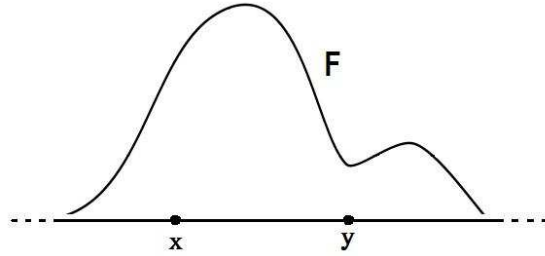


Figura 2.1: Il problema di posizionamento del chiosco di Hotelling (1929).

A posiziona il chiosco per primo? Se il gelataio A posiziona il chiosco nel punto x ed il gelataio B nel punto y e supponiamo $x < y$, il gelataio A si assicurerà i clienti nell'intervallo $(-\infty, (x+y)/2)$ mentre il gelataio B i clienti in $((x+y)/2, \infty)$ e si ripartiranno egualmente, in media, i clienti nel punto $(x+y)/2$. Il gelataio A avrà un profitto proporzionale a

$$P\left(C < \frac{x+y}{2}\right) + \frac{1}{2}P\left(C = \frac{x+y}{2}\right).$$

Se $x > y$ si ottiene un'espressione simile. Il minimo saldo finale possibile per il gelataio A è dato da

$$\min(P(C < x), 1 - P(C \leq x)).$$

Tale quantità non è altro che la funzione HS del punto x rispetto alla distribuzione F ed è massimizzata quando x corrisponde alla mediana della distribuzione F .

Se X è una variabile casuale unidimensionale assolutamente continua, la funzione HS calcolata nel punto $x \in \mathbb{R}$ può essere scritta nel seguente modo

$$\min(P(X < x), P(X > x)).$$

La generalizzazione al caso multidimensionale è data considerando la proiezione unidimensionale del vettore casuale X con distribuzione F . La funzione HS calcolata nel punto $x \in \mathbb{R}^d$ diventa allora

$$\text{HS}(x; F) = \min_{\|u\|=1} P(u^T X \leq u^T x).$$

La dimostrazione che HS è effettivamente una funzione *depth* secondo la Definizione 2.1.1 è fornita da Zuo & Serfling (2000a). Possiamo pertanto

definire la mediana dei semispazi M_F come il punto x in cui HS è massima, ossia

$$M_F = \arg \max_{x \in \mathbb{R}^d} \text{HS}(x; F).$$

Per quanto riguarda il problema di unicità della mediana dei semispazi si ha che se la distribuzione F è centralmente simmetrica (vedi Definizione 2.2.3) allora la funzione HS è unicamente massimizzata in M_F , il centro di simmetria, e $\text{HS}(M_F; F) = 1/2$.

Enunciamo il seguente risultato, presentato in Dutta, Ghosh & Chaudhuri (2011), circa il valore assunto dalla funzione HS nel suo punto di massimo globale nel caso di distribuzioni radialmente simmetriche.

Teorema 2.3.1. *Sia X sia un vettore casuale d -dimensionale e sia la mediana dei semispazi della distribuzione in $\mu \in \mathbb{R}^d$. Allora la funzione HS calcolata in μ sarà $1/2$ se e solo se X è radialmente simmetrica in μ .*

Nei casi in cui la distribuzione F non sia centralmente simmetrica allora le condizioni di unicità sono quelle indicate dal Teorema 3.3 di Small (1987) (che non riportiamo in quanto molto tecniche) e che in \mathbb{R}^2 si riducono ad imporre che la funzione di densità sia non nulla ovunque. Un'importante proprietà della funzione HS è che nel caso di distribuzioni simmetriche sfericamente e che hanno pertanto una funzione di densità f del tipo $f(x) = l(\|x\|)$ per una qualche funzione l da \mathbb{R}^d a \mathbb{R}^+ , la funzione HS risulta essere una funzione decrescente di $\|x\|$. Pertanto se l è una funzione strettamente decrescente, la funzione HS è una funzione crescente di f e viceversa. In questi casi le curve di livello della funzione HS coincidono con le curve di livello della funzione di densità f . Nei casi in cui invece f è più in generale del tipo $f(x) = l(\|x\|_p)$ dove $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$, le curve di livello della funzione HS non coincidono con quelle della funzione di densità se $p \neq 2$. Lo studio del comportamento delle curve di livello della funzione HS nel caso in cui $p \neq 2$ è presentato in Dutta, Ghosh & Chaudhuri (2011). Le curve di livello della funzione HS coincidono con le curve di livello della funzione di densità anche nel caso delle distribuzioni ellittiche, come già precedentemente notato nel §2.1 e dimostrato in Zuo & Serfling (2000b).

Tuttavia anche nei casi in cui le curve di livello della funzione HS non coincidono con le curve di livello della funzione di densità, tale funzione *depth* fornisce utili indicazioni circa l'estremità di un punto rispetto alla distribuzione di probabilità ed in una forma più adatta rispetto alla funzione di densità stessa (Dutta, Ghosh & Chaudhuri, 2011). Mentre quest'ultima rappresenta

soltanto una misura locale della massa di probabilità, la funzione HS, sebbene direttamente derivata dalla funzione di densità, è una misura globale dell'estremità di un punto rispetto alla distribuzione della popolazione. Si pensi ad esempio ad una distribuzione multidimensionale uniforme quando la rispettiva funzione di densità non da alcuna indicazione rispetto a quanto un determinato punto della distribuzione sia periferico o centrale.

2.3.2 Aspetti campionari dello stimatore non parametrico della funzione HS

Due proprietà importanti della mediana dei semispazi, che la rendono più attraente rispetto ad altre possibili mediane multidimensionali, sono le proprietà di equivarianza e robustezza, difficilmente presenti in altre mediane.

Supponiamo di disporre di un campione $\{x_1, \dots, x_n\}$ dove $x_i \in \mathbb{R}$, $\forall i \leq n$; allora avremo che la funzione *depth* campionaria di Tukey, che indicheremo con $\text{HS}(\cdot; \hat{F})$, è data da

$$\text{HS}(x; \hat{F}) = \min \left\{ \sum_i I(x_i \leq x), \sum_i I(x_i \geq x) \right\}, \quad (2.4)$$

dove $I(p)$ è la funzione indicatrice che vale 1 se la condizione p è verificata o 0 altrimenti. La (2.4) non restituisce altro che il minimo numero di punti tra quelli che sono a destra e a sinistra di x .

Nel caso multidimensionale, quando $x_i \in \mathbb{R}^d$, $\forall i \leq n$, dobbiamo considerare la proiezione unidimensionale $u^T x$ del vettore x , dove $u \in \mathbb{R}^d$ e $\|u\| = 1$. In questo caso $\text{HS}(\cdot; \hat{F})$ è definita come

$$\text{HS}(x; \hat{F}) = \min_{\|u\|=1} \left\{ \sum_i I(u^T x_i \geq u^T x) \right\}. \quad (2.5)$$

Nel caso unidimensionale avremo dunque che $\text{HS}(\cdot; \hat{F})$ vale 1 nel minimo e nel massimo delle statistiche d'ordine e circa $n/4$ ed $n/2$ rispettivamente nei quartili e nella mediana (se n è dispari).

Nel caso multidimensionale, come chiariremo meglio più avanti, il massimo valore della funzione campionaria $\text{HS}(\cdot; \hat{F})$ è legato alla simmetria centrale della distribuzione multidimensionale e si avrà un valore massimo tanto più prossimo ad $n/2$ quanto più i dati mostrano simmetria centrale.

Disponiamo ora in ordine decrescente ciascun vettore x_i , corrispondente all' i -esima osservazione nel campione, in base ai valori assunti dalla funzione data dalla (2.5),

$$x_{[1]}, \dots, x_{[n]}.$$

Allora si ha che $x_{[1]}$ corrisponde alla mediana campionaria, ossia il centro della distribuzione campionaria secondo la funzione $\text{HS}(\cdot; \hat{F})$, mentre $x_{[n]}$ rappresenta il punto più estremo della distribuzione campionaria e si ha dunque che ranghi maggiori, corrispondenti a valori più piccoli della funzione *depth*, sono associati ad osservazioni più estreme rispetto alla nuvola campionaria di punti. In Figura 2.2 osserviamo un esempio di quanto appena detto: il centro campionario in questo caso corrisponde al punto con *depth* pari a 10 e muovendoci verso l'esterno della distribuzione campionaria ci spostiamo verso curve di livello corrispondenti a valori della funzione *depth* più bassi. Nel caso mostrato in Figura 2.2 la mediana campionaria risulta essere unica, ma generalmente tale non è nel caso campionario ed appare esserci poca speranza circa le condizioni di unicità della mediana campionaria dei semispazi. Tuttavia, come dimostrato in Small (1987), la mediana campionaria dei semispazi risulta essere un insieme chiuso, limitato e convesso. Ricordiamo inoltre anche che nel caso unidimensionale la mediana non risulta essere unica secondo il criterio dato dalla (2.4) (ed anche secondo altri criteri) nel caso in cui n è pari.

Per il calcolo mediana di Tukey nel caso bidimensionale, Rousseeuw & Ruts (1998) hanno proposto un algoritmo, detto HALFMED, avente una complessità computazionale $O(n^2 \log^2 n)$. In uno spazio in più di due dimensioni, per il calcolo della mediana di Tukey si può impiegare l'algoritmo DEEPLOC proposto da Struyf & Rousseeuw (2000), il quale fornisce un valore approssimato della mediana dei semispazi.

Abbiamo inoltre che $R_{\hat{F}}(k)$, ossia la versione campionaria di $R(k)$ (vedi Definizione 2.1.3), risulta essere secondo la (2.5) l'intersezione di tutti i semispazi contenuti $n + 1 - k$ punti, da cui segue il seguente risultato.

Lemma 2.3.2. *Le curve di livello della funzione HS formano una sequenza di insiemi convessi ed annidati, ossia ogni $R_{\hat{F}}(k)$ è convesso e $R_{\hat{F}}(k+1) \subset R_{\hat{F}}(k)$.*

I vertici delle curve di livello in Figura 2.2 sono pertanto l'intersezione di due rette, ciascuna passante per una coppia di osservazioni. Notiamo inoltre che in questo problema di ottimizzazione non c'è alcun rischio di finire un

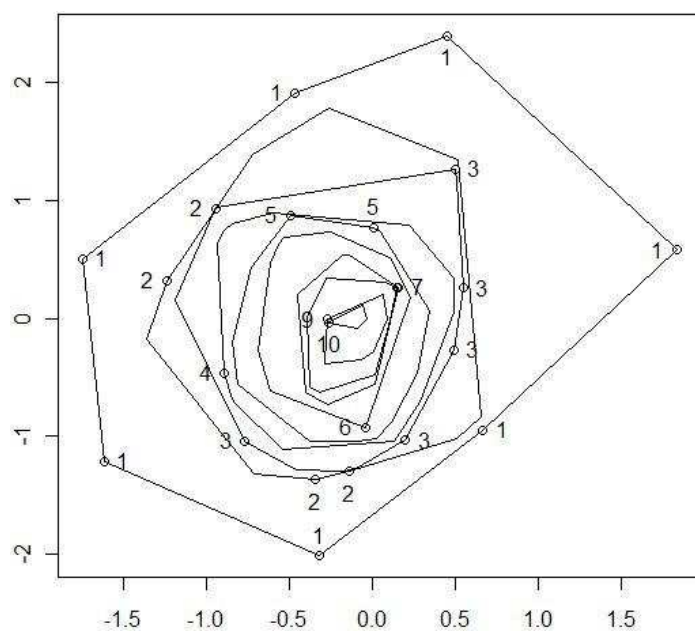


Figura 2.2: Curve di livello campionarie per 24 osservazioni ottenute da una distribuzione normale bidimensionale standard. I valori riportati accanto a ciascun punto rappresentano il valore assunto dalla funzione HS nel punto stesso.

massimo locale in quanto, come abbiamo appena osservato, la funzione HS è costituita da curve di livello convesse.

Un ulteriore proposito di Tukey (1975) era quello di utilizzare le curve di livello pari ad $n/4$. La forma di tali curve di livello indica la variabilità e la correlazione presente nel campione di dati e permette di definire una media troncata nel caso multidimensionale semplicemente facendo la media delle osservazioni per le quali la funzione HS è maggiore o uguale di 10. Risulta pertanto utile nel caso campionario, stabilire quante curve di livello possiamo definire e qual è il punto (o l'insieme) di massima *depth*. Donoho & Gasko (1992) mostrano, nel caso in cui il campione osservato è in *posizione regolare*, ossia quando non più di due punti giacciono su una retta, non più di tre su un piano e così via, ed assumendo che la distribuzione campionaria sia approssimativamente simmetrica centralmente, che il valore di massima *depth* risulta essere prossimo ad $n/2$. Se l'insieme di dati è approssimativamente simmetrico ci possono essere pertanto circa $n/2$ curve di livello e molte meno se l'insieme di dati mostra una pronunciata asimmetria.

Se infatti X è una variabile casuale assolutamente continua con distribuzione F simmetrica centralmente, si può dimostrare che

$$\max \text{HS}(x; \hat{F}_n) \approx n/2 - O_P(n^{1/2}),$$

dove \hat{F}_n è la distribuzione empirica ottenuta da un campione di dimensione n proveniente da F . Donoho & Gasko (1992) dimostrano inoltre che se il campione è in posizione regolare allora

$$\left\lceil \frac{n}{3} \right\rceil \leq \max_x \text{HS}(x; \hat{F}_n) \leq \left\lceil \frac{n}{2} \right\rceil,$$

dove $\lceil x \rceil$ è la parte intera superiore di x . Nel caso in cui la mediana campionaria di Tukey non sia unica possiamo considerare il centroide della regione di massima *depth*, senza in questo modo modificare i successivi risultati, ossia

$$M_{\hat{F}_n} = \text{Ave} \left\{ x : \text{HS}(x; \hat{F}_n) = \max_x \text{HS}(x; \hat{F}_n) \right\}.$$

Il seguente risultato stabilisce il punto di rottura dello stimatore non parametrico per la mediana di Tukey nel caso della simmetria centrale.

Lemma 2.3.3. *Sia $x = \{x_1, \dots, x_n\}$ un campione di dimensione n da una distribuzione assolutamente continua e simmetrica centralmente in \mathbb{R}^d , con $d > 2$, allora il punto di rottura ϵ^* dello stimatore $M_{\hat{F}_n}$ converge quasi certamente a $1/3$ quando $n \rightarrow \infty$.*

Ciò significa, che se circa il 67% delle osservazioni di un campione proviene da una distribuzione centralmente simmetrica, allora la stima delle mediana dei semispazi rimane ragionevole, indipendentemente dai valori assunti dal rimanente 33% delle osservazioni. Se invece X non proviene da una distribuzione simmetrica centralmente, allora senza far uso di probabilità o risultati asintotici abbiamo il seguente risultato.

Lemma 2.3.4. *Sia $x = \{x_1, \dots, x_n\}$ un campione di dimensione n in \mathbb{R}^d ed in posizione regolare. Allora il punto di rottura ϵ^* dello stimatore $M_{\hat{F}_n}$ è maggiore o uguale di $1/(1+d)$.*

Costruire degli stimatori equivarianti e robusti è un compito piuttosto arduo. Esistono infatti diverse procedure per la stima di indici di posizione il cui un punto di rottura non è mai superiore $1/(d+1)$. Approfondimenti sono forniti nel §4 di Donoho & Gasko (1992).

2.3.3 Funzione HS e distribuzione ST

Assumiamo ora che il vettore casuale d -dimensionale X provenga da una variabile di tipo ST. Nel seguito utilizzeremo le espressioni per le forme lineari della ST, facendo riferimento all'espressione (24) in Azzalini (2005) relativa alle analoghe forme lineari della normale asimmetrica.

Consideriamo in prima battuta un insieme di dati, analizzato in Azzalini & Genton (2008), che consiste in velocità medie giornaliere del vento registrate in tre stazioni meteorologiche: Goodnoe Hills (gh), Kennewick (kw) e Vansycle (vs). Il segno delle velocità ci fornisce l'informazione circa la direzione dei venti: il segno positivo corrisponde a venti da ovest e quello negativo a venti da est. Considerazioni circa l'indipendenza, l'identità in distribuzione e la bontà di adattamento della ST sono fornite in Azzalini & Genton (2008). Evidenziamo comunque il fatto che il modello ST multidimensionale porta ad un significativo miglioramento nell'adattamento ai dati rispetto alla distribuzione normale multidimensionale.

Tabella 2.1: Stime ottenute dalla distribuzione stimata con il metodo della massima verosimiglianza ($\hat{M}_i, i \in \{vs, gh, kw\}$) e stime non parametriche ($\tilde{M}_i, i \in \{vs, gh, kw\}$) delle mediane dei semispazi per ciascuna componente (prime tre righe), per coppie di componenti (seconde tre righe) e per la distribuzione congiunta (ultima riga).

\hat{M}_{vs}	\hat{M}_{gh}	\hat{M}_{kw}	HS_{MV}	\tilde{M}_{vs}	\tilde{M}_{gh}	\tilde{M}_{kw}	$HS(.; \hat{F})$
18.960	-	-	0.500	17.621	-	-	0.500
-	15.584	-	0.500	-	15.567	-	0.500
-	-	16.505	0.500	-	-	15.221	0.500
18.895	15.665	-	0.500	19.437	16.121	-	0.471
18.998	-	16.493	0.500	18.749	-	16.178	0.446
-	15.662	16.226	0.500	-	15.158	15.665	0.446
18.941	15.633	16.370	0.499	19.025	15.486	16.294	0.435

In Figura 2.3 osserviamo che la forma delle curve di livello della funzione HS ci fornisce un'informazione analoga alle curve di livello della funzione di densità per quanto riguarda la correlazione e l'asimmetria presente nell'insieme di dati. In Tabella 2.1 sono riportate le stime ottenute dalla distribuzione stimata con il metodo della massima ($\hat{M}_i, i \in \{vs, gh, kw\}$) e le stime non parametriche ($\tilde{M}_i, i \in \{vs, gh, kw\}$) delle mediane dei semispazi per ciascuna componente (prime tre righe), per coppie di componenti (seconde tre righe) e per la distribuzione congiunta (ultima riga). In tale tabella emerge quanto segue: considerando le mediane dei semispazi per coppie di componenti e per la distribuzione congiunta, stimate con il metodo della massima verosimiglianza, la funzione HS calcolata in tali punti è estremamente vicina a $1/2$ (i valori riportati sono arrotondati alla terza cifra decimale ed il più grande scarto da $1/2$ è circa pari a 3.347×10^{-4} che si ha quando si considerano le tre componenti congiuntamente).

Alla luce del Teorema 2.3.1 sembrerebbe di essere molto vicini alla simmetria radiale. Il prossimo risultato sulla Cauchy asimmetrica (SC), corrispondente ad una ST con $\nu = 1$, e successive considerazioni faranno luce su tale questione.

Prima di procedere oltre, notiamo che se X è radialmente simmetrica rispetto ad un punto μ allora $AX + b$ è radialmente simmetrica rispetto ad $A\mu + b$, dove A è una matrice non singolare di dimensioni $d \times d$ e $b \in \mathbb{R}^d$.

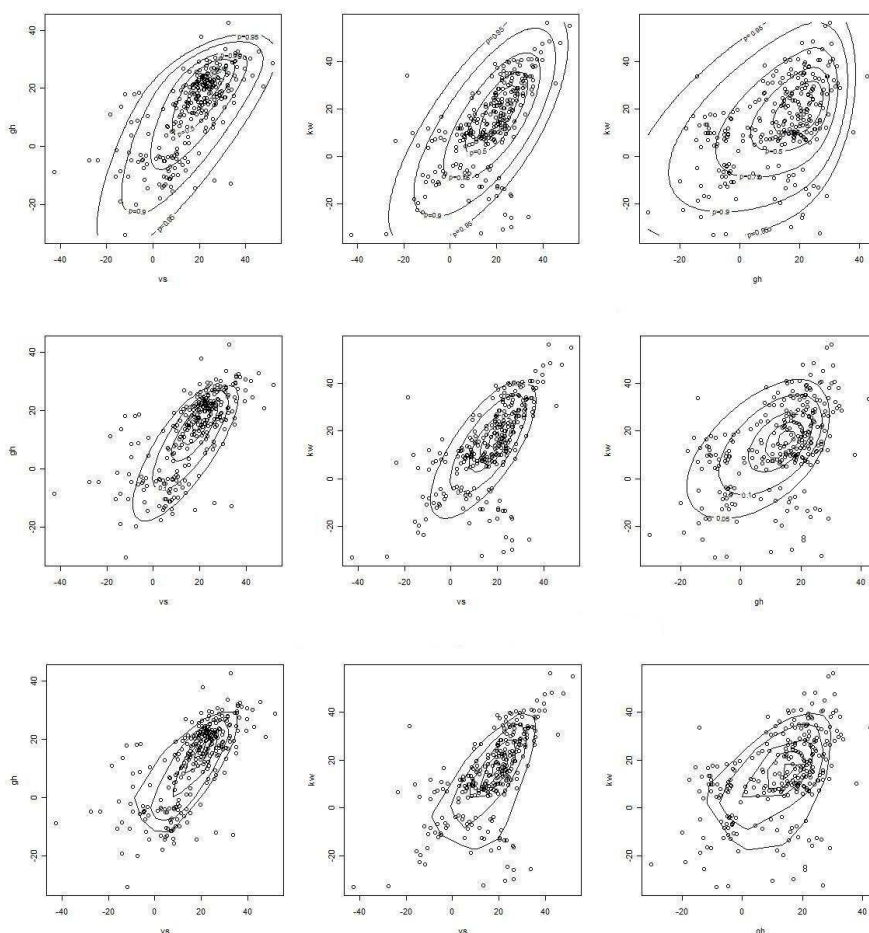


Figura 2.3: Diagrammi di dispersione per le diverse coppie di variabili con le curve di livello della funzione di densità (prime tre grafici in alto) e con le curve di livello della funzione HS (tre grafici nel mezzo) della ST i cui parametri sono stati stimati con il metodo della massima verosimiglianza, ed infine con le curve di livello della funzione HS stimata in modo non parametrico (ultimi tre grafici in basso).

Per ogni $u \in \mathbb{S}^{d-1}$, dove $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$, si ha infatti

$$\begin{aligned} P(u^T X < u^T \mu) &= P(u^T A^{-1} A X < u^T A^{-1} A \mu) \\ &= P(v^T A X < v^T A \mu) \\ &= P(v^T (A X + b) < v^T (A \mu + b)) \\ &= \frac{1}{2} \end{aligned}$$

dove $v = (A^{-1})^T u / \|(A^{-1})^T u\|$. Inoltre $v \neq 0$ in quanto lo spazio nullo della matrice $(A^{-1})^T$ è costituito solamente dal vettore nullo.

Teorema 2.3.5. *Sia $X \sim SC_d(\xi, \Omega, \alpha)$. Allora X è radialmente simmetrica.*

Dimostrazione. Senza perdita di generalità consideriamo $X \sim SC_d(0, I_d, \alpha^*)$ in forma canonica, dove il vettore α^* ha la prima componente non nulla pari ad α e I_d è la matrice identità di dimensioni $d \times d$.

Sia inoltre η la mediana per componenti della variabile casuale X , la quale risulterà avere solamente la prima componente non nulla che indicheremo con η_1 . Utilizzando il risultato di Behboodian, Jamalizadeh & Balakrishnan (2006) per cui la mediana di una variabile casuale $SC_1(0, 1, \alpha)$ è pari a $\alpha/\sqrt{1 + \alpha^2}$, otteniamo un'analogia espressione per η_1 .

Per ogni $u \in \mathbb{S}^{d-1}$, abbiamo che

$$P(u^T X < u^T \eta) = P\left(u_1 X_1 + \sum_{i=2}^d u_i X_i < u_1 \eta_1\right). \quad (2.6)$$

Se $u_1 = 0$ si avrebbe banalmente che la probabilità considerata risulta pari ad $1/2$. Pertanto ipotizziamo, senza perdita di generalità, che $u_1 > 0$.

Per le proprietà delle forme lineari della ST, dato il vettore $(X_1, T) \sim SC_2(0, I_2, (\alpha, 0)^T)$, abbiamo che

$$Y = X_1 + aT \sim SC_1\left(0, 1 + a^2, \frac{\alpha}{\sqrt{1 + a^2(1 + \alpha^2)}}\right),$$

con $T \sim t_1$.

Ponendo $a = \sqrt{1 - u_1^2}/u_1$, la (2.6) può essere riscritta nel seguente modo

$$\begin{aligned} P(u^T X < u^T \eta) &= P\left(X_1 + \frac{1}{u_1} \sum_{i=2}^d u_i X_i < \eta_1\right) \\ &= P(Y < \eta_1), \end{aligned} \quad (2.7)$$

poiché $X_1 + \frac{1}{u_1} \sum_{i=2}^d u_i X_i \stackrel{d}{=} Y$. Notando infine che la mediana di Y risulta essere pari a $\alpha/\sqrt{1+\alpha^2} = \eta_1$ si ha che la (2.6) è pari a $1/2$ per ogni $u \in \mathbb{S}^{d-1}$ e pertanto X è radialmente simmetrica in η . \square

Una volta stabilita la simmetria radiale nel caso in cui $\nu = 1$ rimane da capire se tale proprietà distributiva persiste anche per $\nu > 1$. Evidenza numerica piuttosto forte suggerisce che tale proprietà sia valida in modo esatto solo per $\nu = 1$ e che ci sia un graduale ed estremamente lento allontanamento dalla simmetria radiale per $\nu \rightarrow +\infty$. Ripercorrendo inoltre la stessa strada tracciata dalla dimostrazione del Teorema 2.3.5 ed ipotizzando $\nu \neq 1$, si giunge a questo punto

$$Y = X_1 + aT \sim \text{ST}_1 \left(0, 1 + a^2, \frac{\alpha}{\sqrt{1 + a^2(1 + \alpha^2)}}, \nu \right).$$

Ebbene è immediato verificare per via numerica che la mediana di Y è in generale diversa dalla mediana di X_1 per $\nu \neq 1$. Un'ulteriore prova numerica è data dal calcolo del valore assunto dalla funzione HS nella mediana per componenti η , poiché, se una distribuzione è simmetrica radialmente, il centro di simmetria radiale è dato da η . Nel caso della famiglia ST ed in virtù della proprietà di invarianza della funzione HS, tale valore può essere calcolato limitandosi alla forma canonica della ST nello spazio bidimensionale, in quanto, come mostrato dalla (2.7), tale valore è indipendente dalla dimensione dello spazio multidimensionale per $d \geq 2$.

In Figura 2.4 notiamo che il massimo scarto da $1/2$ ottenibile della funzione HS calcolata nella mediana per componenti η è pari a circa 0.035 (oltre un certo valore di α le curve tendono a sovrapporsi) che corrisponde al caso di una normale bidimensionale troncata. In Figura 2.5, dove è stata considerata tale distribuzione, risulta chiara la notevole vicinanza tra η ed M_F sia in termini della funzione di densità che di HS.

Data la prossimità dei due indici nel caso della ST in forma canonica, una possibile misura di posizione per una generica ST è la seguente. Se C^{-1} è la matrice di passaggio alla rispettiva forma canonica, definiamo il seguente indici di posizione

$$\eta^* = C\eta + \xi, \tag{2.8}$$

dove η è appunto la mediana per componenti della forma canonica. Va chiarito che ad una distribuzione ST corrisponde più di una distribuzione canonica.

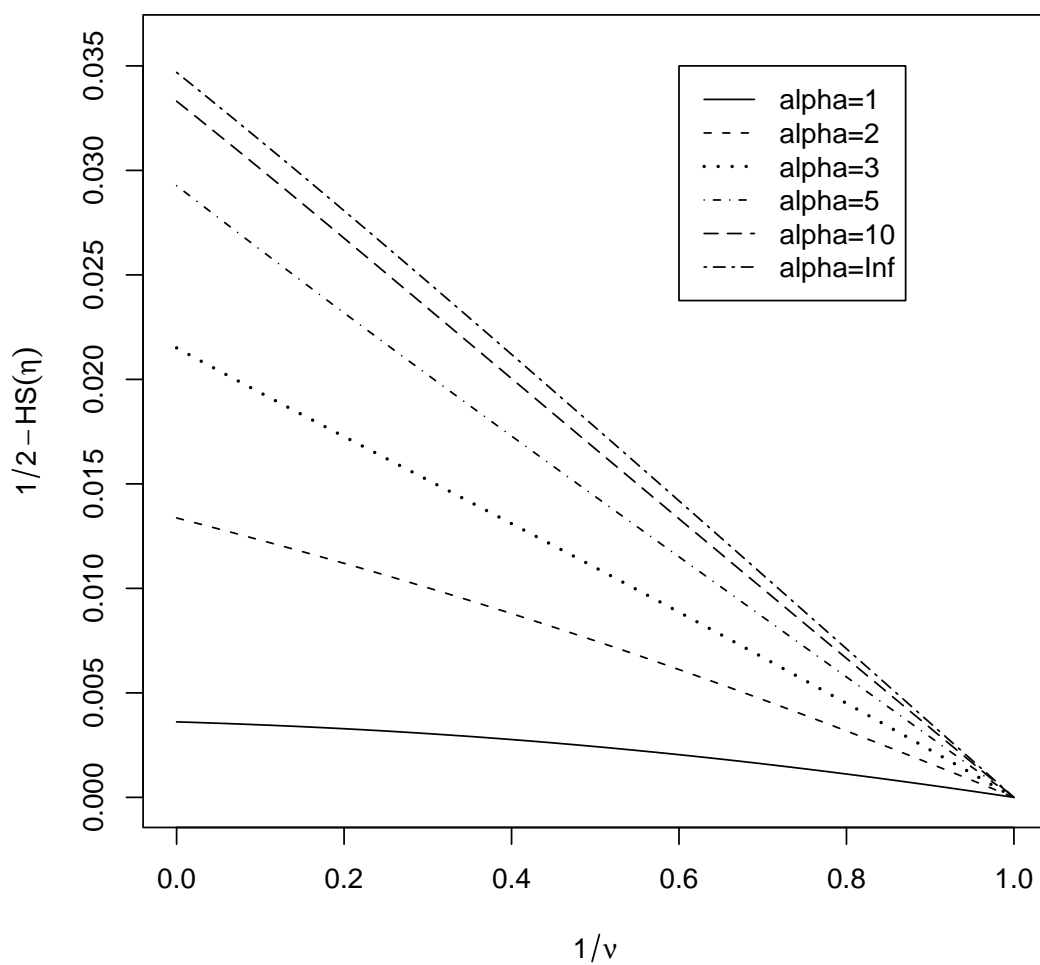


Figura 2.4: Scarto tra $1/2$ ed il valore della funzione HS calcolato nella mediana per componenti rispetto al reciproco dei gradi di libertà ($1/\nu$), per diversi valori del parametro di forma ($\alpha = 1, 2, 3, 5, 10, \infty$).

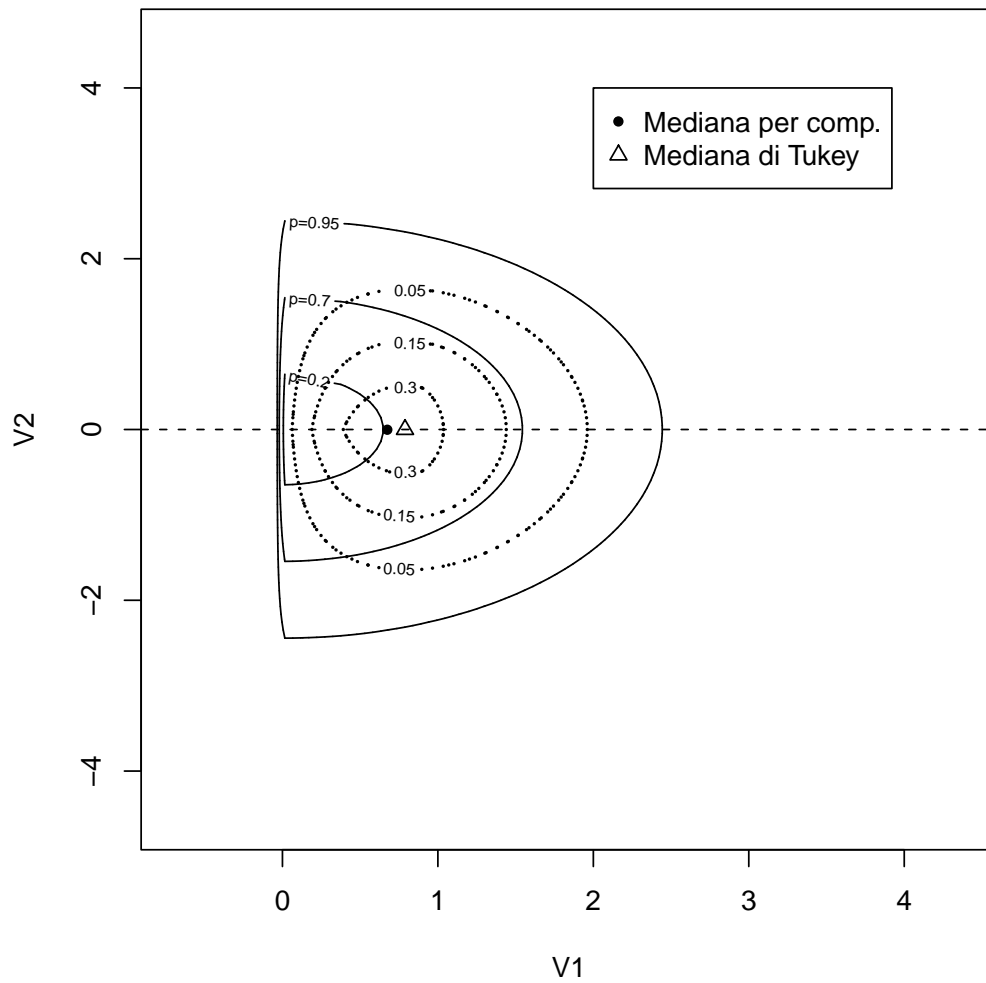


Figura 2.5: Curve di livello della funzione di densità (curve continue) e curve di livello di HS (curve punteggiate) per la distribuzione della variabile casuale $(V_1, V_2) \sim ST_2(0, I_2, (\infty, 0)^T, \infty)$, ossia una normale troncata bidimensionale. Si ha che $\eta = (0.675, 0)^T$ e $M_F = (0.789, 0)^T$.

Tuttavia l'unica differenza esistente tra una distribuzione canonica ed un'altra è la posizione della componente asimmetrica. Ciò non altera l'unicità dell'indice η^* per una data distribuzione ST. Infatti, utilizzando la procedura di costruzione della matrice C^{-1} , così come descritta nella dimostrazione del Lemma 1.2.2, la (2.8) può essere riscritta nel seguente modo

$$\begin{aligned}\eta^* &= \frac{\Omega\omega^{-1}\alpha}{(\alpha^T\bar{\Omega}\alpha)^{1/2}}\eta_0 + \xi \\ &= \left(\frac{1 + \alpha^T\bar{\Omega}\alpha}{\alpha^T\bar{\Omega}\alpha}\right)^{1/2} \eta_0\omega\delta + \xi,\end{aligned}\tag{2.9}$$

dove η_0 è l'unica componente non nulla della mediana per componenti η della distribuzione canonica e

$$\delta = \frac{\bar{\Omega}\alpha}{(1 + \alpha^T\bar{\Omega}\alpha)^{1/2}}.$$

Si nota pertanto che l'indice η^* è indipendente dalla posizione dell'unica componente asimmetrica della distribuzione canonica.

Lo scarto da 1/2 della funzione HS calcolata in η^* segue l'andamento mostrato in Figura 2.4, grazie alla proprietà di invarianza della funzione HS. Un'ulteriore importante proprietà dell'indice η^* è l'equivarianza, come mostrato dal seguente risultato.

Lemma 2.3.6. *Siano $X \sim \text{ST}_d(0, \Omega, \alpha, \nu)$ ed η_X^* l'indice di posizione dato dalla (2.9). Sia inoltre $Y = AX + b$, con A matrice non singolare di dimensioni $d \times d$ e $b \in \mathbb{R}^d$. Allora*

$$\eta_Y^* = A\eta_X^* + b,$$

dove η_Y^* è l'analogo indice di posizione calcolato per Y .

Dimostrazione. Per la scomposizione di Cholesky abbiamo che esiste un'opportuna matrice triangolare inferiore B tale che $\Omega_X = BB^T$. Allora avremo che $\Omega_Y = A\Omega_X A^T = (AB)(AB)^T$. Se C_Y e C_X sono le matrici di passaggio dalla distribuzione di Y ed X alla rispettiva forma canonica, queste allora avranno la seguente forma

$$\begin{aligned}C_Y^{-1} &= P_Y^T(AB)^{-1}, \\ C_X^{-1} &= P_X^T B^{-1}\end{aligned}$$

con P_X e P_Y opportune matrici ortonormali come descritto nella dimostrazione del Lemma 1.2.2.

Notiamo che

$$(AB)^{-1}(Y - b) = B^{-1}X,$$

ma allora $P_X = P_Y = P$. Pertanto alla variabile casuale Y corrisponde la stessa forma canonica di X .

Se η è la mediana della variabile casuale in forma canonica, si ottiene allora che

$$\begin{aligned} \eta_Y^* &= C_Y \eta + b \\ &= AB(P^T)^{-1} \eta + b \\ &= AC_X \eta + b \\ &= A\eta_X^* + b. \end{aligned}$$

□

Approcci simili, in ambito non parametrico, per la costruzione di indici equivarianti sono stati impiegati in Hettmansperger & Randles (2002), Chakraborty, Chaudhuri & Oja (1998) e Chakraborty & Chaudhuri (1996). L'idea è quella di ottenere una misura di posizione sulla scala di una variabile standardizzata del tipo $Z = AX + b$, ottenuta mediante trasformazione affine dei dati, per poi ottenerla sulla scala originaria mediante la trasformazione affine inversa.

Se $\hat{\eta}^*$ è lo stimatore per l'indice η^* ottenuto a partire dalla distribuzione stimata con il metodo della massima verosimiglianza, allora questo avrà distribuzione asintotica normale di media η e varianza $J_{\eta^*} i(\theta)^{-1} J_{\eta^*}^T$ dove J_{η^*} è la matrice jacobiana delle derivate parziali di η^* rispetto a θ , il vettore dei parametri che indicizza la distribuzione ST, e $i(\theta)$ è l'informazione attesa di Fisher per θ . Ponendo $\alpha^* = (\alpha^T \bar{\Omega} \alpha)^{1/2}$ e denotando con η_i^* l' i -esima componente di η^* , la matrice J_{η^*} è costituita dalle seguenti componenti

$$\begin{aligned} \left[\frac{\partial \eta_i^*}{\partial \xi_j} \right]_{i \leq d, j \leq d} &= I_d, \\ \left[\frac{\partial \eta_i^*}{\partial \alpha_j} \right]_{i \leq d, j \leq d} &= \Omega \omega^{-1} \left[I_d \eta_0 (\alpha^*)^{-1} + \alpha \alpha^T \bar{\Omega} \left(\frac{\partial \eta_0}{\partial \alpha^*} (\alpha^*)^{-2} - \eta_0 (\alpha^*)^{-3} \right) \right]. \end{aligned}$$

Sia

$$c(\eta_0, \alpha^*) = -\frac{1}{2} (\alpha^*)^{-2} \left((\alpha^*)^{-1} \eta_0 - \frac{\partial \eta_0}{\partial \alpha^*} \right).$$

Denotando con ω_{ij} l'elemento in posizione (i, j) nella matrice Ω , le espressioni di $\partial\eta_i^*/\partial\omega_{hh}$ e $\partial\eta_i^*/\partial\omega_{hk}$ per $1 \leq h < k \leq d$ sono

$$\begin{aligned} \frac{\partial\eta_i^*}{\partial\omega_{hh}} &= \frac{\alpha_h}{2\alpha^*\sqrt{\omega_{hh}}}\eta_0 - \frac{\alpha_h}{\omega_{hh}^{3/2}}c(\eta_0, \alpha^*) \sum_{j=1}^d \frac{\omega_{ij}}{\sqrt{\omega_{jj}}}\alpha_j \sum_{j \neq h} \frac{\omega_{hj}}{\sqrt{\omega_{jj}}}\alpha_j, \text{ se } h = i, \\ \frac{\partial\eta_i^*}{\partial\omega_{hh}} &= -\frac{\omega_{ih}\alpha_h}{2\alpha^*\omega_{hh}^{3/2}}\eta_0 - \frac{\alpha_h}{\omega_{hh}^{3/2}}c(\eta_0, \alpha^*) \sum_{j=1}^d \frac{\omega_{ij}}{\sqrt{\omega_{jj}}}\alpha_j \sum_{j \neq h} \frac{\omega_{hj}}{\sqrt{\omega_{jj}}}\alpha_j, \text{ se } h \neq i, \\ \frac{\partial\eta_i^*}{\partial\omega_{hk}} &= \frac{\alpha_k}{\alpha^*\sqrt{\omega_{kk}}}\eta_0 + 2\frac{\alpha_h\alpha_k}{\sqrt{\omega_{hh}\omega_{kk}}}c(\eta_0, \alpha^*) \sum_{j=1}^d \frac{\omega_{ij}}{\sqrt{\omega_{jj}}}\alpha_j, \text{ se } h = i, \\ \frac{\partial\eta_i^*}{\partial\omega_{hk}} &= \frac{\alpha_h}{\alpha^*\sqrt{\omega_{hh}}}\eta_0 + 2\frac{\alpha_h\alpha_k}{\sqrt{\omega_{hh}\omega_{kk}}}c(\eta_0, \alpha^*) \sum_{j=1}^d \frac{\omega_{ij}}{\sqrt{\omega_{jj}}}\alpha_j, \text{ se } h \neq i \text{ e } k = i, \\ \frac{\partial\eta_i^*}{\partial\omega_{hk}} &= 2\frac{\alpha_h\alpha_k}{\sqrt{\omega_{hh}\omega_{kk}}}c(\eta_0, \alpha^*) \sum_{j=1}^d \frac{\omega_{ij}}{\sqrt{\omega_{jj}}}\alpha_j, \text{ se } h \neq i \text{ e } k \neq i, \end{aligned}$$

ciascuna per $i \leq d$. Infine abbiamo

$$\frac{\partial\eta_i^*}{\partial\nu} = \alpha^* \frac{\partial\eta_0}{\partial\nu} \sum_{j=1}^d \frac{\omega_{ij}}{\sqrt{\omega_{jj}}}\alpha_j, \text{ per } i \leq d.$$

Come ricavare delle espressioni per $\partial\eta_0/\partial\alpha^*$ e $\partial\eta_0/\partial\nu$, sar  discusso in §2.4.

Analizziamo ora alcune propriet  della funzione HS nel caso della distribuzione ST in forma canonica e dalle quali deriveremo ulteriori propriet  nel caso di una generica distribuzione ST.

Prima di procedere oltre, definiamo la trasformazione di simmetria rispetto agli assi dello spazio \mathbb{R}^d . Supponiamo inoltre che gli assi, rispetto ai quali operiamo la trasformazione di simmetria, siano dati dall'insieme di indici J . Il punto $x \in \mathbb{R}^d$ verr  allora trasformato nel punto $x' \in \mathbb{R}^d$ nel seguente modo

$$\begin{cases} x'_i = x_i, & \text{se } i \notin J \\ x'_i = -x_i, & \text{se } i \in J \end{cases}$$

dove x_i ed x'_i sono rispettivamente l' i -esima componente di x ed x' .

Teorema 2.3.7. *Sia $X \sim \text{ST}_d(0, I_d, \alpha^*, \nu)$ con α^* avente l' i -esima componente non nulla e tutte le altre nulle. Dati i punti $x_0, x'_0 \in \mathbb{R}^d$ tali che l' i -esima componente sia uguale e le altre componenti differiscano al più solamente per il segno, allora si ha che*

$$\text{HS}(x_0) = \text{HS}(x'_0). \quad (2.10)$$

Dimostrazione. Supponiamo, senza perdita di generalità, che la prima componente di α^* sia l'unica non nulla. I punti x_0 ed x'_0 sono in corrispondenza tra loro mediante trasformazione di simmetria rispetto agli assi delle componenti per le quali x'_0 ha segno opposto rispetto ad x_0 .

Consideriamo ora un generico iperpiano π_0 passante per x_0 , ossia

$$\pi_0 : u_1(x_1 - x_{01}) + u_2(x_2 - x_{02}) + \dots + u_d(x_d - x_{0d}) = 0,$$

dove x_{0i} è l' i -esima componente di x_0 ed $u_i \in [-1, 1]$ per ogni $i \leq d$, tale che $\sum_{i=1}^d u_i^2 = 1$.

Supponiamo che J sia l'insieme degli indici delle componenti di x'_0 per le quali avviene il cambio di segno rispetto alle componenti di x_0 . Si noti dunque che $\{1\} \notin J$ e che $J \subseteq \{2, \dots, d\}$, poiché la prima componente di x_0 è uguale a quella di x'_0 per ipotesi. L'equazione dell'iperpiano π'_0 , ottenuto mediante simmetria dell'iperpiano π_0 rispetto agli assi delle componenti in J e passante per x'_0 , è la seguente

$$\pi'_0 : \sum_{i \notin J} u_i(x_i - x_{0i}) - \sum_{i \in J} u_i(x_i + x_{0i}) = 0.$$

Abbiamo dunque che

$$\begin{aligned} \text{HS}(x'_0) &= \min_{\|u\|=1} P \left(\sum_{i \notin J} u_i(X_i - x_{0i}) - \sum_{i \in J} u_i(X_i + x_{0i}) < 0 \right) \\ &= \min_{\|u\|=1} P \left(\sum_{i=1}^d u_i(X_i - x_{0i}) < 0 \right) = \text{HS}(x_0) \end{aligned}$$

poichè

$$\begin{pmatrix} X_{\bar{J}} \\ X_J \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} X_{\bar{J}} \\ -X_J \end{pmatrix},$$

dove $X_{\bar{J}} = (X_i)_{i \notin J}$ e $X_J = (X_i)_{i \in J}$. □

Il successivo risultato sarà in seguito rilevante nel caratterizzare la mediana di Tukey per la distribuzione ST.

Lemma 2.3.8. *Sia $X \sim \text{ST}_d(\xi, \Omega, \alpha, \nu)$, avente moda μ_m , indice η^* e, se $\nu > 1$, media μ . Allora ξ, η^*, μ ed μ_m sono allineati lungo una retta.*

Dimostrazione. Consideriamo dapprima il caso in cui X è in forma canonica ed assumiamo, senza perdita di generalità, che la prima componente di X sia l'unica componente asimmetrica. Pertanto, il vettore delle medie avrà al massimo una componente non nulla, corrispondente alla prima.

Per quanto riguarda μ_m osserviamo che questa, per la Proposizione 7 in Azzalini & Regoli (2012), è unica. Abbiamo allora che μ_m ha al massimo una componente non nulla, corrispondente anche in questo caso alla prima. Se così non fosse allora esisterebbe un punto $x \in \mathbb{R}^d$ con la prima componente uguale a quella di μ_m e le altre componenti non nulle che differiscono al più solamente per il segno rispetto a quelle di μ_m , tale che $f_X(x) = f_X(\mu_m)$, dove f_X è la funzione di densità di X . Si avrebbe pertanto una contraddizione.

Analogamente, l'indice η^* non è altro che la mediana per componenti η della ST in forma canonica, poiché nella (2.9) si ha che $C = I_d$ e $\xi = 0$. Pertanto anche tale indice ha al massimo la prima componente non nulla.

Sfruttando infine la proprietà di equivarianza degli indici ξ, η^*, μ e μ_m , abbiamo che, nel caso di trasformazioni affini, l'allineamento di tali indici si conserva. \square

Grazie a questo risultato possiamo inoltre meglio comprendere se l'indice η^* e la mediana per componenti di una generica distribuzione ST coincidono. Ciò che si osserva è che generalmente tale coincidenza non è verificata. In Figura 2.6 è mostrato un controesempio: data la variabile casuale

$$\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \sim \text{ST}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \end{pmatrix}, 4 \right)$$

si ha non soltanto che i due indici in questione non coincidono ma che la mediana per componenti non è allineata assieme agli altri indici. L'unico caso di coincidenza dei due indici si ha quando consideriamo la distribuzione SC per la quale la mediana di Tukey, l'indice η^* e la mediana per componenti convergono tutti e tre nel centro di simmetria radiale.

L'unicità della mediana di Tukey, nel caso della distribuzione ST, risulta una questione aperta nel caso in cui la dimensione dello spazio è $d > 2$ e $\nu \neq 1$. Nel caso in cui $\nu = 1$, infatti, abbiamo quanto segue.

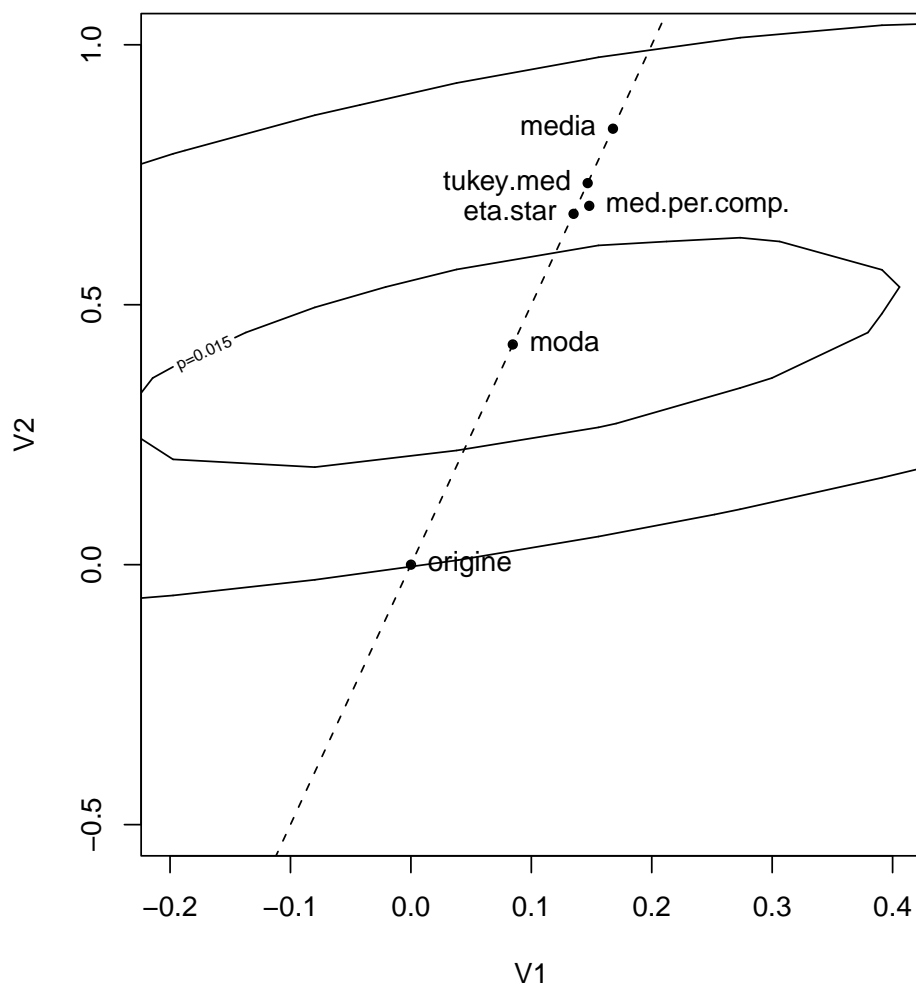


Figura 2.6: Diversi indici di posizione per una distribuzione ST.

Corollario 2.3.9. *Sia $X \sim SC_d(\xi, \Omega, \alpha)$. Allora la mediana di Tukey M_F è unica e giace, insieme al parametro di posizione ξ e la moda μ_m , lungo una stessa retta.*

Dimostrazione. Per il Teorema 2.3.5 abbiamo che X è radialmente simmetrica. La mediana di Tukey è pertanto il centro di simmetria radiale della distribuzione che è unico e coincide con la mediana per componenti. Pertanto abbiamo che $\eta = \eta^* = M_F$. Con una analoga argomentazione del Lemma 2.3.8, si ha che per l'equivarianza dell'indice M_F , allineamento con gli altri indici di posizione si conserva anche per una qualsiasi distribuzione ST. \square

Resta ora da caratterizzare la mediana di Tukey nel caso in cui $\nu \neq 1$. Sappiamo già che la regione di massima *depth* sarà convessa (Small, 1987) e, se la distribuzione ST è in forma canonica, simmetrica rispetto agli assi delle componenti simmetriche della distribuzione, per il Teorema 2.3.7.

Il successivo risultato stabilisce inoltre che la funzione HS è unicamente massimizzata lungo la retta data dal Lemma 2.3.8. Tale punto sarà considerato, nel successivo sviluppo della trattazione, come la mediana di Tukey della distribuzione ST.

Teorema 2.3.10. *Sia X una variabile aleatoria d -dimensionale con distribuzione ST, avente regione di massima *depth**

$$R = \{x \in \mathbb{R}^d : HS(x) \geq HS(y), \forall y \in \mathbb{R}^d\}.$$

Allora esiste un unico punto $v \in \mathbb{R}^d$, che giace lungo la retta data dal Lemma 2.3.8, tale che $v \in R$. Inoltre se $d = 2$ allora la regione R coincide con v .

Dimostrazione. Assumiamo senza perdita di generalità che X sia in forma canonica e che la prima componente sia l'unica asimmetrica con parametro di forma α .

Nel caso in cui $d = 2$ l'unicità della mediana di Tukey M_F è data dalla Proposizione 3.4 di Small (1987), ossia è sufficiente richiedere che la funzione di densità sia strettamente positiva ovunque, condizione soddisfatta dalla funzione di densità della ST. Inoltre M_F , nel caso della distribuzione in forma canonica, ha solamente la prima componente non nulla poiché se avesse più di una componente nulla allora per il Teorema 2.3.7 avremmo che esiste un punto $x \in \mathbb{R}^2$ tale che $HS(x) = HS(M_F)$, giungendo pertanto ad una contraddizione.

Nel caso in cui $d > 2$ consideriamo il problema di massimizzazione della funzione HS lungo gli assi delle componenti simmetriche. Abbiamo che

$$\begin{aligned} & \arg \max_{x \in \mathbb{R}} \left\{ \min_{\|u\|=1} P \left(u_1 X_1 + \sum_{i=2}^d u_i X_i < u_1 x \right) \right\} \\ = & \arg \max_{x \in \mathbb{R}} \left\{ \min_{u_1 \in [-1,1]} P \left(u_1 X_1 + \sqrt{1 - u_1^2} T < u_1 x \right) \right\} \end{aligned} \quad (2.11)$$

poiché $u_1 X_1 + \sum_{i=2}^d u_i X_i \stackrel{d}{=} u_1 X_1 + \sqrt{1 - u_1^2} T$, con

$$(X_1, T) \sim \text{ST}_2(0, I_2, (\alpha, 0)^T, \nu).$$

Pertanto il problema di ottimizzazione, posto dalla (2.11), per $d > 2$ è equivalente al caso in cui $d = 2$, dove l'unicità della mediana di Tukey ed il fatto che questa giaccia lungo l'asse della componente simmetrica ci garantiscono che la (2.11) ha un unico massimo. Tale punto appartiene alla regione di massima *depth*, in virtù della convessità e della simmetria rispetto agli assi delle componenti simmetriche, di tale regione. \square

2.4 La varianza dello stimatore quantile

Consideriamo ora il problema della stima di η e, a tal fine, il problema più generale della stima del quantile r -esimo $F^{-1}(r) = F_r^{-1}$, denotata con \hat{F}_r^{-1} ed ottenuta dalla distribuzione stimata con il metodo della massima verosimiglianza, dove $F^{-1}(\cdot)$ è la funzione quantile della distribuzione F della variabile casuale unidimensionale X assolutamente continua. Supponiamo inoltre, più in generale, di dover fare inferenza sul modello statistico $\mathcal{F} = \{f(x; \lambda) : x \in S_X \subseteq \mathbb{R}^d, \lambda \in \Lambda \subseteq \mathbb{R}^p\}$ dove $f(\cdot; \lambda)$ è la funzione di densità indicizzata dal parametro p -dimensionale λ , S_X è il supporto della variabile casuale X e Λ lo spazio parametrico. Assumiamo che siano valide le usuali condizioni di regolarità (Azzalini, 2001, pp. 76-77). Indicheremo in seguito con $\text{var}(\hat{F}_r^{-1})$ la varianza della distribuzione asintotica dello stimatore \hat{F}_r^{-1} . Abbiamo pertanto, secondo la teoria asintotica della massima verosimiglianza, che

$$\hat{F}_r^{-1} \sim N \left(F_r^{-1}, \text{var} \left(\hat{F}_r^{-1} \right) \right).$$

Abbiamo inoltre che

$$\int_{-\infty}^{F_r^{-1}} f(x; \lambda) dx - p = 0$$

definisce la funzione implicita $F_r^{-1} : \Lambda \rightarrow \mathbb{R}$.

$$\text{var} \left(\hat{F}_r^{-1} \right) = \nabla F_r^{-1}(\lambda)^T i(\lambda)^{-1} \nabla F_r^{-1}(\lambda) \quad (2.12)$$

dove $\nabla F_r^{-1}(\lambda)$ rappresenta il gradiente della funzione implicita F_r^{-1} calcolata in λ ed $i(\lambda)$ l'informazione attesa di Fisher per il parametro λ . Per il Teorema di Dini abbiamo la seguente espressione per il vettore gradiente

$$\nabla F_p^{-1}(\lambda) = - \left(\frac{h(\lambda_i)}{f(F_p^{-1}(\lambda); \lambda)} \right)_{i \leq p} \quad (2.13)$$

dove

$$h(\lambda_i) = \int_{-\infty}^{F_r^{-1}(\lambda)} \frac{\partial f(x; \lambda)}{\partial \lambda_i} dx.$$

Grazie alla (2.13) siamo in grado di un'ottenere un'espressione per la varianza della distribuzione asintotica dello stimatore $\hat{\eta}$, ossia la mediana calcolata sulla distribuzione stimata con il metodo della massima verosimiglianza, ponendo $r = 1/2$.

Capitolo 3

Misure matriciali di concordanza

Discuteremo in questo capitolo come sintetizzare la struttura di dipendenza e la variabilità di una distribuzione di probabilità multidimensionale. Ricorremo a tal fine al concetto di associazione per coppie di variabili (X, Y) ed alle relative misure di concordanza che hanno avuto ampio seguito nella statistica non parametrica. L'obiettivo è quello di pervenire ad una matrice di concordanza che sia in grado di supplire alla mancanza di una matrice di covarianza nel caso di distribuzioni multidimensionali prive dei momenti del secondo ordine. Indicheremo in seguito il dominio di una funzione f con $\text{Dom}f$ e l'insieme immagine di f con $\text{Ran}f$.

3.1 Indici di concordanza

La concordanza è un concetto relativo all'associazione tra coppie di variabili casuali. In termini non formali, una coppia di variabili è *concordante* se “grandi” valori di una variabile sono associati a “grandi” valori dell'altra o, analogamente, “piccoli” valori di una sono associati a “piccoli” valori dell'altra. Viceversa una coppia di variabili è *discordante* se “grandi” valori di una variabile sono associati a “piccoli” valori dell'altra. Questo può essere tradotto in termini matematici nel seguente modo. Siano (x_i, y_i) e (x_j, y_j) due coppie di valori provenienti dalla distribuzione bidimensionale (X, Y) . Allora (x_i, y_i) ed (x_j, y_j) vengono detti concordanti se $(x_i - x_j)(y_i - y_j) > 0$ o discordanti se $(x_i - x_j)(y_i - y_j) < 0$.

Le due più note misure di concordanza sono la τ di Kendall e la ρ di Spearman. La τ di Kendall nella versione della popolazione è definita come la probabilità di concordanza meno la probabilità di discordanza. Siano (X_1, Y_1) e (X_2, Y_2) due vettori casuali indipendenti ed identicamente distribuiti, allora abbiamo che

$$\begin{aligned} \tau(X, Y) &= P[(X_1 - X_2)(Y_1 - Y_2) > 0] \\ &\quad - P[(X_1 - X_2)(Y_1 - Y_2) < 0]. \end{aligned} \quad (3.1)$$

Per quanto riguarda l'espressione della ρ di Spearman nella popolazione, consideriamo ora i vettori casuali (X_1, Y_1) , (X_2, Y_2) e (X_3, Y_3) indipendenti ed identicamente distribuiti. Abbiamo che

$$\begin{aligned} \rho(X, Y) &= 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] \\ &\quad - P[(X_1 - X_2)(Y_1 - Y_3) < 0]). \end{aligned} \quad (3.2)$$

Si noti che la coppia (X_3, Y_2) potrebbe essere ugualmente impiegata in questa espressione al posto della coppia (X_2, Y_3) . Il fattore 3 nella (3.2) è una costante di normalizzazione affinché $\rho(X, Y) \in [-1, 1]$.

Come vedremo questi due indici di concordanza sono esprimibili in termini della cosiddetta *copula* della distribuzione congiunta di (X, Y) . Il termine “copula” è un sostantivo latino che significa “legame” ed in statistica viene infatti usato per indicare quelle funzioni che esprimono il legame esistente tra le funzioni di ripartizione delle distribuzioni marginali di una coppia di variabili casuali e che definiscono, per mezzo di questo legame, la funzione di ripartizione della distribuzione congiunta (Nelsen, 2006, pp. 1-5). La funzione di ripartizione della distribuzione congiunta di un vettore casuale può essere dunque scritta in termini delle funzioni di ripartizione delle distribuzioni marginali ed una copula. In seguito considereremo soltanto copule nel caso bidimensionale, in quanto l'obiettivo di questo capitolo è definire misure di concordanza per coppie di variabili.

Definizione 3.1.1. *Una copula è una funzione $C : I \times I \rightarrow I$, dove $I = [0, 1]$, con le seguenti proprietà.*

1. Per ogni $u, v \in I$,

$$C(u, 0) = C(0, v) = 0$$

e

$$C(u, 1) = u \text{ e } C(1, v) = v.$$

2. Per ogni $u_1, v_1, u_2, v_2 \in I$, tale che $u_1 \leq u_2$ e $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

La seconda proprietà caratteristica di una copula sta ad indicare che questa è una funzione “quasi-monotona”. Le funzioni quasi-monotone sono quelle funzioni bidimensionali H , tali che, per ogni rettangolo $B = [x_1, y_1] \times [x_2, y_2]$ i cui vertici sono in $\text{Dom}H$, si ha

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1) \geq 0,$$

dove $V_H(B)$ è l' H -volume di B . Dalla proprietà di quasi-monotonia, insieme al fatto che $C(u, 0) = C(0, v) = 0$, si ha che la copula C è non decrescente in ogni argomento, ossia per due dati valori $a, b \in [0, 1]$ abbiamo che $C(u, a)$ e $C(b, v)$ sono funzioni non decrescenti, per ogni $a, b \in [0, 1]$.

Il risultato, che dà senso all'impiego del termine “copula” per tali tipi di funzioni, è fornito dal Teorema di Sklar.

Teorema 3.1.2 (Sklar). *Sia H la funzione di ripartizione congiunta del vettore casuale (X, Y) , con funzioni di ripartizione marginali rispettivamente F e G . Allora esiste una copula C tale che per ogni $x, y \in \mathbb{R} \cup \{-\infty, \infty\}$,*

$$H(x, y) = C(F(x), G(x)). \quad (3.3)$$

Se F e G sono continue allora C è unica; altrimenti, C è unicamente determinata in $\text{Ran}F \times \text{Ran}G$. Viceversa, se C è una copula ed F e G sono funzioni di ripartizione, allora la funzione H definita in (3.3) è una funzione di ripartizione congiunta con funzioni di ripartizioni marginali F e G .

Un altro risultato di cui faremo uso per costruire copule a partire da una data distribuzione bidimensionale è il seguente.

Corollario 3.1.3. *Siano H , F , G e C come definite nel Teorema 3.1.2, con F e G continue. Siano inoltre F^{-1} e G^{-1} le funzioni quantile, ossia le funzioni inverse rispettivamente di F e G . Allora per ogni $u, v \in \text{Dom}C$ si ha che*

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)). \quad (3.4)$$

Siamo ora in grado di definire la τ di Kendall e la ρ di Spearman in termini della copula della distribuzione di (X, Y) . Ebbene abbiamo che (Nelsen, 2006, pp. 161-162 e pp. 169-170)

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1, \quad (3.5)$$

$$\rho(X, Y) = 12 \int_0^1 \int_0^1 uv dC(u, v) - 3, \quad (3.6)$$

o equivalentemente

$$\tau(X, Y) = 1 - 4 \int_0^1 \int_0^1 \frac{\partial C(u, v)}{\partial u} \frac{\partial C(u, v)}{\partial v} dudv, \quad (3.7)$$

$$\rho(X, Y) = 12 \int_0^1 \int_0^1 C(u, v) dudv - 3. \quad (3.8)$$

Si noti che il doppio integrale nella (3.5) può essere interpretato come il valore atteso della funzione $C(U, V)$, dove U e V sono variabili casuali uniformi in $(0, 1)$ la cui funzione di ripartizione congiunta è C , ossia

$$\tau(X, Y) = 4E[C(U, V)] - 1. \quad (3.9)$$

Il coefficiente ρ di Spearman nella versione della popolazione può essere equivalentemente definito come il coefficiente di correlazione dei ranghi nel continuo, ossia le variabili $U = F(X)$ e $V = G(Y)$. Ricordando infatti che il valore atteso di una variabile casuale uniforme in $(0, 1)$ è $1/2$ e la varianza è $1/12$, possiamo riscrivere la (3.6) nel seguente modo

$$\begin{aligned} \rho(X, Y) &= 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 = 12E[UV] - 3 \\ &= \frac{E[UV] - 1/4}{1/12} = \frac{E[UV] - E[U]E[V]}{\sqrt{V(U)V(V)}} \\ &= r(U, V), \end{aligned} \quad (3.10)$$

dove $r(., .)$ è il coefficiente di correlazione di Pearson.

La τ di Kendall e la ρ di Spearman posseggono inoltre tutta una serie di proprietà che possono essere sintetizzate nella seguente definizione di indice di concordanza fornita da Scarsini (1984).

Definizione 3.1.4. *Una misura numerica κ di associazione tra due variabili casuali continue X ed Y , la cui copula è C , è una misura di concordanza se soddisfa le seguenti proprietà*

P1 κ è definita per ogni coppia X, Y di variabili casuali continue;

P2 $-1 \leq \kappa \leq 1$, $\kappa(X, X) = 1$ e $\kappa(X, -X) = -1$;

P3 $\kappa(X, Y) = \kappa(Y, X)$;

P4 se X ed Y sono indipendenti allora $\kappa(X, Y) = 0$;

P5 $\kappa(-X, Y) = \kappa(X, -Y) = -\kappa(X, Y)$;

P6 se C_1 e C_2 sono due distinte copule, tali che $C_1(u, v) \leq C_2(u, v)$ per ogni $(u, v) \in [0, 1] \times [0, 1]$ allora $\kappa_1 \leq \kappa_2$;

P7 se $\{(X_n, Y_n)\}$ è una sequenza di variabili casuali continue con copule C_n , e se C_n converge puntualmente a C , allora $\lim_{n \rightarrow \infty} \kappa_{C_n} = \kappa_C$.

Dalla Definizione 3.1.4 si ottengono tre ulteriori proprietà che sono

P8 se Y è una funzione strettamente crescente di X allora $\kappa(X, Y) = 1$;

P9 se Y è una funzione strettamente decrescente di X allora $\kappa(X, Y) = -1$;

P10 se α e β sono funzioni strettamente monotone in $\text{Ran}(X)$ e $\text{Ran}(Y)$, rispettivamente, allora $\kappa(\alpha(X), \beta(Y)) = \kappa(X, Y)$.

I due indici di concordanza τ e ρ , qui considerati, rispettano tali proprietà e possono essere trattati a tutti gli effetti come misure di concordanza; tale risultato è fornito dal Teorema 5.1.9 in Nelsen (2006). In Taylor (2007) la definizione di misura di concordanza fornita da Scarsini (1984) viene generalizzata al caso multidimensionale per $d > 2$ e si ha che per ogni vettore casuale X d -dimensionale viene definita una funzione $\kappa(X)$ a valori reali che soddisfa analoghe proprietà della Definizione 3.1.4 generalizzate nello spazio multidimensionale.

Nel caso della famiglia ellittica gli indici τ di Kendall e ρ di Spearman sono esprimibili in forma esplicita come funzione del coefficiente di correlazione sebbene soltanto l'indice τ mantiene la stessa espressione per ogni distribuzione della famiglia ellittica (Hult & Lindskog, 2002).

Teorema 3.1.5. *Sia $X \sim E_d(\mu, \Sigma, g)$, dove per ogni $i, j \in \{1, \dots, d\}$ X_i ed X_j sono continue e $\text{var}(X_h) < \infty$, per ogni $h \leq d$, allora*

$$\tau(X_i, X_j) = \frac{2}{\pi} \arcsin r(X_i, X_j). \quad (3.11)$$

Corollario 3.1.6. *Sia $X \sim N_d(\mu, \Sigma)$, per ogni $i, j \in \{1, \dots, d\}$ si ha*

$$\rho(X_i, X_j) = \frac{6}{\pi} \arcsin(r(X_i, X_j)/2). \quad (3.12)$$

Dalla (3.11) e dalla (3.12) abbiamo che, nel caso della distribuzione normale, la relazione tra i due coefficienti risulta essere

$$\rho(X, Y) = \frac{6}{\pi} \arcsin \left\{ \frac{1}{2} \sin \left(\frac{\pi}{2} \tau(X, Y) \right) \right\}.$$

Sebbene gli indici τ di Kendall e ρ di Spearman misurino lo stesso attributo distributivo, ossia l'associazione tra variabili, risultano spesso essere non poco differenti tra loro (Capéraà & Genest, 1993). In generale abbiamo le seguenti disuguaglianze (Nelsen, 2006, p. 177)

$$\frac{3\tau - 1}{2} \leq \rho \leq \frac{1 + 2\tau - \tau^2}{2}, \text{ se } \tau \geq 0, \quad (3.13)$$

$$\frac{\tau^2 + 2\tau - 1}{2} \leq \rho \leq \frac{1 + 3\tau}{2}, \text{ se } \tau \leq 0. \quad (3.14)$$

In Figura 3.1 è rappresentata la regione τ - ρ risultante da tali vincoli. Approfondimenti sulla relazione tra τ di Kendall e ρ di Spearman per coppie di variabili continue sono forniti in Fredricks & Nelsen (2007).

Versioni multidimensionali degli indici τ e ρ sono state proposte rispettivamente in Choi & Marden (1998) e Schmid & Schmidt (2007). Choi & Marden (1998) propongono una versione vettoriale della τ di Kendall in un contesto di regressione multipla tra una variabile unidimensionale Y , la variabile risposta, ed il vettore di variabili esplicative X . In Schmid & Schmidt (2007) vengono presentate tre possibili generalizzazioni della ρ di Spearman al caso multidimensionale per $d > 2$ e che coincidono per $d = 2$, presentando inoltre stime non parametriche attraverso la copula empirica.

Consideriamo ora il problema di stima degli indici di associazione fin qui studiati nel caso in cui $(X, Y) \sim \text{ST}_2(\xi, \Omega, \alpha, \nu)$. Notiamo subito che sia τ che ρ non dipendono ξ poiché sono invarianti rispetto a traslazioni.

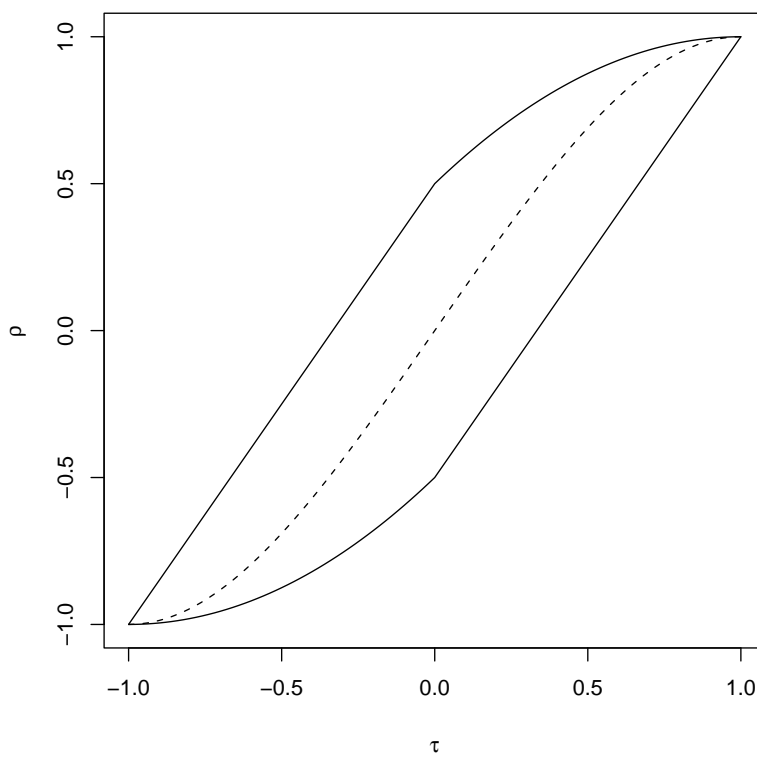


Figura 3.1: Regione τ - ρ risultante dalle disuguaglianze (3.13) e (3.14) (curve continue); riportata anche la relazione funzionale tra i due indici nel caso normale (curva tratteggiata).

Sfruttando l'invarianza rispetto a trasformazioni di scala, abbiamo che τ e ρ sono funzione di $\bar{\omega} = \omega_{12}/\sqrt{\omega_{11}\omega_{22}}$, $\alpha = (\alpha_1, \alpha_2)$ e ν .

Denoteremo inoltre $\theta = (\xi_1, \xi_2, \omega_{11}, \omega_{22}, \bar{\omega}, \alpha_1, \alpha_2, \nu)$, $\lambda = (\bar{\omega}, \alpha_1, \alpha_2, \nu)$ e $\tilde{\lambda} = (\alpha_1, \alpha_2, \bar{\omega})$. La varianza della distribuzione asintotica degli stimatori di ρ e τ , ottenuti a partire dai parametri della distribuzione stimata con il metodo della massima verosimiglianza, sono rispettivamente

$$\begin{aligned} \nabla\rho(\lambda)^T \left[(i(\theta)^{-1})_{ij} \right]_{5 \leq i \leq 8, 5 \leq j \leq 8} & \nabla\rho(\lambda), \\ \nabla\tau(\lambda)^T \left[(i(\theta)^{-1})_{ij} \right]_{5 \leq i \leq 8, 5 \leq j \leq 8} & \nabla\tau(\lambda), \end{aligned}$$

con $i(\theta)$ matrice dell'informazione attesa di Fisher per il parametro θ .

Forniamo ora le espressioni degli indici ρ e τ che verranno utilizzate per la distribuzione ST. A partire dalla (3.5) e dalla (3.6), esprimiamo i due indici in termini di integrali nelle variabili x ed y operando la trasformazione $x = F_X^{-1}(u)$ e $y = F_Y^{-1}(v)$, dove F_X^{-1} ed F_Y^{-1} sono le funzioni quantile delle distribuzioni marginali di X ed Y rispettivamente. Indicheremo in seguito con F , f e l rispettivamente la funzione di ripartizione, la funzione di densità ed il logaritmo di quest'ultima per la variabile casuale $ST_2(0, \bar{\Omega}, (\alpha_1, \alpha_2)^T, \nu)$, con $\bar{\omega}$ elemento sulla diagonale secondaria di $\bar{\Omega}$. Sfruttando il risultato del Corollario 3.1.3 si ha che

$$\begin{aligned} \tau &= 4 \int \int_{\mathbb{R}^2} F(x, y; \lambda) dF(x, y; \lambda) - 1 \\ &= 4 \int \int_{\mathbb{R}^2} F(x, y; \lambda) \frac{\partial^2 F(x, y; \lambda)}{\partial x \partial y} dx dy - 1 \\ &= 4 \int \int_{\mathbb{R}^2} F(x, y; \lambda) f(x, y; \lambda) dx dy - 1 \end{aligned} \quad (3.15)$$

e

$$\begin{aligned} \rho &= 12 \int \int_{\mathbb{R}^2} F_X(x; \alpha_X(\tilde{\lambda}), \nu) F_Y(y; \alpha_Y(\tilde{\lambda}), \nu) dF(x, y; \lambda) - 3 \\ &= 12 \int \int_{\mathbb{R}^2} F_X(x; \alpha_X(\tilde{\lambda}), \nu) F_Y(y; \alpha_Y(\tilde{\lambda}), \nu) \frac{\partial^2 F(x, y; \lambda)}{\partial x \partial y} dx dy - 3 \\ &= 12 \int \int_{\mathbb{R}^2} F_X(x; \alpha_X(\tilde{\lambda}), \nu) F_Y(y; \alpha_Y(\tilde{\lambda}), \nu) f(x, y; \lambda) dx dy - 3 \end{aligned} \quad (3.16)$$

dove $\alpha_X(\tilde{\lambda})$ ed $\alpha_Y(\tilde{\lambda})$, i parametri di forma delle distribuzioni marginali di X ed Y , sono funzioni a valori reali del vettore $\tilde{\lambda}$; tali equazioni sono ottenibili

dalla (20) di Azzalini (2005). Più esattamente abbiamo

$$\begin{aligned}\alpha_X &= \frac{\alpha_1 + \bar{\omega}\alpha_2}{\sqrt{1 + (1 - \bar{\omega}^2)^2\alpha_2^2}} \\ \alpha_Y &= \frac{\alpha_2 + \bar{\omega}\alpha_1}{\sqrt{1 + (1 - \bar{\omega}^2)^2\alpha_1^2}}.\end{aligned}$$

Si noti che in base alle assunzioni fatte $X \sim \text{ST}_1(0, 1, \alpha_X(\tilde{\lambda}), \nu)$ e $Y \sim \text{ST}_1(0, 1, \alpha_Y(\tilde{\lambda}), \nu)$. Le componenti del vettore $\nabla\rho(\lambda)$ risultano essere

$$\begin{aligned}\frac{\partial\rho}{\partial\bar{\omega}} &= 12 \left[\int \int_{\mathbb{R}^2} \left(\frac{\partial F_X(x; \alpha_X(\tilde{\lambda}), \nu)}{\partial\alpha_X(\tilde{\lambda})} \frac{\partial\alpha_X(\tilde{\lambda})}{\partial\bar{\omega}} F_Y(y; \alpha_Y(\tilde{\lambda}), \nu) \right. \right. \\ &\quad \left. \left. + \frac{\partial F_Y(y; \alpha_Y(\tilde{\lambda}), \nu)}{\partial\alpha_Y(\tilde{\lambda})} \frac{\partial\alpha_Y(\tilde{\lambda})}{\partial\bar{\omega}} F_X(x; \alpha_X(\tilde{\lambda}), \nu) \right. \right. \\ &\quad \left. \left. + F_X(x; \alpha_X(\tilde{\lambda}), \nu) F_Y(y; \alpha_Y(\tilde{\lambda}), \nu) \frac{\partial l(x, y; \lambda)}{\partial\bar{\omega}} \right) f(x, y; \lambda) dx dy \right], \quad (3.17)\end{aligned}$$

dove

$$\begin{aligned}\frac{\partial F_X(x; \alpha_X, \nu)}{\partial\alpha_X} &= \int_{-\infty}^x 2s \sqrt{\frac{\nu+1}{s^2+\nu}} t_1(s; \nu) t_1\left(\alpha_X s \sqrt{\frac{\nu+1}{s^2+\nu}}; \nu+1\right) ds \\ \frac{\partial F_Y(y; \alpha_Y, \nu)}{\partial\alpha_Y} &= \int_{-\infty}^y 2t \sqrt{\frac{\nu+1}{t^2+\nu}} t_1(t; \nu) t_1\left(\alpha_Y t \sqrt{\frac{\nu+1}{t^2+\nu}}; \nu+1\right) dt \\ \frac{\partial\alpha_X(\tilde{\lambda})}{\partial\bar{\omega}} &= \frac{\alpha_2(2\alpha_1\alpha_2\bar{\omega}(1-\bar{\omega}^2) + \alpha_2^2(1-\bar{\omega}^4) + 1)}{(1 + (1 - \bar{\omega}^2)^2\alpha_2^2)^{3/2}} \\ \frac{\partial\alpha_Y(\tilde{\lambda})}{\partial\bar{\omega}} &= \frac{\alpha_1(2\alpha_1\alpha_2\bar{\omega}(1-\bar{\omega}^2) + \alpha_1^2(1-\bar{\omega}^4) + 1)}{(1 + (1 - \bar{\omega}^2)^2\alpha_1^2)^{3/2}} \\ \frac{\partial l(x, y; \lambda)}{\partial\bar{\omega}} &= \frac{\bar{\omega}}{1 - \bar{\omega}^2} - \frac{(\nu+2)Q'_{\bar{\omega}}(x, y)}{2(Q_{\bar{\omega}}(x, y) + \nu)} \\ &\quad + \frac{t_1(h(x, y; \lambda); \nu+2)}{T_1(h(x, y; \lambda); \nu+2)} \frac{\partial h(x, y; \lambda)}{\partial\bar{\omega}}\end{aligned}$$

e

$$\begin{aligned}
h(x, y; \lambda) &= (\alpha_1 x + \alpha_2 y) \sqrt{\frac{\nu + 2}{Q_{\bar{\omega}}(x, y) + \nu}} \\
\frac{\partial h(x, y; \lambda)}{\partial \bar{\omega}} &= -(\alpha_1 x + \alpha_2 y) \frac{(\nu + 2)^{1/2}}{2(Q_{\bar{\omega}}(x, y) + \nu)^{3/2}} Q'_{\bar{\omega}}(x, y) \\
Q_{\bar{\omega}}(x, y) &= \frac{x(x - \bar{\omega}y) + y(y - \bar{\omega}x)}{1 - \bar{\omega}^2} \\
Q'_{\bar{\omega}}(x, y) &= \frac{2(\bar{\omega}x^2 - xy(1 + \bar{\omega}^2) + \bar{\omega}y^2)}{(1 - \bar{\omega}^2)^2}.
\end{aligned}$$

La seconda componente di $\nabla \rho(\lambda)$ risulta essere

$$\begin{aligned}
\frac{\partial \rho}{\partial \alpha_1} &= 12 \left[\int \int_{\mathbb{R}^2} \left(\frac{\partial F_X(x; \alpha_X(\tilde{\lambda}), \nu)}{\partial \alpha_X(\tilde{\lambda})} \frac{\partial \alpha_X(\tilde{\lambda})}{\partial \alpha_1} F_Y(y; \alpha_Y(\tilde{\lambda}), \nu) \right. \right. \\
&\quad \left. \left. + \frac{\partial F_Y(y; \alpha_Y(\tilde{\lambda}), \nu)}{\partial \alpha_Y(\tilde{\lambda})} \frac{\partial \alpha_Y(\tilde{\lambda})}{\partial \alpha_1} F_X(x; \alpha_X(\tilde{\lambda}), \nu) \right. \right. \\
&\quad \left. \left. + F_X(x; \alpha_X(\tilde{\lambda}), \nu) F_Y(y; \alpha_Y(\tilde{\lambda}), \nu) \frac{\partial l(x, y; \lambda)}{\partial \alpha_1} \right) f(x, y; \lambda) dx dy \right], \quad (3.18)
\end{aligned}$$

dove

$$\begin{aligned}
\frac{\partial l(x, y; \lambda)}{\partial \alpha_1} &= \frac{t_1(h(x, y; \lambda); \nu + 2)}{T_1(h(x, y; \lambda); \nu + 2)} \frac{\partial h(x, y; \lambda)}{\partial \alpha_1} \\
\frac{\partial \alpha_X(\tilde{\lambda})}{\partial \alpha_1} &= \frac{1}{\sqrt{1 + (1 - \bar{\omega}^2)\alpha_2^2}} \\
\frac{\partial \alpha_Y(\tilde{\lambda})}{\partial \alpha_1} &= \frac{\bar{\omega} - \alpha_1 \alpha_2 (1 - \bar{\omega}^2)}{(1 + \alpha_1^2 (1 - \bar{\omega}^2))^{3/2}}
\end{aligned}$$

e

$$\frac{\partial h(x, y; \lambda)}{\partial \alpha_1} = x \sqrt{\frac{\nu + 2}{Q_{\bar{\omega}}(x, y) + \nu}}.$$

Espressioni del tutto analoghe si ottengono per $\partial \rho / \partial \alpha_2$ scambiando α_1 con α_2 nelle espressioni di $\partial \rho / \partial \alpha_1$, mentre

$$\frac{\partial h(x, y; \lambda)}{\partial \alpha_2} = y \sqrt{\frac{\nu + 2}{Q_{\bar{\omega}}(x, y) + \nu}}.$$

Infine

$$\frac{\partial \rho}{\partial \nu} = 12 \left[\int \int_{\mathbb{R}^2} \left(\frac{\partial F_X(x; \alpha_X, \nu)}{\partial \nu} F_Y(y; \alpha_Y, \nu) + \frac{\partial F_Y(y; \alpha_Y, \nu)}{\partial \nu} F_X(x; \alpha_X, \nu) \right. \right. \\ \left. \left. + F_X(x; \alpha_X, \nu) F_Y(y; \alpha_Y, \nu) \frac{\partial l(x, y; \lambda)}{\partial \nu} \right) f(x, y; \lambda) dx dy \right]$$

dove

$$\frac{\partial F_X(x; \alpha_X, \nu)}{\partial \nu} = \int_{-\infty}^x \left(\psi_0 \left(\frac{\nu+1}{2} \right) - \psi_0 \left(\frac{\nu}{2} \right) - \frac{1}{\nu} - \log \left(1 + \frac{s^2}{\nu} \right) + \frac{s^2(\nu+1)}{2\nu(s^2+\nu)} \right. \\ \left. + 2 \frac{\partial(T_1(\alpha_X s \sqrt{(\nu+1)/(s^2+\nu)}; \nu+1))/\partial \nu}{T_1(\alpha_X s \sqrt{(\nu+1)/(s^2+\nu)}; \nu+1)} \right) t_1(s; \nu) T_1 \left(\alpha_X s \sqrt{\frac{\nu+1}{s^2+\nu}}; \nu+1 \right) ds$$

$$\frac{\partial F_Y(y; \alpha_Y, \nu)}{\partial \nu} = \int_{-\infty}^y \left(\psi_0 \left(\frac{\nu+1}{2} \right) - \psi_0 \left(\frac{\nu}{2} \right) - \frac{1}{\nu} - \log \left(1 + \frac{t^2}{\nu} \right) + \frac{t^2(\nu+1)}{2\nu(t^2+\nu)} \right. \\ \left. + 2 \frac{\partial(T_1(\alpha_Y t \sqrt{(\nu+1)/(t^2+\nu)}; \nu+1))/\partial \nu}{T_1(\alpha_Y t \sqrt{(\nu+1)/(t^2+\nu)}; \nu+1)} \right) t_1(t; \nu) T_1 \left(\alpha_Y t \sqrt{\frac{\nu+1}{t^2+\nu}}; \nu+1 \right) dt$$

$$\frac{\partial l(x, y; \lambda)}{\partial \nu} = \frac{1}{2} \left[\psi_0 \left(\frac{\nu+2}{2} \right) - \psi_0 \left(\frac{\nu}{2} \right) - \frac{2}{\nu} - \log(1 + Q_{\bar{w}}(x, y)/\nu) \right. \\ \left. + \frac{(\nu+2)Q_{\bar{w}}(x, y)}{\nu(Q_{\bar{w}}(x, y) + \nu)} + 2 \frac{\partial T_1(h(x, y; \lambda); \nu+2)/\partial \nu}{T_1(h(x, y; \lambda); \nu+2)} \right]$$

in cui $\psi_0(t)$ è la funzione digamma calcolata in $t \in \mathbb{R}$. Dove sono state lasciate indicate le derivate parziali rispetto a ν , il calcolo viene effettuato per via numerica.

Tenendo conto che

$$\frac{\partial F(x, y; \tilde{\lambda})}{\partial \bar{w}} = \int_{-\infty}^y \int_{-\infty}^x \frac{\partial l(s, t; \lambda)}{\partial \bar{w}} f(s, t; \lambda) ds dt,$$

con analoghe espressioni per le derivate parziali dei parametri α_1 , α_2 e ν , abbiamo che la prima componente di $\nabla\tau(\lambda)$ è

$$\frac{\partial\tau}{\partial\bar{\omega}} = 4 \left[\int \int_{\mathbb{R}^2} \left(\frac{\partial F(x, y; \lambda)}{\partial\bar{\omega}} + F(x, y; \lambda) \frac{\partial l(x, y; \lambda)}{\partial\bar{\omega}} \right) f(x, y; \lambda) dx dy \right]. \quad (3.19)$$

Le altre componenti mantengono un'analogha espressione.

Una stima semiparametrica per τ e ρ , che ha il vantaggio di essere semplice, deriva dalla (3.15) per τ e dalla (3.16) per ρ . La (3.15) e la (3.16) possono essere infatti riscritte nel seguente modo

$$\tau = 4E[F(X, Y)] - 1 \quad (3.20)$$

$$\rho = 12E[F_X(X)F_Y(Y)] - 3. \quad (3.21)$$

Disponendo di un campione casuale semplice $((x_1, y_1), \dots, (x_n, y_n))$ ed una volta ottenuta la distribuzione di (X, Y) stimata con il metodo della massima verosimiglianza, due stimatori per ρ e τ sono

$$\begin{aligned} \hat{\tau} &= 4\bar{m}_1 - 1, \\ \hat{\rho} &= 12\bar{m}_2 - 3, \end{aligned}$$

dove

$$\begin{aligned} \bar{m}_1 &= \frac{1}{n} \sum_{i=1}^n \hat{F}(x_i, y_i), \\ \bar{m}_2 &= \frac{1}{n} \sum_{i=1}^n \hat{F}_X(x_i) \hat{F}_Y(y_i). \end{aligned}$$

Al fine di calcolare intervalli di confidenza per grandi campioni, i rispettivi standard error dei due stimatori possono essere calcolati nel seguente modo

$$\begin{aligned} \text{se}(\hat{\tau}) &= \frac{4}{n} \sqrt{\sum_{i=1}^n (\hat{F}(x_i, y_i) - \bar{m}_1)^2}, \\ \text{se}(\hat{\rho}) &= \frac{12}{n} \sqrt{\sum_{i=1}^n (\hat{F}_X(x_i) \hat{F}_Y(y_i) - \bar{m}_2)^2}. \end{aligned}$$

3.2 Matrici di concordanza

Focalizzeremo ora la nostra attenzione su indici matriciali basati sul concetto di segno e rango, riconducibili alle misure di concordanza presentate nel precedente paragrafo, ossia gli indici τ di Kendall e ρ di Spearman.

Le matrici di concordanza basate sugli indici τ e ρ , di cui discuteremo in questo paragrafo, non risultano essere matrici di dispersione benché siano simmetriche, semi-definite positive ed invarianti rispetto a trasformazioni di scala (Visuri, Koivunen & Oja, 2000). Per poter ottenere misure matriciali della struttura di dipendenza di una distribuzione multidimensionale, che siano equivarianti rispetto ad una generica trasformazione affine, si può ricorrere al concetto di rango e segno spaziali o ad altri tipi di procedure a cui si accennerà nella parte finale di questo paragrafo.

Indichiamo con $\text{sign}(x)$ la funzione dei segni marginali del vettore $x \in \mathbb{R}^d$ che calcola per ogni componente x_i del vettore x il segno, ossia $\text{sign}(x_i) = 1$ se $x_i > 0$, $\text{sign}(x_i) = -1$ se $x_i < 0$ e 0 altrimenti. Siano inoltre i vettori casuali X_i, X_j e X_k indipendenti ed identicamente distribuiti. Abbiamo allora la seguente matrice di concordanza della ρ di Spearman

$$S_\rho = E \left[\text{sign}(X_i - X_j) \text{sign}(X_i - X_k)^T \right]. \quad (3.22)$$

Nella definizione di questa tralasciamo il fattore moltiplicativo 3 per la normalizzazione dell'indice ρ , senza con questo compromettere nulla di quanto verrà successivamente detto. La matrice di concordanza della τ di Kendall è

$$S_\tau = E \left[\text{sign}(X_i - X_j) \text{sign}(X_i - X_j)^T \right]. \quad (3.23)$$

La τ di Kendall e la ρ di Spearman possono infatti essere riscritte, a partire dalla (3.1) e dalla (3.2), nel seguente modo

$$\begin{aligned} \tau(X, Y) &= P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \\ &= E[\text{sign}(X_1 - X_2) \text{sign}(Y_1 - Y_2)], \\ \rho(X, Y) &= 3(P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]) \\ &= 3E[\text{sign}(X_1 - X_2) \text{sign}(Y_1 - Y_3)]. \end{aligned}$$

Si noti che la matrice S_τ è la matrice di covarianza per il vettore $\text{sign}(X_i - X_j)$. Facendo uso della funzione $R^X(y) = E[\text{sign}(X - y)]$, detta funzione dei ranghi centrati in y del vettore casuale X , e notando che $R^{X_j}(y) = R^{X_k}(y) = R(y)$,

la matrice S_ρ può anche essere definita equivalentemente nel seguente modo (Visuri, Koivunen & Oja, 2000)

$$S_\rho = E [R(X_i) R(X_i)^T] . \quad (3.24)$$

La funzione $R(\cdot)$ rappresenta inoltre il gradiente della funzione obiettivo, data dalla (2.1), minimizzata dalla mediana per componenti.

Data la natura delle matrici S_ρ e S_τ queste possono essere effettivamente considerate matrici di concordanza normalizzate, in quanto l'informazione presente in esse è indipendente dalla scala di misurazione delle componenti del vettore X . Se vogliamo pertanto ottenere una misura matriciale, non normalizzata, che esprima il grado di variabilità di ogni componente del vettore casuale X , lo scarto interquartile può essere impiegato come misura di dispersione. Indicando con Δ la matrice diagonale che ha nella posizione (i, i) lo scarto interquartile della componente i -esima del vettore X e che indichiamo con IQ_i , le matrici di concordanza non normalizzate sono le seguenti

$$\begin{aligned} \tilde{S}_\rho &\stackrel{def}{=} \Delta S_\rho \Delta, \\ \tilde{S}_\tau &\stackrel{def}{=} \Delta S_\tau \Delta. \end{aligned}$$

Tali matrici continuano ad essere simmetriche e semi-definite positive. Infatti ponendo $z = \Delta v$ per ogni $v \in \mathbb{R}^d$ si ha, sfruttando il fatto che S_τ e S_ρ sono semi-definite positive, che

$$\begin{aligned} v^T \tilde{S}_\tau v &= v^T \Delta S_\tau \Delta v = z^T S_\tau z \geq 0, \\ v^T \tilde{S}_\rho v &= v^T \Delta S_\rho \Delta v = z^T S_\rho z \geq 0. \end{aligned}$$

Nel caso di trasformazioni lineari $Y = AX + b$ dove A è una matrice diagonale di dimensioni $d \times d$ e $b \in \mathbb{R}^d$ si ha che

$$\begin{aligned} \tilde{S}_\tau^Y &= A \tilde{S}_\tau^X A \\ \tilde{S}_\rho^Y &= A \tilde{S}_\rho^X A. \end{aligned}$$

Notando inoltre che

$$\begin{aligned} \tau(aX, bY) &= \text{sign}(a)\text{sign}(b)\tau(X, Y) \\ \rho(aX, bY) &= \text{sign}(a)\text{sign}(b)\rho(X, Y) \end{aligned}$$

dove a e b sono costanti reali, si ha che

$$\begin{aligned} S_\tau^Y &= DS_\tau^X D, \\ S_\rho^Y &= DS_\rho^X D, \end{aligned}$$

dove D è una matrice diagonale il cui i -esimo elemento diagonale è dato dal segno di a_i , con a_i i -esimo elemento diagonale di A . Nel caso in cui A è una generica matrice non singolare di dimensioni $d \times d$ non possiamo dire nulla di generale sulla forma di S_τ^Y . Assumendo $X \sim E_d(\mu, \Sigma, g)$, sappiamo che il nuovo coefficiente di correlazione per la coppia di variabili Y_i ed Y_j è dato da

$$r(Y_i, Y_j) = \frac{a_i^T \Sigma a_j}{(a_i^T \Sigma a_i a_j^T \Sigma a_j)^{1/2}}$$

dove a_i ed a_j sono rispettivamente l' i -esima e la j -esima riga di A . Infine sfruttando sfruttando la (3.11) otteniamo in modo immediato $\tau(Y_i, Y_j)$. Per quanto riguarda invece la matrice S_ρ non è possibile ottenere una espressione generale per tutte le distribuzioni della famiglia ellittica.

In Croux, Ollila & Oja (2002) e Visuri, Koivunen & Oja (2000) viene inoltre utilizzato il concetto di segno spaziale al fine di costruire misure matriciali di associazione. Se $x \in \mathbb{R}^d$, il segno spaziale di x non è altro che il vettore unitario nella direzione di x , ossia $\text{sign}_s(x) = x/\|x\|$ se $x \neq 0$ altrimenti 0. Si definiscono così delle analoghe matrici come quelle espresse in (3.23) e (3.22) in cui la funzione segno viene sostituita con il segno spaziale

$$T_1 = E [\text{sign}_s(X_i - X_j) \text{sign}_s(X_i - X_k)^T], \quad (3.25)$$

$$T_2 = E [\text{sign}_s(X_i - X_j) \text{sign}_s(X_i - X_j)^T]. \quad (3.26)$$

Definendo inoltre il rango spaziale come la funzione $R_s^X(y) = E[\text{sign}_s(X - y)]$ e notando che $R_s^{X_j}(y) = R_s^{X_k}(y) = R_s(y)$, la matrice T_1 può essere equivalentemente espressa nel seguente modo (Visuri, Koivunen & Oja, 2000)

$$T_1 = E [R_s(X_i) R_s(X_i)^T]. \quad (3.27)$$

Va notato questa volta che la funzione $R_s(\cdot)$ è il gradiente della funzione obiettivo in (2.2), il cui massimo è raggiunto nella mediana spaziale η_s .

Una delle importanti proprietà delle matrici T_1 e T_2 , basate su ranghi e segni spaziali, è l'equivarianza rispetto a rotazioni, ossia se A è una matrice

ortogonale $d \times d$ e $b \in \mathbb{R}^d$ allora abbiamo che le matrici dei ranghi e segni spaziali, rispettivamente T_1^* e T_2^* , per $AX + b$ sono

$$T_1^* = AT_1A^T$$

e

$$T_2^* = AT_2A^T.$$

Assumendo inoltre una distribuzione ellittica per X , gli autovettori delle matrici T_1 e T_2 coincidono con quelli della matrice di covarianza (se esiste) di X , pertanto tali matrici possono essere impiegate nell'analisi delle componenti principali.

Al fine di ottenere delle effettive matrici di dispersione, un ulteriore approccio è stato proposto da Hettmansperger & Randles (2002), in cui viene proposta una matrice di dispersione ed un indice di posizione basati sui segni spaziali che vengono però impiegati in una procedura equivariante. Indicando tale parametro di posizione d -dimensionale con ϵ e la matrice $d \times d$ di dispersione con V , la cui traccia è d , allora si ha che

$$E[\text{sign}_s(Z)] = 0 \text{ e } E[\text{sign}_s(Z) \text{sign}_s(Z)^T] = \frac{1}{d}I_d$$

dove $Z = V^{-1/2}(X - \epsilon)$. Maggiori dettagli al riguardo sono forniti in Hettmansperger & Randles (2002) e Tyler (1987).

Citiamo infine le tecniche equivarianti proposte da Oja (1983). Tuttavia queste difficilmente si prestano per un impiego in un contesto parametrico. Ulteriori approfondimenti sono forniti in Hettmansperger, Möttönen & Oja (1997) e Hettmansperger, Möttönen & Oja (1998).

Capitolo 4

L'indice di asimmetria

In questo capitolo tratteremo diversi concetti di asimmetria unidimensionale che verranno introdotti nel primo paragrafo. Nel secondo paragrafo verrà discusso l'indice di Yule-Bowley, direttamente legato ad un dei concetti di asimmetria unidimensionale più generale rispetto agli altri.

4.1 Concetti di asimmetria unidimensionale

Una variabile casuale unidimensionale X , assolutamente continua, ha distribuzione simmetrica rispetto ad un punto μ se, indicata con f la funzione di densità, $f(x - \mu) = f(-(x - \mu))$, per ogni $x \in \mathbb{R}$. Deviazioni dalla simmetria sono pertanto intuitivamente visualizzabili, poiché la distribuzione “poggia” su un lato. Matematicamente il concetto di asimmetria non è però facilmente esprimibile. Nella descrizione di asimmetria fornita da Benjamini & Krieger (1996) la funzione quantile F^{-1} viene scritta nel seguente modo

$$\begin{aligned} F^{-1}(p) &= \eta + \frac{1}{2}(F^{-1}(p) - F^{-1}(1 - p)) \\ &\quad + \frac{1}{2}(F^{-1}(p) + F^{-1}(1 - p) - 2\eta), \quad p \in (0, 1). \end{aligned} \quad (4.1)$$

Il primo addendo della (4.1), η , è la mediana (assunta essere unica) che rappresenta una misura di posizione. Il secondo addendo è una misura di dispersione per $p > 1/2$ (se infatti $p = 3/4$ tale quantità è proporzionale allo scarto interquartile), ossia

$$SP_F(p) = \frac{1}{2}(F^{-1}(p) - F^{-1}(1 - p)). \quad (4.2)$$

Il terzo addendo è

$$SK_F(p) = \frac{1}{2}(F^{-1}(p) + F^{-1}(1-p) - 2\eta), \quad (4.3)$$

che vale esattamente 0 per distribuzioni simmetriche e pertanto può essere visto come una deviazione dalla simmetria per ogni $p > 1/2$. Poiché $SK_F(p) = SK_F(1-p)$ e $SP_F(p) = -SP_F(1-p)$, è sufficiente focalizzare la nostra attenzione per $p > 1/2$.

Nella letteratura sono state introdotte fino ad oggi quattro definizioni di asimmetria “a destra”. Una distribuzione è detta asimmetrica a destra se:

- r_1 : $SK_F(p) \geq 0$;
- r_2 : $SK_F(p)$ è non decrescente;
- r_3 : $SK_F(p)/SP_F(p)$ è non decrescente;
- r_4 : $(dSK_F(p)/dp)/(dSP_F(p)/dp)$ è non decrescente,

in ciascuna definizione per ogni $p \in (1/2, 1)$. Nel caso delle definizioni di asimmetria “a sinistra”, basta sostituire le condizioni di non negatività e non decrescenza con le condizioni di non positività e non crescita. Indicheremo con l_1 , l_2 , l_3 ed l_4 le corrispettive definizioni di asimmetria a sinistra. La relazione esistente tra le varie definizioni di asimmetria a destra consiste nel fatto che $r_4 \Rightarrow r_3 \Rightarrow r_2 \Rightarrow r_1$, ossia r_4 è più stringente di r_3 e così via. Si veda a tal proposito il Teorema 2.2 in MacGillivray (1964). Nella letteratura r_1 ed r_2 sono equivalenti ai concetti di “asimmetria a destra” e di “forte asimmetria a destra” in Doksum (1975). I concetti r_3 ed r_4 seguono rispettivamente dagli ordinamenti di asimmetria di Oja (1981) e MacGillivray (1964), e di van Zwet (1986). Per ordinamento di asimmetria si intende un criterio, denotato (usando la stessa notazione impiegata da Benjamini & Krieger (1996)) con s_i , che ci permette di stabilire se una distribuzione F è, o non è, più asimmetrica a destra di un'altra distribuzione di probabilità G . La notazione $F_{s_i}G$ sta ad indicare che la distribuzione F è più asimmetrica a destra rispetto a G secondo il criterio s_i . Gli ordinamenti che sono stati finora proposti nella

letteratura (Benjamini & Krieger, 1996) sono i seguenti:

- $F_{s_0}G$: $SK_F(p)/SP_F(P) \geq SK_F(p)/SP_F(P)$, per ogni $p \in (1/2, 1)$;
- $F_{s_1}G$: esiste un a tale che $a[F^{-1}(p) - \eta] - [G^{-1}(p) - G^{-1}(1/2)] \geq 0$, per ogni $p \in (0, 1)$;
- $F_{s_2}G$: esiste un a tale che $a[F^{-1}(p) - \eta] - [G^{-1}(p) - G^{-1}(1/2)]$ è non crescente per $p < 1/2$ e non decrescente per $p > 1/2$;
- $F_{s_3}G$: $[F^{-1}(p) - \eta]/[G^{-1}(p) - G^{-1}(1/2)]$ è non crescente per $p \in (0, 1)$ e $p \neq 1/2$;
- $F_{s_4}G$: $(dF^{-1}(p)/dp)/(dG^{-1}(p)/dp)$ è non crescente per $p \in (0, 1)$.

Gli ordinamenti s_i sono legati alle definizioni di asimmetria r_i dal fatto che, se F è la distribuzione della variabile casuale X e G è la distribuzione della variabile casuale $-X$ e se $F_{s_i}G$, allora la distribuzione F è r_i . L'unica eccezione è che $F_{s_0}G$ implica che F è s_1 .

Le definizioni di asimmetria r_2 , r_3 ed r_4 a destra possono essere reinterpretate in termini di proprietà della funzione di densità.

Teorema 4.1.1. *Sia X una variabile casuale assolutamente continua con funzione di densità f e funzione quantile F^{-1} . Se le condizioni che definiscono r_2 , r_3 ed r_4 sono verificate allora si ha che:*

$$\begin{aligned} r_2 : & f(F^{-1}(1-p)) \geq f(F^{-1}(p)); \\ r_3 : & \frac{SK_F(p)}{SP_F(p)} \leq \frac{f(F^{-1}(1-p)) - f(F^{-1}(p))}{f(F^{-1}(1-p)) + f(F^{-1}(p))}; \\ r_4 : & f(F^{-1}(1-p))/f(F^{-1}(p)) \text{ è non decrescente,} \end{aligned}$$

in ciascuna definizione per ogni $p \in (1/2, 1)$.

Dimostrazione. Notiamo da subito che, per il teorema della derivata della funzione inversa, si ha che

$$\frac{dF^{-1}(p)}{dp} = \frac{1}{f(F^{-1}(p))}.$$

Nel caso di r_2 , imporre che $SK_F(p)$ sia non decrescente implica che

$$\frac{dSK_F(p)}{dp} \geq 0, \quad \forall p \in (1/2, 1),$$

ossia

$$\frac{f(F^{-1}(1-p)) - f(F^{-1}(p))}{f(F^{-1}(p))f(F^{-1}(1-p))} \geq 0, \quad \forall p \in (1/2, 1)$$

da cui segue $f(F^{-1}(1-p)) \geq f(F^{-1}(p))$ per ogni $p \in (1/2, 1)$. In modo analogo per r_3 abbiamo che le condizioni imposte implicano che la derivata prima sia non negativa per ogni $p \in (1/2, 1)$, ossia

$$\begin{aligned} & \left[[f(F^{-1}(1-p)) - f(F^{-1}(p))]SP_F(p) - [f(F^{-1}(1-p)) + f(F^{-1}(p))]SK_F(p) \right] \\ & \quad \times \left[f(F^{-1}(p))f(F^{-1}(1-p))SP_F^2(p) \right]^{-1} \geq 0 \end{aligned}$$

da cui segue, con semplici passaggi algebrici, che

$$\frac{SK_F(p)}{SP_F(p)} \leq \frac{f(F^{-1}(1-p)) - f(F^{-1}(p))}{f(F^{-1}(1-p)) + f(F^{-1}(p))}, \quad \forall p \in (1/2, 1).$$

Infine notando che

$$\frac{dSK_F(p)/dp}{dSP_F(p)/dp} = \frac{f(F^{-1}(1-p)) - f(F^{-1}(p))}{f(F^{-1}(1-p)) + f(F^{-1}(p))} = \frac{f(F^{-1}(1-p))/f(F^{-1}(p)) - 1}{f(F^{-1}(1-p))/f(F^{-1}(p)) + 1}$$

e che la funzione $(x-1)/(x+1)$ è una funzione strettamente crescente in x , abbiamo che $(dSK_F(p)/dp)/(dSP_F(p)/dp)$ è non decrescente se $f(F^{-1}(1-p))/f(F^{-1}(p))$ è non decrescente per ogni $p \in (1/2, 1)$. \square

Dato un campione da una distribuzione F , le quantità SK_F ed SP_F possono essere ottenute in modo non parametrico attraverso la distribuzione empirica \hat{F} . Al fine di indagare quale tipo di asimmetria è espressa nei dati, un utile grafico è quello di $SK_{\hat{F}}$ rispetto a $SP_{\hat{F}}$ (Benjamini & Krieger, 1996), considerando valori di queste funzioni per $p \in (1/2, 1)$. Osservazioni provenienti da una distribuzione simmetrica dovrebbero esibire fluttuazioni casuali attorno alla retta orizzontale $SK_F(p) = 0$. Sempre da questo grafico possiamo capire se SK_F è positiva (o negativa), se la relazione tra SK_F e SP_F è crescente (o decrescente), se il coefficiente angolare della retta passante per l'origine ed un generico punto della curva ha un andamento crescente (o decrescente) al variare del secondo punto sulla curva e se la curva è convessa (o concava). Tutte queste caratteristiche riflettono asimmetria a destra (o a sinistra) rispettivamente secondo le definizioni r_1 , r_2 , r_3 ed r_4 (l_1 , l_2 , l_3 ed l_4).

Notando che

$$\begin{aligned} |SK_F(p)| &\leq \frac{1}{2}(|F^{-1}(1-p) - \eta| + |F^{-1}(p) - \eta|) \\ &= SP_F(p) \end{aligned}$$

abbiamo che $|SK_{\hat{F}}(p)| \leq SK_{\hat{F}}(n/(n+1))$, dove n è la numerosità del campione, per ogni $p \leq n/(n+1)$. I limiti di visualizzazione dell'asse verticale possono essere pertanto fissati tra $[-SK_{\hat{F}}(n/(n+1)), SK_{\hat{F}}(n/(n+1))]$.

In Figura 4.1, in cui vengono considerati nuovamente i dati relativi alla velocità del vento, si è tuttavia preferito mettere a fuoco l'area della curva di interesse al fine di meglio evidenziare le peculiarità di questa. Da questi grafici osserviamo che nel caso delle velocità del vento registrate nelle prime due stazioni meteorologiche, i dati esibiscono asimmetria a sinistra del tipo l_2 , in quanto le curve sono negative e strettamente decrescenti. Sembrerebbe esserci anche asimmetria a sinistra del tipo l_3 : il coefficiente angolare della retta passante per l'origine ed una dato punto della curva ha un andamento approssimativamente decrescente. I dati relativi all'ultima torre meteorologica sono chiaramente asimmetrici ma l'asimmetria rilevata non rientra in alcun tipo tra quelli precedentemente definiti: la funzione $SK_F(p)$ risulta essere positiva all'incirca fino a $p = 0.81$ ed assume poi valori negativi. Possiamo dunque affermare che la distribuzione in questione è asimmetrica verso destra per $p \in (1/2, 0.81)$ secondo r_1 .

Tuttavia va detto che risulta essere piuttosto arduo individuare tra le più note distribuzioni di probabilità asimmetriche unimodali quelle che non siano del tipo r_4 . In Figura 4.2 osserviamo un andamento strettamente crescente della funzione $f(F^{-1}(1-p))/f(F^{-1}(p))$ per varie note distribuzioni di probabilità. Si noti inoltre che le curve in questione non dipendono dai parametri di scala, caratteristici di alcune delle distribuzioni in questione, ma soltanto dai parametri di forma.

Nel caso specifico della distribuzione ST l'asimmetria a destra o a sinistra è direttamente regolata dal parametro di forma α che definisce il grado di asimmetria della distribuzione. Infatti abbiamo che se $X \sim ST_1(0, 1, \alpha)$ allora

$$SK_X(p) = -SK_{-X}(p), \forall p \in (1/2, 1),$$

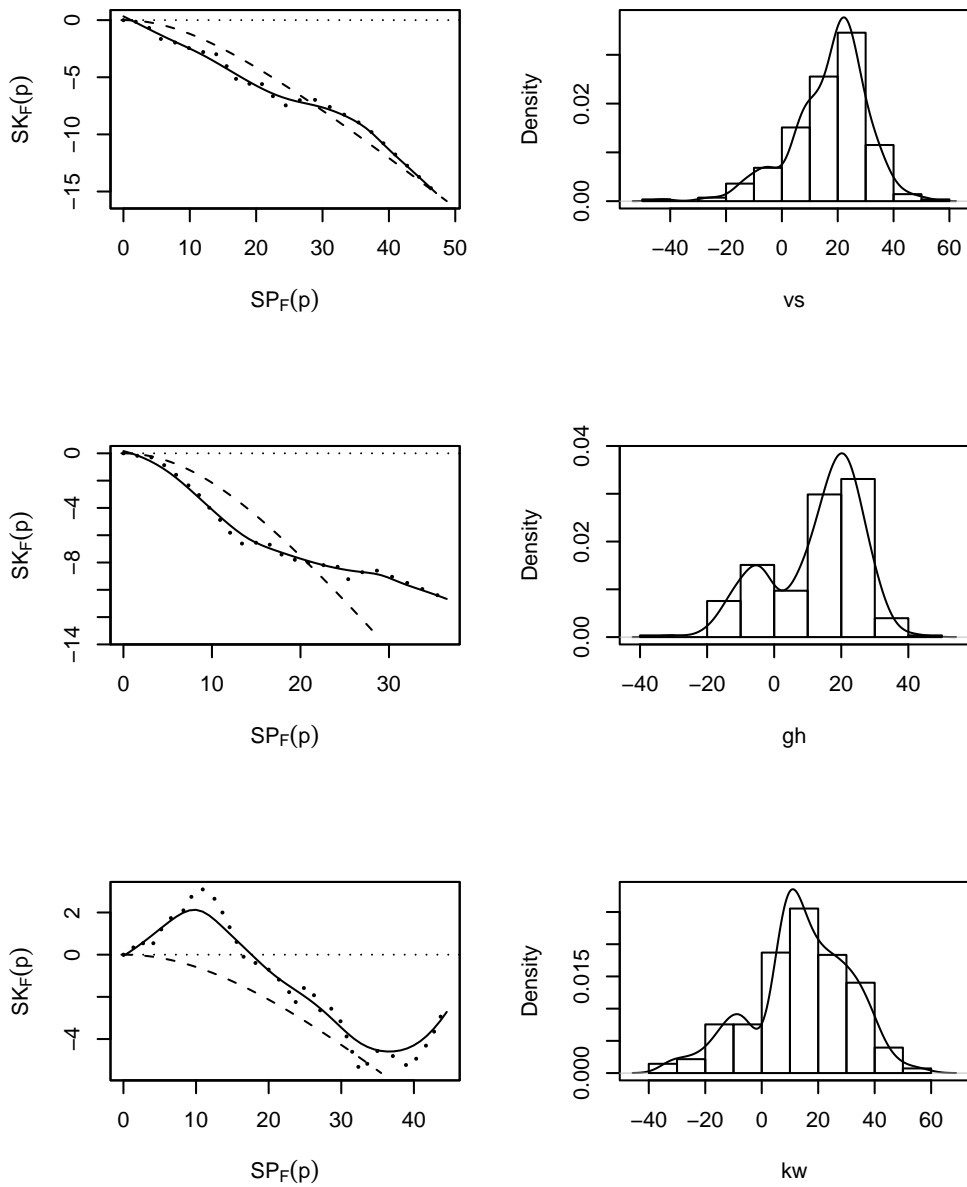


Figura 4.1: Grafici di $SK_{\hat{F}}$ rispetto a $SP_{\hat{F}}$ per i dati relativi alla velocità del vento registrata in tre diverse stazioni meteorologiche. Le curve punteggiate sono ottenute a partire dalla distribuzione empirica; le curve continue sono ottenute mediante lisciamiento non parametrico della precedente curva; le curve tratteggiate sono ottenute dalla distribuzione ST stimata con il metodo della massima verosimiglianza.

dove si noti che $-X \sim \text{ST}_1(0, 1, -\alpha)$. Nel caso in cui $\nu = 1$, si ottiene che

$$SK_F(p) = \frac{2\delta}{\sin(\pi p)} - 2(\sqrt{2} - 1)\delta \quad (4.4)$$

$$SP_F(p) = -2\frac{\cos(\pi p)}{\sin(\pi p)}. \quad (4.5)$$

Pertanto

$$\begin{aligned} \frac{dSK_F(p)}{dp} &= -2\pi\delta \frac{\cos(\pi p)}{\sin^2(\pi p)} \\ \frac{dSP_F(p)}{dp} &= \frac{2\pi}{\sin^2(\pi p)}, \end{aligned}$$

dove $\delta = \alpha/\sqrt{1 + \alpha^2}$. Allora abbiamo

$$\frac{dSK_F(p)/dp}{dSP_F(p)/dp} = -\delta \cos(\pi p)$$

che è chiaramente crescente per $p \in (1/2, 1)$ se $\alpha > 0$ ed allora la distribuzione SC è asimmetrica a destra secondo r_4 . Se invece $\alpha < 0$ allora la SC è asimmetrica a sinistra secondo l_4 . In Figura 4.1 osserviamo che l'asimmetria a sinistra delle diverse ST è di tipo l_4 , evidente dalla concavità delle curve decrescenti tratteggiate. Questa risulta essere troppo forte rispetto al tipo di asimmetria presente nei dati che, come è stato già osservato, mostrano al massimo asimmetria di tipo l_3 .

4.2 L'indice di Yule-Bowley

Il nostro obiettivo è quello di disporre di un indice che sia sensibile al più generale concetto di asimmetria a destra al fine di poter includere un più ampio insieme di possibili famiglie parametriche. Rivolgiamo pertanto la nostra attenzione all'asimmetria del tipo r_1 .

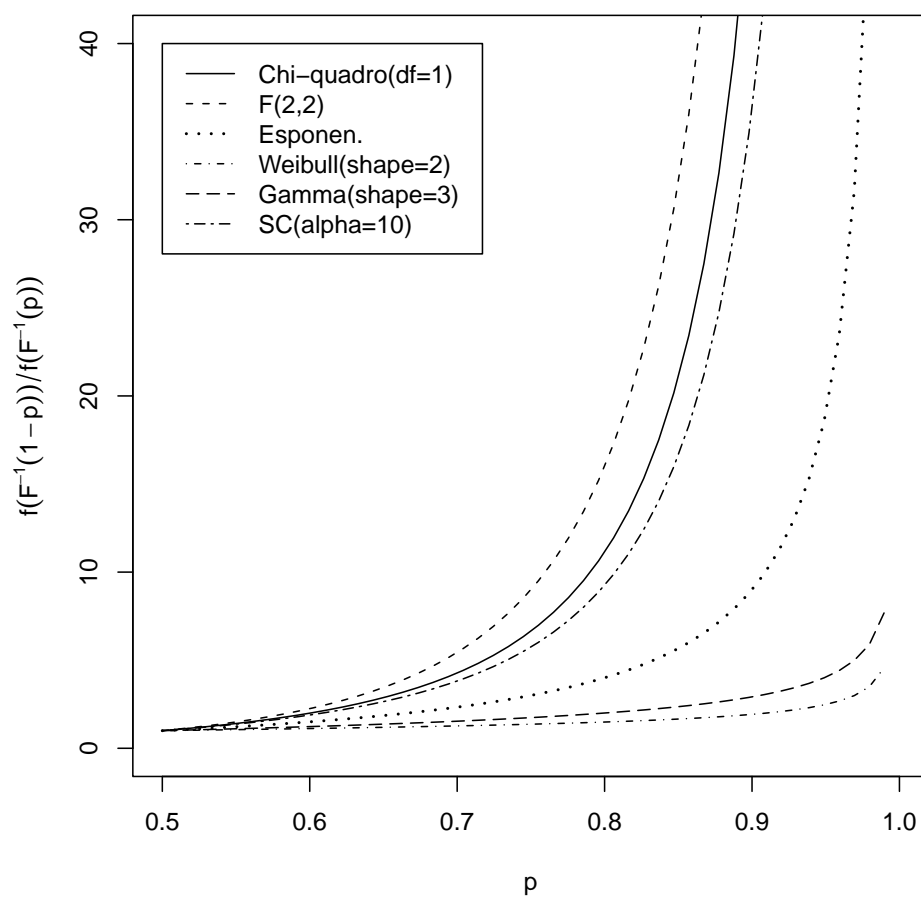


Figura 4.2: Grafici di $f(F^{-1}(1-p))/f(F^{-1}(p))$ per diverse distribuzioni di probabilità.

L'indice di asimmetria proposto da Yule (1911) e successivamente da Bowley (1920), dato dalla seguente espressione

$$\begin{aligned} b_F &= \frac{(F^{-1}(3/4) - \eta) - (\eta - F^{-1}(1/4))}{(F^{-1}(3/4) - \eta) + (\eta - F^{-1}(1/4))} \\ &= \frac{F^{-1}(3/4) + F^{-1}(1/4) - 2\eta}{F^{-1}(3/4) - F^{-1}(1/4)} \\ &= \frac{SK_F(3/4)}{SP_F(3/4)}, \end{aligned} \quad (4.6)$$

risulta essere una misura diretta dell'asimmetria di tipo r_1 . Infatti come si evince dalla (4.6), $b_F \in (-1, 1)$ ed inoltre $b_F > 0$ se la distribuzione è asimmetrica a destra secondo r_1 , mentre $b_F < 0$ se la distribuzione è asimmetrica a sinistra secondo l_1 . Infine se $b_F = 0$ la distribuzione è simmetrica.

Una generalizzazione dell'indice di Yule-Bowley, proposta da David & Johnson (1956), è la seguente

$$\gamma_F(p) = \frac{F^{-1}(p) + F^{-1}(1-p) - 2\eta}{F^{-1}(p) - F^{-1}(1-p)}, \quad p \in (1/2, 1). \quad (4.7)$$

L'indice γ_F è indipendente dalla posizione e dalla scala della distribuzione F . Inoltre MacGillivray (1992) si riferisce a γ_F come la funzione di asimmetria di una variabile casuale con distribuzione F e trova l'espressione di $\gamma_F(p)$ per la famiglia di distribuzioni di Tadikamalla & Johnson (1982), tra le quali è presente anche la distribuzione log-normale. MacGillivray (1992) suggerisce inoltre di impiegare $\sup_{r \leq u \leq 1/2} |\gamma_F(u)|$, per un valore di r fissato, come misura globale dell'asimmetria nella "zona centrale" $100(1 - 2r)\%$. Questa viene definita come "l'asimmetria centrale" della distribuzione.

L'indice γ_F appartiene alla famiglia di misure di asimmetria del tipo

$$S_F(H) = \frac{\int_0^1 F^{-1}(p) dH(p) + \int_0^1 F^{-1}(1-p) dH(p)}{\int_0^1 F^{-1}(p) dH(p) - \int_0^1 F^{-1}(1-p) dH(p)}, \quad p \in (1/2, 1), \quad (4.8)$$

dove $H(p)$ è una misura con segno, ossia una misura che può assumere anche valori negativi. Se $dH(p) = 1$ per $p = p_0 > 1/2$ e $dH(p) = -1$ per $p = 1/2$ allora otteniamo l'indice γ_F . Imponendo diversi vincoli a $dH(p)$ otteniamo diversi indici di asimmetria. Se $p = 1/\alpha$, per $p > 1 - \alpha$ e $dH(p) = -2$ per $1/4 < p < 3/4$, allora la (4.8) diventa

$$h_F(\alpha) = \frac{\frac{1}{\alpha} \int_{1-\alpha}^1 F^{-1}(p) dp - 2 \int_{1/4}^{3/4} F^{-1}(p) dp}{2 \int_{1/4}^{3/4} F^{-1}(p) dp - \frac{1}{\alpha} \int_0^\alpha F^{-1}(p) dp}.$$

L'indice h_F , proposto in Hogg (1974), risulta essere invece sensibile ad asimmetria del tipo r_2 . Altri indici unidimensionali di asimmetria ottenibili dalla (4.8) sono discussi in Benjamini & Krieger (1996).

Lo stimatore di γ_F , ottenuto calcolando tale indice per la distribuzione stimata con il metodo della massima verosimiglianza ed indicato con $\hat{\gamma}_F$, ha distribuzione asintotica $N(\gamma_F, \text{var}(\hat{\gamma}_F))$, dove $\text{var}(\hat{\gamma}_F)$ è la varianza della distribuzione asintotica. Denotando il vettore della mediana, del quantile p -esimo e del quantile $(1-p)$ -esimo con $q_p = (\eta, F^{-1}(p), F^{-1}(1-p))$ abbiamo che

$$\text{var}(\hat{\gamma}_F) = \nabla q_p(\gamma_F) \Sigma_{q_p} \nabla q_p(\gamma_F)^T,$$

dove

$$\nabla q_p(\gamma_F) = 2 \begin{pmatrix} -\frac{1}{F^{-1}(p) - F^{-1}(1-p)} \\ \frac{\eta - F^{-1}(1-p)}{[F^{-1}(p) - F^{-1}(1-p)]^2} \\ \frac{F^{-1}(p) - \eta}{[F^{-1}(p) - F^{-1}(1-p)]^2} \end{pmatrix}$$

e Σ_{q_p} è la matrice di covarianza della distribuzione asintotica dello stimatore $(\tilde{\eta}, \tilde{F}^{-1}(p), \tilde{F}^{-1}(1-p))$, avendo indicato con $\tilde{F}^{-1}(r)$ lo stimatore del quantile r -esimo calcolato a partire dalla distribuzione stimata con il metodo della massima verosimiglianza. Si noti che $\tilde{F}^{-1}(1/2) = \tilde{\eta}$. L'espressione di Σ_{q_p} è facilmente ottenibile a partire dalla (2.12), ossia

$$\Sigma_{q_p} = J_{q_p} i(\lambda)^{-1} J_{q_p}^T,$$

dove λ è il vettore dei parametri che indicizzano la distribuzione F e

$$J_{q_p} = (\nabla \eta(\lambda) \quad \nabla F_p^{-1}(\lambda) \quad \nabla F_{1-p}^{-1}(\lambda))^T.$$

Data una variabile casuale d -dimensionale, un possibile indice di asimmetria vettoriale è dato pertanto dal vettore la cui i -esima componente è data dall'indice b_F dell' i -esima distribuzione marginale. Ulteriori indici di asimmetria vettoriali, la cui definizione non preveda l'esistenza di alcuno dei momenti distributivi, sono stati proposti da Avérous & Meste (1997) che introducono un funzionale statistico vettoriale basato sulla mediana spaziale e le cosiddette "palle mediane", ossia una possibile generalizzazione multidimensionale del "intervallo interquartile" unidimensionale $[F^{-1}((1-r)/2), F^{-1}((1+r)/2)]$, $r \in (0, 1)$. Tuttavia questi indici multidimensionali, oltre ad essere più complessi, mal si prestano ad un utilizzo in ambito parametrico.

Capitolo 5

L'indice di curtosi

In questo capitolo discuteremo un indice di curtosi basato sui quantili. L'estensione di tale indice al caso multidimensionale (proposta da Wang & Serfling (2005) e di cui il §5.1.2 rappresenta una sintesi) è ottenuta mediante le funzioni *depth*. Considerando la funzione *depth* HS, verranno indagati diversi metodi al fine di semplificare il calcolo di tale indice per una distribuzione multidimensionale asimmetrica: nel §5.2.1 si propone un metodo di approssimazione nel caso bidimensionale; nel §5.2.2 verrà discussa una famiglia di ordinamenti di curtosi per distribuzioni multidimensionali, proposta da Wang (2009), al fine di ridurre il calcolo dell'indice ad una distribuzione con simmetria sferica; infine nel 5.2.3 si analizzerà il particolare caso della distribuzione SC sfruttando la particolare forma geometrica delle curve di livello della funzione HS per tale distribuzione.

5.1 Concetti di curtosi e misure basate sui quantili

5.1.1 Caso unidimensionale

La curtosi presenta maggiori difficoltà nella sua caratterizzazione ed interpretazione dovute al suo legame con la variabilità, il grado di appuntimento della distribuzione ed il peso delle code, oltreché con l'asimmetria se presente.

Nel caso unidimensionale la classica nozione di curtosi è basata sul momento standardizzato del quarto ordine $\kappa = E[(X - \mu)^4]/\sigma^4$ o quantità legate a questo, quale $\gamma_2 = \kappa - 3$. Il termine “curtosi” è stato introdotto da Pear-

son (1905) per denotare la forma di una distribuzione simmetrica legata al suo grado di appuntimento, che potrebbe essere minore o maggiore di quello di una distribuzione normale. Pearson introduce inoltre i seguenti termini: “platocurtica”, per una distribuzione simmetrica più “piatta”, ossia con un maggior spessore delle code di una normale, se $\kappa > 3$; “leptocurtica”, per una distribuzione simmetrica più “appuntita”, ossia con un minor spessore delle code di una normale, se $\kappa < 3$; “normocurtica”, per una distribuzione simmetrica tanto “piatta” quanto la distribuzione normale, se $\kappa = 3$. Va notato inoltre che questi termini sono stati utilizzati da Pearson per confrontare la forma di una distribuzione simmetrica unimodale con media μ e deviazione standard σ con la forma di una distribuzione normale con la stessa media μ e stessa deviazione standard σ .

Il momento standardizzato del quarto ordine può essere riscritto nel seguente modo

$$\kappa = \text{var} \left[\left(\frac{X - \mu}{\sigma} \right)^2 \right] + 1, \quad (5.1)$$

ossia l'indice κ può essere visto come la varianza di $(X - \mu)^2/\sigma^2$ attorno alla sua media 1, o equivalentemente come la dispersione della variabile casuale X attorno ai punti $\mu \pm \sigma$ che possono essere visti, utilizzando un linguaggio pittoresco, come le “spalle” della distribuzione. L'indice κ misura pertanto, indipendentemente dalla posizione e dalla scala di una distribuzione, la dispersione della massa di probabilità verso il centro o le code della distribuzione o entrambi, ossia verso regioni lontane dalle spalle della distribuzione. Pertanto tanto più la massa di probabilità in prossimità delle spalle diminuisce ed incrementa in regioni lontane da queste, tanto più la dispersione aumenta.

Come si nota, la classica nozione di curtosi basata sul momento standardizzato del quarto ordine misura la dispersione della massa di probabilità nella regione delle “spalle” ma non caratterizza la forma. Misure di curtosi basate sui quantili forniscono invece informazione relativa alla forma della distribuzione. Nel descrivere la curtosi di una distribuzione asimmetrica, la curtosi diventa necessariamente correlata con l'asimmetria.

Groeneveld & Meeden (1984) mostrano che una ragionevole misura di curtosi, nel caso in cui X abbia una distribuzione simmetrica, può essere generata calcolando l'indice di asimmetria $\gamma_F(p)$, dato dalla (4.7), per la variabile casuale $|X - \eta|$, dove η è la mediana della distribuzione F . Si

ottiene allora il seguente indice di curtosi

$$\phi_F(p) = \frac{F^{-1}\left(\frac{3}{4} - \frac{p}{4}\right) + F^{-1}\left(\frac{3}{4} + \frac{p}{4}\right) - 2F^{-1}\left(\frac{3}{4}\right)}{F^{-1}\left(\frac{3}{4} + \frac{p}{4}\right) - F^{-1}\left(\frac{3}{4} - \frac{p}{4}\right)}, \quad p \in (0, 1). \quad (5.2)$$

Così come l'indice κ , dato dalla (5.1), l'indice ϕ_F misura la dispersione in corrispondenza delle “spalle” della distribuzione, che sono date in questo caso dal primo e terzo quartile e non più da $\mu \pm \sigma$. Nel caso di distribuzioni simmetriche è sufficiente considerare solamente il lato destro della distribuzione, nel quale la “spalla” $F^{-1}(3/4)$ divide la “zona centrale” dalla “coda destra” della distribuzione. Definendo

$$\begin{aligned} l_1(p) &= F^{-1}(3/4 + p/4) - F^{-1}(3/4) \\ l_2(p) &= F^{-1}(3/4) - F^{-1}(3/4 - p/4), \end{aligned} \quad (5.3)$$

le regioni con stessa massa di probabilità $p/4$ prese rispettivamente all'interno ed all'esterno delle “spalle”, possiamo riscrivere la (5.2) nel seguente modo

$$\phi_F(p) = \frac{l_1(p) - l_2(p)}{l_1(p) + l_2(p)}, \quad p \in (0, 1). \quad (5.4)$$

Come si nota $|\phi_F(p)| \leq 1$. Se ϕ_F è vicino a +1 la massa di probabilità è maggiormente concentrata nella “zona centrale”; se ϕ_F è vicino a -1 la distribuzione ha invece una forma ad U; se ϕ_F è prossimo a 0 allora la distribuzione è piuttosto uniforme. Si veda a tal proposito la Figura 5.1.

Wang & Serfling (2005) estendono la (5.2) nel caso di distribuzioni asimmetriche combinando insieme la versione della (5.2) sul lato destro e sul lato sinistro della distribuzione. La (5.3) diventa allora

$$\begin{aligned} l_1(p) &= \left[F^{-1}\left(\frac{3}{4} + \frac{p}{4}\right) - F^{-1}\left(\frac{3}{4}\right) \right] + \left[F^{-1}\left(\frac{1}{4}\right) - F^{-1}\left(\frac{1}{4} - \frac{p}{4}\right) \right], \\ l_2(p) &= \left[F^{-1}\left(\frac{3}{4}\right) - F^{-1}\left(\frac{3}{4} - \frac{p}{4}\right) \right] + \left[F^{-1}\left(\frac{1}{4} + \frac{p}{4}\right) - F^{-1}\left(\frac{1}{4}\right) \right]. \end{aligned}$$

Questa volta $l_1(p)$ è la lunghezza totale della regione esterna alla “zona centrale” con massa di probabilità $p/2$ e $l_2(p)$ è a sua volta la lunghezza totale della regione interna alla “zona centrale” con massa di probabilità pari a $p/2$.

Consideriamo ora il funzionale statistico di dispersione SP_F , dato dalla (4.2). Wang & Serfling (2005), una volta esteso l'indice ϕ_F al caso con asimmetria, esprimono questo in termini di SP_F . L'indice ϕ_F è pertanto esprimibile in termini della dispersione della distribuzione, similmente a quanto

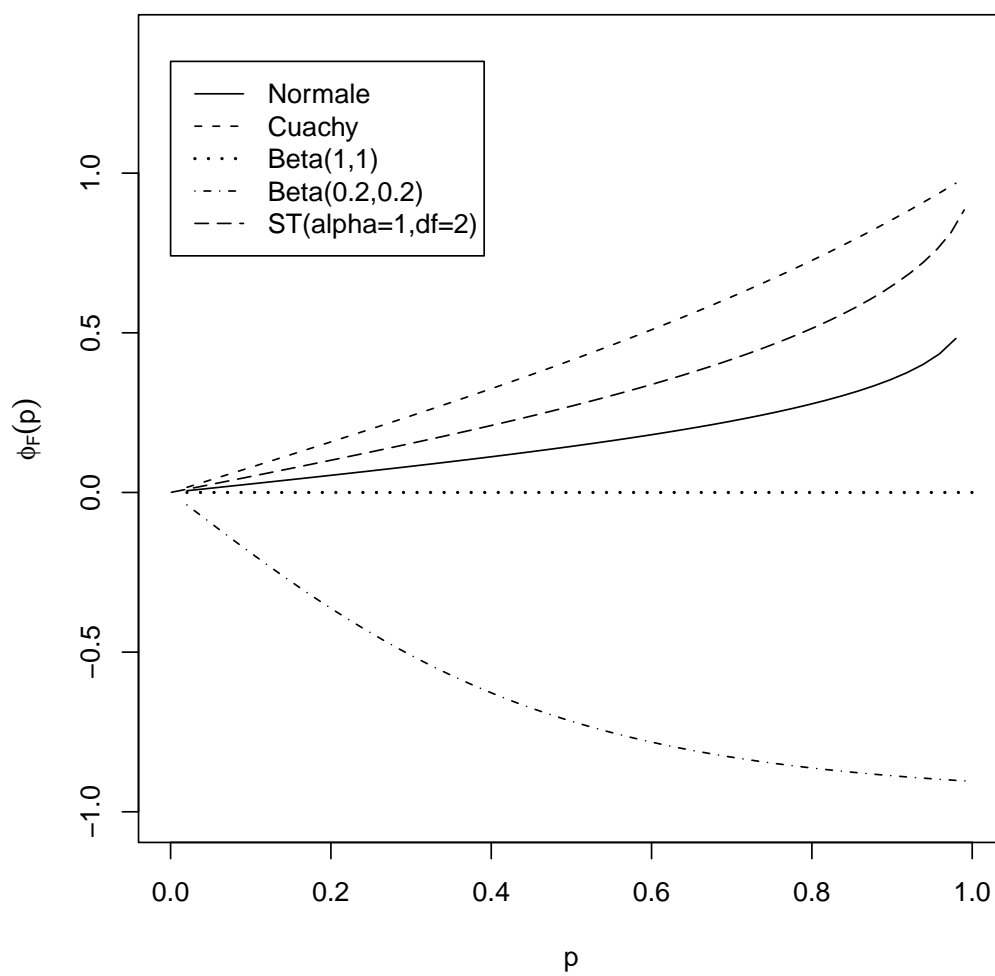


Figura 5.1: Curve di curtosi dell'indice $\phi_F(p)$ per alcune distribuzioni unidimensionali.

avviene per l'indice κ che, come abbiamo osservato, è esprimibile in termini della varianza. Abbiamo infatti che

$$\phi_F(p) = \frac{d_F\left(\frac{1}{2} - \frac{p}{2}\right) + d_F\left(\frac{1}{2} + \frac{p}{2}\right) - 2d_F\left(\frac{1}{2}\right)}{d_F\left(\frac{1}{2} + \frac{p}{2}\right) - d_F\left(\frac{1}{2} - \frac{p}{2}\right)}, \quad (5.5)$$

dove $d_F(p) = SP_F((1+p)/2)$. Dalla (5.5), come si vedrà nel prossimo paragrafo, Wang & Serfling (2005) estendono l'indice ϕ_F al caso multidimensionale.

5.1.2 Estensione al caso multidimensionale

Per l'estensione dell'indice ϕ_F al caso multidimensionale, le funzioni *depth* tornano a coprire un ruolo centrale.

Data una distribuzione F assolutamente continua, consideriamo la famiglia di regioni centrali (vedi Definizione 2.1.4) $\mathcal{C} = \{C_p : p \in [0, 1]\}$ individuata da una funzione *depth* $D(\cdot; F)$ (vedi Definizione 2.1.1). Le regioni centrali C_p risultano essere annidate attorno al punto di massima *depth*, che indichiamo con M_F (utilizziamo lo per il momento lo stesso simbolo adottato precedentemente per la mediana di Tukey), e si riducono a tale punto per $p \rightarrow 0$. Per un dato punto $x \in \mathbb{R}^d$ su $q(p)$, la frontiera di C_p , indichiamo con v il vettore di norma 1 che da M_F punta in direzione di x e definiamo $u = pv$, dove p è la massa di probabilità della regione centrale C_p . Definendo $Q_F(u) = x$, con $Q_F(0) = M_F$, si ha che i punti $x \in \mathbb{R}^d$ generano una funzione quantile $Q_F(u)$ con $u \in \mathbb{B}^d = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$.

Va chiarito che tale funzione quantile è unicamente determinata una volta che è stata definita la funzione *depth* che individua la regione C_p . A differenza del caso unidimensionale in cui la funzione quantile è unicamente individuata dalla distribuzione F , nel caso multidimensionale non c'è più tale univocità, in quanto avremo una diversa funzione quantile $Q_F(\cdot)$ a seconda della funzione *depth* che stiamo utilizzando.

Possiamo ora definire, per una data famiglia di regioni centrali \mathcal{C} individuate da una specifica funzione *depth* $D(\cdot; F)$, un *volume funzionale* definito come

$$V_{F, \mathcal{C}}(p) = \text{volume}(C_p), \quad p \in [0, 1].$$

Tale volume funzionale è una funzione crescente in p e rappresenta la dispersione della distribuzione F in termini dell'espansione della regione centrale C_p . Il volume funzionale gioca, nello spazio multidimensionale, un ruolo analogo a quello della funzione d_F definita nel caso unidimensionale.

Una naturale estensione della (5.5) al caso multidimensionale è pertanto

$$\phi_F(p) = \frac{V_{F,\mathcal{G}}\left(\frac{1}{2} - \frac{p}{2}\right) + V_{F,\mathcal{G}}\left(\frac{1}{2} + \frac{p}{2}\right) - 2V_{F,\mathcal{G}}\left(\frac{1}{2}\right)}{V_{F,\mathcal{G}}\left(\frac{1}{2} + \frac{p}{2}\right) - V_{F,\mathcal{G}}\left(\frac{1}{2} - \frac{p}{2}\right)}, p \in (0, 1). \quad (5.6)$$

Analogamente alla (5.4), l'indice ϕ_F multidimensionale può essere riscritto nel seguente modo

$$\phi_F(p) = \frac{\text{volume}(A) - \text{volume}(B)}{\text{volume}(A) + \text{volume}(B)} \quad (5.7)$$

Le “spalle” della distribuzione sono date in questo caso dalla frontiera regione centrale $C_{1/2}$, ossia $q(1/2)$, e rappresentano la regione di demarcazione tra la “zona centrale” e le “code” della distribuzione. Pertanto l'indice ϕ_F misura la relativa differenza volumetrica tra una regione A interna alla “zona centrale” ed una regione B esterna alla “zona centrale”. Tali regioni sono ottenute sottraendo ed addizionando a $p = 1/2$ una data quantità $p/2$, rispettivamente.

L'interpretazione dei valori assunti da ϕ_F è del tutto analoga al caso unidimensionale: se ϕ_F assume valori vicini a $+1$ allora la massa di probabilità è maggiormente concentrata nella “zona centrale”; se ϕ_F assume valori vicini a -1 allora la massa di probabilità è maggiormente concentrata nelle “code”; se ϕ_F assume valori vicini a 0 allora la distribuzione ha una funzione di densità piuttosto uniforme.

Elenchiamo ora alcune proprietà dell'indice ϕ_F :

1. $|\phi_F(p)| \leq 1$;
2. ϕ_F è definito senza alcuna assunzione sull'esistenza dei momenti;
3. $\phi_F(p) = 0$ nel caso di distribuzioni uniformi;
4. per distribuzioni unimodali si ha che $\phi_F > 0$;
5. per distribuzioni a forma di “ciotola”, ossia se $1 - F$ è unimodale, si ha che $\phi_F < 0$;
6. Se $V_{F,\mathcal{G}}(p)$ è una funzione continua in p allora

$$\lim_{p \rightarrow 1} \phi_F(p) = \begin{cases} \frac{\text{volume}(S_F) - 2V_{F,\mathcal{G}}(1/2)}{\text{volume}(S_F)}, & \text{se } \text{volume}(S_F) < +\infty \\ 1, & \text{altrimenti,} \end{cases}$$

dove S_F è il supporto della distribuzione F .

7. Se $V_{F,\mathcal{E}}$ è derivabile con derivata non nulla in $p = 1/2$, allora

$$\lim_{p \rightarrow 0} \phi_F(p) = 0.$$

Un'importante proprietà della funzione quantile Q_F è l'equivarianza rispetto a trasformazioni affini. Sia F la distribuzione della variabile casuale X e G la distribuzione della variabile casuale $AX + b$, dove A è una matrice $d \times d$ non singolare e $b \in \mathbb{R}^d$, allora se Q_F è equivariante rispetto a trasformazioni affini della variabile casuale X , si ha che

$$Q_G \left(\frac{Au}{\|Au\|} \|u\| \right) = AQ_F(u) + b,$$

dove si noti che $\|u\| = p$, essendo $u = pv$ per come è stato definito all'inizio di questo paragrafo.

Dalla proprietà di equivarianza della funzione quantile Q_F abbiamo il seguente risultato.

Teorema 5.1.1. *Sia $\phi_F(p)$ basata su una funzione quantile $Q_F(\cdot)$ equivariante, allora per ogni $p \in [0, 1)$ l'indice $\phi_F(p)$ è invariante rispetto a trasformazioni affini, ossia per ogni matrice A di dimensioni $d \times d$ non singolare e $b \in \mathbb{R}^d$, si ha che*

$$\phi_F(p) = \phi_G(p),$$

dove F è la distribuzione della variabile casuale X e G è la distribuzione della variabile casuale $AX + b$.

Nel caso delle distribuzioni ellittiche tale risultato è valido anche nella direzione opposta.

Teorema 5.1.2. *Sia $Q_F(\cdot)$ equivariante. Siano X ed Y due variabili casuali aventi rispettivamente distribuzione F e G ellittica. Se*

$$\phi_F(p) = \phi_G(p), \text{ per ogni } p \in [0, 1),$$

allora $Y \stackrel{d}{=} AX + b$ per una qualche matrice A di dimensioni $d \times d$ non singolare e $b \in \mathbb{R}^d$.

5.2 L'indice ϕ_F nel caso della funzione *depth* HS

5.2.1 Un'approssimazione di C_p per $d = 2$

Come già osservato nel Capitolo 2, le curve di livello delle funzioni *depth* coincidono con quelle della funzione di densità nel caso in cui F appartiene alla famiglia delle distribuzioni ellittiche, pertanto le regioni centrali C_p hanno forma ellissoidale ed il calcolo dell'indice ϕ_F risulta essere piuttosto agevole.

Assumiamo ora che F non appartenga alla famiglia delle distribuzioni ellittiche. Più precisamente consideriamo il caso in cui la variabile casuale X sia distribuita come una ST d -dimensionale. Una prima semplificazione che otteniamo, grazie al Teorema 5.1.1, è che l'indice ϕ_F può essere ottenuto a partire dalla forma canonica della distribuzione ST. Utilizzando infatti la funzione *depth* HS, abbiamo che la funzione quantile Q_F è equivariante.

Se $X \sim \text{ST}_d(0, I_d, \alpha^*)$, con α unica componente non nulla di α^* , allora la regione centrale C_p assumerà la seguente forma

$$C_p = \{x \in \mathbb{R}^d : HS(x; F) \geq r_p\}$$

dove

$$HS(x; F) = \min_{\|u\|=1} \int_{-\infty}^{u^T x} 2t_1(s; \nu) T_1 \left(\tilde{\alpha}(\bar{u}) s \left(\frac{\nu + 1}{s^2 + \nu} \right)^{1/2}; \nu + 1 \right) ds$$

e

$$\alpha(\bar{u}) = \frac{\bar{u}\alpha}{\sqrt{1 + \alpha^2(1 - \bar{u}^2)}} \quad (5.8)$$

con \bar{u} componente di u in corrispondenza della componente asimmetrica di X .

Il problema che si pone nel caso della distribuzione ST, come per la maggior parte delle distribuzioni asimmetriche multidimensionali, è che la forma della regione centrale C_p è generalmente una figura geometrica non nota.

Un metodo consiste nell'approssimare il volume della regione centrale C_p mediante la media dei volumi di un poligono inscritto ed di un altro circoscritto a $q(p)$, la frontiera di C_p , secondo il seguente metodo. In Figura 5.2 è illustrato un possibile procedimento per l'approssimazione di C_p . Date $q(p)$ e la mediana di Tukey M_F , individuiamo una serie di punti su $q(p)$

mediante un fascio di rette passanti per M_F . Ciascuna retta è ottenuta ruotando la precedente retta individuata di π/k , dove k è un numero intero scelto a priori. La retta di partenza che è stata scelta in Figura 5.2 è data dall'asse delle ascisse. Individuiamo così $2k$ punti v_1, \dots, v_{2k} su $q(p)$, che rappresentano i vertici del poligono inscritto in $q(p)$ e che denotiamo con $P_k^{I_p}$. Ovviamente $\lim_{k \rightarrow \infty} P_k^{I_p} = C_p$. In corrispondenza di ciascun vertice v_i per ogni $i \leq 2k$ tracciamo la retta tangente a $q(p)$. Individuiamo così ulteriori $2k$ punti v_1^*, \dots, v_{2k}^* , dati dall'intersezione delle rette tangenti in ogni punto v_i , che rappresentano i vertici del poligono $P_k^{L_p}$ circoscritto a $q(p)$. Anche in questo caso $\lim_{k \rightarrow \infty} P_k = C_p$.

Mediante $P_k^{I_p}$ e $P_k^{L_p}$ siamo in grado di approssimare il volume e la massa di probabilità delimitati da C_p . Notiamo inoltre che

$$P_k^{I_p} = \bigcup_{i=1}^{2k} T_i^{I_p}$$

$$P_k^{L_p} = \bigcup_{i=1}^{2k} T_i^{L_p}$$

dove $T_i^{I_p}$ e $T_i^{L_p}$ sono i triangoli che hanno per vertici rispettivamente v_{1i}, v_{2i} e M_F , e v_{1i}, v_{2i}^* e M_F . Le approssimazioni proposte sono allora le seguenti

$$\begin{aligned} \text{volume}(C_p) &\approx \frac{1}{2}(\text{volume}(P_k^{I_p}) + \text{volume}(P_k^{L_p})) \\ &= \frac{1}{2} \left(\sum_{i=1}^{2k} \text{volume}(T_i^{I_p}) + \sum_{i=1}^{2k} \text{volume}(T_i^{L_p}) \right) \\ P(X \in C_p) &\approx \frac{1}{2}(P(X \in P_k^{L_p}) + P(X \in P_k^{I_p})) \\ &= \frac{1}{2} \left(\sum_{i=1}^{2k} P(X \in T_i^{L_p}) + \sum_{i=1}^{2k} P(X \in T_i^{I_p}) \right) \end{aligned}$$

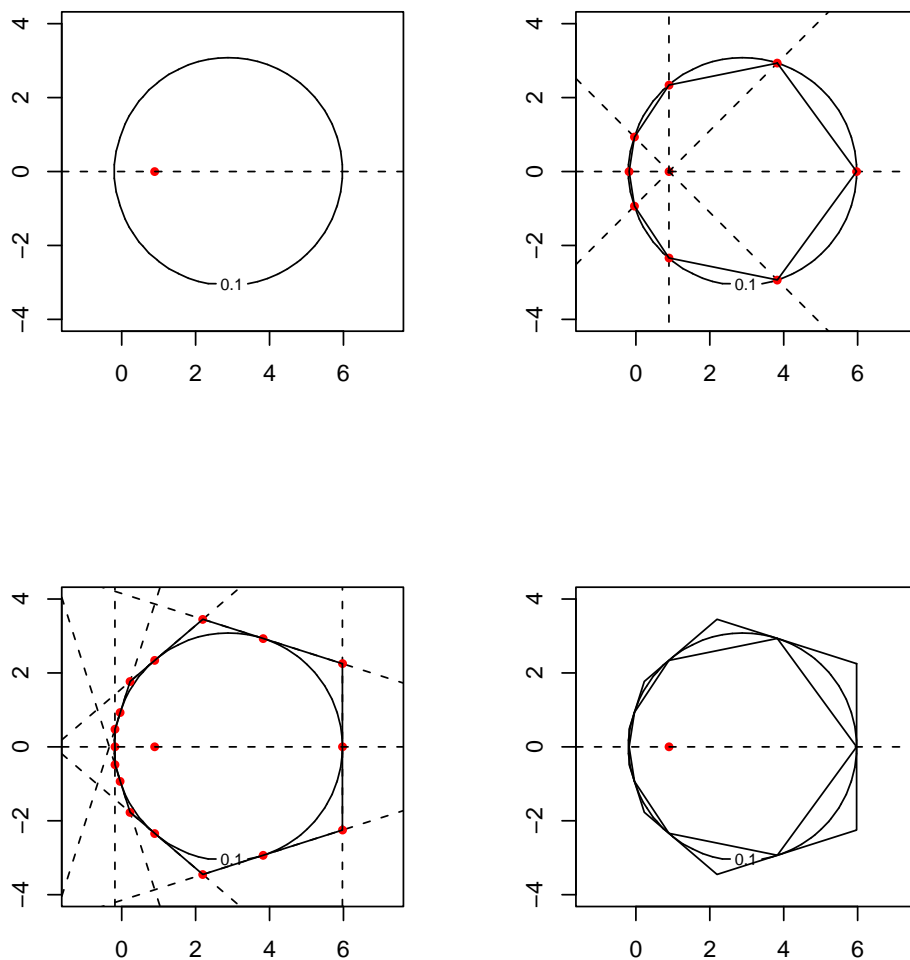


Figura 5.2: Costruzione di due poligoni rispettivamente inscritto e circoscritto ad un particolare curva di livello della funzione HS nel caso della distribuzione SC in forma canonica.

dove abbiamo che

$$\begin{aligned} \text{volume}(T_i^{I_p}) &= \frac{1}{2} |\det\{(v_{1i} - M_F, v_{2i} - M_F)\}| \\ &= \frac{1}{2} |(v_{11} - M_1)(v_{12} - M_2) - (v_{21} - M_1)(v_{22} - M_1)| \\ \text{volume}(T_i^{L_p}) &= \frac{1}{2} |\det\{(v_{1i}^* - M_F, v_{2i}^* - M_F)\}| \\ &= \frac{1}{2} |(v_{11}^* - M_1)(v_{12}^* - M_2) - (v_{21}^* - M_1)(v_{22}^* - M_1)| \end{aligned}$$

con $v_{1i} = (v_{11}, v_{12})$, $v_{2i} = (v_{12}, v_{22})$, $v_{1i}^* = (v_{11}^*, v_{12}^*)$, $v_{2i}^* = (v_{12}^*, v_{22}^*)$, $M_F = (M_1, M_2)$ e $\det(\cdot)$ è la funzione determinante di una matrice quadrata.

Il metodo di approssimazione appena proposto solleva il problema di come valutare la bontà dell'approssimazione, in quanto il calcolo dell'indice ϕ_F per una distribuzione F asimmetrica è piuttosto complesso e computazionalmente oneroso. Nel prossimo paragrafo si cercherà di fornire un'interpretazione della curtosi multidimensionale che consente di ridurre il calcolo di tale indice ad una distribuzione con simmetria sferica centrata nell'origine.

5.2.2 Una famiglia di ordinamenti di curtosi

Nel caso unidimensionale van Zwet (1986) propone un ordinamento di curtosi, denotato con \leq_s , per due distribuzioni simmetriche F_X e G_Y , ossia

$$F_X \leq_s G_Y \Leftrightarrow G_Y^{-1}(F_X(x)) \text{ è convessa per } x > \mu_F \quad (5.9)$$

dove μ_F è il punto di simmetria di F . L'estensione al caso asimmetrico della (5.9) è fornito da Balanda & MacGillivray (1990).

Nel seguito di questo paragrafo considereremo l'ordinamento di curtosi di Wang (2009), nel quale viene anche proposto una generalizzazione al caso multidimensionale. La definizione dell'ordinamento di curtosi, indicato in questo caso con \leq_{k_s} , per due distribuzioni unidimensionali F_X e G_Y è la seguente:

$$\text{per ogni } s > 0, F_X \leq_{k_s} G_Y \Leftrightarrow G_Y^{-1}(F_X(z)) \text{ è convessa per } z \geq 0, \quad (5.10)$$

dove $\tilde{X} = |X - \mu_X|/\sigma_X$ e $\tilde{Y} = |Y - \mu_Y|/\sigma_Y$ con μ_X e σ_X due opportuni indici di posizione e variabilità. Nel caso in cui μ_X e σ_X sono la media e lo scarto quadratico medio di X , allora la (5.10) coincide con l'ordinamento proposto

da Balanda & MacGillivray (1990). Se X è simmetrica allora la (5.10) è equivalente all'ordinamento dalla (5.9) proposto da van Zwet (1986).

Nel caso multidimensionale, Wang (2009) considera la distanza di Mahalanobis $R_X = [(X - \mu_X)^T \Sigma_X^{-1} (X - \mu_X)]^{1/2}$ di X da μ_X , dove μ_X e Σ_X sono rispettivamente un opportuno indice multidimensionale di posizione ed una opportuna matrice di dispersione. Va detto che i diversi risultati presentati in Wang (2009) fanno uso della media e della matrice di covarianza della variabile casuale X . Nel caso delle distribuzioni ellittiche vengono invece utilizzati il parametro di posizione e la matrice di dispersione che parametrizzano la distribuzione ellittica. Affinché gli stessi risultati siano ancora validi quando consideriamo diversi indici di posizione e matrici di dispersione, alcune condizioni, che però non sono state esplicitate in Wang (2009), devono essere rispettate.

L'estensione al caso multidimensionale dell'ordinamento di curtosi unidimensionale \leq_{k_s} è basato su $F_{R_X^s}$ e $G_{R_Y^s}$, che sono rispettivamente le funzioni di ripartizione della variabili casuali $R_X^s = [(X - \mu_X)^T \Sigma_X^{-1} (X - \mu_X)]^{s/2}$ e $R_Y^s = [(Y - \mu_Y)^T \Sigma_Y^{-1} (Y - \mu_Y)]^{s/2}$. Utilizzando la stessa notazione di Wang (2009) per l'ordinamento di curtosi multidimensionale, ossia \leq_{k_s} , abbiamo che

$$\text{per } s > 0, \text{ se } G_{R_Y^s}^{-1}(F_{R_X^s}(r)) \text{ è convessa per } r > 0 \Rightarrow F_X \leq_{k_s} G_Y. \quad (5.11)$$

Inoltre abbiamo che se $F_X \leq_{k_s} G_Y$ e $G_Y \leq_{k_s} F_X$, allora $F_X =_{k_s} G_Y$, ossia F_X e G_Y hanno lo stesso ordinamento di curtosi.

Un ordinamento di curtosi per la famiglia ST

Nel §4 di Wang (2009), notevoli considerazioni vengono sviluppate riguardo la curtosi di una distribuzione asimmetrica multidimensionale.

La curtosi di una distribuzione asimmetrica multidimensionale può essere interpretata nel seguente modo. Per ogni distribuzione asimmetrica F_X in \mathbb{R}^d , le distribuzioni che sono $=_{k_s}$, ossia k_s equivalenti ad F_X , consistono di una classe di equivalenza in curtosi $\mathcal{C}_{F_X} = \{G_Y \text{ in } \mathbb{R}^d : G_Y =_{k_s} F_X\}$. Quando studiamo la curtosi, in termini dell'ordinamento k_s , ogni distribuzione nella classe d'equivalenza può servire da rappresentante. L'unica condizione è che la classe contenga un'unica distribuzione con simmetria sferica con il centro nell'origine. La curtosi di F_X può essere allora studiata per mezzo della curtosi di questa distribuzione con simmetria sferica. Secondo questo

approccio, asimmetria e curtosi sono due componenti distinte della forma di una distribuzione.

Consideriamo ora la variabile casuale $X \sim \text{ST}_d(\xi, \Omega, \alpha, \nu)$ ed il caso in cui $s = 2$. Nel definire l'ordinamento di curtosi k_2 , considereremo come indici di posizione e dispersione rispettivamente ξ ed Ω . A tal fine risulta essere importante la proprietà di *invarianza rispetto a perturbazioni* (Azzalini & Capitanio, 2003) della distribuzione ST. Se $Z \sim T_d(\xi, \Omega, \nu)$, allora

$$t(Z - \xi) \stackrel{d}{=} t(X - \xi), \tag{5.12}$$

per ogni funzione t a valori reali tale che $t(x) = t(-x)$ per ogni $x \in \mathbb{R}^d$.

In virtù della (5.12) e delle forme quadratiche della t di Student multidimensionale (Kotz & Nadarajah (2004), p. 19), abbiamo pertanto che

$$R_X^2 = \frac{(X - \xi)^T \Omega^{-1} (X - \xi)}{d} \sim F(d, \nu), \tag{5.13}$$

dove $F(d, \nu)$ è un distribuzione F di Snedecor con d gradi di libertà al numeratore e ν gradi di libertà al denominatore. Inoltre, poiché $R_X^2 \stackrel{d}{=} R_{\tilde{Z}}^2$, con $\tilde{Z} \sim T_d(0, I_d, \nu)$, si ha che $X =_{k_2} \tilde{Z}$. Abbiamo dunque che secondo il criterio di ordinamento di curtosi k_2 appena definito, la famiglia di distribuzioni ST viene suddivisa in classi di equivalenza a ciascuna della quali corrisponde un determinato valore dei gradi di libertà ν .

Tuttavia in ciascuna classe di equivalenza abbiamo che l'indice di curtosi ϕ_F non coincide necessariamente per tutti i membri di tale insieme. Nella particolare classe di equivalenza in cui $\nu = 1$ e $d = 1$, abbiamo che l'indice ϕ_F è identico per ogni distribuzione in tale classe. Infatti, come si può desumere dalla (4.5), il funzionale statistico SP_F non dipende dal parametro di forma responsabile dell'asimmetria distributiva ed assume un'identica forma per la distribuzione Cauchy.

In generale, data una classe di equivalenza \mathcal{C}_ν , corrispondente a ν gradi di libertà, disponiamo di una famiglia di indici di curtosi ϕ_F per ogni $F \in \mathcal{C}_\nu$. L'idea è quella di scegliere una distribuzione rappresentante della classe di equivalenza e calcolare per questa l'indice di curtosi ϕ_F . Notiamo inoltre che ciascuna classe di equivalenza contiene un'unica distribuzione con simmetria sferica centrata nell'origine, ossia una $T_d(0, I_d, \nu)$, che rappresenta una delle possibili ed anche delle più convenienti scelte per rappresentare la classe di equivalenza, in quanto matematicamente più facile da trattare.

Indicheremo l'indice di curtosi, dato dalla (5.6) e calcolato per la distribuzione $T_d(0, I_d, \nu)$, con $\phi_{T_d, \nu}$ e tale indice sarà usato per descrivere la curtosi di ciascuna distribuzione appartenente a tale classe di equivalenza.

Il seguente risultato presentato in Wang (2009) ci permette di meglio caratterizzare l'ordinamento k_2 per la distribuzione ST.

Corollario 5.2.1. *Siano $X \sim T_d(0, I_d, \nu_X)$ e $Y \sim T_d(0, \Omega_Y, \nu_Y)$. Se $\nu_X > \nu_Y$ allora $X \geq_{k_2} Y$.*

Pertanto date due distribuzioni ST_ν con ν gradi di libertà ed un'altra $ST_{\nu'}$ con ν' gradi di libertà, tale che $\nu' > \nu$, abbiamo che $ST_\nu \leq_{k_2} ST_{\nu'}$. Allora avremo anche che $\phi_{T_d, \nu} \leq \phi_{T_d, \nu'}$, come è mostrato in Figura 5.3.

Va detto tuttavia che l'impiego dell'indice $\phi_{T_d, \nu}$ per confrontare la curtosi di due diverse distribuzioni ST necessita delle dovute cautele. Curtosi ed asimmetria sono infatti caratteristiche non indipendenti tra di loro (una discussione al riguardo nel caso unidimensionale è presentata in MacGillivray & Balanda (1988)). Come si evince da Figura 5.4, in cui sono riportate alcune curve dell'indice ϕ_F , nel caso in cui $\alpha \neq 0$ c'è un incremento positivo dell'indice di curtosi, ossia c'è uno spostamento della massa di probabilità dalle "spalle" verso la "zona centrale" della distribuzione. L'indice $\phi_{T_d, \nu}$ può essere allora utilizzato nel confrontare la curtosi di distribuzioni ST che hanno una grado di asimmetria nella rispettiva forma canonica prossimo tra di loro, ossia per valori di $(\alpha^T \bar{\Omega} \alpha)^{1/2}$ simili.

5.2.3 Il caso della SC

In questo paragrafo analizzeremo il caso in cui $X \sim SC_d(\xi, \Omega, \alpha)$. Una caratteristica rilevante della distribuzione SC riguarda la forma delle curve di livello della funzione HS.

Teorema 5.2.2. *Sia $X \sim SC_d(\xi, \Omega, \alpha)$, allora le curve di livello della funzione HS hanno forma ellissoidale.*

Dimostrazione. Consideriamo, senza perdita di generalità, la distribuzione di X in forma canonica ed assumiamo che la prima componente X_1 sia asimmetrica, con parametro di forma α . Ci limitiamo al caso in cui $x_i > 0$, $\forall 2 \leq i \leq d$, poiché per il Teorema 2.3.7 le curve di livello sono simmetriche rispetto agli assi delle componenti simmetriche di X .

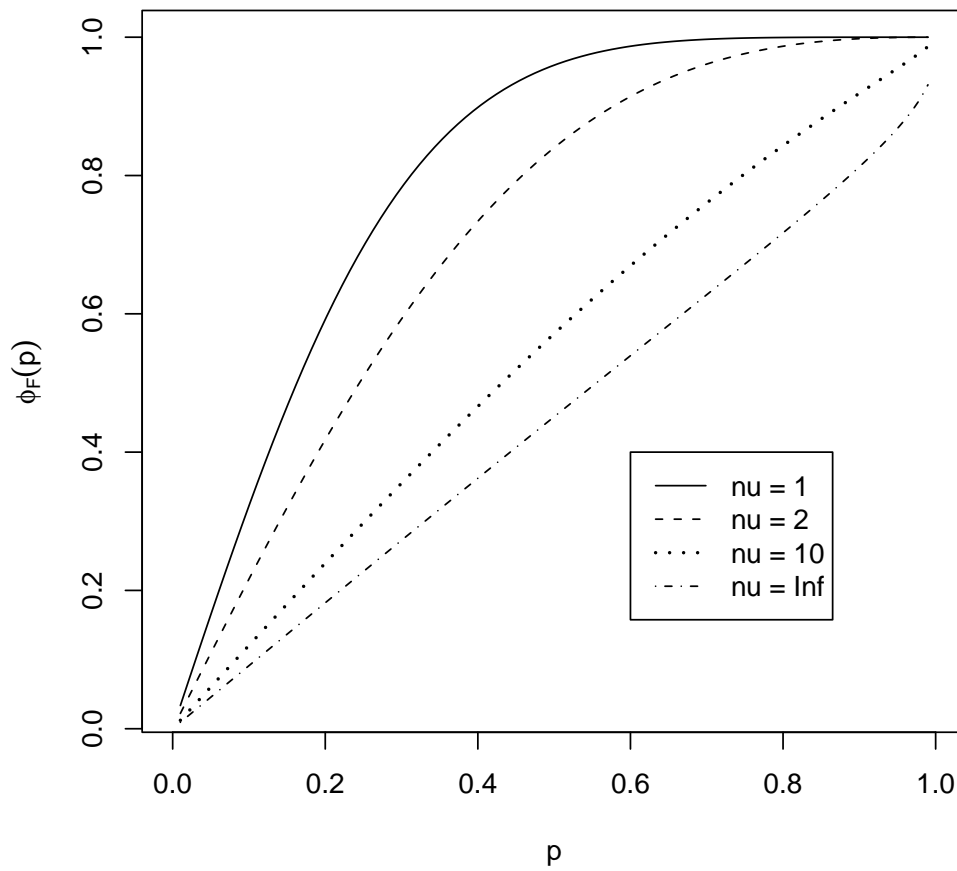


Figura 5.3: Curve dell'indice ϕ_F , utilizzando la funzione *depth* HS, per la distribuzione T_d nello spazio di dimensione $d = 5$, con gradi di libertà $\nu = 1, 2, 10, \infty$. Nel caso in cui $\nu = \infty$, la distribuzione è una normale multidimensionale a componenti indipendenti standardizzate.

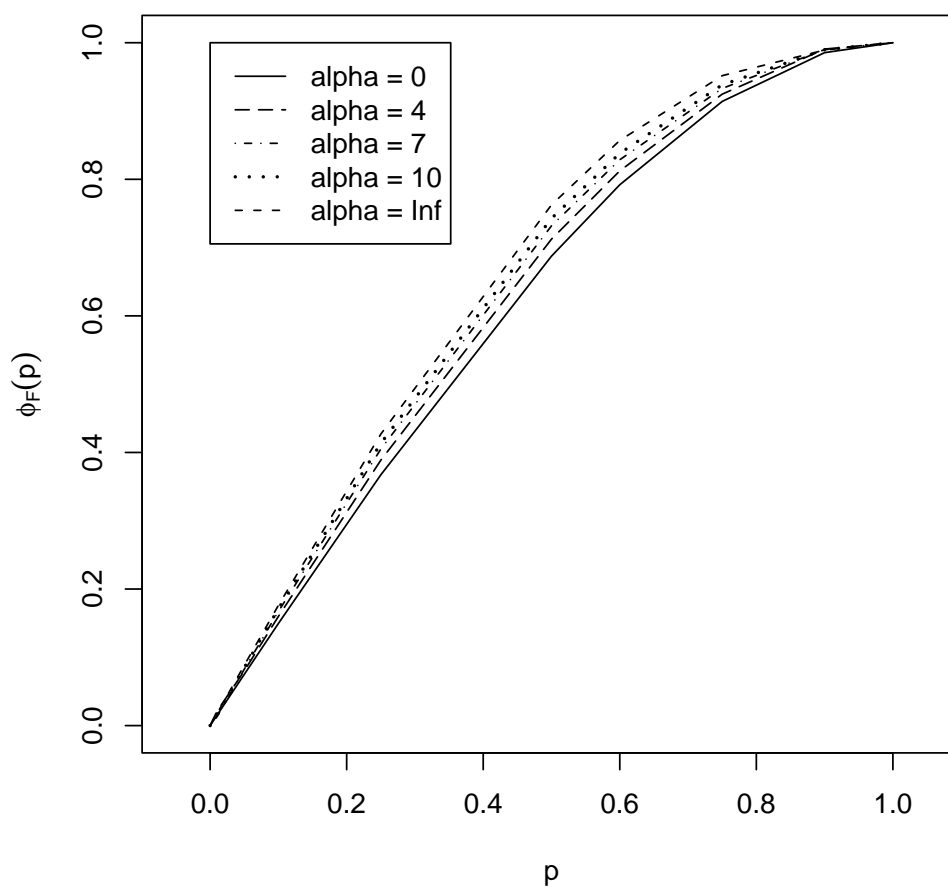


Figura 5.4: Alcune curve dell'indice ϕ_F , utilizzando la funzione *depth* HS, per la distribuzione $SC_2(0, I_2, (\alpha, 0)^T)$ con $\alpha = 0, 4, 7, 10, \infty$.

Dato un punto $x \in \mathbb{R}^d$ appartenente alla curva di livello r ed utilizzando la stessa argomentazione usata per la (2.11), abbiamo che

$$\begin{aligned}
 \text{HS}(x) &= \min_{\|u\|=1} P \left(u_1 X_1 + \sum_{i=2}^d u_i X_i > \sum_{i=1}^d u_i x_i \right) \\
 &= \min_{\|u\|=1} P \left(u_1 X_1 + \sqrt{1 - u_1^2} T > \sum_{i=1}^d u_i x_i \right) \\
 &= P \left(\tilde{u}_1 X + \sqrt{1 - \tilde{u}_1^2} T > \sum_{i=1}^d \tilde{u}_i x_i \right) \\
 &= r \in (0, 1/2)
 \end{aligned} \tag{5.14}$$

dove $(X_1, T) \sim SC_2(0, I_2, (\alpha, 0)^T)$ e $\tilde{u} = (\tilde{u}_1, \dots, \tilde{u}_d)$ è il valore di u che minimizza la probabilità del semipiano.

Abbiamo inoltre che

$$\tilde{u}_1 X_1 + \sqrt{1 - \tilde{u}_1^2} T \sim SC_1(0, 1, \alpha(\tilde{u}_1)),$$

con $\alpha(\tilde{u}_1)$ dato dalla (5.8). Utilizzando le espressioni per i quantili di una SC, fornite da Behboodian, Jamalizadeh & Balakrishnan (2006), abbiamo che

$$\tilde{u}_1 x_1 + \sum_{i=2}^d \tilde{u}_i x_i = \tilde{u}_1 s(r) + t(r),$$

che è equivalente a

$$\tilde{u}_1 (x_1 - s(r)) + \sum_{i=2}^d \tilde{u}_i x_i = t(r) \tag{5.15}$$

dove

$$s(r) = \frac{\alpha}{\sqrt{1 + \alpha^2}} \sec \left\{ \left(\frac{1}{2} - r \right) \pi \right\}, \tag{5.16}$$

$$t(r) = \tan \left\{ \left(\frac{1}{2} - r \right) \pi \right\}. \tag{5.17}$$

Si noti che $t(r) > 0$ ed $s(r) > 0$ per ogni $r \in (0, 1/2)$.

Consideriamo ora l'insieme di punti

$$\mathbb{S}_r^{d-1} = \left\{ x \in \mathbb{R}^d : (x_1 - s(r))^2 + \sum_{i=2}^d x_i^2 = t(r)^2 \right\}.$$

Allora abbiamo che

$$\begin{aligned} t(r) &= \frac{t(r)^2}{t(r)} \\ &= \frac{(x_1 - s(r))^2 + \sum_{i=2}^d x_i^2}{t(r)} \\ &= \frac{x_1 - s(r)}{t(r)}(x_1 - s(r)) + \sum_{i=2}^d \frac{x_i}{t(r)}x_i. \end{aligned}$$

Pertanto ponendo

$$\tilde{u}_1 = \frac{x_1 - s(r)}{t(r)} \text{ e } \tilde{u}_i = \frac{x_i}{t(r)}, \forall 2 \leq i \leq d$$

si ha che ogni punto $x \in \mathbb{S}_r^{d-1}$ soddisfa la (5.15) e dunque \mathbb{S}_r^{d-1} rappresenta la curva di livello r -esima.

Infine nel caso in cui X non sia in forma canonica, sfruttando l'invarianza della funzione HS, si ha direttamente la tesi. \square

Consideriamo ora la distribuzione SC in forma canonica poiché l'indice ϕ_F sarà uguale per tutte le altre SC ottenibili per trasformazione affine da questa. Abbiamo allora che

$$C_p = \left\{ x \in \mathbb{R}^d : (x_1 - s[r(p)])^2 + \sum_{i=2}^d x_i^2 \leq t[r(p)]^2 \right\},$$

dove

$$r(p) = \left\{ r \in (0, 1/2) : P \left((X_1 - s(r))^2 + \sum_{i=2}^d X_i^2 \leq t(r)^2 \right) = p \right\},$$

mentre $s(r)$ e $t(r)$ sono dati rispettivamente dalla (5.16) e dalla (5.17). Il calcolo di $r(p)$ viene effettuato per via numerica trasformando le variabili casuali in coordinate ipersferiche. Si ottiene allora che

$$\text{volume}(C_p) = \frac{\pi^{d/2} (t[r(p)])^d}{\Gamma(d/2 + 1)},$$

pertanto

$$\phi_{\text{SC}_d} = \frac{(t[r(\frac{1-p}{2})])^d + (t[r(\frac{1+p}{2})])^d - 2(t[r(\frac{1}{2})])^d}{(t[r(\frac{1+p}{2})])^d - (t[r(\frac{1-p}{2})])^d}. \quad (5.18)$$

Capitolo 6

Considerazioni finali

6.1 Una sintesi

Nel corso della trattazione sono stati considerati di volta in volta i principali aspetti che caratterizzano una distribuzione di probabilità, nell'ordine posizione, variabilità e struttura di dipendenza, asimmetria e curtosi. Per ognuno di questi sono stati individuati degli opportuni indici basati sul concetto di quantile, e di concordanza nel caso della struttura di dipendenza. Particolare attenzione è stata inoltre riservata alla distribuzione ST.

Nel Capitolo 2 si è fatto ricorso al concetto di funzione *depth*, una generalizzazione del concetto di rango al caso multidimensionale. Le funzioni *depth*, tra le diverse proprietà possedute, sono invarianti rispetto a trasformazioni affini della variabile casuale. Ci siamo così focalizzati sulla funzione *depth* HS proposta da Tukey (1975). Per un dato punto $x \in \mathbb{R}^d$, la funzione HS restituisce la probabilità del semipiano a minima probabilità sulla cui frontiera giace x , ossia

$$\text{HS}(x) = \min_{\|u\|=1} P(u^T X \leq u^T x).$$

Si sono analizzati, mediante la funzione HS, alcuni aspetti della distribuzione ST legati alla simmetria radiale. Si è osservato che nel caso particolare in cui $\nu = 1$ la simmetria radiale è esatta. Un naturale indice di posizione per tale caso distributivo è dato allora dalla mediana per componenti η , che rappresenta il centro di simmetria radiale. Per $\nu \rightarrow \infty$ si ha un allontanamento, seppur estremamente lento, dalla simmetria radiale. Due possibili indici di posizione sono allora la mediana di Tukey M_F e l'indice η^* dato dalla (2.9).

Tali indici sono equivarianti ed estremamente vicini tra di loro sia in termini di funzione HS che di densità (Figura 2.4, Figura ??).

Nel Capitolo 3 è stato indagato il concetto di concordanza al fine di ottenere una misura di associazione per coppie di variabili casuali (X, Y) . Date due realizzazioni (x_1, y_1) e (x_2, y_2) provenienti da (X, Y) , queste sono concordanti se $(x_1 - x_2)(y_1 - y_2) > 0$ o discordanti se $(x_1 - x_2)(y_1 - y_2) < 0$. Sono stati così considerati gli indici τ di Kendall e ρ di Spearman. Siano (X_1, Y_1) , (X_2, Y_2) e (X_3, Y_3) tre variabili casuali bidimensionali indipendenti ed identicamente distribuite. L'indice τ è definito come la differenza tra la probabilità di concordanza e la probabilità di discordanza, ossia

$$\tau(X, Y) = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0).$$

L'indice ρ è invece dato dalla seguente espressione

$$\rho(X, Y) = 3 [P((X_1 - X_2)(Y_1 - Y_3) > 0) - P((X_1 - X_2)(Y_1 - Y_3) < 0)].$$

Tali indici permettono di costruire matrici, simmetriche e definite positive, di concordanza (Visuri, Koivunen & Oja, 2000) che esprimono il grado di associazione in termini appunto della concordanza tra coppie di componenti del vettore casuale. Date le variabili casuali d -dimensionali X_1 , X_2 e X_3 , indipendenti ed identicamente distribuite, definiamo la funzione $\text{sign}(\cdot)$, ossia il segno marginale delle componenti del vettore $x \in \mathbb{R}^d$. Allora la matrice di concordanza della τ di Kendall è

$$S_\tau = E [\text{sign}(X_1 - X_2) \text{sign}(X_1 - X_2)^T].$$

La matrice di concordanza della ρ di Spearman è data invece dalla seguente espressione

$$S_\rho = E [\text{sign}(X_1 - X_2) \text{sign}(X_1 - X_3)^T].$$

Quando si vuol tener conto anche della variabilità delle componenti del vettore casuale, lo scarto interquartile può essere considerato come misura di dispersione, ottenendo le seguenti matrici di concordanza non normalizzate

$$\begin{aligned} \tilde{S}_\tau &= \Delta S_\tau \Delta, \\ \tilde{S}_\rho &= \Delta S_\rho \Delta, \end{aligned}$$

dove $\Delta = \text{diag}([IQ_i]_{i \leq d})$, con IQ_i scarto interquartile dell' i -esima componente del vettore casuale X . Tuttavia tali matrici di concordanza non sono

equivarianti e pertanto non possono essere trattate come matrici di dispersione.

Nel Capitolo 4 sono stati discussi diversi concetti di asimmetria unidimensionale, legati tra loro da un preciso ordine di generalità. Ci si è così focalizzati sul concetto più generale di asimmetria tra quelli discussi, secondo il quale una distribuzione F è detta asimmetrica a destra se

$$SK_F(p) = \frac{1}{2}(F^{-1}(1-p) + F^{-1}(p) - 2F^{-1}(1/2)) > 0, \forall p \in (1/2, 1).$$

Nel caso di asimmetria a sinistra è solamente necessario invertire il predicato della disuguaglianza. Al fine di utilizzare una misura di asimmetria che fosse sensibile a tale concetto di asimmetria, è stato scelto l'indice di Yule-Bowley, dato dalla seguente espressione

$$b_F = \frac{F^{-1}(3/4) + F^{-1}(1/4) - 2F^{-1}(1/2)}{F^{-1}(3/4) + F^{-1}(1/4)}.$$

Nel Capitolo 5 è stata discussa la famiglia di indici multidimensionali di curtosi ϕ_F di Wang & Serfling (2005), basata sulle funzioni *depth*. Date una distribuzione F assolutamente continua con funzione di densità non nulla ovunque ed una funzione *depth* $D(\cdot; F)$, abbiamo che la p -esima regione centrale è

$$C_p = \{x \in \mathbb{R}^d : D(x; F) \geq t_p\},$$

per un opportuno $t_p > 0$ tale che $P(C_p) = p$. Le “spalle” della distribuzione possono essere interpretate come la frontiera della regione centrale $C_{1/2}$ che separa la “zona centrale” dalle “code”. L'indice ϕ_F misura la relativa differenza volumetrica tra una regione interna alla “zona centrale” ed una regione esterna alla “zona centrale”. Queste due regioni sono ottenute sottraendo ed addizionando a $p = 1/2$ una data quantità $p/2$. Indicando la famiglia di regioni centrali con $\mathcal{C} = \{C_p : p \in [0, 1]\}$ ed il volume funzionale con $V_{F,\mathcal{C}}(p) = \text{volume}(C_p)$, si ha che

$$\phi_F(p) = \frac{V_{F,\mathcal{C}}\left(\frac{1}{2} - \frac{p}{2}\right) + V_{F,\mathcal{C}}\left(\frac{1}{2} + \frac{p}{2}\right) - 2V_{F,\mathcal{C}}\left(\frac{1}{2}\right)}{V_{F,\mathcal{C}}\left(\frac{1}{2} + \frac{p}{2}\right) - V_{F,\mathcal{C}}\left(\frac{1}{2} - \frac{p}{2}\right)}, p \in (0, 1).$$

L'attenzione è stata poi rivolta all'impiego della funzione *depth* HS per la costruzione di tale indice. Il calcolo dell'indice per una distribuzione multidimensionale asimmetrica presenta notevoli difficoltà. Due diverse vie sono

state percorse per risolvere il problema della computazione. Nel caso bidimensionale si è proposta un'approssimazione del volume funzionale $V_{F,\mathcal{C}(p)}$ mediante la costruzione di un poligono inscritto alla frontiera della regione centrale C_p , denotato con P^{I_p} , ed uno circoscritto alla stessa, denotato con P^{C_p} . Allora otteniamo la seguente approssimazione

$$V_{F,\mathcal{C}(p)} \approx \frac{1}{2} (\text{volume}(P^{I_p}) + \text{volume}(P^{C_p})).$$

Un differente approccio consiste nel considerare la famiglia di ordinamenti di curtosi di Wang (2009). Date due variabili casuali d -dimensionali X ed Y , tale ordinamento di curtosi è basato su $F_{R_X^s}$ e $G_{R_Y^s}$ che sono rispettivamente le funzioni di ripartizione della variabili casuali $R_X^s = [(X - \mu_X)^T \Sigma_X^{-1} (X - \mu_X)]^{s/2}$ e $R_Y^s = [(Y - \mu_Y)^T \Sigma_Y^{-1} (Y - \mu_Y)]^{s/2}$, dove μ_X e Σ_X sono rispettivamente un opportuno indice di posizione ed una opportuna matrice di dispersione per la variabile X , ed analogamente per Y . Utilizzando la stessa notazione di Wang (2009) per l'ordinamento di curtosi, ossia \leq_{k_s} , abbiamo che

$$\text{per } s > 0, \text{ se } G_{R_Y^s}^{-1}(F_{R_X^s}(r)) \text{ è convessa per } r > 0 \Rightarrow F_X \leq_{k_s} G_Y. \quad (6.1)$$

Nel caso della distribuzione ST e per $s = 2$, un possibile ordinamento di curtosi è ottenuto considerando $\mu_X = \xi$ e $\Sigma_X = \Omega$, ossia

$$R_X^2 = (X - \xi)^T \Omega^{-1} (X - \xi).$$

Poiché $R_X^2 \sim F(d, \nu)$, otteniamo una serie di classi di equivalenza corrispondenti a diversi gradi di libertà ν ognuna delle quali contiene un'unica distribuzione con simmetria sferica centrata nell'origine. Secondo quanto argomentato in Wang (2009), ogni membro della una classe di equivalenza può essere scelto come rappresentante di questa, pertanto la curtosi può essere reinterpretata come una caratteristica distributiva indipendente dall'asimmetria. Il calcolo dell'indice ϕ_F può allora essere effettuato in modo agevole per la distribuzione con simmetria sferica appartenente alla classe di equivalenza ed utilizzare tale indice per descrivere la curtosi di tutte le distribuzioni nella classe di equivalenza. Tuttavia nel caso in cui la regione centrale C_p ha una forma riconducibile ad una forma geometrica nota, il calcolo dell'indice ϕ_F può essere effettuato direttamente: questo è il caso della distribuzione SC per la quale la funzione *depth* HS risulta avere curve di livello di forma ellissoidale. Va detto però che le due procedure seguite presentano entrambe alcune problematiche: nel primo caso è necessario fornire una qualche stima

l'errore di approssimazione; nel secondo l'interpretazione dell'indice viene limitata a distribuzioni ST che hanno un valore simile per l'indice $(\alpha^T \bar{\Omega} \alpha)^{1/2}$ (l'unica componente non nulla del parametro responsabile dell'asimmetria distributiva nella rispettiva forma canonica) in quanto asimmetria e curtosi sono in realtà interrelate tra di loro.

6.2 Applicazione: dati sui rendimenti finanziari

In questo paragrafo analizzeremo i rendimenti delle azioni della Deutsche Bank (DBK) e della Siemens (SIE), quotate a Francoforte. I rendimenti vanno dal 2 Gennaio 2003 al 23 Marzo 2010 ed abbiamo in tutto 1840 osservazioni giornaliere.

In Figura 6.1 è riportato il diagramma di dispersione per le due variabili. Come mostrato in Figura 6.2, il modello ST bidimensionale sembrerebbe adattarsi in modo più che soddisfacente ai dati. Il grafico pp-plot di Figura 6.2 è stato costruito sfruttando la forma distributiva delle forme quadratiche della distribuzione ST, data dalla (5.13). Le stime di massima verosimiglianza dei parametri della distribuzione sono le seguenti.

$$\begin{aligned}\hat{\xi} &= (0.185 \quad 0.101)^T, \\ \hat{\Omega} &= \begin{pmatrix} 1.815 & 1.146 \\ 1.146 & 1.555 \end{pmatrix}, \\ \hat{\alpha} &= (-0.161 \quad 0.062)^T, \\ \hat{\nu} &= 2.450.\end{aligned}$$

Notiamo che il valore del parametro di asimmetria dell'unica componente asimmetrica della corrispondente forma canonica è pari a $(\hat{\alpha}^T \hat{\omega}^{-1} \hat{\Omega} \hat{\omega}^{-1} \hat{\alpha})^{1/2} = 0.127$, pertanto i dati non mostrano una forte asimmetria.

Di seguito è riportata la stima di alcuni indici discussi nei precedenti capitoli e la rispettiva stima delle matrici di covarianza della distribuzione

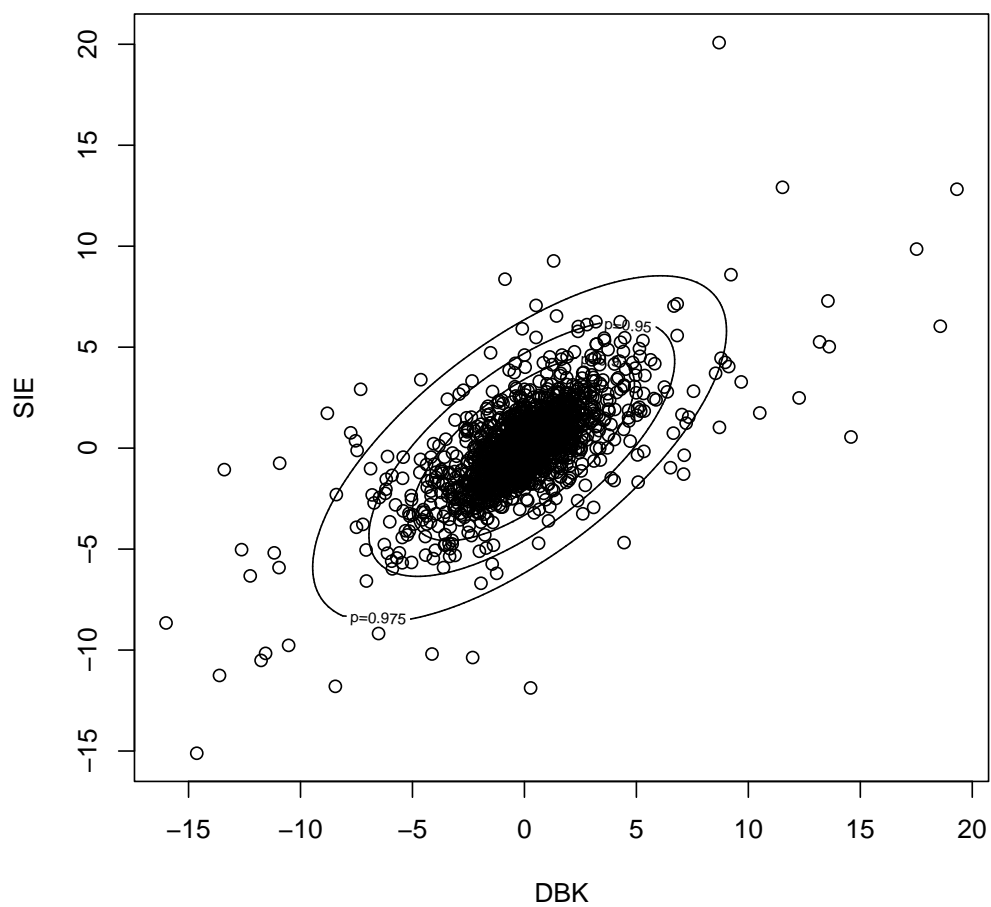


Figura 6.1: Diagrammi di dispersione delle variabili DBK e SIE con alcune curve di livello della funzione di densità.

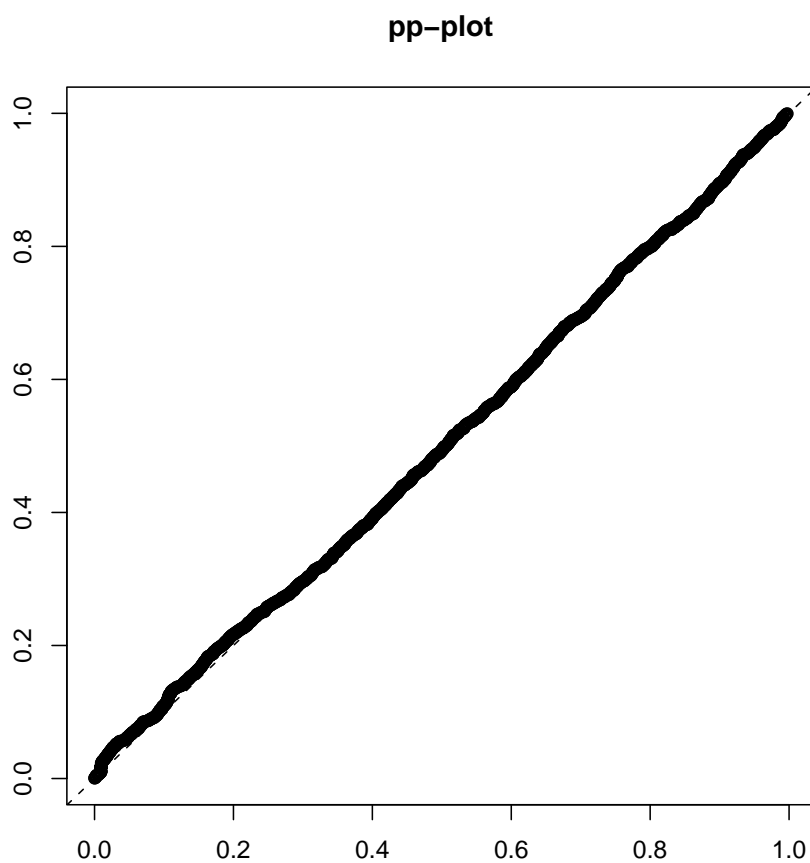


Figura 6.2: Diagramm pp-plot per la verifica della bontà di adattamento del modello ST.

asintotica per componenti di indici.

$$\begin{aligned} \begin{pmatrix} \hat{\eta}_{\text{DBK}}^* \\ \hat{\eta}_{\text{SIE}}^* \end{pmatrix} &= \begin{pmatrix} 0.046 \\ 0.049 \end{pmatrix}, \quad \text{var} \begin{pmatrix} \hat{\eta}_{\text{DBK}}^* \\ \hat{\eta}_{\text{SIE}}^* \end{pmatrix} = \begin{pmatrix} 0.001 & 0.001 \\ 0.001 & 0.001 \end{pmatrix}, \\ \hat{\tau}_{\text{DBK,SIE}} &= 0.481, \quad \text{var}(\hat{\tau}_{\text{DBK,SIE}}) = 0.053, \\ \hat{\rho}_{\text{DBK,SIE}} &= 0.644, \quad \text{var}(\hat{\rho}_{\text{DBK,SIE}}) = 0.161, \\ \begin{pmatrix} \hat{b}_{\text{DBK}} \\ \hat{b}_{\text{SIE}} \end{pmatrix} &= \begin{pmatrix} 0.019 \\ 0.006 \end{pmatrix}, \quad \text{var} \begin{pmatrix} \hat{b}_{\text{DBK}} \\ \hat{b}_{\text{SIE}} \end{pmatrix} = \begin{pmatrix} 0.006 & 0.006 \\ 0.006 & 0.034 \end{pmatrix}, \\ \hat{\phi}_{T_2,\nu}(0.5) &= 0.460, \quad \text{var}(\hat{\phi}_{T_2,\nu}(0.5)) = 9.657 \times 10^{-5}. \end{aligned}$$

Indicando con IQ_{DBK}^2 e IQ_{SIE}^2 rispettivamente il quadrato dello scarto interquartile al quadrato per le variabili DBK e SIE abbiamo che

$$\begin{pmatrix} \hat{IQ}_{\text{DBK}}^2 \\ \hat{IQ}_{\text{SIE}}^2 \end{pmatrix} = \begin{pmatrix} 4.473 \\ 3.853 \end{pmatrix}, \quad \text{var} \begin{pmatrix} \hat{IQ}_{\text{DBK}}^2 \\ \hat{IQ}_{\text{SIE}}^2 \end{pmatrix} = \begin{pmatrix} 0.212 & 0.201 \\ 0.201 & 0.194 \end{pmatrix}.$$

6.3 Software: le principali routine

In questo paragrafo descriveremo le principali funzioni, presentate in appendice, che sono state sviluppate utilizzando il software R 2.14.1 e che sono state utilizzate nel corso della trattazione.

- `TukMed.mst`: tale funzione calcola la mediana di Tukey per una data distribuzione multidimensionale ST.

Argomenti

`f` un oggetto `SEDistrMultiv` con `family=ST`.

Output

Il vettore corrispondente alla mediana di Tukey della distribuzione ST data da `f`.

Note

Sfruttando il risultato del Teorema 2.3.10, tale funzione calcola prima la mediana di Tukey per la rispettiva distribuzione ST in forma canonica,

risolvendo pertanto un problema di ottimizzazione in una variabile reale. Mediante trasformazione affine della mediana di Tukey della forma canonica viene così ottenuta sulla scala della distribuzione f .

- `etas.mst`: tale funzione calcola le stime dell'indice η^* per ciascuna componente di una data distribuzione bidimensionale ST stimata con il metodo della massima verosimiglianza, e rispettiva matrice di covarianza.

Argomenti

`f` un oggetto `SEDistrMultiv` con `family=ST`.

`info` matrice dell'informazione osservata di Fisher per la distribuzione `f`.

Output

Una lista di componenti:

`etas` l'indice η^* per la distribuzione `f`.

`etas.var` matrice di covarianza della distribuzione asintotica congiunta dello stimatore di η^* .

- `tau.mst`: tale funzione calcola la stima dell'indice τ di Kendall e rispettiva varianza per una data distribuzione bidimensionale ST stimata con il metodo della massima verosimiglianza.

Argomenti

`f` un oggetto `SEDistrMultiv` con `family=ST`.

`info` matrice dell'informazione osservata di Fisher per la distribuzione `f`.

... ulteriori argomenti da passare alla funzione `cuhre` della libreria `R2Cuba`.

Output

Una lista di componenti:

`tau` l'indice τ di Kendall per la distribuzione `f`.

`tau.var` varianza della distribuzione asintotica dello stimatore.

Note

La computazione dell'indice τ e rispettiva stima della varianza dello stimatore possono richiedere anche più di dieci minuti, si consiglia pertanto di aumentare, ragionevolmente, il valore relativo tolleranza dell'errore assoluto `abs.tol`.

- `rho.mst`: tale funzione calcola la stima dell'indice ρ di Spearman e rispettiva varianza per una data distribuzione bidimensionale ST stimata con il metodo della massima verosimiglianza.

Argomenti

`f` un oggetto `SEDistrMultiv` con `family=ST`.

`info` matrice dell'informazione osservata di Fisher per la distribuzione `f`.

... ulteriori argomenti da passare alla funzione `cuhre` della libreria `R2Cuba`.

Output

Una lista di componenti:

`rho` l'indice ρ di Spearman per la distribuzione `f`.

`rho.var` varianza della distribuzione asintotica dello stimatore.

Note

La computazione dell'indice ρ e rispettiva stima della varianza dello stimatore possono richiedere anche più di dieci minuti, si consiglia pertanto di aumentare, ragionevolmente, il valore relativo tolleranza dell'errore assoluto `abs.tol`.

- `yb.st`: tale funzione calcola le stime dell'indice Yule-Bowley per ciascuna componente di una data distribuzione bidimensionale ST stimata con il metodo della massima verosimiglianza, e rispettiva matrice di covarianza.

Argomenti

`f` un oggetto `SEDistrMultiv` con `family=ST`.

`info` matrice dell'informazione osservata di Fisher per la distribuzione `f`.

Output

Una lista di componenti:

`yb` il vettore degli indici di Yule-Bowley per ciascuna componente della distribuzione `f`.

`yb.var` matrice di covarianza della distribuzione asintotica congiunta dei due stimatori.

- `phi.Td`: tale funzione calcola la stima dell'indice di curtosi $\phi_{T_d,\nu}$ e rispettiva varianza per una data distribuzione bidimensionale ST stimata con il metodo della massima verosimiglianza.

Argomenti

`p` un valore compreso tra 0 ed 1.

`nu` valori dei gradi di libertà.

`info` matrice dell'informazione osservata di Fisher per la distribuzione ST.

Output

Una lista di componenti:

`phi` l'indice $\phi_{T_d, \nu}$ per la distribuzione bidimensionale ST.

`phi.var` varianza della distribuzione asintotica dello stimatore.

- `IQ2.st`: tale funzione calcola le stime del quadrato dello scarto interquartile per ciascuna componente di una data distribuzione bidimensionale ST stimata con il metodo della massima verosimiglianza, e rispettiva matrice di covarianza.

Argomenti

`f` un oggetto `SEDistrMultiv` con `family=ST`.

`info` matrice dell'informazione osservata di Fisher per la distribuzione `f`.

Output

Una lista di componenti:

`IQ2` il vettore dei quadrati degli scarti interquartili di ciascuna componente della distribuzione `f`.

`IQ2.var` matrice di covarianza della distribuzione asintotica congiunta dei due stimatori.

Ringraziamenti

Ringrazio in prima persona il prof. Azzalini per aver ripetutamente corretto diverse bozze preliminari di questa tesi e per gli svariati suggerimenti riguardo lo stile di stesura.

Ringrazio anche Robert Serfling (Department of Mathematics, University of Texas at Dallas Richardson), Probal Chaudhuri (Theoretical Statistics and Mathematics Unit, Indian Statistical Institute) e Jin Wang (Department of Mathematics and Statistics, Northern Arizona University) per i diversi suggerimenti che hanno consentito di migliorare il Capitolo 5.

Appendice A

Codice R

```
#####  
##### STIME E VARIANZE DEI QUANTILI #####  
#####  
  
#### Funzione per il calcolo del gradiente  
  
comp.quant <- function(r,f,comp) {  
  dp <- f@dp; dp.m <- marginalDistribution(f,comp=comp)@dp  
  xi <- dp$xi[comp]; w1 <- diag(dp$Omega)[comp];  
  w2 <-diag(dp$Omega)[-comp]; w12 <- dp$Omega[2,1]  
  a <- dp$alpha[comp]; b <- dp$alpha[-comp]; nu <- dp$df  
  a.marg <- dp.m[3]  
  quant <- qst(r,dp=dp.m)  
  alpha.m <- function(w1,w2,w12,a1,a2) {  
    w.bar <- w12/sqrt(w1*w2)  
    (a1+a2*w.bar)/sqrt(1+(a2^2)*(1-w.bar^2))  
  }  
  c2 <- grad(function(x) qst(r,xi=dp.m[1],  
    omega=sqrt(x),alpha=alpha.m(x,w2,w12,a,b),  
    df=dp.m[4]),w1)  
  c3 <- grad(function(x) qst(r,xi=dp.m[1],  
    omega=sqrt(w1),alpha=alpha.m(w1,x,w12,a,b),  
    df=dp.m[4]),w2)
```

```

c4 <- grad(function(x) qst(r,xi=dp.m[1],
  omega=sqrt(w1),alpha=alpha.m(w1,w2,x,a,b),
  df=dp.m[4]),w12)
c5 <- grad(function(x) qst(r,xi=dp.m[1],
  omega=sqrt(w1),alpha=alpha.m(w1,w2,w12,x,b),
  df=dp.m[4]),a)
c6 <- grad(function(x) qst(r,xi=dp.m[1],
  omega=sqrt(w1),alpha=alpha.m(w1,w2,w12,a,x),
  df=dp.m[4]),b)
c7 <- grad(function(x) qst(r,xi=dp.m[1],
  omega=sqrt(w1),alpha=alpha.m(w1,w2,w12,a,b),
  df=x),dp.m[4])
c(1,c2,c3,c4,c5,c6,c7)
}

mat.med12 <- function(r1,r2,f,info) {
  grad1 <- matrix(NA,nrow=length(r1),ncol=7)
  grad2 <- matrix(NA,nrow=length(r2),ncol=7)
  for(i in 1:length(r1)) {
    grad1[i,] <- comp.quant(r1[i],f,1)
  }
  for(i in 1:length(r2)) {
    grad2[i,] <- comp.quant(r2[i],f,2)
  }
  gr1 <- matrix(c(grad1[,1],rep(0,length(r1)),
    grad1[,-1]),ncol=8,byrow=FALSE)
  gr2 <- matrix(c(rep(0,length(r2)),
    grad2),ncol=8,byrow=FALSE)
  g <- rbind(gr1,gr2)
  g%%solve(info)%*%t(g)
}

#####
##### MEDIANA DI TUKEY #####
#####

### Funzione per il calcolo della matrice per il passaggio

```

```

### alla forma canonica della ST

matrixA <- function(f) {
  dp <- f@dp
  d <- length(dp$xi)
  B <- solve(t(chol(dp$Omega)))
  alphas <- affineTransformation(f, A = B)@dp$alpha
  P <- t(orthonormalization(cbind(alphas, matrix(rnorm(d *
    (d - 1)), , ncol = d - 1))))
  P %*% B
}

### Funzione per il calcolo della funzione HS
### lungo l'asse delle componenti simmetriche
### della distribuzione ST in forma canonica
library(far)

HS0 <- function(x,alpha,df=1) {
  int.u <- function(u) {
    up <- u*x
    alpha.tilde <- u*alpha/sqrt(1+alpha^2*(1-u^2))
    p <- pst(up,alpha=alpha.tilde,df=df)
    min(p,1-p)
  }
  optimize(function(u) int.u(u),lower=-1,upper=1)
}

### Funzione per il calcolo della mediana di Tukey

TukMed.mst <- function(f) {
  dp <- f@dp
  d <- length(dp$xi)
  A <- matrixA(f)
  w <- solve(sqrt(diag(diag(dp$Omega))))
  alphaC <- (t(dp$alpha)%*%w)%*%dp$Omega%*%w)%*%dp$alpha)^0.5
  TukMedC <- c(optimize(function(x)
    HS0(x,alphaC,df=dp$df)$objective,maximum=TRUE,
    lower=-2,upper=3)$maximum,rep(0,d-1))
}

```

```

      as.vector(as.vector(solve(A)%*%TukMedC)+dp$xi)
    }
  TukMed.mst(f)
  w <- solve(sqrt(diag(diag(f@dp$Omega))))
  alphaC <- as.vector(t(f@dp$alpha)%*%
    w%*%f@dp$Omega%*%w%*%f@dp$alpha)^0.5
  medC <- qst(0.5,alpha=alphaC,df=f@dp$df)
  HSO(medC,alphaC,f@dp$df)

#####
##### INDICE ETA* (vedi la 2.9)#####
#####

etas.comp <- function(dp) {
  d <- length(dp$xi)
  O <- dp$Omega
  df <- dp$df
  w <- diag(sqrt(diag(O))); u <- diag(w)
  w.i <- solve(w)
  O.b <- w.i%*%O%*%w.i
  a <- dp$alpha
  a.s <- as.numeric(sqrt(t(a)%*%O.b%*%a))
  theta <- function(x) {
    qst(0.5,alpha=x,df=df)
  }
  Dtheta.alpha <- function(x) {
    grad(function(v) qst(0.5,alpha=v,df=df),x)
  }
  Dtheta.nu <- function(x) {
    grad(function(nu) qst(0.5,alpha=a.s,df=nu),x)
  }
  t1 <- theta(a.s); t2 <- Dtheta.alpha(a.s)
  comp.a <- t1*(O%*%w.i*(a.s^2)-O%*%w.i%*%a%*%
    t(a)%*%O.b)/(a.s^3)+
    t2*O%*%w.i%*%a%*%t(a)%*%O.b/(a.s^2)
  const <- -0.5*a.s^(-3)*t1+0.5*a.s^(-2)*t2

  comp.w <- matrix(NA,ncol=d,nrow=d)

```

```

for(i in 1:d) {
  for(h in 1:d) {
    if(h==i) {
      comp.w[i,h] <- (a[h]/(2*a.s*u[h]))*t1-
      sum(O[h,]*a/u)*(a[h]/u[h]^3)*
      sum(O[h,-h]*a[-h]/u[-h])*const
    } else {
      comp.w[i,h] <- -O[i,h]*(a[h]/(2*a.s*u[h]^3))*t1-
      sum(O[i,]*a/u)*(a[h]/u[h]^3)*
      sum(O[h,-h]*a[-h]/u[-h])*const
    }
  }
}

comp.wd <- matrix(NA,nrow=d,ncol=d*(d-1)/2)
for(i in 1:d) {
  j <- 1
  for(h in 1:(d-1)) {
    if(h==i) {
      for(k in (h+1):d) {
        comp.wd[i,j] <- a[k]*t1/(a.s*u[k])+
        sum(O[i,]*a/u)*(2*a[h]*a[k]/(u[h]*u[k]))*
        const
        j <- j + 1
      }
    } else {
      for(k in (h+1):d) {
        if (k==i) {
          comp.wd[i,j] <- a[h]*t1/(a.s*u[h])+
          sum(O[i,]*a/u)*(2*a[h]*a[k]/(u[h]*u[k]))*
          const
          j <- j + 1
        } else {
          comp.wd[i,j] <- sum(O[i,]*a/u)*(2*a[h]*
          a[k]/(u[h]*u[k]))*
          const
          j <- j + 1
        }
      }
    }
  }
}

```

```

    }
  }
}

t3 <- Dtheta.nu(df)

comp.nu <- rep(NA,d)

for(i in 1:d) {
  comp.nu[i] <- sum(O[i,]*a/u)*t3/a.s
}
cbind(diag(1,d),comp.w,comp.wd,comp.a,comp.nu)
}

etas.mst <- function(f,info) {
  dp <- f@dp
  O <- dp$Omega
  df <- dp$df
  w <- diag(sqrt(diag(O))); u <- diag(w)
  w.i <- solve(w)
  O.b <- w.i%*%O%*%w.i
  a <- dp$alpha
  a.s <- as.numeric(sqrt(t(a)%*%O.b%*%a))
  eta.0 <- qst(0.5,alpha=a.s,df=df)
  ans <- list()
  ans$etas <- as.numeric((O%*%w.i%*%a)*
    eta.0/a.s+dp$xi)
  J <- etas.comp(dp)
  index <- (1:d)*d+(1:d)-(1:d)*(1:d-1)/2
  info.inv <- solve(info)
  info.inv <- cbind(info.inv[,1:d],info.inv[,index],
    info.inv[,-c(1:d,index)])
  ans$etas.var <- J%*%info.inv%*%t(J)
  ans
}

#####

```

```
##### INDICE TAU DI KENDALL #####
#####
```

```
comp.omega.tau <- function(f,...) {
  logit <- function(p) log(abs(p/(1-p)))
  dp <- f@dp
  nu <- dp$df; a1 <- dp$alpha[1]; a2 <- dp$alpha[2];
  w <- dp$Omega[1,2]
  Dlxy.Dw <- function(x,y) {
    w/(1-w^2)-(nu+2)*Q.star(x,y)/(2*(Q(x,y)+nu))+
    dt(h(x,y),df=nu+2)*Dh.Dw(x,y)/pt(h(x,y),df=nu+2)
  }
  Dlxy.dw <- Vectorize(Dlxy.Dw,c("x","y"))
  DFxy.dw <- function(x,y) {
    u <- exp(x)/(1+exp(x)); v <- exp(y)/(1+exp(y))
    cuhre(2,1,function(vec) Dlxy.Dw(logit(vec[1]),
      logit(vec[2]))*
      dmst(c(logit(vec[1]),
        logit(vec[2])),dp=dp)*
      abs((1-vec[1])*vec[1])^(-1)*
      abs((1-vec[2])*vec[2])^(-1),
      lower=c(0,0),upper=c(u,v),abs.tol=1)$value
  }
  Fxy <- function(x,y) {
    u <- exp(x)/(1+exp(x))
    v <- exp(y)/(1+exp(y))
    cuhre(2,1,function(vec) dmst(c(logit(vec[1]),
      logit(vec[2])),dp=dp)*
      abs((1-vec[1])*vec[1])^(-1)*
      abs((1-vec[2])*vec[2])^(-1),
      lower=c(0,0),upper=c(u,v),abs.tol=1)$value
  }
  Q <- function(x,y) {
    (x*(x-w*y)+y*(y-w*x))/(1-w^2)
  }
  Q <- Vectorize(Q,c("x","y"))
}
```

```

Q.star <- function(x,y) {
  2*(w*x^2-x*y*(1+w^2)+w*y^2)/((-w^2)^2)
}
Q.star <- Vectorize(Q.star,c("x","y"))
h <- function(x,y) {
  (a1*x+a2*y)*sqrt((nu+2)/(Q(x,y)+nu))
}
h <- Vectorize(h,c("x","y"))
Dh.Dw <- function(x,y) {
  -((a1*x+a2*y)*sqrt(nu+2)/
    (2*(Q(x,y)+nu)^(1.5)))*Q.star(x,y)
}
Dh.Dw <- Vectorize(Dh.Dw,c("x","y"))
integrand <- function(u,v) {
  x <- logit(u); y <- logit(v)
  dx <- abs((1-u)*u)^(-1); dy <- abs((1-v)*v)^(-1)
  (DFxy.dw(x,y)+Fxy(x,y)*Dlxy.Dw(x,y))*
  dmst(c(x,y),dp=dp)[1]*dx*dy
}
4*cuhre(2,1,function(vec)
  integrand(vec[1],vec[2]),lower=c(0,0),upper=c(1,1),
  ...)$value
}

comp.alpha1.tau <- function(f,...) {
  logit <- function(p) log(abs(p/(1-p)))
  dp <- f@dp
  nu <- dp$df; a1 <- dp$alpha[1]; a2 <- dp$alpha[2];
  w <- dp$Omega[1,2]
  Dlxy.Da1 <- function(x,y) {
    (dt(h(x,y),df=nu+2)/pt(h(x,y),df=nu+2))*Dh.Da1(x,y)
  }

  Dlxy.da1 <- Vectorize(Dlxy.Da1,c("x","y"))
  DFxy.da1 <- function(x,y) {
    u <- exp(x)/(1+exp(x))
    v <- exp(y)/(1+exp(y))
    cuhre(2,1,function(vec) Dlxy.Da1(logit(vec[1])),

```



```

        logit(vec[2]))*
dmst(c(logit(vec[1]),
        logit(vec[2])), dp=dp)*
abs((1-vec[1])*vec[1])^(-1)*
abs((1-vec[2])*vec[2])^(-1),
lower=c(0,0), upper=c(u,v), abs.tol=0.6)$value
}

Fxy <- function(x,y) {
  u <- exp(x)/(1+exp(x))
  v <- exp(y)/(1+exp(y))
  cuhre(2,1,function(vec) dmst(c(logit(vec[1]),
        logit(vec[2])), dp=dp)*
        abs((1-vec[1])*vec[1])^(-1)*
        abs((1-vec[2])*vec[2])^(-1),
        lower=c(0,0), upper=c(u,v), abs.tol=0.1)$value
}

Q <- function(x,y) {
  (x*(x-w*y)+y*(y-w*x))/(1-w^2)
}
Q <- Vectorize(Q,c("x","y"))
Q.star <- function(x,y) {
  2*(w*x^2-x*y*(1+w^2)+w*y^2)/((-w^2)^2)
}
Q.star <- Vectorize(Q.star,c("x","y"))
h <- function(x,y) {
  (a1*x+a2*y)*sqrt((nu+2)/(Q(x,y)+nu))
}
h <- Vectorize(h,c("x","y"))
Dh.Da1 <- function(x,y) {
  x*sqrt((nu+2)/(Q(x,y)+nu))
}
Dh.Da1 <- Vectorize(Dh.Da1,c("x","y"))
integrand <- function(u,v) {
  x <- logit(u); y <- logit(v)
  dx <- abs((1-u)*u)^(-1); dy <- abs((1-v)*v)^(-1)
  (DFxy.da1(x,y)+Fxy(x,y)*Dlxy.Da1(x,y))*

```

```

      dmst(c(x,y),dp=dp)[1]*dx*dy
    }
    4*cuhre(2,1,function(vec)
      integrand(vec[1],vec[2]),lower=c(0,0),upper=c(1,1),
        ...)$value
  }

comp.alpha2.tau <- function(f,...) {
  logit <- function(p) log(abs(p/(1-p)))
  dp <- f@dp
  nu <- dp$df; a1 <- dp$alpha[1]; a2 <- dp$alpha[2];
  w <- dp$Omega[1,2]
  Dlxy.Da2 <- function(x,y) {
    (dt(h(x,y),df=nu+2)/pt(h(x,y),df=nu+2))*Dh.Da2(x,y)
  }

  Dlxy.da2 <- Vectorize(Dlxy.Da2,c("x","y"))
  DFxy.da2 <- function(x,y) {
    u <- exp(x)/(1+exp(x)); v <- exp(y)/(1+exp(y))
    cuhre(2,1,function(vec) Dlxy.Da2(logit(vec[1]),
      logit(vec[2]))*
      dmst(c(logit(vec[1]),
      logit(vec[2])),dp=dp)*
      abs((1-vec[1])*vec[1])^(-1)*
      abs((1-vec[2])*vec[2])^(-1),
      lower=c(0,0),upper=c(u,v),abs.tol=0.6)$value
  }

  Fxy <- function(x,y) {
    u <- exp(x)/(1+exp(x)); v <- exp(y)/(1+exp(y))
    cuhre(2,1,function(vec) dmst(c(logit(vec[1]),
      logit(vec[2])),dp=dp)*
      abs((1-vec[1])*vec[1])^(-1)*
      abs((1-vec[2])*vec[2])^(-1),
      lower=c(0,0),upper=c(u,v),abs.tol=0.1)$value
  }

  Q <- function(x,y) {

```

```

      (x*(x-w*y)+y*(y-w*x))/(1-w^2)
    }
    Q <- Vectorize(Q,c("x","y"))
    Q.star <- function(x,y) {
      2*(w*x^2-x*y*(1+w^2)+w*y^2)/((-w^2)^2)
    }
    Q.star <- Vectorize(Q.star,c("x","y"))
    h <- function(x,y) {
      (a1*x+a2*y)*sqrt((nu+2)/(Q(x,y)+nu))
    }
    h <- Vectorize(h,c("x","y"))
    Dh.Da2 <- function(x,y) {
      y*sqrt((nu+2)/(Q(x,y)+nu))
    }
    Dh.Da2 <- Vectorize(Dh.Da2,c("x","y"))
    integrand <- function(u,v) {
      x <- logit(u); y <- logit(v)
      dx <- abs((1-u)*u)^(-1); dy <- abs((1-v)*v)^(-1)
      (DFxy.da2(x,y)+Fxy(x,y)*Dlxy.Da2(x,y))*
      dmst(c(x,y),dp=dp)[1]*dx*dy
    }
    4*cuhre(2,1,function(vec)
    integrand(vec[1],vec[2]),lower=c(0,0),upper=c(1,1),
    ...)$value
  }

comp.nu.tau <- function(f,...) {
  logit <- function(p) log(abs(p/(1-p)))
  dp <- f@dp
  nu <- dp$df; a1 <- dp$alpha[1]; a2 <- dp$alpha[2];
  w <- dp$Omega[1,2]
  Dlxy.Dnu <- function(x,y) {
    0.5*(digamma((nu+2)/2)-digamma(nu/2)-
    2/nu-log(1+Q(x,y)/nu)+
    ((nu+2)*Q(x,y))/(nu*(Q(x,y)+nu))+
    2*grad(function(gdl) pt(h(x,y,gdl),df=gdl+2),nu)/
    pt(h(x,y,nu),df=nu+2))
  }

```

```

}

Dlxy.dnu <- Vectorize(Dlxy.Dnu,c("x","y"))
DFxy.dnu <- function(x,y) {
  u <- exp(x)/(1+exp(x)); v <- exp(y)/(1+exp(y))
  cuhre(2,1,function(vec) Dlxy.Dnu(logit(vec[1]),
    logit(vec[2]))*
    dmst(c(logit(vec[1]),
    logit(vec[2])),dp=dp)*
    abs((1-vec[1])*vec[1])^(-1)*
    abs((1-vec[2])*vec[2])^(-1),
    lower=c(0,0),upper=c(u,v),abs.tol=1)$value
}

Fxy <- function(x,y) {
  u <- exp(x)/(1+exp(x))
  v <- exp(y)/(1+exp(y))
  cuhre(2,1,function(vec) dmst(c(logit(vec[1]),
    logit(vec[2])),dp=dp)*
    abs((1-vec[1])*vec[1])^(-1)*
    abs((1-vec[2])*vec[2])^(-1),
    lower=c(0,0),upper=c(u,v),abs.tol=0.7)$value
}

Q <- function(x,y) {
  (x*(x-w*y)+y*(y-w*x))/(1-w^2)
}
Q <- Vectorize(Q,c("x","y"))
Q.star <- function(x,y) {
  2*(w*x^2-x*y*(1+w^2)+w*y^2)/((-w^2)^2)
}
Q.star <- Vectorize(Q.star,c("x","y"))
h <- function(x,y,gdl) {
  (a1*x+a2*y)*sqrt((gdl+2)/(Q(x,y)+gdl))
}
h <- Vectorize(h,c("x","y"))
integrand <- function(u,v) {
  x <- logit(u); y <- logit(v)

```

```

    dx <- abs((1-u)*u)^(-1); dy <- abs((1-v)*v)^(-1)
    (DFxy.dnu(x,y)+Fxy(x,y)*Dlxy.Dnu(x,y))*
    dmst(c(x,y),dp=dp)[1]*dx*dy
  }
  4*cuhre(2,1,function(vec)
  integrand(vec[1],vec[2]),lower=c(0,0),upper=c(1,1),
  ...) $value
}

se.tau <- function(f.hat,info,...) {
  omega.bar <- f.hat@dp$Omega[1,2]/
    sqrt(f.hat@dp$Omega[1,1]*f.hat@dp$Omega[2,2])
  dp.norm <- list(xi=c(0,0),
    Omega=matrix(c(1,omega.bar,
      omega.bar,1),2,2,byrow=TRUE),
    alpha=f.hat@dp$alpha,
    df=f.hat@dp$df)
  w1 <- sqrt(dp$Omega[1,1]); w2 <- sqrt(dp$Omega[2,2])
  w12 <- dp$Omega[2,1];
  f.norm <- SEDistr(dp=dp.norm,family="ST")
  grad.tau <- c(comp.omega.tau(f.norm,...),
    comp.alpha1.tau(f.norm,...),
    comp.alpha2.tau(f.norm,...),
    comp.nu.tau(f.norm,...))
  der.w <- c(1/(w1*w2),-w12/(w1^2*w2),-w1/(w1*w2^2))
  Jw <- diag(1,8); Jw[5,3:5] <- der.w
  new.info <- Jw%*%info%*%t(Jw)
  t(grad.tau)%*%(solve(new.info))[5:8,5:8])%*%grad.tau
}

tau.mst <- function(f,info,...) {
  dp <- f@dp
  dp$df <- 50
  integrand <- function(u,v) {
    x <- log(u/(1-u)); y <- log(v/(1-v))
    dx <- (u*(1-u))^(-1); dy <- (v*(1-v))^(-1)
    pmsn(c(x,y),dp=f@dp)*dmsn(c(x,y),dp=dp)*dx*dy
  }
}

```

```

val <- cuhre(2,1,
  function(vec) integrand(vec[1],vec[2]),
  lower=c(0,0))$value
tau <- 4*val-1
tau.var <- se.tau(f,info,...)
list(tau=tau, tau.var=tau.var)
}

#####
##### INDICE RHO DI SPEARMAN #####
#####

comp.omega.rho <- function(f) {
  dp <- f@dp
  nu <- dp$df; a1 <- dp$alpha[1]; a2 <- dp$alpha[2];
  w <- dp$Omega[1,2]
  dpX <- marginalDistribution(f,comp=1)@dp
  dpY <- marginalDistribution(f,comp=2)@dp
  DFx.Dax <- function(x) {
    integrate(function(t) t*sqrt((nu+1)/
      (t^2+nu))*dst(t,dp=dpX),
      lower=-Inf,upper=x)$value
  }
  DFy.Day <- function(y) {
    integrate(function(t) t*sqrt((nu+1)/
      (t^2+nu))*dst(t,dp=dpY),
      lower=-Inf,upper=y)$value
  }
  Dax.Dw <- (a2*(2*a1*a2*w*(1-w^2)+a2^2*
    (1-w^4)+1))/((1+(1-w^2)*a2^2)^1.5)
  Day.Dw <- (a1*(2*a1*a2*w*(1-w^2)+a1^2*
    (1-w^4)+1))/((1+(1-w^2)*a1^2)^1.5)
  Dlxy.Dw <- function(x,y) {
    w/(1-w^2)-(nu+2)*Q.star(x,y)/(2*(Q(x,y)+nu))+
    dt(h(x,y),df=nu+2)*Dh.Dw(x,y)/pt(h(x,y),df=nu+2)
  }
}

```

```

Q <- function(x,y) {
  (x*(x-w*y)+y*(y-w*x))/(1-w^2)
}
Q.star <- function(x,y) {
  2*(w*x^2-x*y*(1+w^2)+w*y^2)/((-w^2)^2)
}
h <- function(x,y) {
  (a1*x+a2*y)*sqrt((nu+2)/(Q(x,y)+nu))
}
Dh.Dw <- function(x,y) {
  -((a1*x+a2*y)*sqrt(nu+2)/
    (2*(Q(x,y)+nu)^(1.5)))*Q.star(x,y)
}
integrand <- function(u,v) {
  x <- log(abs((1-u)/u)); y <- log(abs((1-v)/v))
  dx <- abs((1-u)*u)^(-1); dy <- abs((1-v)*v)^(-1)
  (DFx.Dax(x)*Dax.Dw*pst(y,dp=dpY)+
   DFy.Day(y)*Day.Dw*pst(x,dp=dpX)+
   pst(x,dp=dpX)*pst(y,dp=dpY)*
   Dlx.Dw(x,y))*dmst(c(x,y),dp=dp)*dx*dy
}
12*cuhre(2,1,function(vec)
  integrand(vec[1],vec[2]),lower=c(0,0),
  upper=c(1,1),abs.tol=0.5)$value
}

comp.alpha1.rho <- function(f) {
  dp <- f@dp
  nu <- dp$df; a1 <- dp$alpha[1]; a2 <- dp$alpha[2];
  w <- dp$Omega[1,2]
  dpX <- marginalDistribution(f,comp=1)@dp
  dpY <- marginalDistribution(f,comp=2)@dp
  DFx.Dax <- function(x) {
    integrate(function(t) t*sqrt((nu+1)/
      (t^2+nu))*dst(t,dp=dpX),
      lower=-Inf,upper=x)$value
  }
}

```

```

DFy.Day <- function(y) {
  integrate(function(t) t*sqrt((nu+1)/
    (t^2+nu))*dst(t,dp=dpY),
    lower=-Inf,upper=y)$value
}
Dax.Da1 <- (1+a2^2*(1-w^2))^-0.5
Day.Da1 <- (w-a1*a2*(1-w^2))*((1+a2^2*(1-w^2))^-1.5)
Q <- function(x,y) {
  (x*(x-w*y)+y*(y-w*x))/(1-w^2)
}
h <- function(x,y) {
  (a1*x+a2*y)*sqrt((nu+2)/(Q(x,y)+nu))
}
Dh.Da1 <- function(x,y) {
  x*sqrt((nu+2)/(Q(x,y)+nu))
}
Dlxy.Da1 <- function(x,y) {
  (dt(h(x,y),df=nu+2)/pt(h(x,y),df=nu+2))*Dh.Da1(x,y)
}
integrand <- function(u,v) {
  x <- log(abs((1-u)/u)); y <- log(abs((1-v)/v))
  dx <- abs((1-u)*u)^(-1); dy <- abs((1-v)*v)^(-1)
  (DFx.Dax(x)*Dax.Da1*
  pst(y,dp=dpY)+DFy.Day(y)*Day.Da1*pst(x,dp=dpX)+
  pst(x,dp=dpX)*pst(y,dp=dpY)*
  Dlxy.Da1(x,y))*dmst(c(x,y),dp=dp)*dx*dy
}
12*cuhre(2,1,function(vec)
  integrand(vec[1],vec[2]),lower=c(0,0),
  upper=c(1,1),abs.tol=0.5)$value
}

comp.alpha2.rho <- function(f) {
  dp <- f@dp
  nu <- dp$df; a1 <- dp$alpha[1]; a2 <- dp$alpha[2];
  w <- dp$Omega[1,2]
  dpX <- marginalDistribution(f,comp=1)@dp
  dpY <- marginalDistribution(f,comp=2)@dp
}

```



```

DFx.Dax <- function(x) {
  integrate(function(t)
    t*sqrt((nu+1)/(t^2+nu))*dst(t,dp=dpX),
    lower=-Inf,upper=x)$value
}
DFy.Day <- function(y) {
  integrate(function(t)
    t*sqrt((nu+1)/(t^2+nu))*dst(t,dp=dpY),
    lower=-Inf,upper=y)$value
}
Day.Da2 <- (1+a1^2*(1-w^2))^( -0.5)
Dax.Da2 <- (w-a1*a2*(1-w^2))*((1+a1^2*(1-w^2))^( -1.5))
Q <- function(x,y) {
  (x*(x-w*y)+y*(y-w*x))/(1-w^2)
}
h <- function(x,y) {
  (a1*x+a2*y)*sqrt((nu+2)/(Q(x,y)+nu))
}
Dh.Da2 <- function(x,y) {
  y*sqrt((nu+2)/(Q(x,y)+nu))
}
Dlxy.Da2 <- function(x,y) {
  (dt(h(x,y),df=nu+2)/pt(h(x,y),df=nu+2))*Dh.Da2(x,y)
}
integrand <- function(u,v) {
  x <- log(abs((1-u)/u)); y <- log(abs((1-v)/v))
  dx <- abs((1-u)*u)^(-1); dy <- abs((1-v)*v)^(-1)
  (DFx.Dax(x)*Dax.Da2*pst(y,dp=dpY)+
  DFy.Day(y)*Day.Da2*pst(x,dp=dpX)+
  pst(x,dp=dpX)*pst(y,dp=dpY)*
  Dlxy.Da2(x,y))*dmst(c(x,y),dp=dp)*dx*dy
}
12*cuhre(2,1,function(vec)
  integrand(vec[1],vec[2]),lower=c(0,0),
  upper=c(1,1),abs.tol=0.5)$value
}

```

```

comp.nu.rho <- function(f) {
  dp <- f@dp
  nu <- dp$df; a1 <- dp$alpha[1]; a2 <- dp$alpha[2];
  w <- dp$Omega[1,2]
  dpX <- marginalDistribution(f,comp=1)@dp
  dpY <- marginalDistribution(f,comp=2)@dp
  aX <- dpX[3]
  aY <- dpY[3]
  pt.int1 <- function(s,gdl) {
    pt(aX*s*sqrt((gdl+1)/(s^2+gdl)),df=gdl)
  }
  int1 <- function(s) {
    0.5*(digamma((nu+1)/2)-digamma(nu/2)-
      1/nu-log(1+s^2/nu)+
      (s^2*(nu+1))/(2*nu*(s^2+nu))+
      2*grad(function(gdl) pt.int1(s,gdl),nu)/
      pt.int1(s,nu))*dst(s,dp=dpX)
  }
  DFx.Dnu <- function(x) {
    integrate(function(s) sapply(s, function(x)
      int1(x)),lower=-Inf,upper=x)$value
  }
  pt.int2 <- function(s,gdl) {
    pt(aY*s*sqrt((gdl+1)/(s^2+gdl)),df=gdl)
  }
  int2 <- function(s) {
    0.5*(digamma((nu+1)/2)-digamma(nu/2)-
      1/nu-log(1+s^2/nu)+
      (s^2*(nu+1))/(2*nu*(s^2+nu))+
      2*grad(function(gdl) pt.int2(s,gdl),nu)/
      pt.int2(s,nu))*dst(s,dp=dpY)
  }
  DFy.Dnu <- function(y) {
    integrate(function(s) sapply(s, function(x)
      int2(x)),lower=-Inf,upper=y)$value
  }
  Q <- function(x,y) {
    (x*(x-w*y)+y*(y-w*x))/(1-w^2)
  }
}

```

```

}
h <- function(x,y,gdl) {
  (a1*x+a2*y)*sqrt((gdl+2)/(Q(x,y)+gdl))
}
Dlxy.Dnu <- function(x,y) {
  0.5*(digamma((nu+2)/2)-digamma(nu/2)-
    2/nu-log(1+Q(x,y)/nu)+
    ((nu+2)*Q(x,y))/(nu*(Q(x,y)+nu))+
    2*grad(function(gdl) pt(h(x,y,gdl),df=gdl+2),nu)/
    pt(h(x,y,nu),df=nu+2))
}
integrand <- function(u,v) {
  x <- log(abs((1-u)/u)); y <- log(abs((1-v)/v))
  dx <- abs((1-u)*u)^(-1); dy <- abs((1-v)*v)^(-1)
  (DFx.Dnu(x)*pst(y,dp=dpY)+DFy.Dnu(y)*pst(x,dp=dpX)+
  pst(x,dp=dpX)*pst(y,dp=dpY)*Dlxy.Dnu(x,y))*
  dmst(c(x,y),dp=dp)*dx*dy
}
12*cuhre(2,1,function(vec)
integrand(vec[1],vec[2]),lower=c(0,0),
  upper=c(1,1),abs.tol=0.5)$value
}

se.rho <- function(f.hat,info) {
  omega.bar <- f.hat@dp$Omega[1,2]/
    sqrt(f.hat@dp$Omega[1,1]*f.hat@dp$Omega[2,2])
  dp.norm <- list(xi=c(0,0),
    Omega=matrix(c(1,omega.bar,
      omega.bar,1),2,2,byrow=TRUE),
    alpha=f.hat@dp$alpha,
    df=f.hat@dp$df)
  w1 <- sqrt(dp$Omega[1,1]); w2 <- sqrt(dp$Omega[2,2])
  w12 <- dp$Omega[2,1];
  f.norm <- SEDistr(dp=dp.norm,family="ST")
  grad.rho <- c(comp.omega.rho(f.norm),comp.alpha1.rho(f.norm),
    comp.alpha2.rho(f.norm),comp.nu.rho(f.norm))
  der.w <- c(1/(w1*w2),-w12/(w1^2*w2),-w1/(w1*w2^2))
  Jw <- diag(1,8); Jw[5,3:5] <- der.w
}

```

```

new.info <- Jw%%info%%t(Jw)
t(grad.rho)%%((solve(new.info))[5:8,5:8])%%grad.rho
}

```

```

#####
##### SCARTO INTERQUARTILE #####
#####

```

```

IQ2.st.cov <- function(f,data,info) {
  dpX <- marginalDistribution(f,comp=1)@dp
  dpY <- marginalDistribution(f,comp=2)@dp
  IQX <- (qst(0.75,dp=dpX)-qst(0.25,dp=dpX))^2
  IQY <- (qst(0.75,dp=dpY)-qst(0.25,dp=dpY))^2
  der.IQ <- rbind(c(-2*IQX,2*IQX,0,0),
                 c(0,0,-2*IQY,2*IQY))
  quant.var <- mat.med12(c(0.25,0.75),c(0.25,0.75),
                        f,data,info)
  der.IQ%%quant.var%%t(der.IQ)
}

```

```

IQ2.st <- function(f) {
  dpX <- marginalDistribution(f,comp=1)@dp
  dpY <- marginalDistribution(f,comp=2)@dp
  IQX <- (qst(0.75,dp=dpX)-qst(0.25,dp=dpX))^2
  IQY <- (qst(0.75,dp=dpY)-qst(0.25,dp=dpY))^2
  IQ <- c(IQX,IQY)
  IQ2.var <- IQ2.st.cov(f34,data[,3:4],info.Mult34)
  list(IQ=IQ2,IQ.var=IQ2.var)
}

```

```

#####
##### INDICE DI YULE-BOWLEY #####
#####

```

```

var.yb.st <- function(f,info) {
  dp1 <- marginalDistribution(f,comp=1)@dp

```

```

dp2 <- marginalDistribution(f,comp=2)@dp
r1 <- c(0.25,0.5,0.75)
r2 <- c(0.25,0.5,0.75)
Sqp <- mat.med12(r1,r2,f,data,info)
q1_0.25 <- qst(0.25,dp=dp1); q1_0.5 <- qst(0.5,dp=dp1)
q1_0.75 <- qst(0.75,dp=dp1); IQ1 <- q1_0.75-q1_0.25
g1 <- 2*c(-1/IQ1,(q1_0.5-q1_0.75)/IQ1^2,(q1_0.25-q1_0.5)/
          IQ1^2,0,0,0)
q2_0.25 <- qst(0.25,dp=dp2); q2_0.5 <- qst(0.5,dp=dp2)
q2_0.75 <- qst(0.75,dp=dp2); IQ2 <- q2_0.75-q2_0.25
g2 <- 2*c(0,0,0,-1/IQ2,(q2_0.5-q2_0.75)/IQ2^2,(q2_0.25-q2_0.5)/
          IQ2^2)
J <- rbind(g1,g2)
J%*%Sqp%*%t(J)
}

YB.st <- function(f) {
  dp <- f@dp
  q0.75 <- qst(0.75,dp=dp)
  q0.25 <- qst(0.25,dp=dp)
  q0.5 <- qst(0.5,dp=dp)
  yb <- (q0.75+q0.25-q0.5)/(q0.75-q0.25)
  yb.var <- var.yb.st(f,info)
  list(yb=yb,yb.var=yb.var)
}

#####
##### INDICE DI CURTOSI WANG-SERFLING (2005) #####
#####(calcolo per la t di Student)#####
#####

p.to.r <- function(p,d,nu) {
  r <- sqrt(d*qf(p,d,nu))
  volume <- (pi^(d/2)/gamma(d/2+1))*r^d
  HS <- HS0(r,0,df=nu)$objective
  ans <- list(volume=volume,HS=HS)
  ans
}

```

```
}

MKurtosis <- function(p,d,nu) {
  V1 <- p.to.r((1-p)/2,d,nu)$volume
  V2 <- p.to.r((1+p)/2,d,nu)$volume
  V3 <- p.to.r(1/2,d,nu)$volume
  (V1+V2-2*V3)/(V2-V1)
}

phi.Td <- function(p,nu,info) {
  phi <- CMKurtosis(p,d,nu)
  phi.var <- (grad(function(nu) MKurtosis(p,2,nu),nu)^2)*
             (solve(info)[8,8])
  list(phi=phi, phi.var=phi.var)
}
```

Bibliografia

- AVÉROUS, J. & MESTE, M. (1997). Skewness for multivariate distributions: two approaches. *The Annals of Statistics* **25**, 1984–1997.
- AZZALINI, A. (2001). *Inferenza statistica, una presentazione basata sul concetto di verosimiglianza*. Springer-Verlag Italia, Milano, 2nd ed.
- AZZALINI, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics* **32**, 159–188.
- AZZALINI, A. & CAPITANIO, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of Royal Statistics. Series B* **61**, 579–602. La versione completa dell'articolo è disponibile a [arXiv.org:0911.2093v1](https://arxiv.org/abs/0911.2093v1).
- AZZALINI, A. & CAPITANIO, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of Royal Statistical Society. Series B* **65**, 367–389. La versione completa dell'articolo è disponibile a [arXiv.org:0911.2342v1](https://arxiv.org/abs/0911.2342v1).
- AZZALINI, A. & GENTON, M. G. (2008). Robust likelihood methods based on the skew- t and related distributions. *International Statistical Review* **76**, 106–129.
- AZZALINI, A. & REGOLI, G. (2012). Some properties of skew-symmetric distributions. *Annals of the Institute of Statistical Mathematics* **64**, 857–879.
- BALANDA, K. P. & MACGILLIVRAY, H. L. (1990). Kurtosis and spread. *The Canadian Journal of Statistics* **18**, 17–30.

- BEHBOODIAN, J., JAMALIZADEH, A. & BALAKRISHNAN, N. (2006). A new class of skew-Cauchy distributions. *Statistics and Probability Letters* **76**, 1488–1493.
- BENJAMINI, Y. & KRIEGER, M. (1996). Concepts and measures for skewness with data-analytic implications. *The Canadian Journal of Statistics* **24**, 131–140.
- BOWLEY, A. L. (1920). *Elements of statistics*. C. Scribner's Sons, New York, 4th ed.
- BROWN, B. M. (1983). Statistical uses of the spatial median. *Journal of the Royal Statistical Society. Series B (Methodological)* **45**, 25–30.
- BRYSON, M. C. (1974). Heavy-tailed distributions: Properties and tests. *Technometrics* **16**, 61–68.
- CAPÉRAÀ, P. & GENEST, C. (1993). Spearman's rho is larger than Kendall's tau for positively dependent random variables. *Journal of Nonparametric Statistics* **2**, 183–194.
- CHAKRABORTY, B. & CHAUDHURI, P. (1996). On a transformation and re-transformation technique for constructing an affine equivariant multivariate median. *Proceedings of the American Mathematical Society* **124**, 2539–2547.
- CHAKRABORTY, B., CHAUDHURI, P. & OJA, H. (1998). Operation transformation retransformation on spatial median and angle test. *Statistica Sinica* **8**, 767–784.
- CHOI, K. & MARDEN, J. (1998). A multivariate version of Kendall's tau. *Nonparametric Statistics* **9**, 261–293.
- CROUX, C., OLLILA, E. & OJA, H. (2002). Sign and rank covariance matrices: statistical properties and application to principal components analysis. In *Statistical data analysis based on the L1-norm and related methods*. D. Yedolah, Birkhäuser, Basel, pp. 257–271.
- DAVID, F. N. & JOHNSON, N. L. (1956). Some tests of significance with ordered variables. *Journal of the Royal Statistical Society. Series B* **18**, 1–31.

- DOKSUM, K. A. (1975). Measures of location and asymmetry. *Scandinavian Journal of Statistics* **2**, 11–22.
- DONOHO, D. L. & GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics* **24**, 1803–1827.
- DUTTA, S., GHOSH, A. K. & CHAUDHURI, P. (2011). Some intriguing properties of Tukey’s half-space depth. *Bernoulli* **17**, 1420–1434.
- FAMA, E. F. (1963). Mandelbrot and the stable Paretian hypothesis. *Journal of Business* **36**, 420–429.
- FERNÁNDEZ, C. & STEEL, M. F. J. (1999). Multivariate Student- t regression models: pitfalls and inference. *Biometrika* **86**, 153–167.
- FREDRICKS, G. A. & NELSEN, R. B. (2007). On the relationship between Spearman’s rho and Kendall’s tau for pairs of continuous random variables. *Journal of Statistical Planning and Inference* **137**, 2143–2150.
- GROENEVELD, R. A. & MEEDEN, G. (1984). Measuring skewness and kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician)* **33**, 391–399.
- GROWER, J. C. (1974). The mediancenter. *Journal of Applied Statistics* **23**, 466–470.
- GUMBEL, E. J. (1962). Statistical theory of extreme values (main results). In *Contributions to order statistics*, chap. 6. A. E. Sarhan and B. G. Greenberg, New York: John Wiley, pp. 406–431.
- HETTMANSPERGER, T. P., MÖTTÖNEN, J. & OJA, H. (1997). Affine invariant multivariate one-sample signed-rank tests. *Journal of the American Statistical Association* **2**, 1591–1600.
- HETTMANSPERGER, T. P., MÖTTÖNEN, J. & OJA, H. (1998). Affine invariant multivariate rank tests for several samples. *Statistica Sinica* **8**, 785–800.
- HETTMANSPERGER, T. P. & RANGLES, R. H. (2002). A practical affine equivariant multivariate median. *Biometrika* **98**, 935–951.

- HOGG, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *Journal of American Statistical Association* **69**, 909–923.
- HOTELLING, H. (1929). Stability in competition. *Economic Journal* **39**, 41–57.
- HULT, H. & LINDSKOG, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability* **34**, 587–608.
- KAFADAR, K. (1988). Slash distribution. In *Encyclopedia of statistical sciences*, vol. 8. S. Kotz, N. Balakrishnan, C. B. Read e B. Vidakovic, Wiley, New York, 2nd ed., pp. 510–511.
- KOTZ, S. & NADARAJAH, S. (2004). *Multivariate t distributions and their applications*. Cambridge University Press, New York.
- LIU, R. Y. (1988). On a notion of simplicial depth. *Proceedings of the National Academy of Sciences of the United States* **85**, 1732–1734.
- LIU, R. Y. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics* **18**, 405–414.
- LIU, R. Y., PRELIUS, J. M. & SINGH, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics* **27**, 783–858.
- MACGILLIVRAY, H. L. (1964). Skewness and asymmetry: measures and orderings. *The Annals of Statistics* **14**, 994–1011.
- MACGILLIVRAY, H. L. (1992). Shape properties of the g-and-h and Johnson families. *Communications in Statistics. Theory and Methods* **21**, 1233–1250.
- MACGILLIVRAY, H. L. & BALANDA, K. P. (1988). The relationship between skewness and kurtosis. *Australian Journal of Statistics* **30**, 319–337.
- MALKOVITCH, J. F. & AFIFI, A. A. (1973). On tests for multivariate normality. *Journal of the American Statistical Association* **68**, 176–179.

- MANDELBROT, B. B. (1963). The variation of certain speculative prices. *Journal of Business* **36**, 392–417.
- MANTEL, N. (1976). Tails of distributions. *The American Statistician* **30**, 14–17.
- MARDIA, K. V. (1970). Measure of multivariate skewness and kurtosis with application. *Biometrika* **57**, 519–530.
- MARDIA, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā: The Indian Journal of Statistics. Series B* **36**, 115–128.
- NELSEN, B. R. (2006). *An introduction to copulas*. Springer, New York, 2nd ed.
- OJA, H. (1981). On location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of Statistics* **8**, 154–168.
- OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statistics and Probability Letters* **58**, 263–277.
- PEARSON, K. R. (1905). Skew variation: a rejoinder. *Biometrika* **4**, 169–212.
- ROGERS, W. H. & TUKEY, J. W. (1972). Understanding some long-tailed symmetrical distributions. *Statistica Neerlandica* **26**, 211–226.
- ROUSSEEUW, P. J. & RUTS, I. (1998). Constructing the bivariate Tukey median. *Statistica Sinica* **8**, 827–839.
- SCARSINI, M. (1984). On measures of concordance. *Stochastica* **8**, 208–218.
- SCHMID, F. & SCHMIDT, R. (2007). Multivariate extensions of Spearman's rho and related statistics. *Statistics and Probability Letters* **77**, 407–416.
- SERFLING, R. (2006). Multivariate symmetry and asymmetry. In *Encyclopedia of statistical sciences*, vol. 8. S. Kotz, N. Balakrishnan, C. B. Read e B. Vidakovic, Wiley, New York, 2nd ed., pp. 5538–5345.
- SMALL, C. G. (1987). Measures of centrality of multivariate and directional distribution. *The Canadian Journal of Statistics* **15**, 31–39.

- SMALL, C. G. (1990). A survey of multidimensional medians. *International Statistical Review* **58**, 327–332.
- SRIVASTAVA, M. S. (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statistics and Probability Letters* **2**, 263–267.
- STRUYF, A. & ROUSSEEUW, P. J. (2000). High-dimensional computation of the deepest location. *Computational Statistics and Data Analysis* **34**, 415–426.
- STUDENT (1908). The probable error of a mean. *Biometrika* **6**, 1–25.
- TADIKAMALLA, P. R. & JOHNSON, N. L. (1982). Systems of frequency curves generated by transformations of logistic variables. *Biometrika* **69**, 461–465.
- TAYLOR, M. D. (2007). Multivariate measures of concordance. *Annals of the Institute of Statistical Mathematics* **59**, 789–806.
- TUKEY, J. W. (1975). Mathematics and picturing data. In *International Congress of Mathematicians*, vol. 2. Vancouver, pp. 523–531.
- TYLER, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics* **15**, 234–251.
- VAN ZWET, W. R. (1986). *Convex transformation of random variables*. Mathematical Centrum, Amsterdam.
- VISURI, S., KOIVUNEN, V. & OJA, H. (2000). Sign and rank covariance matrices. *Journal of Statistical Planning and Inference* **91**, 557–575.
- WANG, J. (2009). A family of kurtosis orderings for multivariate distributions. *Journal of Multivariate Analysis* **100**, 509–517.
- WANG, J. & GENTON, M. G. (2006). The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference* **136**, 209–220.
- WANG, J. & SERFLING, R. (2005). Nonparametric multivariate kurtosis and tailweight measure. *Nonparametric Statistics* **17**, 441–456.

- YULE, G. U. (1911). *An introduction to the theory of statistics*. C. Griffin and Company, London.
- ZELLNER, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* error terms. *Journal of the American Statistical Association* **71**, 400–405.
- ZUO, Y. & SERFLING, R. (1999). On the performance of some robust non-parametric location measures relative to a general notion of multivariate symmetry. *Journal of Statistical Planning and Inference* **84**, 55–79.
- ZUO, Y. & SERFLING, R. (2000a). General notions of statistical depth function. *The Annals of Statistics* **28**, 461–482.
- ZUO, Y. & SERFLING, R. (2000b). Structural properties and convergence results for contours of of sample statistical depth functions. *The Annals of Statistics* **28**, 483–499.