



Università degli Studi di Padova

Facoltà di Scienze Statistiche

Tesi di Laurea Magistrale in Statistica e Informatica

**La distribuzione  $t$  asimmetrica:  
analisi discriminante e regioni di tolleranza**

A.A. 2006/2007

Relatore: Chiar.mo Prof. AZZALINI Adelchi

Candidato: SORIANI Nicola - 529832

21 settembre 2007



*Perchè la vita  
non è sempre  
una retta.*



# Indice

<b>Prefazione</b>	<b>I</b>
<b>1 Le distribuzioni in gioco</b>	<b>1</b>
1.1 La distribuzione normale asimmetrica . . . . .	1
1.2 La distribuzione $t$ asimmetrica . . . . .	3
<b>2 Nell'analisi discriminante</b>	<b>7</b>
2.1 Le leggi discriminanti . . . . .	7
2.1.1 La regola discriminante di verosimiglianza . . . . .	8
2.1.2 La regola lineare di Fisher . . . . .	9
2.2 Confronto fra le due regole . . . . .	10
2.2.1 Nella normale asimmetrica . . . . .	10
2.2.2 Nella $t$ asimmetrica . . . . .	12
2.3 Alcune applicazioni su dati reali . . . . .	18
2.3.1 Dati <i>Hepatic</i> . . . . .	18
2.3.2 Dati <i>Vowel</i> . . . . .	21
2.3.3 Dati <i>Olive</i> . . . . .	26
<b>3 Nelle regioni di tolleranza</b>	<b>37</b>
3.1 Per la normale asimmetrica . . . . .	37
3.2 Per la $t$ asimmetrica . . . . .	41
<b>Conclusioni e sviluppi futuri</b>	<b>51</b>
<b>Bibliografia</b>	<b>53</b>



# Elenco delle figure

2.1	Curve di livello per normale asimmetrica e regole discriminanti: ( - - ) regola di verosimiglianza, (—) regola di Fisher. . . . .	12
2.2	Quattro esempi di $t$ asimmetrica e relative funzioni discriminanti per varie scelte dei gradi di libertà $\nu = 5, 10, 30, 50$ : ( - - - ) regola di verosimiglianza, (—) regola di Fisher. . . . .	14
2.3	[a] Curve di livello per le funzioni discriminanti per una $t$ asimmetrica con 10 gradi di libertà, sotto la condizione (2.7): ( - - - ) regola di verosimiglianza, (—) regola di Fisher. [b] Evoluzione della regola dicriminante di verosimiglianza per la $t$ asimmetrica all'aumentare dei gradi di libertà, sotto la condizione (2.7):(—) $\nu = 10$ ; ( - - - ) $\nu = 100$ , (....) $\nu = 1000$ . . . . .	15
2.4	Diagrammi di dispersione dati <i>Hepatic</i> , suddivisi per gruppi. . . . .	19
2.5	Dati <i>Hepatic</i> : Allocazioni con analisi discriminante classica (Fisher). . . . .	20
2.6	Dati <i>Hepatic</i> . [a] Distribuzioni $SN$ per il modello stimato, suddivisi per gruppi. [b] Allocazioni con modello $SN$ . . . . .	21
2.7	Dati <i>Hepatic</i> . [a] Distribuzioni $St$ per il modello stimato, suddivisi per gruppi. [b] Allocazioni con modello $St$ . . . . .	22
2.8	Allocazioni con la regola discriminante di Fisher per i dati <i>Vowel</i> . [a] Dati <i>Training</i> . [b] Dati <i>Test</i> . . . . .	30
2.9	Allocazioni con la regola discriminante di verosimiglianza per $SN$ dei dati <i>Vowel</i> . [a] Dati <i>Training</i> . [b] Dati <i>Test</i> . . . . .	31
2.10	Allocazioni con la regola discriminante di verosimiglianza per $St$ dei dati <i>Vowel</i> . [a] Dati <i>Training</i> . [b] Dati <i>Test</i> . . . . .	32

2.11	Allocazioni per la regola discriminante lineare di Fisher dati <i>Olive</i> . [a] Allocazioni per Zona. [b] Allocazioni per Macro aree. . . . .	33
2.12	Allocazioni per la regola discriminante di verosimiglianza per <i>SN</i> dati <i>Olive</i> . [a] Allocazioni per Zona. [b] Allocazioni per Macro aree. . . . .	34
2.13	Allocazioni per la regola discriminante di verosimiglianza per <i>St</i> dati <i>Olive</i> . [a] Allocazioni per Zona. [b] Allocazioni per Macro Aree. . . . .	35
3.1	Percentili empirici, triangoli rossi, ed empirici con termine di correzione, triangoli blu, per $SN_2$ di parametri $\alpha_1 = 2$ , $\alpha_2 = 6$ e $\omega = -0.5$ . Linea nera: bisettrice I° quadrante. Linea blu: retta per il termine di correzione $h$ stimato. . . . .	39
3.2	Casi <i>SN</i> . [a] Punti $h$ osservati contro $\alpha^*$ , con funzione interpolante. [b] Punti $1/\ln(\exp(h/2) - 1)$ osservati contro $\alpha^*$ e retta interpolante. . . . .	41
3.3	Percentili empici, triangoli rossi, ed empirici con termine di correzione, triangoli blu, per $St_2$ di parametri $\alpha_1 = 2$ , $\alpha_2 = 6$ , $\omega = -0.5$ , $\nu = 7$ . Linea nera: bisettrice I° quadrante. Linea blu: retta di coefficiente angolare $\beta$ e intercetta $h$ stimati per il termine di correzione $K(\beta, h)$ . . . . .	43
3.4	Casi $St_2$ . [a] Punti $\beta$ osservati contro $\alpha^*$ . [b] Punti $\beta\alpha^*$ contro $\alpha^*$ e rette interpolanti per i dispettivi gradi di libertà $\nu$ . [c] Termini di aggiustamento $\tau(\nu)$ contro $\nu$ e funzione interpolante. [d] Punti $\beta\alpha^*\tau(\nu)^{-1}$ contro $\alpha^*$ e retta interpolante. . . . .	45
3.5	Casi $St_2$ . [a] Punti $h$ osservati contro $\alpha^*$ . [b] Punti $[\ln(\exp(h/100) - 1)\alpha^* + 1.5]$ contro $\alpha^*$ e rette interpolanti per i diversi valori di $\nu$ . [c] Termine di aggiustamento $\lambda(\nu)$ contro $\nu$ e funzione interpolante. [d] Punti $[\ln(\exp(h/100) - 1)\alpha^* + 1.5]\lambda(\nu)^{-1}$ contro $\alpha^*$ e retta interpolante. . . . .	47

# Elenco delle tabelle

2.1	Frequenze relative di errata specificazione per le regole, e della coincidenza delle suddette, per varie scelte della probabilità a priori per popolazione $G1$ e all'aumentare dei gradi di libertà.	17
2.2	Matrici di confusione e tassi di errata allocazione per i dati <i>Hepatic</i> .	23
2.3	Matrici di confusione e tassi di errata allocazione per i dati <i>Training</i> di <i>Vowel</i> .	24
2.4	Matrici di confusione e tassi di errata allocazione per i dati <i>Test</i> di <i>Vowel</i> .	25
2.5	Suddivisione della variabile zona per le varie macro aree.	26
2.6	Matrici di confusione e tassi di errata allocazione per i dati <i>Olive</i> , classificando rispetto la variabile <i>Zona</i> .	28
2.7	Matrici di confusione e percentuali di errata allocazione per i dati <i>Olive</i> , classificando rispetto la variabile <i>Macro Area</i> .	29
3.1	Probabilità assegnate di copertura, empiriche ed empiriche con termine di correzione, ognuna $\times 100$ , per una $SN_2$ di parametri $\alpha_1 = 2$ , $\alpha_2 = 6$ e $\omega = -0.5$ .	39
3.2	Probabilità assegnate di copertura, empiriche e empiriche con termine di correzione, ognuna $\times 100$ , per una $St_2$ di parametri $\alpha_1 = 2$ , $\alpha_2 = 6$ , $\omega = -0.5$ e $\nu = 7$ .	47
3.3	<i>Summary</i> per le probabilità empiriche di copertura $\tilde{p} \times 100$ .	49

3.4	<i>Summary</i> per le probabilità empiriche di copertura corrette $\hat{p}$ $\times 100$ . . . . .	49
-----	---	----

# Prefazione

Le distribuzioni normale e  $t$  di Student rivestono un ruolo centrale in molte tecniche dell'analisi multivariata, spesso ne costituiscono lo scheletro formale fornendo basi e giustificazioni teoriche a supporto dei procedimenti e dei risultati. La rigidità di tali condizioni però rappresenta spesso una strada obbligata che talvolta conduce ad operare forzando realtà a cui i dati non appartengono.

L'introduzione delle varianti asimmetriche per le distribuzioni normale e  $t$  di Student segue la tendenza della letteratura statistica a ricercare metodi più flessibili volti a ridurre la rigidità di tali condizioni e ampliarne il campo di applicabilità alle tecniche.

Lo scopo di questa tesi è indagare sulla possibilità e la convenienza o meno dell'adozione della distribuzione  $t$  asimmetrica per:

- l'analisi discriminante;
- la costruzione di regioni con assegnata probabilità  $p$  e misura geometrica minima.

Per entrambi gli obiettivi si è partiti da risultati noti per la normale asimmetrica cercando le analogie e le possibili differenze con il passaggio alla  $t$  asimmetrica. Questo procedimento ha portato a buoni esiti, permettendo di ricavare le condizioni sotto le quali i risultati per la normale asimmetrica possono essere estesi alla  $t$  asimmetrica.

Per l'analisi discriminante si sono confrontate la regola classica con quella di verosimiglianza scoprendo che il passaggio alla  $t$  asimmetrica limita i casi

in cui le due coincidono. Inoltre attraverso l'applicazione a dati reali si è appurato che tale operazione riduce, o comunque non peggiora, il tasso di errata allocazione sui dati.

Per il secondo punto, l'adozione della  $t$  asimmetrica permette ancora la costruzione di regioni di tolleranza basate sulla tecnica per la  $t$  di Student, ma ad essa dovremo applicare un termine di correzione dipendente dall'asimmetria e dai gradi di libertà della distribuzione considerata.

# Capitolo 1

## Le distribuzioni in gioco

In questo capitolo si darà una breve descrizione delle distribuzioni usate per arrivare ai risultati che verranno presentati nei capitoli successivi.

### 1.1 La distribuzione normale asimmetrica

Una variabile casuale  $d$ -dimensionale  $Z$ , ha distribuzione **Normale Asimmetrica  $d$ -dimensionale**, da Azzalini & Dalla Valle (1996), e la indicheremo con  $Z \sim SN_d(\Omega, \alpha)$ , se è continua con funzione di densità di probabilità del tipo

$$\begin{aligned} f_Z(z; \Omega_z, \alpha) &= 2\phi_d(z; \Omega_z)\Phi(\alpha^T z) \\ &= 2\frac{1}{(2\pi)^{\frac{d}{2}} |\Omega_z|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \|z\|_{\Omega_z}^2\right\} \Phi(\alpha^T z) \quad (z \in \mathbb{R}^d) \end{aligned} \quad (1.1)$$

dove  $\phi_d(z; \Omega_z)$  sta per la densità di una variabile casuale normale  $d$ -dimensionale con vettore delle medie pari a 0 e matrice di correlazione  $\Omega_z$ , calcolata nel punto  $z$ ; mentre  $\Phi(\alpha^T z)$  è la funzione di ripartizione di una normale standard calcolata in  $\alpha^T z$ .

Valore atteso e varianza per una variabile normale asimmetrica sono pari

a:

$$\mu_z = E[Z] = \left(\frac{2}{\pi}\right)^{1/2} \delta, \quad (1.2)$$

$$Var[Z] = \Omega - \mu_z \mu_z^T, \quad (1.3)$$

dove

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

Il parametro  $\alpha$  può essere calcolato come:

$$\alpha = \frac{1}{1 - \delta^T \Omega^{-1} \delta} \Omega^{-1} \delta. \quad (1.4)$$

La distribuzione può essere estesa attraverso l'introduzione dei parametri di posizione ( $\xi$ ) e di scala ( $\omega$ ), infatti se definiamo la quantità  $Y = \xi + \omega Z$  con  $Z \sim SN_d(\Omega_z, \alpha)$ ,  $\xi = (\xi_1, \dots, \xi_d)$  e  $\omega = diag(\omega_1, \dots, \omega_d)$ ,  $\omega_i > 0$  per  $i = 1, \dots, d$ , allora diremo che  $Y$  si distribuirà come una  $SN_d(\xi, \Omega, \alpha)$ , la cui funzione di densità sarà:

$$\begin{aligned} f_Y(y; \xi, \Omega, \alpha) &= 2\phi_d(y - \xi; \Omega) \Phi(\alpha^T \omega^{-1}(y - \xi)) \\ &= 2 \frac{1}{(2\pi)^{\frac{d}{2}} |\Omega_z|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \|y - \xi\|_{\Omega_z}^2\right\} \Phi(\alpha^T \omega^{-1}(y - \xi)) \end{aligned} \quad (1.5)$$

con  $\Omega = \omega \Omega_z \omega$ . I risultati per il valore atteso e la varianza diverranno di conseguenza:

$$E[Y] = \xi + \mu_z, \quad (1.6)$$

$$Var[Y] = \Omega - \omega \mu_z \mu_z^T \omega. \quad (1.7)$$

Una proprietà importante della normale asimmetrica, che risulterà utile nel Capitolo 3, è quella legata alle forme quadratiche studiate da Azzalini & Capitanio (1999), infatti

$$Z^T \Omega^{-1} Z \sim \chi_d^2 \quad (1.8)$$

con  $Z$  distribuita come (1.1) e  $\Omega$  matrice di covarianza. Analogamente il risultato vale per  $(Y - \xi)^T \Omega^{-1} (Y - \xi)$  nel caso in cui la distribuzione sia

del tipo (1.5). La proprietà appena presentata è un caso particolare di un risultato più generale, il quale afferma che se  $B$  è una matrice  $k \times k$  simmetrica di rango  $h$ , semidefinita positiva, e tale che  $B\Omega B = B$ , allora la forma quadratica  $Z^T B Z$ , o ugualmente  $(Y - \xi)^T B (Y - \xi)$ , si distribuisce come un  $\chi_h^2$ .

La funzione di log-verosimiglianza per una singola osservazione  $y$ , realizzazione di una variabile aleatoria  $Y \sim SN_d(\xi, \Omega, \alpha)$  è:

$$\begin{aligned} l(\xi, \Omega, \alpha; y) &= \ln(f_Y(y; \xi, \Omega, \alpha)) & (1.9) \\ &= -\frac{d}{2} \ln(2\pi) - \frac{1}{2} \ln |\Omega| + \\ &\quad -\frac{1}{2} (y - \xi)^T \Omega^{-1} (y - \xi) + \zeta_0(w) \end{aligned}$$

con  $w = \alpha^T \omega^{-1} (y - \xi)$  e  $\zeta_0(x) = \ln(2\Phi(x))$ .

## 1.2 La distribuzione $t$ asimmetrica

La distribuzione  $t$  di Student con  $\nu$  gradi di libertà nasce dal rapporto tra una distribuzione normale standard e una quantità  $V^{\frac{1}{2}}$  con  $V \sim \frac{\chi_\nu^2}{\nu}$ , trasformazione di un Chi-quadrato con gradi di libertà pari a  $\nu$ . Passando al caso multidimensionale, si estende semplicemente la definizione, definendola come il rapporto di una distribuzione Normale  $d$ -dimensionale con la quantità  $V^{\frac{1}{2}}$ . La distribuzione  $t$  asimmetrica da Azzalini & Capitanio (2003) è ottenuta ricalcando le definizioni della  $t$  di Student e della normale asimmetrica, infatti consideriamo la quantità  $S = V^{-\frac{1}{2}} Z$ , dove  $Z$  è una variabile aleatoria con distribuzione del tipo (1.1). Diciamo che la variabile  $S$  si distribuisce come una  **$t$  asimmetrica  $d$ -dimensionale con  $\nu$  gradi di libertà**, e la indicheremo con  $S \sim St_d(\Omega_z, \alpha, \nu)$ , se è continua con funzione di densità

$$\begin{aligned} f_S(s; \Omega, \nu) &= 2t_d(s; \Omega, \nu) T_1(U_s, \nu + d) & (1.10) \\ &= 2 \frac{\Gamma(\frac{\nu+d}{2})}{|\Omega_z|^{\frac{1}{2}} (\nu\pi)^{\frac{d}{2}} \Gamma(\frac{\nu}{2})} \left(1 + \frac{\|s\|_\Omega^2}{\nu}\right)^{-\frac{\nu+d}{2}} T_1(U_s, \nu + d) \end{aligned}$$

dove con  $t_d(s; \Omega, \nu)$  intendiamo la funzione di densità di una  $t$  di Student  $d$ -dimensionale con matrice di correlazione  $\Omega$  e  $\nu$  gradi di libertà, mentre  $T_1(U_s; \nu)$  rappresenta la funzione di distribuzione di una  $t$  di Student unidimensionale con  $\nu$  gradi di libertà, calcolata nel punto  $U_s = \alpha^T s \left( \frac{\nu+d}{\nu + \|s\|_\Omega^2} \right)^{-\frac{1}{2}}$ .

Analogamente a quanto visto nel caso della normale asimmetrica possiamo introdurre parametri di posizione e scala,  $X = \xi + V^{-\frac{1}{2}}Y$ , in cui  $Y$  ha distribuzione del tipo (1.5) con parametro di posizione nullo. In questo caso diremo che  $X \sim St_d(\xi, \Omega, \alpha, \nu)$  e la funzione di densità diverrà:

$$\begin{aligned} f_X(x; \xi, \Omega, \alpha, \nu) &= 2t_d(x - \xi; \Omega, \nu)T_1(U_x, \nu + d) & (1.11) \\ &= 2 \frac{\Gamma(\frac{\nu+d}{2})}{|\Omega|^{\frac{1}{2}} (\nu\pi)^{\frac{d}{2}} \Gamma(\frac{\nu}{2})} \left(1 + \frac{Q_x}{\nu}\right)^{-\frac{\nu+d}{2}} T_1(U_x, \nu + d) \end{aligned}$$

dove  $Q_x = \|x - \xi\|_\Omega^2$  e  $U_x = \alpha^T \omega^{-1}(x - \xi) \left( \frac{\nu+d}{\nu+Q_x} \right)^{\frac{1}{2}}$ .

Il valore atteso e la varianza di una  $St_d(\xi, \Omega, \alpha, \nu)$  sono pari a:

$$E[X] = \xi + \omega \mu_x, \quad (1.12)$$

$$Var[X] = \frac{\nu}{\nu-2} \Omega - \omega \mu_x \mu_x^T \omega \quad (\nu > 2), \quad (1.13)$$

dove

$$\mu_x = \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \left(\frac{\nu}{\pi}\right)^{1/2} \delta \quad (\nu > 1), \quad (1.14)$$

e  $\delta$  è lo stesso del caso della normale asimmetrica.

La proprietà legata al parametro di asimmetria  $\alpha$  descritta nel paragrafo precedente continua a valere per la  $t$  asimmetrica, per cui  $\alpha$  può ancora essere calcolato come nella (1.4).

Anche nel caso della  $t$  asimmetrica esiste un risultato legato alle forme quadratiche del tipo visto nella (1.8), da Azzalini & Capitanio (2003); se consideriamo una variabile  $X \sim St(\xi, \Omega, \alpha, \nu)$  allora per la forma quadratica  $Q = (X - \xi)^T B (X - \xi)$ , con  $B$  matrice  $k \times k$  simmetrica di rango  $h$ , vale la seguente proprietà:

$$Q = \frac{Z^T B Z}{V} \sim \frac{\chi_h^2}{\chi_\nu^2 / \nu},$$

## 1.2. LA DISTRIBUZIONE $T$ ASIMMETRICA

---

da cui è semplice ricavare che  $Q/h \sim F_{h,\nu}$ . Scegliendo opportunamente  $B = \Omega^{-1}$  otteniamo che:

$$\frac{Q}{d} = \frac{Z^T \Omega^{-1} Z / d}{V} \sim \frac{\chi_d^2 / d}{\chi_\nu^2 / \nu} \sim F_{d,\nu}. \quad (1.15)$$

La funzione di log-verosimiglianza per una singola osservazione  $x$ , realizzazione di una variabile aleatoria  $X \sim St_d(\xi, \Omega, \alpha, \nu)$ , sarà:

$$\begin{aligned} l(\xi, \Omega, \alpha, \nu; x) &= \ln(f_X(x; \xi, \Omega, \alpha, \nu)) & (1.16) \\ &= \ln\left(2 \frac{\Gamma(\frac{\nu+d}{2})}{(\pi\nu)^{d/2} \Gamma(\frac{\nu}{2})}\right) - \frac{1}{2} \ln |\Omega| - \frac{\nu+d}{2} \left(1 + \frac{Q_x}{\nu}\right) + \\ &\quad + \ln T_1(U_x, \nu+d). \end{aligned}$$



# Capitolo 2

## Nell'analisi discriminante

In questo capitolo, dopo un breve richiamo ai principali concetti dell'analisi discriminante, si presenteranno e poi si confronteranno due regole di classificazione, quella basata sulla verosimiglianza e quella lineare di Fisher, per studiarne il comportamento nel caso di dati normali asimmetrici e  $t$  asimmetrici.

In seguito, riporteremo alcune applicazioni svolte su dati piuttosto noti in letteratura, per valutare l'efficacia, e quindi la convenienza o meno, nell'utilizzo di tale tecnica nella pratica.

### 2.1 Le leggi discriminanti

Considerando  $K$  popolazioni o gruppi:  $G_1, \dots, G_K$  per  $K \geq 2$ , lo scopo dell'analisi discriminante è di poter costruire una regola, detta **regola di classificazione**, capace di allocare un individuo ad uno di questi gruppi, sulla base delle informazioni a disposizione da un campione di  $n$  osservazioni. Le informazioni sono rappresentate dalla variabile che descrive la popolazione di origine dell'osservazione (*variabile risposta*) e da misurazioni su  $p$  variabili esplicative. La regola di classificazione corrisponde nel dividere lo spazio campionario  $\mathbb{R}^p$  in  $K$  regioni disgiunte, cioè  $R_1, \dots, R_K$  tali che  $R_i \cap R_j =$

$\emptyset \quad \forall i \neq j$  e  $\bigcup R_i = \mathbb{R}^p$  con  $i, j = 1, \dots, K$ ; all'interno delle quali allocare correttamente l'osservazione alla popolazione da cui proviene.

Fra le molteplici scelte per la regola discriminante, ne considereremo principalmente due:

- regola discriminante basata sulla verosimiglianza;
- regola discriminante lineare di Fisher.

Le ipotesi su cui lavoreremo per la trattazione teorica degli argomenti di questo capitolo saranno queste: disporremo di dati provenienti da  $K = 2$  popolazioni distinte,  $G_1$  e  $G_2$ , all'interno delle quali i dati si distribuiranno secondo distribuzioni note che differiranno esclusivamente rispetto al parametro di posizione, cioè la distribuzione sotto  $G_2$  sarà una traslazione di quella sotto  $G_1$ . I risultati a cui giungeremo saranno comunque validi ed estendibili al caso di  $K > 2$  popolazioni.

### 2.1.1 La regola discriminante di verosimiglianza

Per introdurre la regola discriminante basata sulla verosimiglianza, partiremo dal presupposto di dover lavorare con una variabile  $Y_j \sim H(\xi_j, -)$ , dove  $H(\xi_j, -)$  è una generica distribuzione di probabilità governata tra l'altro da un parametro di posizione  $\xi_j$ , per  $j = 1, 2$ ; siano  $\pi_j$  rispettivamente le probabilità a priori di un'osservazione di appartenere alla popolazione  $j$ -esima. La regola discriminante basata sulla verosimiglianza afferma allora che alloco una nuova osservazione  $y$  alla popolazione  $G_1$  se:

$$\frac{L_1(y)}{L_2(y)} \times \frac{\pi_1}{\pi_2} \geq 1 \quad (2.1)$$

o equivalentemente, basandola sulla log verosimiglianza:

$$l_1(y) - l_2(y) + \ln \frac{\pi_1}{\pi_2} \geq 0 \quad (2.2)$$

dove  $L_j(y)$  e  $l_j(y)$  sono rispettivamente la verosimiglianza e la log-verosimiglianza associata all'osservazione condizionatamente al fatto di appartenere alla popolazione  $j$ -esima.

Nel caso le due popolazioni si distribuissero secondo la (1.5), la regola calcolata per la log-verosimiglianza diverrebbe:

$$(\xi_1 - \xi_2)^T \Omega^{-1} \left\{ x - \frac{1}{2} (\xi_1 + \xi_2) \right\} + \zeta_0(w_1) - \zeta_0(w_2) + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0 \quad (2.3)$$

dove  $w_1$  e  $w_2$  sono la quantità  $w$  della (1.9) per le due popolazioni.

Invece, se le distribuzioni fossero del tipo (1.11), la regola sarebbe:

$$\left( \frac{\nu + d}{\nu} \right) \ln \left\{ \frac{\nu + Q_2}{\nu + Q_1} \right\} + \ln \left( \frac{T_1(U_1, \nu + d)}{T_1(U_2, \nu + d)} \right) + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0 \quad (2.4)$$

dove  $Q_j$  e  $U_j$ , per  $j = 1, 2$ , sono le rispettive quantità calcolate per le due popolazioni.

### 2.1.2 La regola lineare di Fisher

Supponiamo sempre che  $Y_j \sim H(\xi_j, -)$ , e che  $H(\xi_j, -)$  sia dotata di varianza, matrice di varianza nel caso multidimensionale. Allora la regola discriminante lineare di Fisher dice che alloco una nuova osservazione  $y$  alla popolazione  $G_1$  se:

$$(E[Y_1] - E[Y_2])^T (Var[Y])^{-1} \left( y - \frac{1}{2} (E[Y_1] + E[Y_2]) \right) + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0.$$

dove  $E[Y_j]$  sta per il valore atteso della distribuzione per la popolazione  $j$ -esima.

Nel caso di due popolazioni da una normale asimmetrica questa regola, tenuto conto delle (1.6) e (1.7), prende la forma:

$$(\xi_1 - \xi_2)^T (\Omega - \omega \mu_z \mu_z^T \omega)^{-1} \left( y - \frac{1}{2} (\xi_1 + \xi_2 + 2\omega \mu_z) \right) + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0, \quad (2.5)$$

mentre nella  $t$  asimmetrica la regola, per le (1.12) e (1.13), assume la forma:

$$(\xi_1 - \xi_2)^T (\Omega - \omega \mu_x \mu_x^T \omega)^{-1} \left( y - \frac{1}{2} (\xi_1 + \xi_2 + 2\omega \mu_x) \right) + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0. \quad (2.6)$$

## 2.2 Confronto fra le due regole

La cosa interessante è che sotto alcune condizioni le due regole che stiamo studiando possono coincidere. Per esempio sappiamo che nelle ipotesi in cui ci siamo posti, nel caso in cui la due distribuzioni siano normali, le due regole coincidono perfettamente. Quello che ci chiediamo è quale sarebbe il comportamento delle due regole nel caso in cui nelle due popolazioni i dati si distribuissero secondo una normale asimmetrica o come una  $t$  asimmetrica.

### 2.2.1 Nella normale asimmetrica

Come sappiamo da Azzalini & Capitanio (1999), le due regole discriminanti per la normale asimmetrica si equivalgono nel caso in cui sia verificata la condizione:

$$\alpha^T \omega^{-1} (\xi_1 - \xi_2) = 0. \quad (2.7)$$

Dimostriamolo. La (2.7) implica che

$$\alpha^T \omega^{-1} \xi_1 = \alpha^T \omega^{-1} \xi_2,$$

e quindi

$$\zeta_0(w_1) = \zeta_0(w_2);$$

allora abbiamo che la regola discriminante di verosimiglianza si riduce a:

$$(\xi_1 - \xi_2)^T \Omega^{-1} \left( y - \frac{1}{2} (\xi_1 + \xi_2) \right) + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0.$$

Per la seconda parte della dimostrazione sviluppiamo la regola di Fisher calcolando l'inversa della matrice di varianza (1.13) secondo la regola, da Mardia et al. (1979):

$$(A_{(n \times n)} + B_{(n \times p)} C_{(p \times p)} D_{(p \times n)})^{-1} = A^{-1} - A^{-1} B (C^{-1} + D A^{-1} B)^{-1} D A^{-1} \quad (2.8)$$

## 2.2. CONFRONTO FRA LE DUE REGOLE

---

ponendo  $A = \Omega$ ,  $B = -\omega\mu_z$ ,  $D = \mu_z^T\omega$  e  $C=1$ . Otteniamo

$$(\xi_1 - \xi_2)^T \left( \Omega^{-1} + \frac{1}{1 - \mu_z^T\omega\Omega^{-1}\omega\mu_z} \Omega^{-1}\omega\mu_z\mu_z^T\omega\Omega^{-1} \right) \times \\ \left( y - \frac{1}{2} (\xi_1 + \xi_2 + 2\omega\mu_z) \right) + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0$$

da cui:

$$(\xi_1 - \xi_2)^T \Omega^{-1} \left( y - \frac{1}{2} (\xi_1 + \xi_2) \right) - (\xi_1 - \xi_2)^T \Omega^{-1}\omega\mu_z + (\xi_1 - \xi_2)^T \times \\ \left( \frac{1}{1 - \mu_z^T\omega\Omega^{-1}\omega\mu_z} \Omega^{-1}\omega\mu_z\mu_z^T\omega\Omega^{-1} \right) \left( y - \frac{1}{2} (\xi_1 + \xi_2 + 2\omega\mu_z) \right) + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0$$

e infine:

$$(\xi_1 - \xi_2)^T \Omega^{-1} \left( y - \frac{1}{2} (\xi_1 + \xi_2) \right) + (\xi_1 - \xi_2)^T \Omega^{-1}\omega\mu_z \times \\ \left[ \frac{1}{1 - \mu_z^T\omega\Omega^{-1}\omega\mu_z} \mu_z^T\omega\Omega^{-1} \left( y - \frac{1}{2} (\xi_1 + \xi_2 + 2\omega\mu_z) \right) + 1 \right] + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0.$$

È dimostrabile inoltre che i vettori  $\Omega^{-1}\omega\mu_z$  e  $\omega^{-1}\alpha$  hanno la medesima direzione. Utilizzando la (1.4) e la (1.6) otteniamo infatti:

$$\omega^{-1}\alpha = \frac{1}{1 - \delta^T\Omega_z^{-1}\delta} \omega^{-1}\Omega_z^{-1}\delta, \\ \Omega^{-1}\omega\mu_z = (\omega\Omega_z\omega)^{-1} \omega \left( \frac{2}{\pi} \right)^{\frac{1}{2}} \delta = \left( \frac{2}{\pi} \right)^{\frac{1}{2}} \omega^{-1}\Omega_z^{-1}\delta;$$

come possiamo vedere la direzione delle due quantità, che abbiamo scritto in grassetto, è la stessa. Per cui se  $\alpha^T\omega^{-1}(\xi_1 - \xi_2) = 0$  anche  $(\xi_1 - \xi_2)^T \Omega^{-1}\omega\mu_z = 0$  e la regola discriminante di Fisher si riduce allora:

$$(\xi_1 - \xi_2)^T \Omega^{-1} \left( y - \frac{1}{2} (\xi_1 + \xi_2) \right) + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0$$

I grafici presenti in Figura 2.1 sono i medesimi osservabili in Azzalini & Capitanio (1999). Essi rappresentano le due funzioni discriminanti, con linea continua la regola di Fisher e con linea tratteggiata quella di verosimiglianza, per 4 coppie di densità normali asimmetriche bidimensionali con  $\alpha = (3, 3)^T$ , matrice di correlazione  $\Omega$  con elemento fuori dalla diagonale pari a 0.4,  $\omega =$

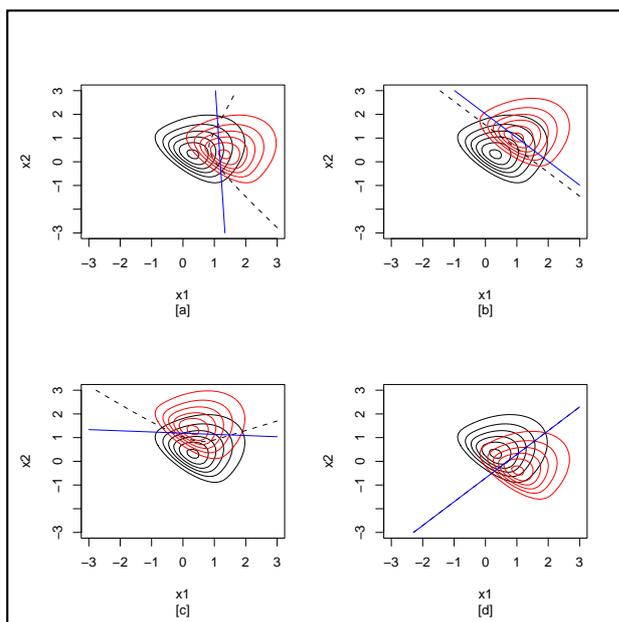


Figura 2.1: Curve di livello per normale asimmetrica e regole discriminanti: (---) regola di verosimiglianza, (—) regola di Fisher.

$I_2$ ,  $\xi_1 = (0, 0)$  e considerando l'equiprobabilità a priori delle due popolazioni, cioè con  $\pi_1 = \pi_2$ . I quattro casi differiscono per la scelta del parametro di posizione della popolazione  $G_2$ , nello specifico, nel grafico [a]  $\xi_2 = (1, 0)^T$ , nel [b]  $\xi_2 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$ , nel [c]  $\xi_2 = (1, 0)^T$  e nel [d]  $\xi_2 = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})^T$ . Come si può vedere il grafico [d] corrisponde al caso in cui la condizione (2.7) è soddisfatta, infatti le due curve coincidono perfettamente.

### 2.2.2 Nella $t$ asimmetrica

Per la  $t$  asimmetrica, proviamo a considerare il caso in cui sia verificata la condizione (2.7) e vediamo cosa avviene alle due regole discriminanti. Partiamo dalla regola lineare di Fisher, sviluppando l'inversa della matrice di varianza (1.13) sempre secondo la (2.8), ponendo in questo caso  $A = \frac{\nu}{\nu-2}\Omega$ ,

## 2.2. CONFRONTO FRA LE DUE REGOLE

---

$B = -\omega\mu_x$ ,  $D = \mu_x^T\omega$ , e  $C = 1$  otteniamo:

$$(\xi_1 - \xi_2)^T \left( \frac{\nu - 2}{\nu} \Omega^{-1} + \frac{\left(\frac{\nu-2}{\nu}\right)^2}{1 - \frac{\nu-2}{\nu} \mu_z^T \omega \Omega^{-1} \omega \mu_x} \Omega^{-1} \omega \mu_x \mu_x^T \omega \Omega^{-1} \right) \times \\ \left( y - \frac{1}{2} (\xi_1 + \xi_2 + 2\omega\mu_x) \right) + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0$$

da cui

$$\frac{\nu - 2}{\nu} (\xi_1 - \xi_2)^T \Omega^{-1} \left( y - \frac{1}{2} (\xi_1 + \xi_2) \right) + \frac{\nu - 2}{\nu} (\xi_1 - \xi_2)^T \Omega^{-1} \omega \mu_x \times \\ \left[ \frac{\frac{\nu-2}{\nu}}{1 - \mu_x^T \omega \Omega^{-1} \omega \mu_x} \mu_x^T \omega \Omega^{-1} \left( y - \frac{1}{2} (\xi_1 + \xi_2 + 2\omega\mu_x) \right) - 1 \right] + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0$$

Come nel paragrafo precedente anche in questo caso è dimostrabile mediante la (1.14) che  $\Omega^{-1}\omega\mu_x$  ha la stessa direzione di  $\alpha^T\omega^{-1}$ , infatti:

$$\Omega^{-1}\omega\mu_x = (\omega\Omega_z\omega)^{-1}\omega \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \left(\frac{\nu}{\pi}\right)^{\frac{1}{2}} \delta = \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \left(\frac{\nu}{\pi}\right)^{\frac{1}{2}} \omega^{-1}\Omega_z^{-1}\delta.$$

Allora la regola discriminante di Fisher diviene:

$$(\xi_1 - \xi_2)^T \Omega^{-1} \left( y - \frac{1}{2} (\xi_1 + \xi_2) \right) + \ln \left( \frac{\pi_1}{\pi_2} \right) \geq 0. \quad (2.9)$$

Per quanto concerne la regola di verosimiglianza sotto la condizione (2.7) non si presentano apparenti semplificazioni algebriche come avveniva invece nel caso della normale asimmetrica, per cui la regola discriminante rimane nella forma (2.4). Tuttavia dai grafici di Figura 2.2, rappresentanti gli stessi casi e condizioni di Figura 2.1 per quattro valori del parametro  $\nu$ , vediamo la coincidenza delle due regole nei casi [d], cioè quando la (2.7) è soddisfatta, indipendentemente dal valore dei gradi di libertà.

Una risposta ci viene osservando il grafico [a] di Figura 2.3 rappresentante, tramite curve di livello, l'andamento delle due regole discriminanti per delle  $t$  asimmetriche, con gli stessi parametri dei grafici di Figura 2.2 e 10 gradi di libertà, nel caso sia soddisfatta la (2.7), al variare del termine  $\ln\left(\frac{\pi_1}{\pi_2}\right)$ . Il grafico mostra, infatti, come le due regole si equivalgano solo nel caso di equiprobabilità a priori delle popolazioni, cioè quando  $\pi_1 = \pi_2$ . Questo

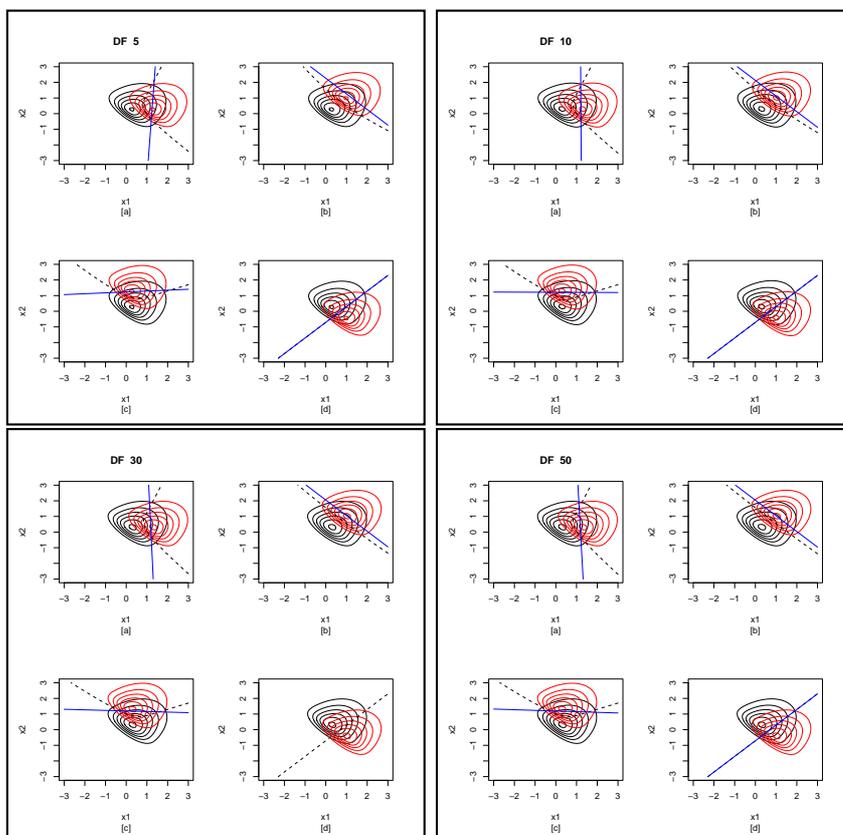


Figura 2.2: Quattro esempi di  $t$  asimmetrica e relative funzioni discriminanti per varie scelte dei gradi di libertà  $\nu = 5, 10, 30, 50$ : ( - - ) regola di verosimiglianza, ( — ) regola di Fisher.

avviene perché, sotto le condizioni (2.7) e che  $\pi_1 = \pi_2$ , entrambe le funzioni discriminanti si annullano nei punti in cui  $Q_1 = Q_2$ . Notiamo che:

$$\begin{aligned}
 Q_1 - Q_2 &= (x - \xi_1)^T \Omega^{-1} (x - \xi_1) - (x - \xi_2)^T \Omega^{-1} (x - \xi_2) \\
 &= \xi_1^T \Omega^{-1} \xi_1 - \xi_2^T \Omega^{-1} \xi_2 - 2\xi_1^T \Omega^{-1} x + 2\xi_2^T \Omega^{-1} x \\
 &= (\xi_1 - \xi_2)^T \Omega^{-1} (\xi_1 + \xi_2) - (\xi_1 - \xi_2)^T \Omega^{-1} x \\
 &= (\xi_1 - \xi_2)^T \Omega^{-1} \left( x - \frac{1}{2} (\xi_1 + \xi_2) \right);
 \end{aligned}$$

quindi quando  $Q_1 = Q_2$  la funzione discriminante di Fisher si annulla.

Invece per la regola di verosimiglianza, il fatto che  $Q_1 = Q_2$  implica che

## 2.2. CONFRONTO FRA LE DUE REGOLE

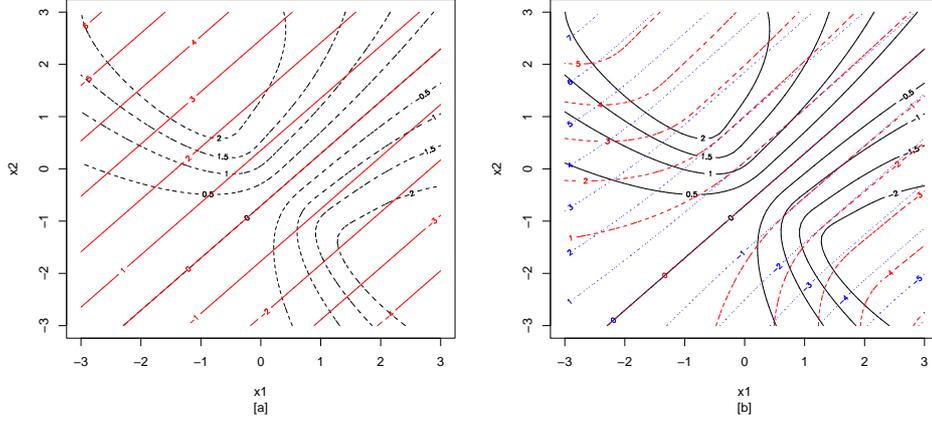


Figura 2.3: [a] Curve di livello per le funzioni discriminanti per una  $t$  asimmetrica con 10 gradi di libert , sotto la condizione (2.7): (- - -) regola di verosimiglianza, (—) regola di Fisher. [b] Evoluzione della regola discriminante di verosimiglianza per la  $t$  asimmetrica all'aumentare dei gradi di libert , sotto la condizione (2.7): (—)  $\nu = 10$ ; (- - -)  $\nu = 100$ , (...)  $\nu = 1000$ .

$U_1 = U_2$ , infatti :

$$\begin{aligned}
 U_1 - U_2 &= \alpha^T \omega^{-1} (x - \xi_1) \left( \frac{\nu + d}{\nu + Q_1} \right)^{\frac{1}{2}} - \alpha^T \omega^{-1} (x - \xi_2) \left( \frac{\nu + d}{\nu + Q_2} \right)^{\frac{1}{2}} \\
 &= \alpha^T \omega^{-1} x \left( \frac{\nu + d}{\nu + Q_1} \right)^{\frac{1}{2}} - \alpha^T \omega^{-1} \xi_1 \left( \frac{\nu + d}{\nu + Q_1} \right)^{\frac{1}{2}} + \\
 &\quad - \alpha^T \omega^{-1} x \left( \frac{\nu + d}{\nu + Q_2} \right)^{\frac{1}{2}} + \alpha^T \omega^{-1} \xi_2 \left( \frac{\nu + d}{\nu + Q_2} \right)^{\frac{1}{2}},
 \end{aligned}$$

e l'uguaglianza   assicurata dalla (2.7), ricordando che sotto tale condizione  $\alpha^T \omega^{-1} \xi_1 = \alpha^T \omega^{-1} \xi_2$ .

Quanto detto ci porta alla seguente conclusione.

**Proposizione 1** *Nel caso della  $t$  asimmetrica e se  $K = 2$ , le regole discriminanti (2.4) e (2.6) coincidono solo se sono soddisfatte le condizioni (2.7) e di equiprobabilit  delle popolazioni, cio   $\pi_1 = \pi_2$ .*

Dal grafico [b] di Figura 2.3 vediamo come al crescere dei gradi di libert  la funzione discriminante di verosimiglianza per la  $t$  asimmetrica, nel caso

sia soddisfatta la (2.7), si riduca a coincidere con quella di Fisher, infatti per  $\nu \rightarrow \infty$  otteniamo il caso della normale asimmetrica in cui le due regole coincidono indipendentemente dalle probabilità a priori delle due popolazioni.

A conferma delle affermazioni fatte sia per la normale asimmetrica sia per la  $t$  asimmetrica, è stato condotto uno studio di simulazione di cui la Tabella 2.1 riassume le principali informazioni. La tabella contiene le frequenze relative di errata classificazione per le due regole discriminanti e la percentuale per cui vi è coincidenza delle stesse, cioè quando entrambe allocano l'osservazione alla medesima popolazione sia che essa sia quella corretta che quella errata. I vari casi da cui abbiamo simulato riguardano sempre il confronto tra due distribuzioni  $t$  asimmetriche bivariate, di cui quella per la popolazione  $G_1$  sempre con parametro di posizione  $\xi_1 = (0, 0)$ , con stesso parametro di asimmetrica  $\alpha = (3, 3)$  e termine fuori dalla diagonale nella matrice di correlazione  $\Omega$  pari a 0.4. I casi però differiscono o per diverso valore del parametro di posizione per  $G_2$ , i cui valori sono i medesimi considerati nei casi [a], [b], [c] e [d] dei grafici in Figura 2.1 e 2.2, o per diverso valore del parametro dei gradi di libertà  $\nu = (5, 10, 30, \infty)$ , ovviamente il caso con  $\nu = \infty$  coincide con lo studiare la normale asimmetrica, e infine per i valori  $\pi_1 = (0.3, 0.5, 0.6)$  della probabilità a priori per la popolazione  $G_1$ . Per ognuno dei casi sono state simulate complessivamente 100000 osservazioni provenienti da entrambe le popolazioni in proporzione tale da rispettare il valore delle probabilità a priori.

$\xi_2$	Df.	Probabilità a priori G1 30%			Probabilità a priori G1 50%			Probabilità a priori G1 60%		
		Verosim.	Fisher	Coinc. Regole	Verosim.	Fisher	Coinc. Regole	Verosim.	Fisher	Coinc. Regole
(1, 0)	5	0.197	0.225	0.928	0.259	0.282	0.875	0.260	0.295	0.828
	10	0.191	0.208	0.939	0.247	0.265	0.890	0.252	0.271	0.876
	30	0.189	0.205	0.933	0.240	0.255	0.901	0.243	0.255	0.900
	SN	0.186	0.202	0.930	0.239	0.251	0.907	0.240	0.249	0.912
$(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$	5	0.169	0.186	0.946	0.248	0.303	0.778	0.270	0.320	0.721
	10	0.164	0.164	0.995	0.239	0.277	0.812	0.259	0.291	0.783
	30	0.150	0.161	0.983	0.230	0.262	0.833	0.250	0.271	0.822
	SN	0.158	0.161	0.976	0.226	0.253	0.844	0.246	0.266	0.834
(0, 1)	5	0.197	0.227	0.927	0.255	0.279	0.874	0.261	0.294	0.828
	10	0.190	0.207	0.938	0.251	0.270	0.890	0.249	0.269	0.875
	30	0.188	0.204	0.934	0.242	0.257	0.902	0.241	0.254	0.903
	SN	0.186	0.203	0.932	0.239	0.253	0.907	0.238	0.250	0.912
$(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$	5	0.237	0.250	0.907	0.273	0.273	1.000	0.266	0.273	0.937
	10	0.228	0.230	0.96	0.267	0.267	1.000	0.259	0.259	0.972
	30	0.223	0.223	0.987	0.264	0.264	1.000	0.253	0.253	0.992
	SN	0.220	0.220	1.000	0.258	0.258	1.000	0.248	0.248	1.000

Tabella 2.1: Frequenze relative di errata specificazione per le regole, e della coincidenza delle suddette, per varie scelte della probabilità a priori per popolazione  $G1$  e all'aumentare dei gradi di libertà.

## 2.3 Alcune applicazioni su dati reali

Verranno ora presentati i risultati ottenuti dall'analisi di tre insiemi di dati molto noti in letteratura: *Hepatic* tratto da Albert & Harris (1987), *Vowel* da Hastie et al. (1994), e *Olive* da Forina et al. (1983). Tutti i risultati sono stati ottenuti tramite l'utilizzo del software statistico R, vers. 2.5.1, dei pacchetti 'sn' per le applicazioni sulle distribuzioni normali e *t* asimmetriche, e 'MASS' per l'analisi discriminante.

Le matrici, dette di **confusione**, e i tassi di errata classificazione riportati come indici per saggiare la bontà dei vari modelli stimati, sono stati ricavati, salvo altra indicazione, tramite il metodo *Cross-Validation* nella variante del *Leave One Out*. Questo metodo consiste nell'utilizzare  $n - 1$  osservazioni delle  $n$  disponibili per determinare la regola discriminante, e in base ad essa, allocare l'osservazione esclusa, ripetendo poi tale procedimento per ogni osservazione del campione. La stima del tasso di errore è ottenuto dal rapporto tra le osservazioni allocate erroneamente con la numerosità campionaria.

### 2.3.1 Dati *Hepatic*

I dati *Hepatic*, presentati da Albert & Harris (1987), si riferiscono a 218 pazienti affetti da problemi epatici, suddivisi in 4 gruppi corrispondenti al tipo di problema:

- **G1**: epatite virale, 57 pazienti;
- **G2**: epatite persistente cronica, 44 pazienti;
- **G3**: epatite persistente aggressiva, 40 pazienti;
- **G4**: cirrosi post-necrotica, 77 pazienti.

Le variabili disponibili si riferiscono a quattro enzimi epatici. Le regole di classificazione sono basate sulla trasformata logaritmica di tre di queste variabili, in Figura 2.4 possiamo vedere i relativi diagrammi di dispersione:

### 2.3. ALCUNE APPLICAZIONI SU DATI REALI

---

- aspartate aminotransferase (*AST*);
- alanine aminotrasferase (*ALT*);
- glutamate dehydrogenase (*GLDH*).

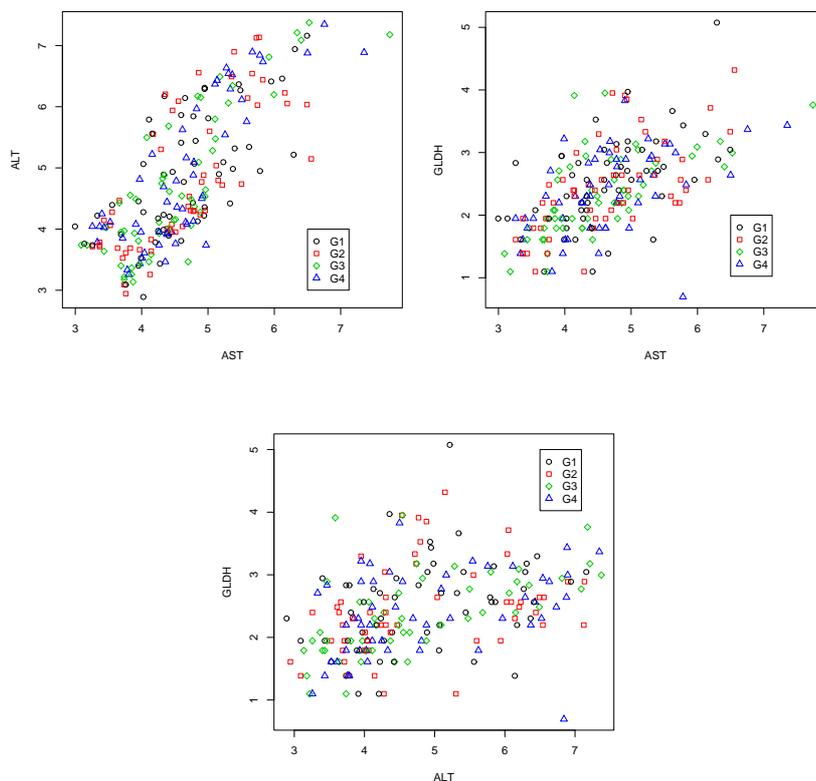


Figura 2.4: Diagrammi di dispersione dati *Hepatic*, suddivisi per gruppi.

I grafici che seguono rappresentano i risultati derivanti dall'analisi discriminante con modellazione classica in Figura 2.5, modellazione con normale asimmetrica in Figura 2.6, e modellazione con  $t$  asimmetrica in Figura 2.7. Per quanto riguarda Figura 2.6 e Figura 2.7, i grafici [a] riportano le distribuzioni stimate per i vari gruppi; nei grafici [b], così come in quelli di Figura 2.5, i punti portano una duplice informazione: il numero indica la

popolazione a cui la regola ha allocato l'osservazione, il colore rappresenta il gruppo di provenienza dell'osservazione, così da poter già avere una prima superficiale idea della bontà o meno della regola discriminante.

Nella Tabella 2.2 sono riportate le matrici di confusione dei tre metodi, cioè il numero di individui allocati ad ogni gruppo dalla regola discriminante di Fisher, dalla regola di verosimiglianza per  $SN$ , e dalla regola di verosimiglianza per  $St$ ; nell'ultima riga sono riportati invece i tassi di errata classificazione per le tre regole discriminanti con la *cross-validation*. Come possiamo vedere, sembra che l'utilizzo della  $t$  asimmetrica porti un irrisorio miglioramento nella bontà allocativa.

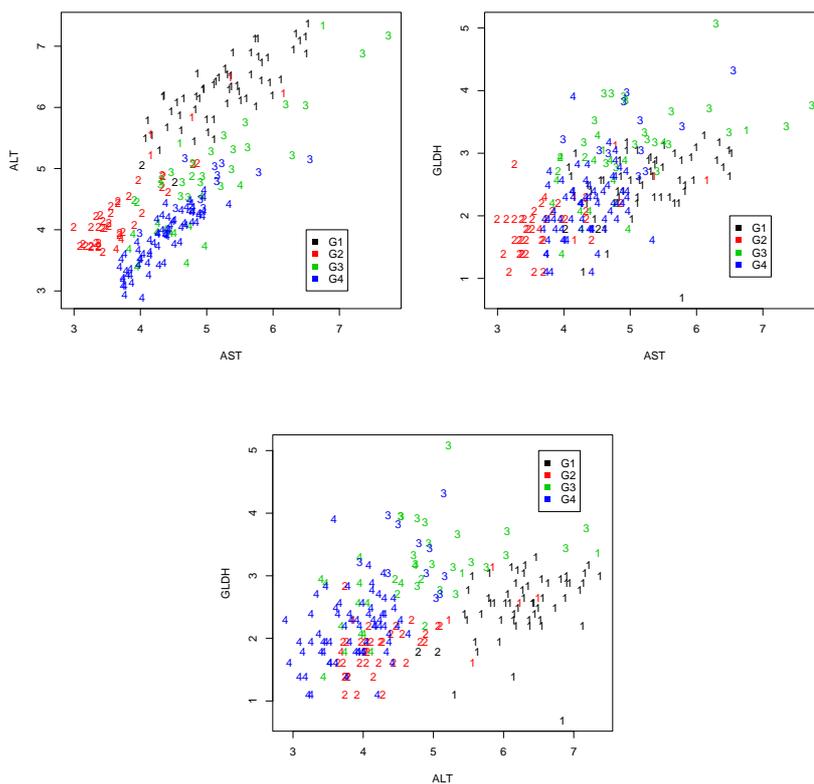


Figura 2.5: Dati *Hepatic*: Allocations con analisi discriminante classica (Fisher).

## 2.3. ALCUNE APPLICAZIONI SU DATI REALI

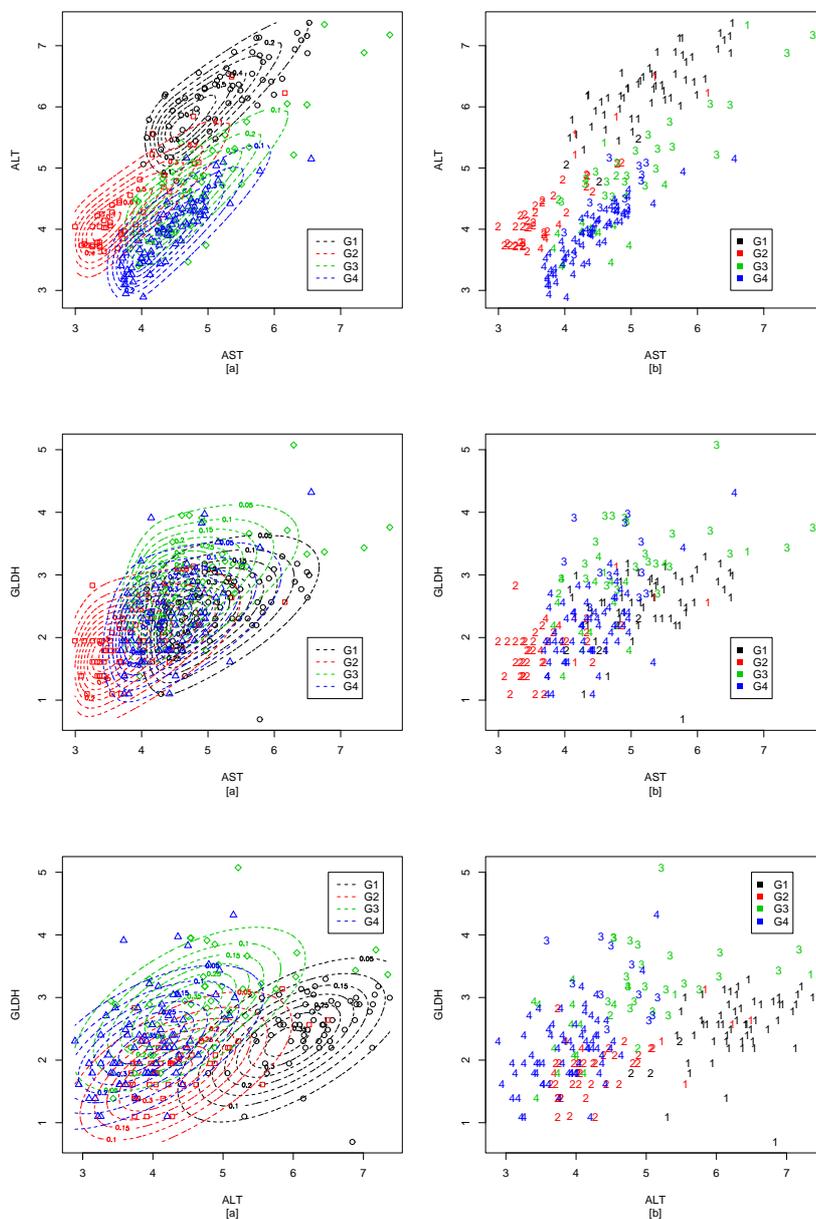


Figura 2.6: Dati *Hepatic*. [a] Distribuzioni  $SN$  per il modello stimato, suddivisi per gruppi. [b] Allocazioni con modello  $SN$ .

### 2.3.2 Dati *Vowel*

I dati *Vowel* riportati da Hastie et al. (1994) riguardano la pronuncia di 11 suoni fonetici della lingua inglese. Le informazioni sono contenute in 10 vari-

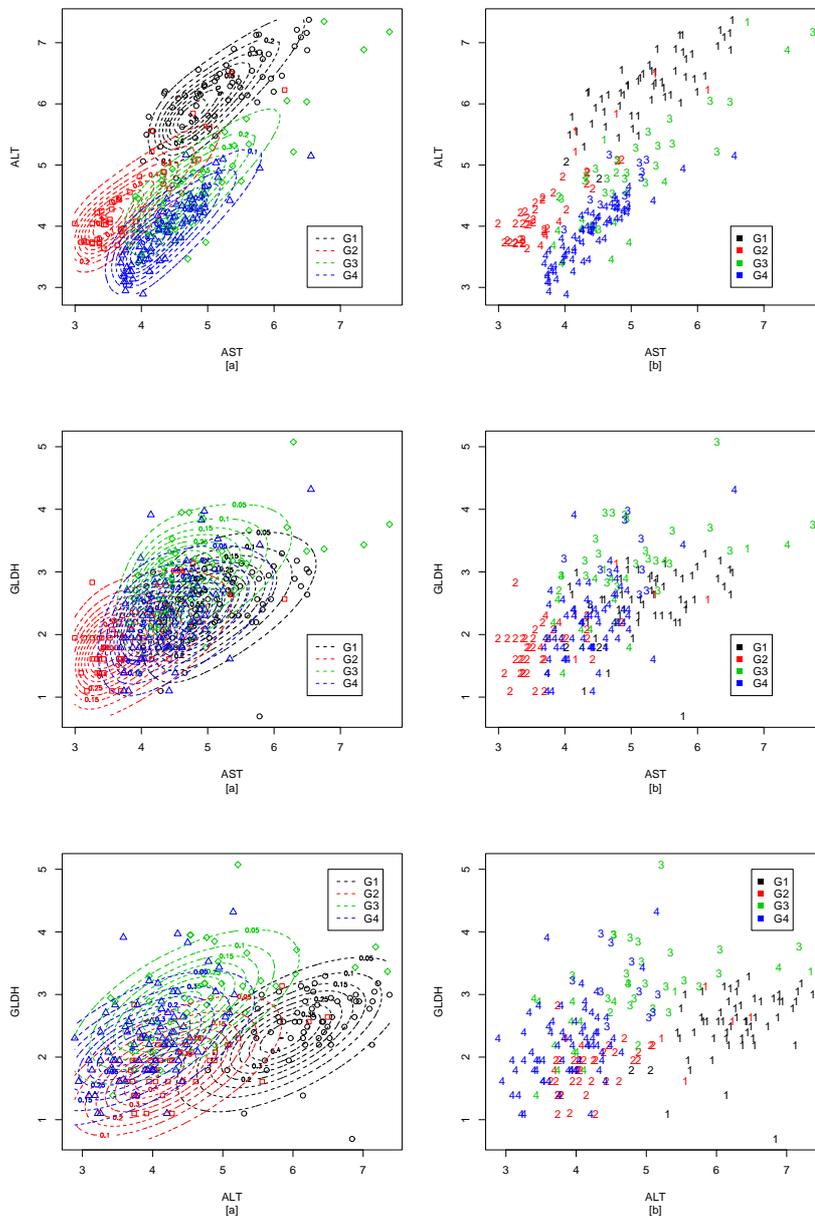


Figura 2.7: Dati *Hepatic*. [a] Distribuzioni  $St$  per il modello stimato, suddivisi per gruppi. [b] Allocazioni con modello  $St$ .

abili continue, denominate  $X.1, \dots, X.10$ , ottenute dalla registrazione della pronuncia dei suoni da vari lettori. I dati sono divisi in due insiemi rispet-

### 2.3. ALCUNE APPLICAZIONI SU DATI REALI

Metodo	LDA				SN				St			
Popolazione	G1	G2	G3	G4	G1	G2	G3	G4	G1	G2	G3	G4
G1	55	5	2	0	53	5	2	0	55	5	2	0
G2	2	37	4	0	3	36	2	0	2	37	2	0
G3	0	0	20	12	1	0	22	11	0	0	20	11
G4	0	2	14	65	0	3	14	66	0	2	16	66
Errata allocazione	18.81				18.81				18.35			

Tabella 2.2: Matrici di confusione e tassi di errata allocazione per i dati *Hepatic*.

tivamente di 528, che useremo come insieme di *Training* su cui stimeremo i modelli, e di 426 righe, che useremo come insieme di *Test* per verificare le *performance* delle regole discriminanti.

Il fatto di avere  $K = 11$  sottopopolazioni in  $d = 10$  dimensioni richiede di stimare un elevato numero di parametri,  $11 \times 10 + 10 \times \frac{11}{2} + 10 = 175$  per il caso della *SN* e 176 per la *St*; senza contare le stime delle 11 probabilità a priori per le popolazioni, stimate mediante le frequenze relative dell'insieme di *Training*. I parametri stimati sulla base dell'insieme di *Training* sono stati inseriti nelle funzioni discriminanti e applicate all'insieme di *Test*.

Inoltre l'elevato numero di combinazioni tra le variabili hanno reso difficile una rappresentazione grafica dei risultati che potesse essere chiara, ma che non risultasse pesante. Per questo motivo i grafici di Figura 2.8, Figura 2.9, e Figura 2.10 presentano le allocazioni per l'insieme di *Training* nei grafici [a] e per l'insieme di *Test* nei grafici [b], per tre combinazioni delle variabili esplicative:  $X.1$  con  $X.2$ ,  $X.5$  con  $X.6$ , e  $X.9$  con  $X.10$ , evidenziano in rosso le osservazioni allocate in maniera errata.

Nelle Tabelle 2.3 e 2.4 troviamo le matrici di confusione e i tassi percentuali di errata allocazione delle varie tecniche per gli insiemi di *Training* e *Test*. In questo caso l'adozione della  $t$  porta un miglioramento in termini di errata allocazione, sia per i dati di *Training*, ma soprattutto per i dati di *Test*, per i quali il miglioramento rispetto all'analisi classica è di 9 punti percentuali.

CAPITOLO 2. NELL'ANALISI DISCRIMINANTE

Metodo	LDA										
Popolazione	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
G1	32	10	0	0	0	0	0	0	0	0	0
G2	8	28	4	0	0	2	0	0	0	0	0
G3	3	9	42	0	0	0	0	0	0	0	2
G4	0	0	0	36	0	1	0	0	0	0	5
G5	0	0	0	0	33	11	8	0	0	0	0
G6	0	0	0	6	6	23	1	0	0	0	2
G7	0	0	0	0	8	3	33	6	3	0	0
G8	0	0	0	0	0	0	0	34	5	6	0
G9	1	0	0	0	0	0	3	1	29	9	1
G10	5	0	0	0	0	0	0	7	11	33	0
G11	0	1	2	6	1	8	3	0	0	0	38
Errata allocazione	31.63										
Metodo	SN										
Popolazione	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
G1	40	9	0	0	0	0	0	0	0	0	0
G2	2	30	3	0	0	0	0	0	0	0	0
G3	0	9	43	0	0	0	0	0	0	0	1
G4	0	0	0	36	0	1	0	0	0	0	5
G5	0	0	0	0	34	12	7	0	0	0	0
G6	0	0	0	6	3	23	0	0	0	0	2
G7	0	0	0	0	10	3	35	6	3	0	0
G8	1	0	0	0	0	0	1	35	6	5	0
G9	3	0	0	0	0	0	2	4	24	9	1
G10	2	0	0	0	0	0	0	3	14	34	0
G11	0	0	2	6	1	9	3	0	1	0	39
Errata allocazione	29.36										
Metodo	St										
Popolazione	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
G1	42	7	0	0	0	0	0	0	0	0	0
G2	0	33	3	0	0	0	0	0	0	0	0
G3	0	7	43	0	0	0	0	0	0	0	1
G4	0	0	0	36	0	1	0	0	0	0	5
G5	0	0	0	0	37	11	7	0	0	0	0
G6	0	0	0	6	2	29	0	0	0	0	3
G7	0	0	0	0	7	0	34	5	3	0	0
G8	1	0	0	0	0	0	1	36	6	6	0
G9	3	0	0	0	0	0	2	5	24	9	0
G10	2	0	0	0	0	0	0	1	15	33	1
G11	0	1	2	6	2	7	4	1	0	0	38
Errata allocazione	27.08										

Tabella 2.3: Matrici di confusione e tassi di errata allocazione per i dati *Training* di *Vowel*.

### 2.3. ALCUNE APPLICAZIONI SU DATI REALI

Metodo	LDA										
Popolazione	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
G1	28	23	0	0	0	0	0	0	0	8	0
G2	10	16	11	0	0	0	0	0	2	1	1
G3	1	2	16	2	0	5	1	0	0	5	2
G4	0	0	11	33	1	3	0	0	0	0	0
G5	0	0	0	0	7	8	9	1	0	0	0
G6	0	1	4	6	22	19	12	0	0	0	11
G7	0	0	0	0	9	1	11	2	0	0	1
G8	0	0	0	0	0	0	4	23	6	0	0
G9	3	0	0	0	0	0	4	8	15	9	2
G10	0	0	0	0	0	0	0	8	14	13	1
G11	0	0	0	1	3	6	1	0	5	6	24
Errata allocazione	55.62										
Metodo	SN										
Popolazione	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
G1	31	21	0	0	0	0	0	0	0	10	0
G2	8	19	9	0	0	0	0	0	2	4	2
G3	0	1	16	2	0	5	1	0	0	2	1
G4	0	0	15	33	2	3	0	0	0	0	0
G5	0	0	0	0	12	5	7	1	0	0	0
G6	0	0	2	6	14	23	12	0	0	0	9
G7	0	0	0	0	6	0	18	2	0	0	1
G8	0	0	0	0	0	0	0	24	10	1	0
G9	3	0	0	0	4	0	3	7	16	8	2
G10	0	0	0	0	0	0	0	8	9	11	2
G11	0	1	0	1	4	6	1	0	5	6	25
Errata allocazione	50.64										
Metodo	St										
Popolazione	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
G1	29	16	0	0	0	0	0	0	0	8	0
G2	10	20	6	0	0	0	0	0	2	1	0
G3	0	2	21	3	0	6	1	0	0	5	2
G4	0	0	13	32	0	2	1	0	0	0	0
G5	0	0	0	0	15	5	11	2	0	0	0
G6	0	0	2	6	14	26	7	0	0	0	7
G7	0	0	0	0	6	0	18	2	0	0	1
G8	0	0	0	0	0	0	0	23	11	1	0
G9	3	0	0	0	3	0	3	7	18	9	0
G10	0	0	0	0	0	0	0	8	6	12	2
G11	0	4	0	1	4	3	1	0	5	6	30
Errata allocazione	47.18										

Tabella 2.4: Matrici di confusione e tassi di errata allocazione per i dati *Test* di *Vowel*.

### 2.3.3 Dati Olive

I dati *Olive* riportati da Forina et al. (1983) si riferiscono a misurazioni per otto componenti chimiche (*acidi oleici*) condotte su  $n = 572$  tipi di oli d'oliva provenienti da nove zone d'Italia. A loro volta le nove zone possono essere raccolte in tre macro aree: Sud, Sardegna e Centro-Nord. Le variabili a disposizione sono gli otto tipi di acidi, denominate: *palmitic*, *palmitoleic*, *stearic*, *oleic*, *linoleic*, *linolenic*, *arachidic*, *eicosenoic*; inoltre abbiamo due variabili qualitative che descrivono rispettivamente la zona e la macro area di provenienza dell'olio.

Macro Area	Zona
1:Sud	1:Puglia Nord 2:Puglia Sud 3:Calabria 4:Sicilia
2:Sardegna	5:Sardegna Entro-terra 6:Sardegna Costa
3:Centro-Nord	7:Umbria 8:Liguria Est 9:Liguria Ovest

Tabella 2.5: Suddivisione della variabile zona per le varie macro aree.

I dati presentano un problema di trattabilità dovuto al fatto di trovarci di fronte a dati di composizione; le variabili rappresentano infatti la concentrazione (*parti su 10000*) dell'acido nell'olio. Per ovviare a questo problema si è adottata la trasformazione log-rapporto additivo (**ALR**: *additive log-ratio*) di Aitchison (1986). Se  $z_j$  ( $j = 1, \dots, 8$ ) denota la  $j$ -ima componente chimica, la trasformata ALR è data da:

$$y_j = \ln \frac{z_j}{z_k}, \quad (j \neq k),$$

dove  $k$  rappresenta una variabile arbitraria, ma fissata, la scelta della variabile

per la trasformazione è essenzialmente irrilevante; nel nostro caso per la trasformazione abbiamo utilizzato la variabile *palmitoleic*.

Una complicazione pratica, dovuta ad errori di misurazione, è che le  $z_j$  non sommano esattamente 10000 in tutti i casi. Una seconda complicazione è la presenza di qualche 0; questo avviene quando il valore effettivo è al di sotto della sensibilità dello strumento. Per ovviare a questi problemi, abbiamo aggiunto 1 a tutti i dati e normalizzato, dividendo ogni numero per la rispettiva somma della riga  $\sum_j (z_j + 1)$ .

Della matrice dei dati  $y$ , risultante dalla trasformazione, considereremo come variabili esplicative le prime cinque componenti principali, che denoteremo con  $X1, \dots, X5$ , contenenti circa il 96% della variabilità totale. Anche in questo caso, dato l'elevato numero di combinazioni tra le variabili, abbiamo rappresentato i risultati delle allocazioni con i tre metodi, solo per tre combinazioni delle variabili esplicative:  $X1$  con  $X2$ ,  $X3$  con  $X4$ , e  $X4$  con  $X5$ . I grafici in Figura 2.11, Figura 2.12, Figura 2.13 mostrano rispettivamente i risultati dell'analisi discriminante classica, di verosimiglianza per  $SN$ , e di verosimiglianza per  $St$ . Nei grafici [a] troviamo le allocazioni per la variabile Zona, mentre nei grafici [b] troviamo le stesse allocazioni, ma accorpate per le tre macro aree di appartenenza. Dato il bassissimo numero di errori in queste ultime, le osservazioni allocate erroneamente sono state evidenziate nei grafici cerchiandole in rosso. Le Tabelle 2.6 e 2.7 contengono le matrici di confusione e i tassi di errata classificazione per le vari tecniche rispetto alla variabile Zona, e accorpando le Zone nelle rispettive macro aree. Come possiamo vedere con questa operazione l'errore dei vari metodi praticamente si azzerava, ciò può significare che gli errori di allocazione che le regole commettono riguardano zone appartenenti alla stessa macro area, per cui plausibilmente con caratteristiche simili quindi più difficili da distinguere. La cosa più importante da notare dalla Tabella 2.6 è che, anche in questo caso, l'utilizzo della  $t$  asimmetrica porta ad una diminuzione del tasso di errata allocazione.

CAPITOLO 2. NELL'ANALISI DISCRIMINANTE

Metodo	LDA								
Popolazione	G1	G2	G3	G4	G5	G6	G7	G8	G9
G1	17	1	0	15	0	0	0	0	0
G2	0	48	3	6	0	0	0	0	0
G3	0	2	203	7	0	0	0	0	0
G4	8	5	0	8	0	0	0	0	0
G5	0	0	0	0	64	17	1	0	0
G6	0	0	0	0	1	16	0	0	0
G7	0	0	0	0	0	0	39	5	11
G8	0	0	0	0	0	0	2	45	0
G9	0	0	0	0	0	0	8	0	40
Errata allocazione	16.08								
Metodo	SN								
Popolazione	G1	G2	G3	G4	G5	G6	G7	G8	G9
G1	17	1	0	15	0	0	0	0	0
G2	0	49	3	5	0	0	0	0	0
G3	0	2	203	7	0	0	0	0	0
G4	8	4	0	9	0	0	0	0	0
G5	0	0	0	0	63	14	0	1	0
G6	0	0	0	0	2	19	0	0	0
G7	0	0	0	0	0	0	41	5	15
G8	0	0	0	0	0	0	0	43	0
G9	0	0	0	0	0	0	9	1	36
Errata allocazione	16.08								
Metodo	St								
Popolazione	G1	G2	G3	G4	G5	G6	G7	G8	G9
G1	18	1	0	16	0	0	0	0	0
G2	0	50	4	6	0	0	0	0	0
G3	0	2	202	6	0	0	0	0	0
G4	7	3	0	8	0	0	0	0	0
G5	0	0	0	0	62	5	0	2	0
G6	0	0	0	0	3	28	0	0	0
G7	0	0	0	0	0	0	44	2	11
G8	0	0	0	0	0	0	0	45	0
G9	0	0	0	0	0	0	6	1	40
Errata allocazione	13.11								

Tabella 2.6: Matrici di confusione e tassi di errata allocazione per i dati *Olive*, classificando rispetto la variabile *Zona*.

### 2.3. ALCUNE APPLICAZIONI SU DATI REALI

---

Metodo	LDA			SN			St		
Popolazione	G1	G2	G3	G1	G2	G3	G1	G2	G3
G1	323	0	0	323	0	0	323	0	0
G2	0	98	1	0	98	1	0	98	2
G3	0	0	150	0	0	150	0	0	149
Errata allocazione	0.17			0.17			0.35		

Tabella 2.7: Matrici di confusione e percentuali di errata allocazione per i dati *Olive*, classificando rispetto la variabile *Macro Area*

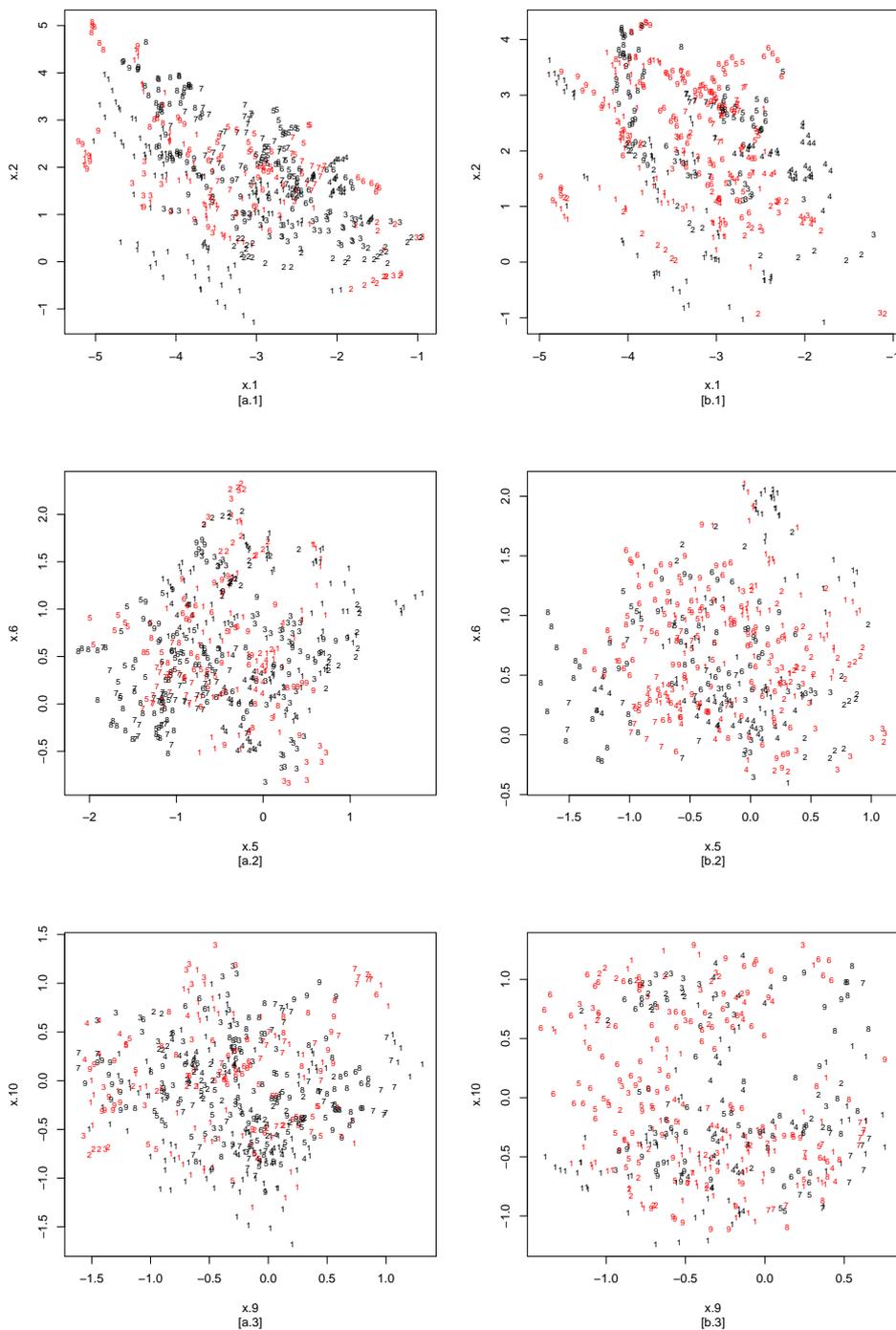


Figura 2.8: Allocazioni con la regola discriminante di Fisher per i dati *Vowel*. [a] Dati *Training*. [b] Dati *Test*.

### 2.3. ALCUNE APPLICAZIONI SU DATI REALI

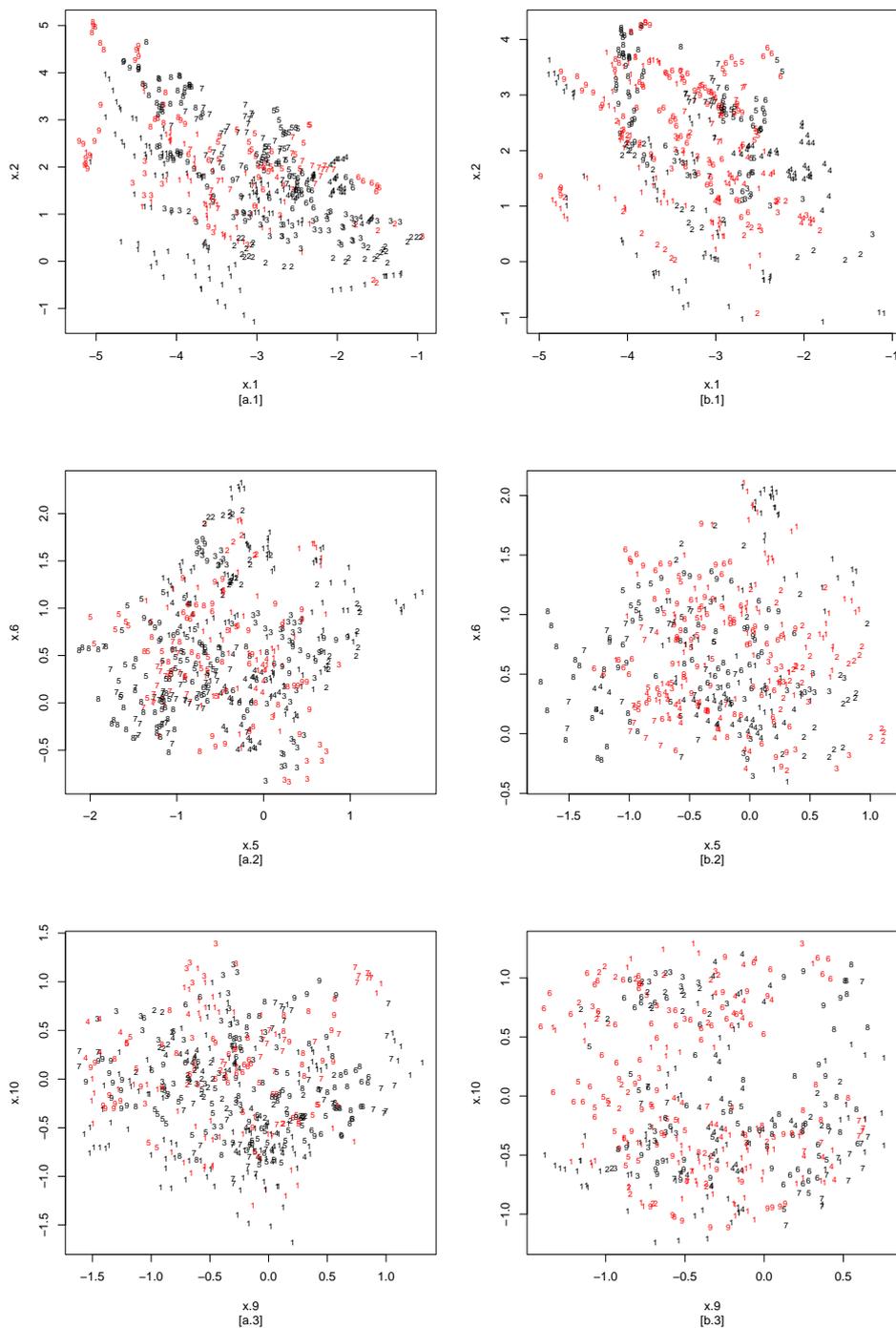


Figura 2.9: Allocations con la regola discriminante di verosimiglianza per  $SN$  dei dati *Vowel*. [a] Dati *Training*. [b] Dati *Test*.

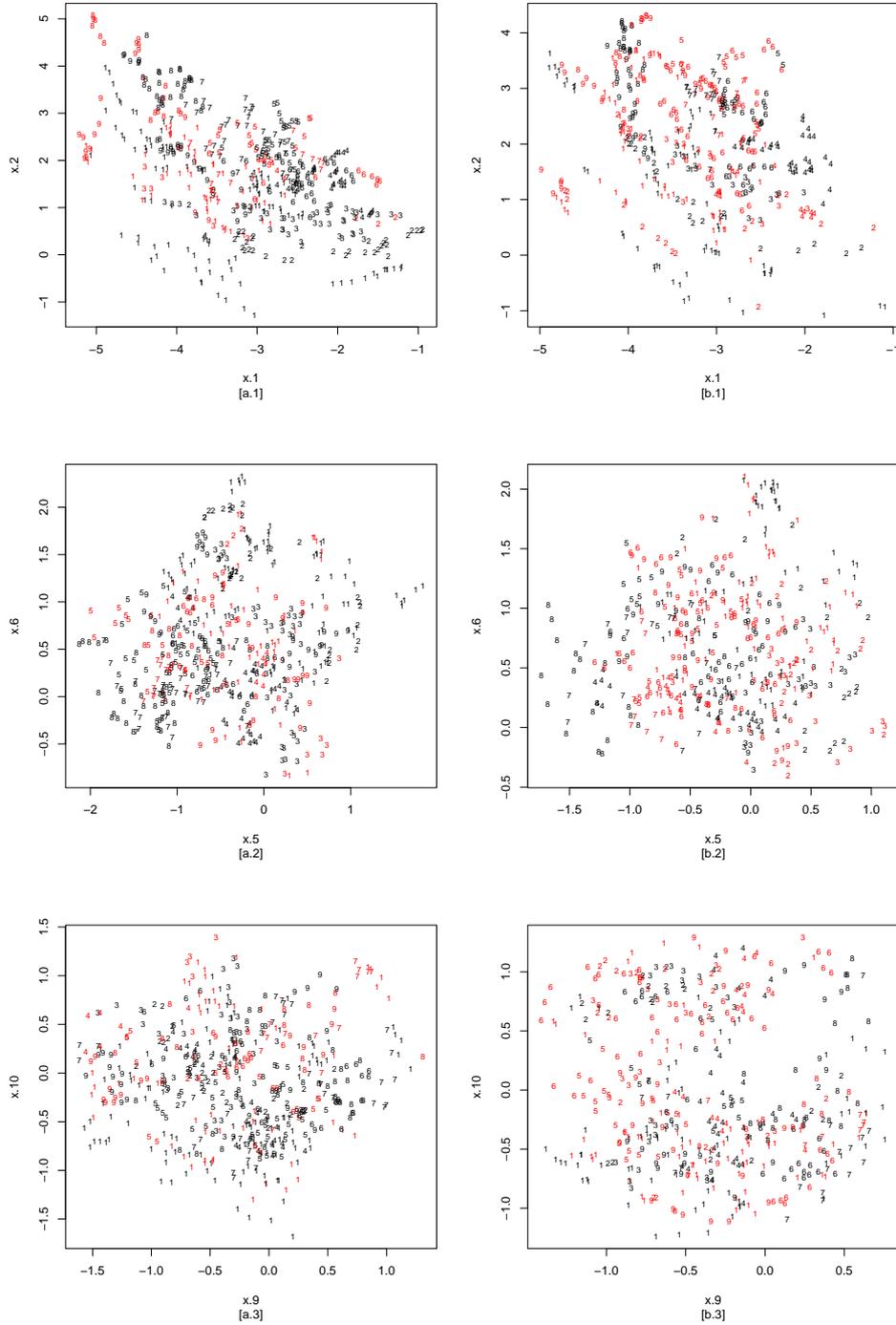


Figura 2.10: Allocations con la regola discriminante di verosimiglianza per *St* dei dati *Vowel*. [a] Dati *Training*. [b] Dati *Test*.

### 2.3. ALCUNE APPLICAZIONI SU DATI REALI

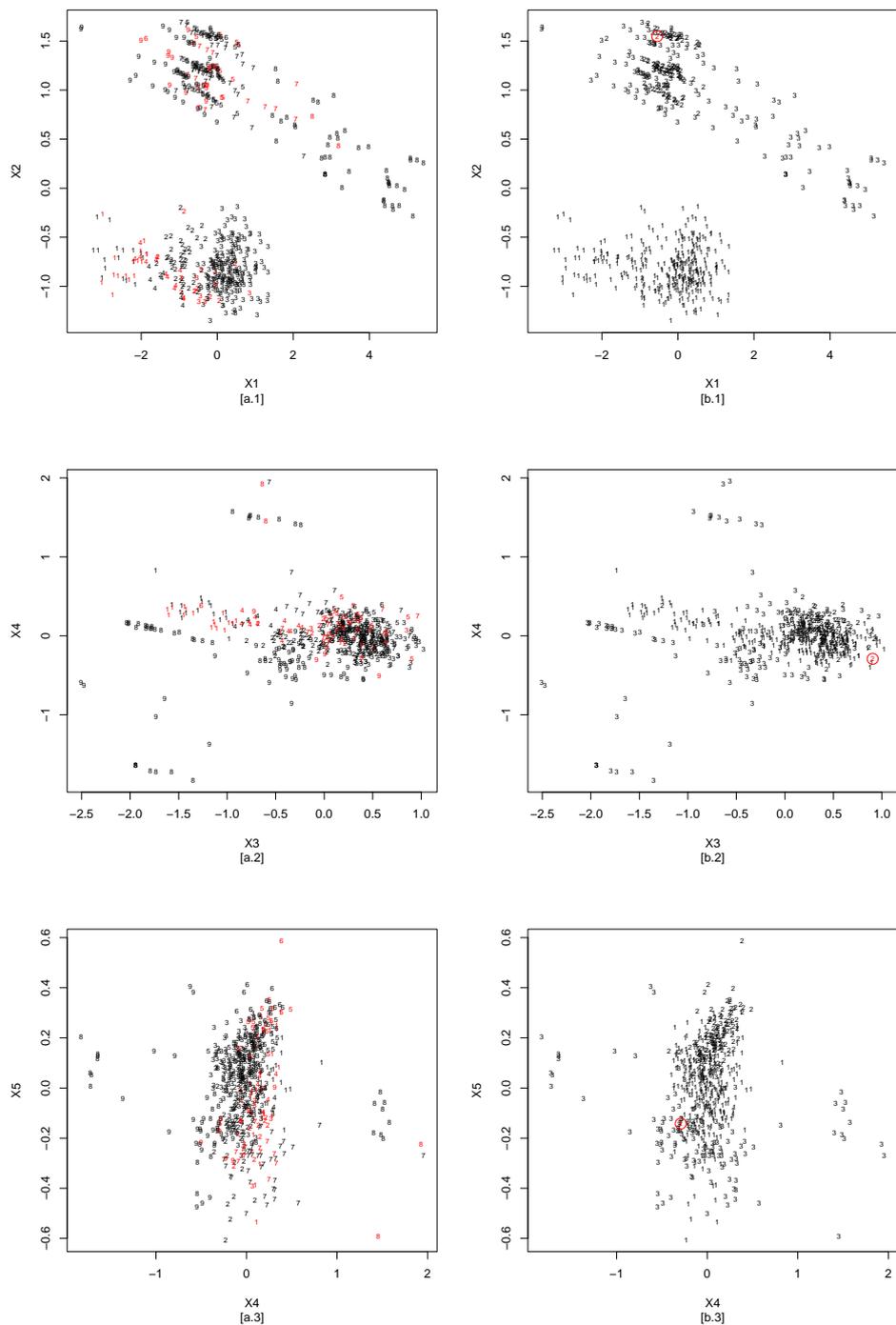


Figura 2.11: Allocations per la regola discriminante lineare di Fisher dati *Olive*. [a] Allocations per Zona. [b] Allocations per Macro aree.

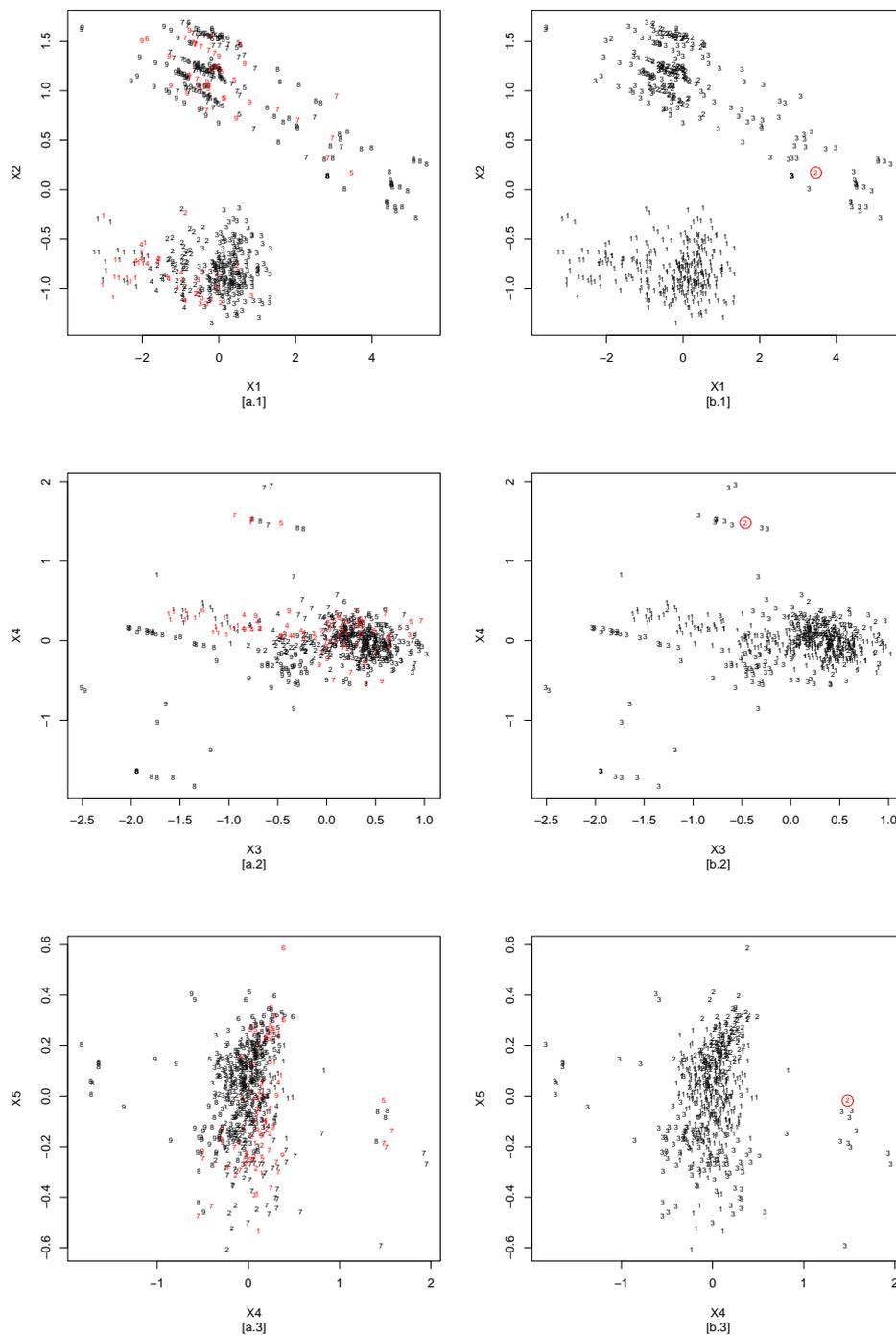


Figura 2.12: Allocazioni per la regola discriminante di verosimiglianza per *SN* dati *Olive*. [a] Allocazioni per Zona. [b] Allocazioni per Macro aree.

### 2.3. ALCUNE APPLICAZIONI SU DATI REALI

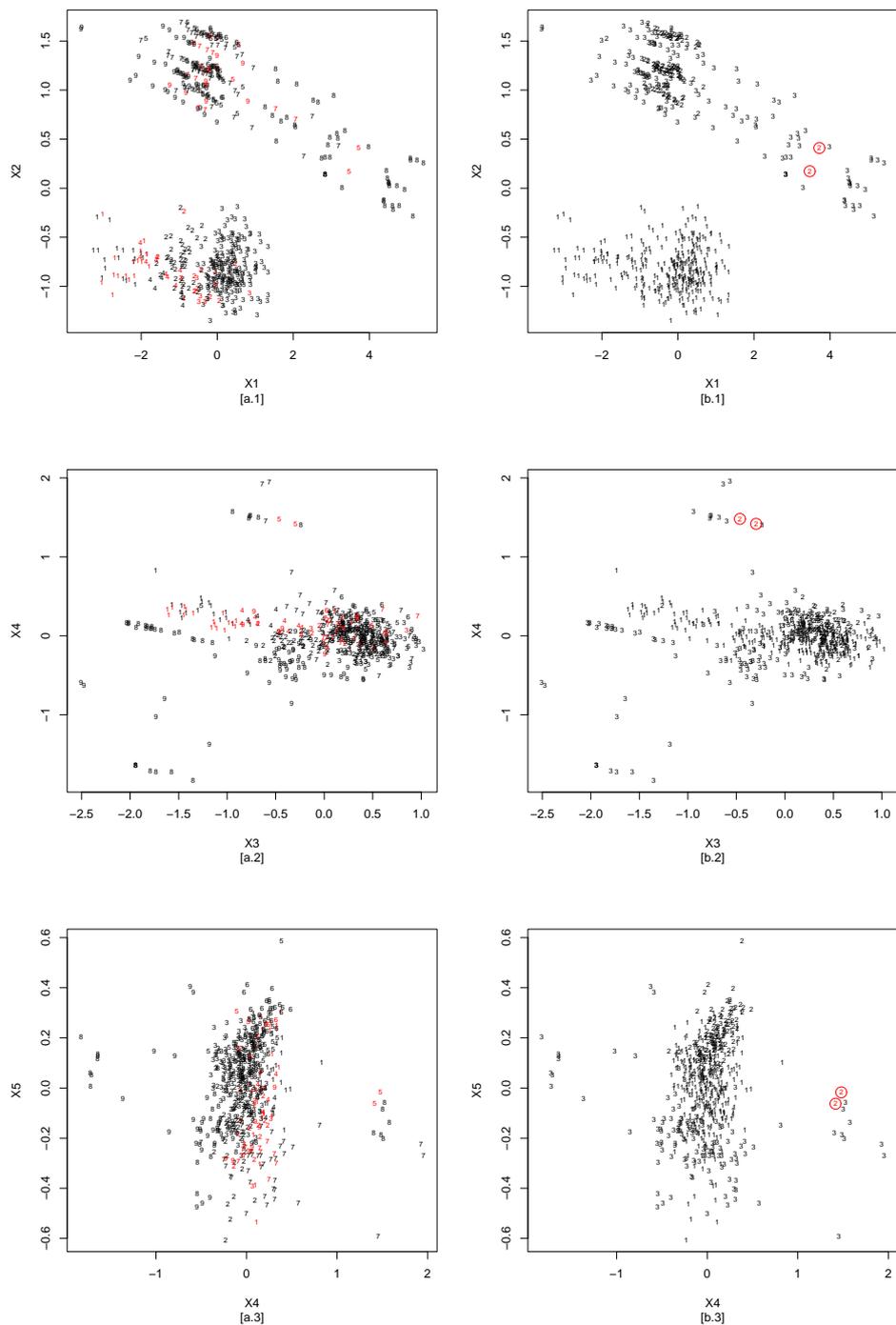


Figura 2.13: Allocazioni per la regola discriminante di verosimiglianza per *St* dati *Olive*.  
[a] Allocazioni per Zona. [b] Allocazioni per Macro Aree.



# Capitolo 3

## Nelle regioni di tolleranza

Il problema della costruzione di regioni con assegnata probabilità  $p$  e di misura geometrica minima è molto comune in molti problemi teorici e pratici, ad esempio la costruzione di regioni di tolleranza. L'uso di distribuzioni asimmetriche del tipo (1.1) e (1.10) introduce complicazioni sul versante della minima misura.

In questo capitolo si cercherà di estendere alcuni risultati già noti in letteratura per la normale asimmetrica, da Azzalini (2001), al caso della  $t$  asimmetrica.

Quello che scopriremo è che, con i dovuti aggiustamenti del caso, molti di questi risultati si conserveranno anche passando alla  $t$  asimmetrica. In entrambi i casi, le regioni cercate si otterranno partendo dalle rispettive tecniche di costruzione per le distribuzioni simmetriche più un termine di correzione dipendente da un indice di asimmetria  $\alpha^*$ , ma nella  $St$  questo termine di correzione dipenderà, logicamente, anche dal valore assunto dai gradi di libertà.

### 3.1 Per la normale asimmetrica

Per una variabile  $Z \sim N_d(0, \Omega)$ , con  $\Omega$  matrice di covarianza, una regione appropriata è data da:

$$R_N = \{x : x^T \Omega^{-1} x \leq c_p\} \quad (3.1)$$

dove  $c_p$  è il quantile di livello  $p$  di una distribuzione  $\chi_d^2$ .

Come sappiamo dalla (1.8), se la distribuzione della variabile  $Z$  viene rimpiazzata con quella di una normale asimmetrica, allora la forma quadratica  $Z^T \Omega^{-1} Z$  continua a distribuirsi come un  $\chi_d^2$ , di conseguenza una regione del tipo (3.1) è ancora di esatta probabilità  $p$ , ma non ha minimo volume, in quanto non corrisponde all'insieme dei punti con più alto valore della funzione di densità. In questo caso, un insieme più appropriato, che chiameremo  $R_{SN}$ , sarà nella forma:

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

per un opportuno valore  $f_0$  dipendente da  $p$ ,  $\Omega$  e  $\alpha$ ; e che soddisfi la condizione  $\mathbb{P}\{R_{SN}\} = p$ .

La proposta fatta in Azzalini (2001) è basata sul fatto che la forma quadratica  $Z^T \Omega^{-1} Z$  nel caso della normale può essere ottenuta partendo dalla funzione di densità, come:

$$2 \ln \phi_d(z; \Omega) - d \ln(2\pi) - \ln |\Omega|.$$

Se sostituiamo la  $\phi_d(z; \Omega)$ , funzione di densità di una normale  $d$ -dimensionale, con la funzione di densità della normale asimmetrica  $f_Z(z)$ , allora otterremo la nostra regione candidata:

$$\{x : 2 \ln f(x) \geq -c_p - d \ln(2\pi) - \ln |\Omega|\} \quad (3.2)$$

I punti triangolari rossi del grafico in Figura 3.1 rappresentano i percentili calcolati sulle probabilità empiriche di copertura  $\tilde{p}$ , i cui valori sono riportati nella rispettiva riga della Tabella 3.1, ottenute dall'applicazione della regione (3.2) per varie scelte della probabilità assegnata:

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

### 3.1. PER LA NORMALE ASIMMETRICA

---

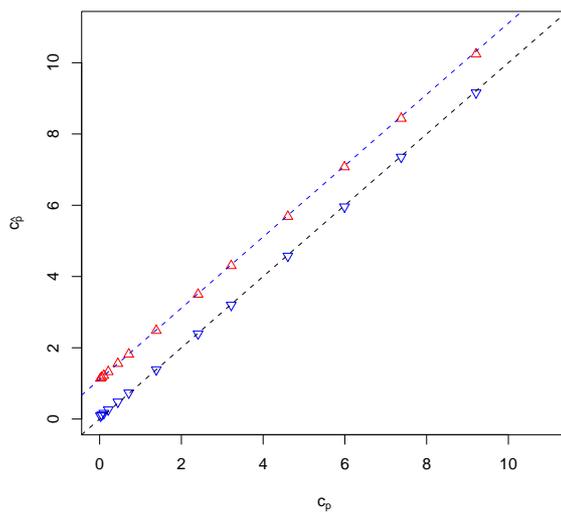


Figura 3.1: Percentili empirici, triangoli rossi, ed empirici con termine di correzione, triangoli blu, per  $SN_2$  di parametri  $\alpha_1 = 2$ ,  $\alpha_2 = 6$  e  $\omega = -0.5$ . Linea nera: bisettrice I° quadrante. Linea blu: retta per il termine di correzione  $h$  stimato.

$p$	99.00	97.50	95.00	90.00	80.00	70.00	50.00	30.00	20.00	10.00	5.00	2.50	1.00
$\tilde{p}$	99.40	98.52	97.09	94.16	88.36	82.60	71.09	59.74	54.10	48.50	45.68	44.28	43.45
$\hat{p}$	98.97	97.42	94.90	89.86	79.84	69.81	50.07	30.75	21.39	12.29	7.86	5.72	4.46

Tabella 3.1: Probabilità assegnate di copertura, empiriche ed empiriche con termine di correzione, ognuna  $\times 100$ , per una  $SN_2$  di parametri  $\alpha_1 = 2$ ,  $\alpha_2 = 6$  e  $\omega = -0.5$ .

ad un vettore di  $10^6$  osservazioni generate da una distribuzione  $SN_2(\Omega, \alpha)$ , dove  $\Omega$  matrice di correlazione con termine fuori diagonale pari a  $-0.5$  e vettore del parametro di asimmetria  $\alpha = (2, 6)^T$ . Possiamo vedere come la differenza di questi percentili empirici dalla bisettrice del I° quadrante, retta lungo la quale dovrebbero giacere i punti nel caso normale, sia costante, nel senso che i punti sembrano ben allineati rispetto ad una retta apparentemente parallela alla bisettrice. Questo ci porta a pensare che il valore  $f_0$ , da inserire nella regione  $R_{SN}$ , potrebbe essere pensato come:

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

dove  $h$  è un termine di correzione.

Come abbiamo detto, la differenza osservata in Figura 3.1 è dovuta all'uso della normale asimmetrica; ci viene quindi naturale supporre che il termine correttivo  $h$  dipenda in qualche modo dall'asimmetria della distribuzione considerata. Un indice di facile calcolo e comunemente usato per valutare globalmente l'asimmetria di una distribuzione normale asimmetrica è

$$\alpha^* = \sqrt{\alpha^T \Omega \alpha} \tag{3.3}$$

come evidenziato da Azzalini & Capitanio (1999).

Il grafico in Figura 3.2 riporta circa 2700 valori di  $h$  in funzione di  $\alpha^*$ , per altrettante combinazioni di  $\alpha$  e del termine di correlazione  $\omega$ . Il singolo  $h$  è l'intercetta dell'interpolazione lineare per i percentili empirici ottenuti come in precedenza per il grafico in Figura 3.1, per varie scelte del parametro di asimmetria  $\alpha$  e del termine di correlazione  $\omega$ , quindi al variare di  $\alpha^*$ . Dal grafico [a] in Figura 3.2 vediamo che  $h$  è una funzione monotona in  $\alpha^*$ . Dal grafico [b] in Figura 3.2 vediamo come la relazione tra  $\alpha^*$  e il termine  $\{\ln(e^{h/2} - 1)\}^{-1}$  sia di proporzionalità con rapporto  $-0.6373$ . Questo si traduce nella funzione interpolante per  $h$ :

$$\hat{h} = 2 \ln \{1 + \exp(-b/\alpha^*)\}. \tag{3.4}$$

### 3.2. PER LA $T$ ASIMMETRICA

dove  $b = 1.569$ ; la funzione interpolante appena trovata corrisponde alla linea rossa tratteggiata nel grafico [a] in Figura 3.2, che sembra ben interpolare i valori di  $h$ . Applicando ora l'opportuno termine di correzione  $h$  al caso di Figura 3.1, otteniamo i nuovi percentili corretti, rappresentati dai triangoli blu, per le probabilità stimate  $\hat{p}$  osservabili nella rispettiva riga della Tabella 3.1. Come possiamo vedere il risultato è molto più che soddisfacente, infatti i nuovi percentili empirici sono molto vicini a quelli assegnati, escludendo i casi di  $p$  prossimo a 0 dove comunque vi è un sostanziale miglioramento rispetto ai percentili empirici calcolati sulle  $\tilde{p}$ .

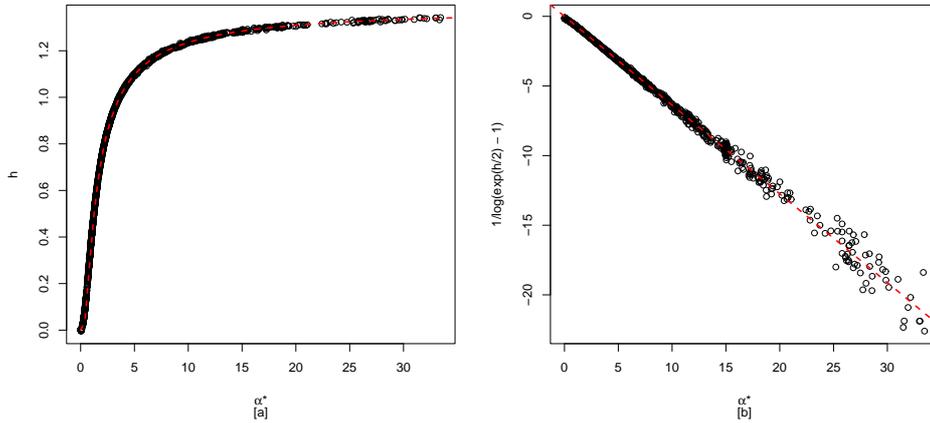


Figura 3.2: Casi  $SN$ . [a] Punti  $h$  osservati contro  $\alpha^*$ , con funzione interpolante. [b] Punti  $1/\ln(\exp(h/2) - 1)$  osservati contro  $\alpha^*$  e retta interpolante.

### 3.2 Per la $t$ asimmetrica

Per studiare il caso della  $t$  asimmetrica, partiamo dal fatto che da Azzalini & Capitanio (2003) sappiamo che se  $X \sim t_d(\xi, \Omega, \nu)$  allora per la quantità  $Q_x = X^T \Omega^{-1} X$  vale la seguente proprietà:

$$\frac{Q_x}{d} = \frac{(x - \xi)^T \Omega^{-1} (x - \xi)}{d} \sim F_{d, \nu}. \quad (3.5)$$

Una regione con probabilità assegnata e volume minimo per una variabile  $t$  di Student sarà del tipo:

$$R_T = \{x : Q_x \leq d \cdot F_{p;d,\nu}\},$$

dove  $F_{p;d,\nu}$  è il percentile  $p$  di una distribuzione  $F$  di Fisher con  $d$  gradi di libertà a numeratore e  $\nu$  a denominatore.

Passando alla distribuzione  $t$  asimmetrica (1.10), come avveniva nella normale asimmetrica, la proprietà (3.5) continua a valere. Potremmo quindi costruire una regione simile alla precedente del tipo:

$$R_{ST} = \{x : f(x) \geq f_0\}$$

dove  $f(x)$  è la funzione di densità di una  $t$  asimmetrica, e  $f_0$  è ancora un opportuno valore dipendente dalla probabilità  $p$ , da  $\Omega$ ,  $\alpha$ , e probabilmente anche dal  $\nu$ . La regione  $R_{ST}$  presenterà però gli stessi problemi riscontrati nel caso della normale asimmetrica.

Parallelamente a quanto fatto per la normale asimmetrica, possiamo ottenere la quantità  $Q_x$  partendo dalla funzione di densità di una  $t$  di Student, come:

$$Q_x = \left\{ \left[ t_d(x; \Omega, \nu) \cdot \frac{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{d}{2}} |\Omega|^{\frac{1}{2}}}{\Gamma(\frac{\nu+d}{2})} \right]^{-\frac{2}{\nu+d}} - 1 \right\} \nu,$$

allora una regione candidata potrebbe essere ottenuta sostituendo la funzione di densità della  $t$  di Student  $t_d(x; \Omega, \nu)$ , con la densità della  $t$  asimmetrica  $f(x)$ :

$$R_{\tilde{S}t} = \left\{ x : \ln f(x) \geq \ln \left[ \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{d}{2}} |\Omega|^{\frac{1}{2}}} \right] - \frac{\nu+d}{2} \ln \left[ \frac{d}{\nu} F_{p;d,\nu} + 1 \right] \right\}. \quad (3.6)$$

Il passaggio al logaritmo ha lo scopo di semplificare i calcoli rendendo somme e differenze ciò che altrimenti sarebbero prodotti e rapporti senza modificare di fatto la regione.

I punti triangolari rossi nel grafico in Figura 3.3 rappresentano i percentili corrispondenti alle probabilità empiriche  $\tilde{p}$  calcolate con la regione (3.6) per

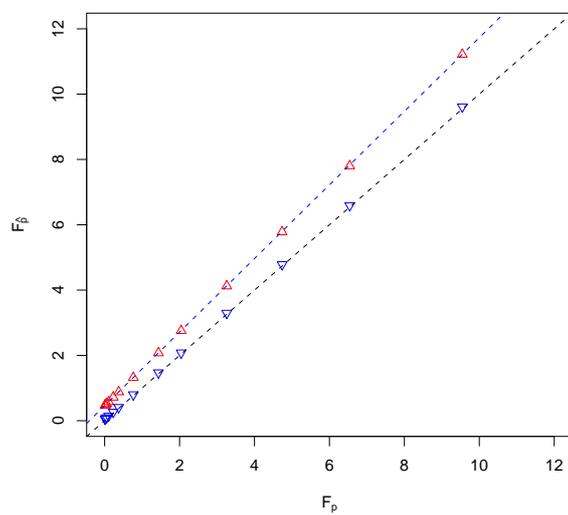


Figura 3.3: Percentili empirici, triangoli rossi, ed empirici con termine di correzione, triangoli blu, per  $St_2$  di parametri  $\alpha_1 = 2$ ,  $\alpha_2 = 6$ ,  $\omega = -0.5$ ,  $\nu = 7$ . Linea nera: bisettrice I° quadrante. Linea blu: retta di coefficiente angolare  $\beta$  e intercetta  $h$  stimati per il termine di correzione  $K(\beta, h)$ .

$10^6$  osservazioni da una distribuzione  $St_2$  con 7 gradi di libertà, parametro di asimmetria  $\alpha = (2, 6)^T$  e termine di correlazione  $\omega = -0.5$ ; e per i consueti valori di

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

Come possiamo vedere i punti sono ben allineati, ma in questo caso rispetto ad una retta che non sembra essere parallela alla bisettrice del primo quadrante. Questo fatto ci porta a pensare che le regioni per la  $t$  asimmetrica necessiteranno di un termine di correzione del tipo:

$$F_{\tilde{p};d,\nu} = F_{p;d,\nu} + K(h, \beta),$$

dove  $K(h, \beta)$  è una funzione di  $h$  e  $\beta$ , rispettivamente intercetta e coefficiente angolare della retta dei percentili empirici.

È naturale supporre che sia  $h$  che  $\beta$  dipendano in qualche modo dall'asimmetria della distribuzione considerata. A questo scopo abbiamo simulato 1200 casi, ciascuno di  $10^6$  osservazioni, per 100 valori di  $\alpha^*$ , ottenuti agendo sul termine di correlazione e sul parametro di asimmetria, per i seguenti 12 valori del parametro dei gradi di libertà  $\nu$ :

$$\nu = (2, 5, 7, 10, 15, 20, 25, 30, 40, 50, 70, 100).$$

Gli  $h$  e i  $\beta$  saranno rispettivamente le intercette e i coefficienti angolari delle rette stimate sulla base dei percentili per le probabilità empiriche  $\tilde{p}$  nei diversi casi.

Partiamo con lo studiare i termini  $\beta$ : possiamo vedere dai grafici [a] e [b] in Figura 3.4 come i punti seguano curve diverse a seconda del valore di  $\nu$ . Soprattutto nel grafico [b] i punti  $[\beta \cdot \alpha^*]$  si allineano lungo delle rette le cui pendenze sembrano essere influenzate dai valori di  $\nu$ . Nel grafico [c] in Figura 3.4 sono rappresentati i coefficienti angolari stimati, che abbiamo chiamato  $\tau$ , delle 12 rette, per gli altrettanti valori di  $\nu$ , del grafico [b]. Non vanno però confusi questi ultimi con i  $\beta$  che rappresentano la componente per il

### 3.2. PER LA $T$ ASIMMETRICA

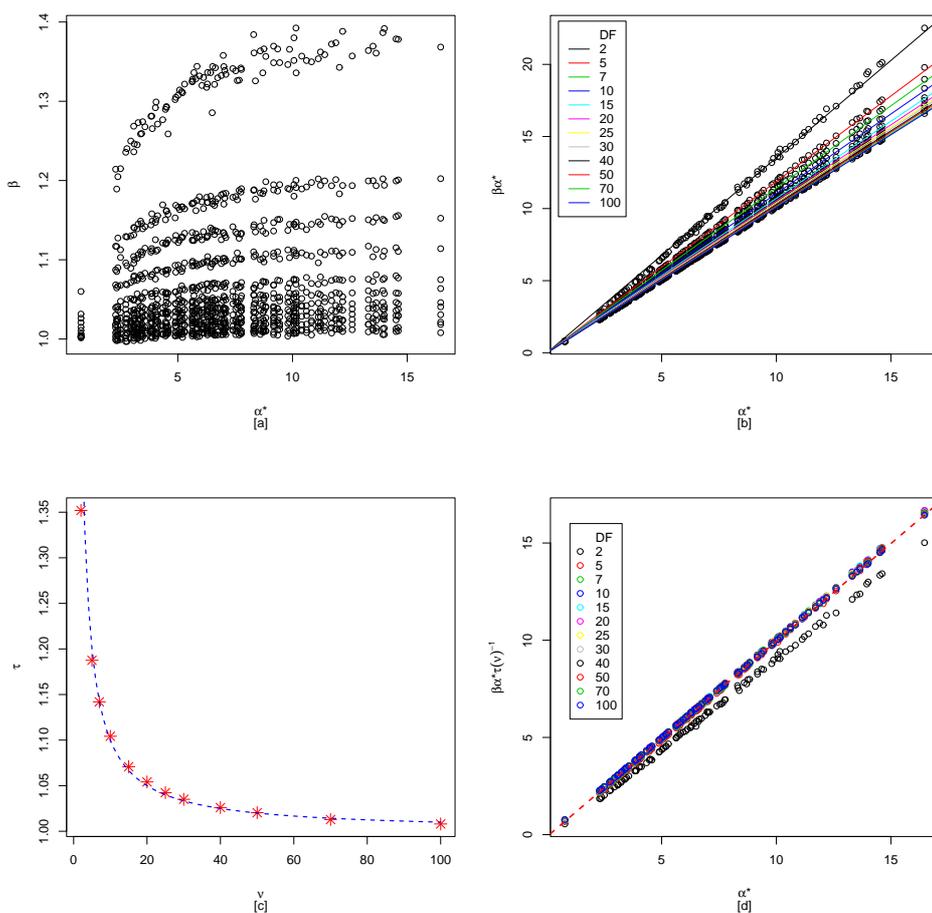


Figura 3.4: Casi  $St_2$ . [a] Punti  $\beta$  osservati contro  $\alpha^*$ . [b] Punti  $\beta\alpha^*$  contro  $\alpha^*$  e rette interpolanti per i dispettivi gradi di libertà  $\nu$ . [c] Termini di aggiustamento  $\tau(\nu)$  contro  $\nu$  e funzione interpolante. [d] Punti  $\beta\alpha^*\tau(\nu)^{-1}$  contro  $\alpha^*$  e retta interpolante.

coefficiente angolare nel termine di correzione delle regioni (3.6). I punti  $\tau$  sembrano essere ben interpolati dall'iperbole di equazione:

$$\tau(\nu) = \frac{1 + \nu}{\nu}.$$

Otteniamo allora un termine di aggiustamento per i  $\beta$ , infatti come possiamo vedere nel grafico [d] i punti  $[\beta \cdot \alpha^* \cdot \tau(\nu)^{-1}]$  sembrano ora indipendenti, ad eccezione che per  $\nu = 2$ , dal valore assunto da  $\nu$ . Interpolando ora questi ultimi punti, che dipenderanno ora solamente da  $\alpha^*$ , otterremo la seguente funzione per la stima della componente  $\beta$ :

$$\hat{\beta} = \frac{(1.005\alpha^* - 0.1)}{\alpha^*} \tau(\nu) = \frac{(1.005\alpha^* - 0.1)}{\alpha^*} \frac{(1 + \nu)}{\nu}. \quad (3.7)$$

Analogamente passando ai termini  $h$ , dai grafici [a] e [b] in Figura 3.5 osserviamo come anch'essi si allineino su curve diverse al variare del valore di  $\nu$ . Nel grafico [b] vediamo che anche in questo caso  $\nu$  influenza i coefficienti angolari, che abbiamo chiamato  $\lambda$ , delle rette interpolanti i punti  $[\alpha^* \ln(\exp(\frac{h}{100}) - 1) + 1.5]$ . Dal grafico [c] vediamo che i  $\lambda$  possono essere interpolati con l'iperbole:

$$\lambda(\nu) = \frac{-1.3 - 4.93\nu}{\nu}.$$

Anche in questo caso otteniamo un termine di aggiustamento per i  $\nu$  da applicare agli  $h$ , da cui nel grafico [d] i punti  $[(\alpha^* \ln(\exp(\frac{h}{100}) - 1) + 1.5) \cdot \lambda(\nu)^{-1}]$  sono ora indipendenti da  $\nu$  e interpolandoli otteniamo la seguente funzione per la stima della componente  $h$ :

$$\hat{h} = 100 \ln \left\{ \exp \left[ \frac{(1.005\alpha^* - 0.045)\lambda(\nu) - 1.5}{\alpha^*} \right] + 1 \right\} \quad (3.8)$$

Per la funzione del termine di correzione per la regione (3.6), la scelta è caduta su

$$K(\beta, h) = \beta \cdot h, \quad (3.9)$$

### 3.2. PER LA $T$ ASIMMETRICA

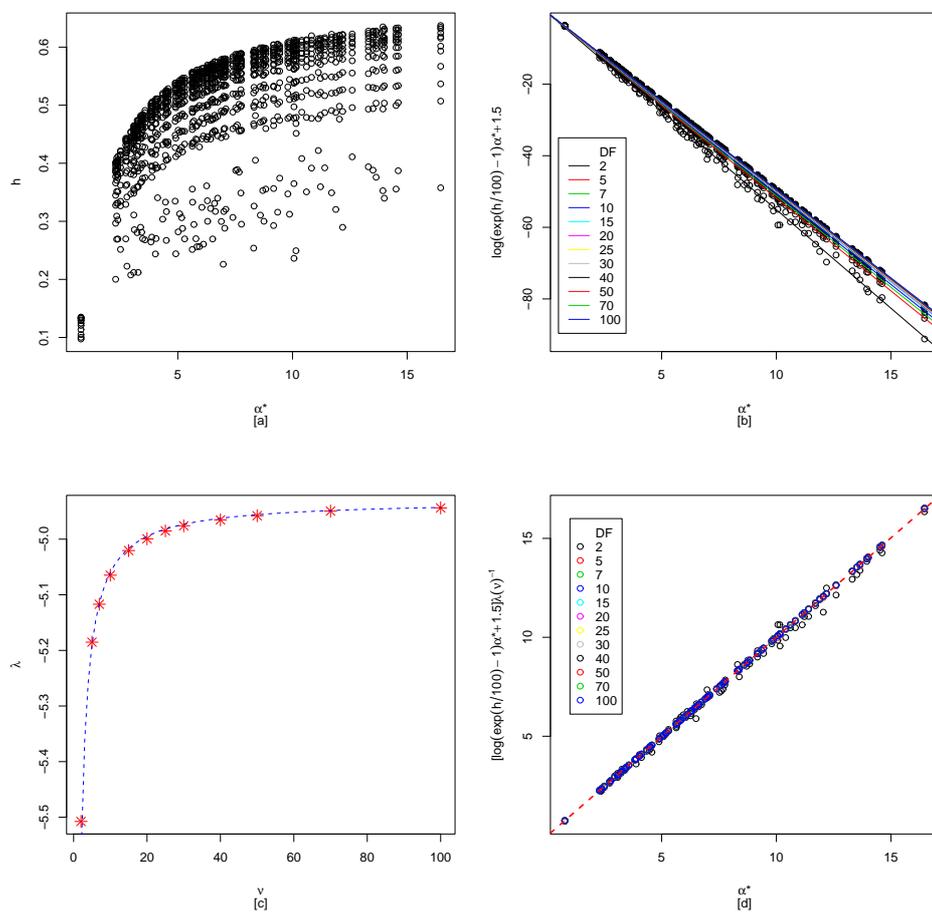


Figura 3.5: Casi  $St_2$ . [a] Punti  $h$  osservati contro  $\alpha^*$ . [b] Punti  $[\ln(\exp(h/100) - 1)\alpha^* + 1.5]$  contro  $\alpha^*$  e rette interpolanti per i diversi valori di  $\nu$ . [c] Termine di aggiustamento  $\lambda(\nu)$  contro  $\nu$  e funzione interpolante. [d] Punti  $[\ln(\exp(h/100) - 1)\alpha^* + 1.5]\lambda(\nu)^{-1}$  contro  $\alpha^*$  e retta interpolante.

$p$	99.00	97.50	95.00	90.00	80.00	70.00	50.00	30.00	20.00	10.00	5.00	2.50	1.00
$\tilde{p}$	99.34	98.35	96.71	93.44	86.91	80.39	67.38	54.40	47.94	41.52	38.32	36.71	35.76
$\hat{p}$	99.02	97.54	95.10	90.19	80.45	70.73	51.32	32.16	22.68	13.43	8.92	6.70	5.37

Tabella 3.2: Probabilità assegnate di copertura, empiriche e empiriche con termine di correzione, ognuna  $\times 100$ , per una  $St_2$  di parametri  $\alpha_1 = 2$ ,  $\alpha_2 = 6$ ,  $\omega = -0.5$  e  $\nu = 7$ .

per cui la regione corretta è:

$$R_{\hat{S}_t} = \left\{ x : \ln f(x) \geq \ln \left[ \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{d}{2}} |\Omega|^{\frac{1}{2}}} \right] - \frac{\nu+d}{2} \ln \left[ \frac{d}{\nu} F_{p,d,\nu} + 1 \right] + K(\hat{\beta}, \hat{h}) \right\}. \quad (3.10)$$

I risultati di tale correzione sono osservabili dai triangoli blu nel grafico in Figura 3.3, inoltre la linea blu tratteggiata rappresenta la retta di intercetta  $\hat{h}$  e coefficiente angolare  $\hat{\beta}$  stimati per il caso considerato.

Per valutare ulteriormente la tecnica appena costruita, abbiamo simulato 200 ulteriori casi lasciando libere di variare le componenti di  $\alpha^*$ , e il valore di  $\nu$ . Le Tabelle 3.3 e 3.4 riassumono il comportamento delle probabilità empiriche di copertura  $\tilde{p}$  ed empiriche con termine di correzione  $K(h, \beta)$  stimato per i 200 casi. Come possiamo vedere il comportamento della regione (3.10), sempre se escludiamo i valori di  $p$  prossimi allo 0, è molto buono. Se per esempio consideriamo un  $p = 0.95$ , il valore più comunemente usato nelle applicazioni pratiche, allora nel 50% dei casi simulati la probabilità empirica con termine di correzione,  $\hat{p}$ , non si discosta dalla  $p$  assegnata per più di 0.001, e nel 75% per non più di 0.0013, cioè significa che nel 50% (75%) dei casi simulati, la tecnica costruisce regioni con probabilità di copertura del 0.951 (0.9513) anziché del 0.95 assegnato. È da evidenziare che i valori peggiori per  $\hat{p}$  sono riscontrabili per valori di  $\nu$  inferiori a 5, stiamo parlando comunque di scostamenti che al massimo non superano 0.003 per una probabilità  $p$  assegnata di 0.95.

$p$	99.00	97.50	95.00	90.00	80.00	70.00	50.00	30.00	20.00	10.00	5.00	2.500	1.000
Min.	99.08	97.67	95.33	90.66	81.37	72.08	53.47	34.85	25.49	16.18	11.56	9.228	7.824
1st Qu.	99.32	98.30	96.61	93.24	86.48	79.74	66.23	52.83	46.18	39.55	36.24	34.580	33.600
Median	99.36	98.40	96.80	93.60	87.23	80.84	68.10	55.43	49.12	42.87	39.74	38.200	37.260
Mean	99.34	98.36	96.72	93.44	86.88	80.34	67.28	54.28	47.80	41.35	38.13	36.520	35.560
3rd Qu.	99.39	98.47	96.93	93.87	87.74	81.65	69.49	57.35	51.32	45.30	42.29	40.800	39.900
Max.	99.43	98.57	97.13	94.25	88.52	82.79	71.34	59.87	54.16	48.48	45.65	44.230	43.400

Tabella 3.3: *Summary* per le probabilità empiriche di copertura  $\tilde{p} \times 100$ .

49

$p$	99.00	97.50	95.00	90.00	80.00	70.00	50.00	30.00	20.00	10.00	5.00	2.500	1.000
Min.	98.98	97.47	94.99	90.03	80.13	70.25	50.56	30.91	21.09	11.19	6.261	3.800	2.313
1st Qu.	99.01	97.52	95.05	90.11	80.26	70.44	50.95	31.62	22.10	12.75	8.180	5.942	4.632
Median	99.01	97.54	95.08	90.16	80.34	70.55	51.09	31.95	22.51	13.28	8.727	6.492	5.160
Mean	99.02	97.55	95.10	90.21	80.43	70.68	51.28	32.07	22.58	13.24	8.659	6.406	5.067
3rd Qu.	99.03	97.57	95.13	90.25	80.53	70.80	51.44	32.25	22.77	13.45	8.884	6.651	5.332
Max.	99.07	97.65	95.30	90.58	81.17	71.84	53.03	34.32	25.02	15.73	11.130	8.845	7.492

Tabella 3.4: *Summary* per le probabilità empiriche di copertura corrette  $\hat{p} \times 100$ .



# Conclusioni e sviluppi futuri

Giunti alla fine di questo lavoro di tesi possiamo trarre le seguenti conclusioni.

- Per l'analisi discriminante, sulla base dei casi considerati esiste un vantaggio nell'utilizzo delle regole basate sulla distribuzione  $t$  asimmetrica rispetto all'analisi classica, ma non possiamo dire che tale vantaggio sia generalizzabile. Nei nostri tre esempi il miglioramento in termini di bontà allocativa va dall'irrisorietà nei dati *Hepatic* a valori più confortanti per i dati *Vowel*, soprattutto il +9% sull'insieme di *Test* in quanto calcolato su nuovi dati, quindi rappresentativo della vera bontà allocativa della regola.

Tuttavia, sia per la normale asimmetrica, che per la  $t$  asimmetrica, si sono riscontrati problemi nella scelta del modello migliore su cui costruire le regole. Nel concreto i problemi sono legati agli algoritmi numerici per il calcolo delle stime dei parametri. L'algoritmo migliore non era sempre quello che conduceva al valore massimo per la verosimiglianza in quanto il modello stimato si traduceva in densità che assegnavano probabilità nulle ad un numero non trascurabile di osservazioni. La scelta del modello è stata un compromesso tra massima verosimiglianza e sensatezza dei modelli ottenuti, valutata a livello grafico tramite le curve di livello delle distribuzioni per i valori stimati dei parametri dei vari gruppi.

Il sospetto è che il problema sia legato ad una non corretta specificazione dei modelli legata alle assunzioni di omoschedasticità e/o di equi-asimmetria dei gruppi; tali assunzioni conducono infatti a stimare

gli stessi parametri per  $\alpha$  e  $\Omega$  per ogni gruppo, probabilmente in almeno uno dei gruppi tali assunzioni non sono soddisfatte.

- Per quanto riguarda la costruzione di regioni di tolleranza, nonostante i buoni risultati ottenuti, la tecnica è senz'altro assoggettabile a migliorie, soprattutto dal lato della determinazione delle funzioni interpolanti le componenti  $h$  e  $\beta$  per il termine di correzione. In secondo luogo il lavoro svolto in questa tesi si è focalizzato sulla determinazione di una tecnica per la costruzione di regioni di tolleranza sulla base di una distribuzione  $t$  asimmetrica bidimensionale. Ora, anche se la tecnica ottenuta parte da un risultato generale indipendente dalla dimensione della distribuzione, ed è quindi facilmente estendibile ai casi con dimensione superiore a 2, tale estensione richiede naturalmente lo studio del comportamento delle componenti  $\beta$  e  $h$  nei casi specifici.

# Bibliografia

Aitchison J.(1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.

Albert, A. & Harris, E.K. (1987). *Multivariate Interpretation of Clinical Laboratory Data*. Dekker, New York.

Azzalini, A. (1985). A class of distribution which includes the normal ones. *Scand.J.Statistics*, **12**, 171-178.

Azzalini, A. (1986). Further results on a class of distribution which includes the normal ones. *Statistica*, **XLVI**, 2, 199-208.

Azzalini, A. & Dalla Valle, A (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 4, 715-726.

Azzalini, A. & Capitanio, A. (1999). Statistical applications of the multivariate skew normal. *J.Royal Stat. Soc.*, B, **61**, 579-602.  
(Full Length Version: <http://azzalini.stat.unipd.it/SN>)

Azzalini, A.(2001). A note on regions of given probability of the skew-normal ditribution. *METRON*, **LIX**, 27-34.

Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of simmetry with emphasis on a multivariate skew *t* distribution. *J.Royal Stat. Soc.*, B, **65**, 367-389.  
(Full Length Version: <http://azzalini.stat.unipd.it/SN>)

- Azzalini, A. & Genton, M.G., (2007). Robust likelihood method based on the skew- $t$  and related distributions. *In corso di stampa su 'Intern.stat.Review'*.
- Azzalini, A. & Torelli, N. (2007). Clustering via nonparametric density estimation. *Stat. Comput.*, **17**, 71-80.
- Forina, M., Armanino C., Lanteri S. and Tiscornia E. (1983). Classification of olive oils from their fatty acid composition. In: H. Martens and H. J. Russwurm (Eds.), *Food Research and Data Analysis*, Applied Science Publishers, London, 189-214.
- Hastie, T., Tibshirani, R., and Buja, A. (1994). Flexible discriminant analysis by optimal scoring. *Journal of American Statistical Association*, **89**, 1255-1270.
- Johnson, N.L & Kotz, S. (1972). *Distributions in Statistics: Continuous Univariate Distributions - 2*. Wiley, New York.
- Johnson, N.L & Kotz, S. (1972). *Distributions in Statistics: Continuous Univariate Distributions - 3*. Wiley, New York.
- Johnson, N.L & Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley, New York.
- Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press, London.

# Ringraziamenti

A conclusione non solo di questa tesi, ma credo di una fase importante della mia vita, non posso non ringraziare le persone che mi hanno accompagnato lungo questi anni, che sono volati.

Primi fra tutti i miei genitori per avermi sempre sostenuto, non solo dal punto di vista economico che non è poco, in tutte le mie scelte, come il trasferirmi a Padova o l'andare in Erasmus. Sono le persone più importanti della mia vita e alle quali, anche se non lo dico o non lo dimostro quanto vorrei, voglio un mondo di bene.

Gli amici. Annapaola, Maria Giovanna, Marta, Alessandra, Anna, Davide, Tamara Sclero, Roberta, Francesca, Ilaria, Dario, Valeria, Paola, Mayra, Roberta, Silvia, Tamara e tanti altri. Per aver condiviso le gioie, i momenti difficili, le vittorie e le mezze sconfitte, mezze perchè esistono sempre i secondi, i terzi e anche i quarti appelli. Un ringraziamento particolare a Silvia di Ravenna per aver salvato il mio Erasmus.

L'E.S.U e l'Università degli Studi di Padova per avermi accolto a casse aperte. Avermi dato un tetto sulla testa e anche dei soldini da mettere in banca. E dandomi la possibilità di conoscere gente straordinaria.

I coinquilini di questi anni. Sacco, Elia, Alberto, Manuel, Oliver, Daniela, Tukam, Federico, Domenico, Benedetto, Luca, Sara, Bruna Maria Teresa Petracci ecc, le tre Marie che poi sono quattro, Concetta, Lila, Camilla e Ale. Perchè mi hanno sopportato e spero continuino a farlo.

Ai professori dei corsi che ho seguito, per avermi trasmesso la loro passione per una materia che spero potrà diventare non solo il mio lavoro ma anche

la mia vita.

Un ringraziamento particolare va al professor Azzalini per avermi seguito in questo lavoro di tesi e per la pazienza che ha portato durante questi mesi.

Ma soprattutto grazie a Dio.....ho finito!!