

Elementi di statistica inferenziale classica

Lucio Barabesi

Indice

| | | |
|----------|---|------------|
| 1 | Il campionamento | 1 |
| 1.1 | Il modello statistico | 1 |
| 1.2 | Le statistiche | 4 |
| 1.3 | La verosimiglianza statistica | 8 |
| 2 | La stima per punti | 19 |
| 2.1 | Introduzione | 19 |
| 2.2 | Gli stimatori corretti | 21 |
| 2.3 | Gli stimatori coerenti | 22 |
| 2.4 | Gli stimatori efficienti | 23 |
| 2.5 | Gli stimatori sufficienti | 33 |
| 3 | Le stime di massima verosimiglianza | 41 |
| 3.1 | Il metodo della massima verosimiglianza | 41 |
| 3.2 | Le proprietà degli stimatori di massima verosimiglianza | 51 |
| 3.3 | Le proprietà degli stimatori di massima verosimiglianza per grandi campioni | 53 |
| 3.4 | Alcune esemplificazioni del metodo della massima verosimiglianza | 59 |
| 4 | La verifica di ipotesi | 67 |
| 4.1 | Il test statistico | 67 |
| 4.2 | Il test del rapporto delle verosimiglianze | 70 |
| 4.3 | Alcune applicazioni del test del rapporto delle verosimiglianze | 77 |
| 5 | La stima per intervalli | 99 |
| 5.1 | Gli intervalli di confidenza | 99 |
| 5.2 | La stima per intervalli con grandi campioni | 104 |
| 5.3 | Le regioni di confidenza | 107 |
| 5.4 | Alcuni esempi di intervalli di confidenza | 109 |
| 6 | L'inferenza multivariata | 113 |
| 6.1 | Il campionamento con osservazioni multivariate | 113 |
| 6.2 | La stima per punti con osservazioni multivariate | 117 |
| 6.3 | La verifica di ipotesi con osservazioni multivariate | 122 |
| 6.4 | Le regioni di confidenza con osservazioni multivariate | 146 |
| | Appendice | 149 |
| A.1 | Alcune distribuzioni continue | 149 |
| A.2 | Alcune distribuzioni discrete | 155 |
| A.3 | Le distribuzioni Chi-quadrato, t ed F | 158 |
| A.4 | Alcune distribuzioni multivariate | 160 |
| | Bibliografia | 165 |

Capitolo 1

Il campionamento

1.1. Il modello statistico

1.1.1. Il campione. Si supponga che un esperimento casuale abbia prodotto un vettore di osservazioni (x_1, \dots, x_n) . Questo vettore è detto campione, mentre la quantità n è detta numerosità campionaria. Ogni campione, una volta osservato, costituisce un risultato acquisito. Tuttavia, prima del rilevamento, il campione è un vettore di variabili casuali (X_1, \dots, X_n) con una funzione di ripartizione congiunta F_n .

Se ogni osservazione viene ottenuta nelle medesime condizioni sperimentali e se il campionamento è effettuato in modo da assicurare anche l'indipendenza delle osservazioni, il campione è la realizzazione di un vettore di variabili casuali indipendenti e ugualmente distribuite. In questo caso, il campione è detto casuale e questa è la situazione campionaria a cui verrà fatto riferimento più spesso nel seguito. Se il campionamento è casuale, la funzione di ripartizione congiunta di (X_1, \dots, X_n) risulta

$$F_n(x_1, \dots, x_n) = \prod_{i=1}^n F(x_i),$$

dove F è la funzione di ripartizione marginale comune alle X_i . Dunque, la funzione di ripartizione congiunta di (X_1, \dots, X_n) dipende solamente da F . Di conseguenza, un campione casuale può essere pensato convenientemente come un insieme di n realizzazioni indipendenti di una variabile casuale X con funzione di ripartizione F .

1.1.2. Il modello statistico classico. L'informazione relativa all'esperimento casuale è interamente contenuta nella funzione di ripartizione congiunta F_n . L'insieme delle funzioni di ripartizione ammissibili per l'esperimento casuale delimita una classe \mathcal{F} detta modello statistico. Obiettivo dell'inferenza statistica è quello di fare affermazioni sulla distribuzione di \mathcal{F} che ha effettivamente generato il campione (x_1, \dots, x_n) sulla base del campione stesso. L'inferenza sarà maggiormente accurata quanto più è possibile delimitare la classe \mathcal{F} , compatibilmente con la condizione che \mathcal{F} contenga effettivamente la distribuzione che ha generato il campione.

Nella statistica inferenziale classica si assume nota la morfologia funzionale delle funzioni di ripartizione congiunte del modello statistico e si assume che queste siano dello stesso tipo a meno di un insieme di parametri. Se θ è un vettore di k parametri in uno spazio parametrico $\Theta \subseteq \mathbb{R}^k$, tale che F_n è una funzione di ripartizione congiunta per ogni $\theta \in \Theta$, allora il modello statistico classico dipende solamente da θ , ovvero

$$\mathcal{F}_\theta = \{F_n : F_n(x_1, \dots, x_n; \theta), \theta \in \Theta\},$$

dove nella notazione si è enfatizzato la dipendenza di F_n da θ . L'obiettivo dell'inferenza si riduce a fare affermazioni su $\theta_0 \in \Theta$ il "vero valore" del parametro della distribuzione in \mathcal{F}_θ che ha generato il campione (x_1, \dots, x_n) .

Quando si adotta un campionamento casuale, il modello statistico dipende solo dalla struttura delle funzioni di ripartizione marginali. Se $F(\cdot; \theta)$ è una funzione di ripartizione per ogni $\theta \in \Theta$, nel caso di un campionamento casuale il modello statistico classico risulta

$$\mathcal{F}_\theta = \left\{ F_n : F_n(x_1, \dots, x_n; \theta) = \prod_{i=1}^n F(x_i; \theta), \theta \in \Theta \right\}.$$

• **Esempio 1.1.1.** Dato un campione casuale da $X \sim N(\mu, \sigma^2)$, il modello statistico dipende dal vettore di parametri $\theta = (\mu, \sigma)^T$ con $\Theta = \mathbb{R} \times \mathbb{R}^+$, ovvero

$$\mathcal{F}_{\mu, \sigma} = \left\{ F_n : F_n(x_1, \dots, x_n; \mu, \sigma) = \prod_{i=1}^n \Phi\left(\frac{x_i - \mu}{\sigma}\right) \right\}. \quad \square$$

• **Esempio 1.1.2.** Dato un campione casuale da $X \sim E(0, \sigma)$, il modello statistico dipende dal parametro $\theta = \sigma$ con $\Theta = \mathbb{R}^+$, ovvero

$$\mathcal{F}_\sigma = \left\{ F_n : F_n(x_1, \dots, x_n; \sigma) = \prod_{i=1}^n (1 - e^{-\frac{x_i}{\sigma}}) \mathbf{1}_{[0, \infty)}(x_i) \right\}. \quad \square$$

A meno di casi che coinvolgono distribuzioni singolari o misture di distribuzioni, si può usualmente supporre che ad ogni funzione di ripartizione F_n corrisponda una funzione di densità (definita a meno di insiemi di misura di Lebesgue nulla) o funzione di probabilità congiunta, denotata con f_n . Dunque, per semplicità il modello statistico classico può essere specificato in termini di f_n , ovvero

$$\mathcal{F}_\theta = \{f_n : f_n(x_1, \dots, x_n; \theta), \theta \in \Theta\},$$

dove nella notazione si è enfatizzato la dipendenza di f_n da θ . Nel caso di un campionamento casuale, il modello statistico classico può essere espresso in termini della funzione di densità o della funzione di probabilità $f(\cdot; \theta)$ relativa a $F(\cdot; \theta)$, ovvero

$$\mathcal{F}_\theta = \left\{ f_n : f_n(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta), \theta \in \Theta \right\}.$$

• **Esempio 1.1.3.** Dato un campione casuale da $X \sim N(\mu, \sigma^2)$, il modello statistico dell'Esempio 1.1.1 può essere scritto in maniera equivalente come

$$\mathcal{F}_{\mu, \sigma} = \left\{ f_n : f_n(x_1, \dots, x_n; \mu, \sigma) = \prod_{i=1}^n \frac{1}{\sigma} \phi\left(\frac{x_i - \mu}{\sigma}\right) \right\}. \quad \square$$

• **Esempio 1.1.4.** Dato un campione casuale da $X \sim E(0, \sigma)$, il modello statistico dell'Esempio 1.1.2 può essere scritto in maniera equivalente come

$$\mathcal{F}_\sigma = \left\{ f_n : f_n(x_1, \dots, x_n; \sigma) = \prod_{i=1}^n \frac{1}{\sigma} e^{-\frac{x_i}{\sigma}} \mathbf{1}_{[0, \infty)}(x_i) \right\}. \quad \square$$

1.1.3. Lo spazio campionario. Connesso al concetto di modello statistico vi è quello di spazio campionario, ovvero l'insieme \mathcal{C}_n di tutte le possibili realizzazioni del campione compatibili con il modello statistico stesso. Se $\mathcal{S}_{n, \theta}$ è il supporto di f_n , allora lo spazio campionario è dato da

$$\mathcal{C}_n = \bigcup_{\theta \in \Theta} \mathcal{S}_{n,\theta}.$$

Nel caso di un campionamento casuale si ha $\mathcal{S}_{n,\theta} = \mathcal{S}_\theta \times \cdots \times \mathcal{S}_\theta$, dove \mathcal{S}_θ è il supporto di f .

• **Esempio 1.1.5.** Dato un campione casuale da $X \sim N(\mu, 1)$, il modello statistico dipende dal parametro $\theta = \mu$ con $\Theta = \mathbb{R}$, ovvero

$$\mathcal{F}_\mu = \left\{ f_n : f_n(x_1, \dots, x_n; \mu) = \prod_{i=1}^n \phi(x_i - \mu) \right\}.$$

Dal momento che $\mathcal{S}_\mu = \mathbb{R}$, allora si ha $\mathcal{S}_{n,\mu} = \mathbb{R} \times \cdots \times \mathbb{R} = \mathbb{R}^n$. Di conseguenza, lo spazio campionario risulta

$$\mathcal{C}_n = \bigcup_{\mu \in \mathbb{R}} \mathcal{S}_{n,\mu} = \mathbb{R}^n. \quad \square$$

• **Esempio 1.1.6.** Dato un campione casuale da $X \sim E(\lambda, 1)$, il modello statistico dipende dal parametro $\theta = \lambda$ con $\Theta = \mathbb{R}$, ovvero

$$\mathcal{F}_\lambda = \left\{ f_n : f_n(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n e^{-(x_i - \lambda)} \mathbf{1}_{[0, \infty)}(x_i - \lambda) \right\}.$$

Dal momento che $\mathcal{S}_\lambda = (\lambda, \infty)$, allora si ha $\mathcal{S}_{n,\lambda} = (\lambda, \infty) \times \cdots \times (\lambda, \infty)$. Di conseguenza lo spazio campionario risulta

$$\mathcal{C}_n = \bigcup_{\lambda \in \mathbb{R}} \mathcal{S}_{n,\lambda} = \mathbb{R}^n. \quad \square$$

1.1.4. La parametrizzazione. Un modello statistico può avere varie formulazioni equivalenti, ovvero si possono avere diverse parametrizzazioni. Sia $\mathbf{g} : \Theta \rightarrow \Gamma$ con $\Gamma \subseteq \mathbb{R}^k$ una funzione vettoriale biunivoca. Se $\boldsymbol{\gamma} = \mathbf{g}(\boldsymbol{\theta})$ con $\boldsymbol{\gamma} \in \Gamma$, il modello statistico \mathcal{F}_θ può essere espresso come

$$\mathcal{F}_\theta = \{f_n : f_n(x_1, \dots, x_n; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\} = \{f_n : f_n(x_1, \dots, x_n; \mathbf{g}^{-1}(\boldsymbol{\gamma})), \boldsymbol{\gamma} \in \Gamma\} = \mathcal{G}_\gamma,$$

dove $\boldsymbol{\gamma} \in \Gamma$ e Γ è l'immagine di Θ attraverso \mathbf{g} , ovvero \mathcal{G}_γ e \mathcal{F}_θ sono modelli statistici equivalenti.

La preferenza per una certa parametrizzazione è legata all'interpretazione statistica del parametro scelto. In ogni caso sussiste una certa arbitrarietà nella parametrizzazione, che può dipendere anche dalla facilità di trattazione matematica del modello statistico selezionato.

• **Esempio 1.1.7.** Dato un campione casuale da $X \sim N(0, \sigma^2)$, il modello statistico dipende dal parametro $\theta = \sigma$ con $\Theta = \mathbb{R}^+$, ovvero

$$\mathcal{F}_\sigma = \left\{ f_n : f_n(x_1, \dots, x_n; \sigma) = \prod_{i=1}^n \frac{1}{\sigma} \phi\left(\frac{x_i}{\sigma}\right) \right\}.$$

La trasformazione $v = g(\sigma) = \sigma^2$ per $\sigma > 0$ è biunivoca. Quindi, il modello statistico può essere riparametrizzato in termini della varianza piuttosto che dello scarto quadratico medio, ovvero

$$\mathcal{G}_v = \left\{ f_n : f_n(x_1, \dots, x_n; v) = \prod_{i=1}^n \frac{1}{\sqrt{v}} \phi\left(\frac{x_i}{\sqrt{v}}\right) \right\}.$$

Generalizzando questo esempio, quando si dispone di un campione casuale da $X \sim N(\mu, \sigma^2)$, anche il modello statistico dell'Esempio 1.1.3 può essere scritto in termini della media e della varianza, ovvero

$$\mathcal{G}_{\mu, \nu} = \left\{ f_n : f_n(x_1, \dots, x_n; \mu, \nu) = \prod_{i=1}^n \frac{1}{\sqrt{\nu}} \phi\left(\frac{x_i - \mu}{\sqrt{\nu}}\right) \right\}.$$

Per questo modello statistico è conveniente considerare la precedente parametrizzazione, in quanto permette di semplificare la trattazione matematica. \square

• **Esempio 1.1.8.** Si supponga che il campione (y_1, \dots, y_n) sia la determinazione di un vettore di variabili casuali (Y_1, \dots, Y_n) con componenti indipendenti e tali che $Y_i \sim N(\alpha + \beta w_i, \sigma^2)$. Si assuma inoltre che le componenti del vettore (w_1, \dots, w_n) siano quantità note e fisse. Il campione non è casuale, in quanto le Y_i non sono ugualmente distribuite. Il modello statistico risultante dipende dal vettore di parametri $\theta = (\alpha, \beta, \sigma)^T$ con $\Theta = \mathbb{R}^2 \times \mathbb{R}^+$ ed è dato da

$$\mathcal{F}_{\alpha, \beta, \sigma} = \left\{ f_n : f_n(y_1, \dots, y_n; \alpha, \beta, \sigma) = \prod_{i=1}^n \frac{1}{\sigma} \phi\left(\frac{y_i - \alpha - \beta w_i}{\sigma}\right) \right\}.$$

Si può porre

$$E[Y_i] = \alpha + \beta w_i = \alpha + \beta \bar{w} + \beta(w_i - \bar{w}) = a + b z_i,$$

dove $a = \alpha + \beta \bar{w}$, $b = \beta$ e $z_i = w_i - \bar{w}$, mentre $\bar{w} = n^{-1} \sum_{i=1}^n w_i$. Questa riformulazione di $E[Y_i]$ è conveniente, dal momento che le nuove quantità fisse (z_1, \dots, z_n) sono tali che $\bar{z} = n^{-1} \sum_{i=1}^n z_i = 0$. Inoltre, la trasformazione $(a, b, \nu) = (\alpha + \beta \bar{w}, \beta, \sigma^2)$ è biunivoca, e quindi il modello può essere riparametrizzato come segue

$$\mathcal{G}_{a, b, \nu} = \left\{ f_n : f_n(y_1, \dots, y_n; a, b, \nu) = \prod_{i=1}^n \frac{1}{\sqrt{\nu}} \phi\left(\frac{y_i - a - b z_i}{\sqrt{\nu}}\right) \right\}.$$

Questo modello statistico è fondamentale e costituisce il cosiddetto modello di regressione lineare (vedi Seber, 1977). \square

1.2. Le statistiche

1.2.1. La definizione di statistica. Piuttosto che considerare il campione originale nella sua interezza può essere conveniente talvolta sintetizzarlo in qualche maniera. Dato un modello statistico \mathcal{F}_θ , si dice statistica una trasformazione misurabile

$$T : \mathcal{C}_n \rightarrow \mathbb{R}$$

che non dipende da θ . Un valore $t = T(x_1, \dots, x_n)$ è detto realizzazione campionaria di T relativa al campione (x_1, \dots, x_n) . Prima del rilevamento la statistica $T = T(X_1, \dots, X_n)$ è una variabile casuale, la cui distribuzione dipende dal modello statistico considerato.

La statistica induce una partizione dello spazio campionario in insiemi del tipo

$$A(t) = \{(x_1, \dots, x_n) : (x_1, \dots, x_n) \in \mathcal{C}_n, T(x_1, \dots, x_n) = t\},$$

ovvero in insiemi di campioni per cui il relativo valore della statistica risulta costante.

La definizione di statistica può essere estesa ad un vettore di statistiche \mathbf{T} , dove

$$\mathbf{T} : \mathcal{C}_n \rightarrow \mathbb{R}^k$$

è un vettore di trasformazioni misurabili a k componenti che non dipende da θ .

• **Esempio 1.2.1.** Dato un campione casuale da $X \sim N(\mu, v)$, si consideri la statistica

$$T = \sum_{i=1}^n X_i .$$

Dal momento che $X_i \sim N(\mu, v)$, tenendo presente che T è una somma di variabili casuali indipendenti, in base alla proprietà ii) della §A.1.2 si ottiene che $T \sim N(n\mu, nv)$. \square

• **Esempio 1.2.2.** Dato un campione casuale $X \sim E(0, \sigma)$, si consideri di nuovo la statistica T dell'Esempio 1.2.1. Dal momento che $X_i \sim E(0, \sigma)$, tenendo presente che T è una somma di variabili casuali indipendenti, in base alla proprietà ii) della §A.1.4 si ha $T \sim G(0, \sigma; n)$. \square

1.2.2. La media campionaria. Dato un modello statistico \mathcal{F}_θ relativo ad un campionamento casuale da una variabile casuale X tale che $\text{Var}[X] = v < \infty$, si consideri la statistica media campionaria, la cui realizzazione campionaria è data da

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i .$$

Da un punto di vista probabilistico, la media campionaria è data dalla variabile casuale

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i .$$

Dal momento che μ e v sono finiti, si ha

$$E[\bar{X}] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \mu ,$$

e, tenendo presente l'indipendenza delle variabili casuali (X_1, \dots, X_n) ,

$$\text{Var}[\bar{X}] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{v}{n} .$$

• **Esempio 1.2.3.** Dato un campione casuale da $X \sim N(\mu, v)$, si consideri la media campionaria \bar{X} . Dal momento che $X_i \sim N(\mu, v)$, tenendo presente che \bar{X} è una combinazione lineare di variabili casuali indipendenti, in base alle proprietà ii) della §A.1.2 si ha $\bar{X} \sim N(\mu, v/n)$. Risultano quindi verificati i risultati di questa sezione, in quanto $E[\bar{X}] = \mu$ e $\text{Var}[\bar{X}] = v/n$. \square

• **Esempio 1.2.4.** Dato un campione casuale da $X \sim E(0, \sigma)$, si consideri la media campionaria \bar{X} . Dal momento che $X_i \sim E(0, \sigma)$, tenendo presente che \bar{X} è una combinazione lineare di variabili casuali indipendenti, in base alle proprietà ii) della §A.1.4 si ha $\bar{X} \sim G(0, \sigma/n; n)$. Inoltre, per la proprietà i) della §A.1.4, si ha $E[\bar{X}] = \sigma$ e $\text{Var}[\bar{X}] = \sigma^2/n$. Dunque, risultano verificati i risultati di questa sezione, in quanto per la $E(0, \sigma)$ si ha $E[X] = \sigma$ e $\text{Var}[X] = \sigma^2$. \square

1.2.3. La varianza campionaria. Dato un modello statistico \mathcal{F}_θ relativo ad un campionamento casuale da una variabile casuale X tale che $\text{Var}[X] = v < \infty$, si consideri la statistica varianza campionaria, la cui realizzazione campionaria è data da

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 .$$

Da un punto di vista probabilistico, la varianza campionaria è data dalla variabile casuale

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Al fine di derivare le proprietà della varianza campionaria, risulta conveniente esprimere s^2 nel modo seguente

$$\begin{aligned} s^2 &= \frac{1}{n} \sum_{i=1}^n ((x_i - \mu) - (\bar{x} - \mu))^2 \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - \frac{2}{n} (\bar{x} - \mu) \sum_{i=1}^n (x_i - \mu) + (\bar{x} - \mu)^2 \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - (\bar{x} - \mu)^2. \end{aligned}$$

Di conseguenza, la varianza campionaria può essere scritta come

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - (\bar{X} - \mu)^2.$$

Sulla base della precedente relazione, si ottiene che

$$E[S^2] = \frac{1}{n} \sum_{i=1}^n E[(X_i - \mu)^2] - E[(\bar{X} - \mu)^2] = \frac{1}{n} \sum_{i=1}^n \text{Var}[X_i] - \text{Var}[\bar{X}] = \frac{n-1}{n} v.$$

Piuttosto che la varianza campionaria S^2 si adotta spesso la cosiddetta varianza campionaria corretta data da

$$S_c^2 = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

La varianza campionaria corretta S_c^2 viene spesso preferita a S^2 , dal momento che gode della proprietà $E[S_c^2] = v$.

• **Esempio 1.2.5.** Dato un campione casuale da $X \sim N(\mu, v)$, si vuole determinare la distribuzione della varianza campionaria corretta S_c^2 . Si consideri le cosiddette statistiche scarto

$$D_i = X_i - \bar{X} = \sum_{j=1}^n c_{ij} X_j,$$

dove

$$c_{ij} = \begin{cases} -\frac{1}{n} & j \neq i \\ 1 - \frac{1}{n} & j = i. \end{cases}$$

Le statistiche scarto sono variabili casuali Normali, essendo combinazioni lineari di variabili casuali Normali indipendenti in base alla proprietà *ii*) della §A.1.2. Inoltre, dal momento che

$$\bar{X} = \sum_{j=1}^n d_j X_j,$$

dove

$$d_j = \frac{1}{n},$$

anche \bar{X} è una combinazione lineare delle medesime variabili casuali Normali indipendenti. Infine, si ha

$$\sum_{j=1}^n c_{ij}d_j = -\frac{n-1}{n^2} + \frac{1}{n} \left(1 - \frac{1}{n}\right) = 0$$

e dunque D_i e \bar{X} sono indipendenti per la proprietà *iii*) della §A.1.2. Di conseguenza, essendo

$$S_c^2 = \frac{1}{n-1} \sum_{i=1}^n D_i^2$$

una trasformata delle statistiche scarto, anche S_c^2 e \bar{X} sono indipendenti. Si può dimostrare che questo risultato è valido solo per questo particolare modello statistico (vedi Wilks, 1962). Dividendo ora ambo i membri dell'espressione della varianza corretta per v e tenendo presente la relazione ottenuta per S^2 , si ottiene la seguente relazione

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{v} = (n-1) \frac{S_c^2}{v} + \frac{(\bar{X} - \mu)^2}{v/n}.$$

Le variabili casuali $X_i \sim N(\mu, v)$ sono indipendenti e dunque risulta

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{v} \sim \chi_n^2$$

(vedi §A.3.2). Inoltre nell'Esempio 1.2.3 è stato verificato che $\bar{X} \sim N(\mu, v/n)$, da cui

$$\frac{(\bar{X} - \mu)^2}{v/n} \sim \chi_1^2.$$

Dal momento che S_c^2 e \bar{X} sono indipendenti, per la proprietà *i*) della §A.3.2, in base alle relazioni precedenti si deve concludere che

$$(n-1) \frac{S_c^2}{v} \sim \chi_{n-1}^2.$$

Dunque, si ha

$$E\left[(n-1) \frac{S_c^2}{v}\right] = n-1, \text{Var}\left[(n-1) \frac{S_c^2}{v}\right] = 2(n-1)$$

(vedi §A.3.2), da cui

$$E[S_c^2] = v, \text{Var}[S_c^2] = \frac{2v}{n-1}.$$

Inoltre, anche S^2 è indipendente da \bar{X} , mentre

$$\frac{nS^2}{v} \sim \chi_{n-1}^2,$$

da cui

$$E[S^2] = \frac{n-1}{n} v, \text{Var}[S^2] = \frac{2(n-1)}{n^2} v^2. \quad \square$$

1.2.4. La statistica ordinata. Dato un modello statistico \mathcal{F}_θ relativo ad un campionamento casuale da una variabile casuale X , si consideri il vettore $(x_{(1)}, \dots, x_{(n)})$ i cui elementi sono gli elementi ordinati del campione (x_1, \dots, x_n) . Da un punto di vista probabilistico il vettore $(x_{(1)}, \dots, x_{(n)})$ è la realizzazione campionaria del vettore di statistiche $(X_{(1)}, \dots, X_{(n)})$, detto statistica ordinata. Inoltre, la statistica $X_{(i)}$ è detta i -esima statistica ordinata.

La mediana campionaria è funzione della statistica ordinata. In effetti, se n è dispari, ovvero $n = 2l + 1$ dove l è un intero, la mediana campionaria è definita come $\tilde{X} = X_{(l+1)}$. Alternativamente, se n è pari, ovvero $n = 2l$, allora la mediana campionaria \tilde{X} viene usualmente definita come $\tilde{X} = (X_{(l)} + X_{(l+1)})/2$. Una realizzazione della mediana campionaria viene indicata con \tilde{x} .

Se la variabile casuale X da cui si sta campionando è continua, allora la distribuzione della statistica ordinata può essere ottenuta in una forma semplice. La funzione di densità congiunta di $(X_{(1)}, \dots, X_{(n)})$ risulta (vedi Wilks, 1962)

$$f_{(X_{(1)}, \dots, X_{(n)})}(x_{(1)}, \dots, x_{(n)}) = n! \mathbf{1}_D(x_{(1)}, \dots, x_{(n)}) \prod_{i=1}^n f(x_{(i)}; \theta),$$

dove $D = \{(x_{(1)}, \dots, x_{(n)}) : -\infty < x_{(1)} < \dots < x_{(n)} < \infty\}$, mentre la funzione di densità marginale di $X_{(i)}$ è data da

$$f_{X_{(i)}}(x_{(i)}) = n \binom{n-1}{i-1} F(x_{(i)}; \theta)^{i-1} (1 - F(x_{(i)}; \theta))^{n-i} f(x_{(i)}; \theta).$$

• **Esempio 1.2.6.** Dato un campione casuale da $X \sim U(0, \delta)$, il modello statistico dipende dal parametro $\theta = \delta$ con $\Theta = \mathbb{R}^+$, ovvero

$$\mathcal{F}_\delta = \left\{ f_n : f_n(x_1, \dots, x_n; \delta) = \prod_{i=1}^n \frac{1}{\delta} \mathbf{1}_{[0,1]} \left(\frac{x_i}{\delta} \right) \right\}.$$

Si consideri la massima statistica ordinata, ovvero $X_{(n)}$. Dal momento che il campione è casuale e la variabile casuale da cui si sta campionando è continua, la funzione di densità di $X_{(n)}$ risulta

$$f_{X_{(n)}}(x_{(n)}) = \frac{n}{\delta} \left(\frac{x_{(n)}}{\delta} \right)^{n-1} \mathbf{1}_{[0,1]} \left(\frac{x_{(n)}}{\delta} \right),$$

ovvero $X_{(n)} \sim Be(0, \delta; n, 1)$. Analogamente, $X_{(1)} \sim Be(0, \delta; 1, n)$. □

• **Esempio 1.2.7.** Dato un campione casuale da $X \sim U(0, \delta)$, supponendo la numerosità campionaria dispari con $n = 2l + 1$, si vuole determinare la distribuzione della mediana campionaria \tilde{X} . Dal momento che il campione è casuale e la variabile casuale da cui si sta campionando è continua, la funzione di densità di $\tilde{X} = X_{(l+1)}$ risulta

$$f_{\tilde{X}}(\tilde{x}) = \frac{2l+1}{\delta} \binom{2l}{l} \left(\frac{\tilde{x}}{\delta} \right)^l \left(1 - \frac{\tilde{x}}{\delta} \right)^l \mathbf{1}_{[0,1]} \left(\frac{\tilde{x}}{\delta} \right),$$

ovvero $\tilde{X} \sim Be(0, \delta; l+1, l+1)$. □

1.3. La verosimiglianza statistica

1.3.1. La funzione di verosimiglianza. Dato un modello statistico \mathcal{F}_θ , una volta che il campione (x_1, \dots, x_n) è stato osservato, la quantità $f_n(x_1, \dots, x_n; \theta)$ è funzione solo di θ . Questa funzione

rappresenta la (densità di) probabilità di osservare *a priori* esattamente il campione (x_1, \dots, x_n) che è stato estratto e contiene tutta l'informazione relativa al campione stesso.

Anche una generica funzione del tipo $c f_n(x_1, \dots, x_n; \theta)$, con $c = c(x_1, \dots, x_n)$ costante positiva che non dipende da θ , offre una informazione equivalente a quella fornita da $f_n(x_1, \dots, x_n; \theta)$. Infatti, se si deve scegliere fra due valori θ' e θ'' in Θ , il criterio fondamentale su cui si basa la preferenza è dato dal rapporto

$$r = \frac{f_n(x_1, \dots, x_n; \theta')}{f_n(x_1, \dots, x_n; \theta'')},$$

che risulta inalterato moltiplicando numeratore e denominatore per una stessa quantità.

Si dice funzione di verosimiglianza (o verosimiglianza) la funzione $L : \Theta \rightarrow \mathbb{R}^+ \cup \{0\}$ data da

$$L(\theta) = L(\theta; x_1, \dots, x_n) = c f_n(x_1, \dots, x_n; \theta), \theta \in \Theta.$$

La notazione $L(\theta; x_1, \dots, x_n)$ viene adottata perchè in alcuni casi si vuole enfatizzare che la verosimiglianza è riferita proprio al campione (x_1, \dots, x_n) . Inoltre, viene considerata anche la funzione di log-verosimiglianza, data da

$$l(\theta) = l(\theta; x_1, \dots, x_n) = \log L(\theta) = \log c + \log f_n(x_1, \dots, x_n; \theta), \theta \in \Theta,$$

con la convenzione che $l(\theta) = -\infty$ se $L(\theta) = 0$.

Nel caso di un campionamento casuale la funzione di verosimiglianza risulta

$$L(\theta) = L(\theta; x_1, \dots, x_n) = c \prod_{i=1}^n f(x_i; \theta), \theta \in \Theta,$$

mentre la funzione di log-verosimiglianza si riduce a

$$l(\theta) = l(\theta; x_1, \dots, x_n) = \log c + \sum_{i=1}^n \log f(x_i; \theta), \theta \in \Theta.$$

È sempre opportuno rappresentare graficamente la verosimiglianza o la log-verosimiglianza, almeno quando $k = 1$ o $k = 2$. Se $k > 2$, è possibile ottenere una rappresentazione grafica parziale della verosimiglianza attraverso il concetto di verosimiglianza profilo. Questo argomento verrà analizzato nella §3.1.1. Dal momento che la scelta di c è arbitraria, per la rappresentazione grafica è conveniente standardizzare la verosimiglianza rispetto al suo massimo (se questo è finito), ovvero scegliere $c = 1/\max_{\theta \in \Theta} L(\theta)$, in modo tale da ottenere una verosimiglianza normata in $[0, 1]$.

Di seguito viene data una serie di esempi che considerano una gamma di funzioni di verosimiglianza al variare del modello statistico.

• **Esempio 1.3.1.** Dato un campione casuale da $X \sim N(\mu, 1)$, la verosimiglianza risulta

$$L(\mu) = c \prod_{i=1}^n (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}(x_i - \mu)^2} = c e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2} = c e^{-\frac{n}{2}(s^2 + (\bar{x} - \mu)^2)}, \mu \in \mathbb{R},$$

dove nell'ultimo passaggio si è adoperato la relazione ottenuta nella §1.2.3. Il grafico della verosimiglianza standardizzata per $n = 5$, $\bar{x} = 1$ e $s^2 = 2$ è riportato in Figura 1.3.1.

Nella precedente espressione è stato effettuato un abuso di notazione, nel senso che la costante c non è la stessa fra i vari passaggi algebrici. Si adopera questa convenzione per indicare che la costante c contiene una quantità che non dipende dai parametri e nel seguito verrà adottata senza ulteriore commento.

Inoltre, la verosimiglianza è stata espressa in funzione delle realizzazioni campionarie delle statistiche \bar{X} e S^2 , al fine di ottenere un'espressione più compatta. Questo fatto ha importanti risvolti teorici, come sarà visto nei successivi capitoli. Infine, la log-verosimiglianza risulta

$$l(\mu) = \log c - \frac{n}{2} (s^2 + (\bar{x} - \mu)^2), \mu \in \mathbb{R}. \quad \square$$

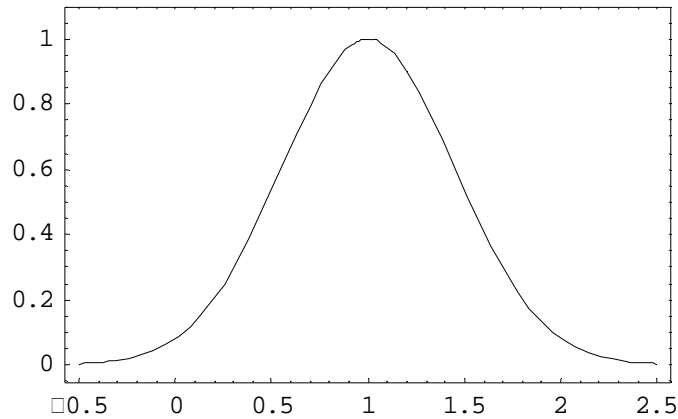


Figura 1.3.1. Verosimiglianza standardizzata per $n = 5$, $\bar{x} = 1$ e $s^2 = 2$.

• **Esempio 1.3.2.** Generalizzando l'esempio precedente, dato un campione casuale da $X \sim N(\mu, \nu)$, tenendo presente i risultati della §1.2.3, si ottiene la seguente verosimiglianza

$$L(\mu, \nu) = c \prod_{i=1}^n (2\pi\nu)^{-\frac{1}{2}} e^{-\frac{1}{2\nu}(x_i - \mu)^2} = c \nu^{-\frac{n}{2}} e^{-\frac{1}{2\nu} \sum_{i=1}^n (x_i - \mu)^2} = c \nu^{-\frac{n}{2}} e^{-\frac{n}{2\nu} (s^2 + (\bar{x} - \mu)^2)},$$

dove $(\mu, \nu) \in \mathbb{R} \times \mathbb{R}^+$. Il grafico della verosimiglianza standardizzata per $n = 5$, $\bar{x} = 1$ e $s^2 = 2$ è riportato in Figura 1.3.2. Infine, la log-verosimiglianza risulta

$$l(\mu, \nu) = \log c - \frac{n}{2} \log \nu - \frac{n}{2\nu} (s^2 + (\bar{x} - \mu)^2), (\mu, \nu) \in \mathbb{R} \times \mathbb{R}^+. \quad \square$$

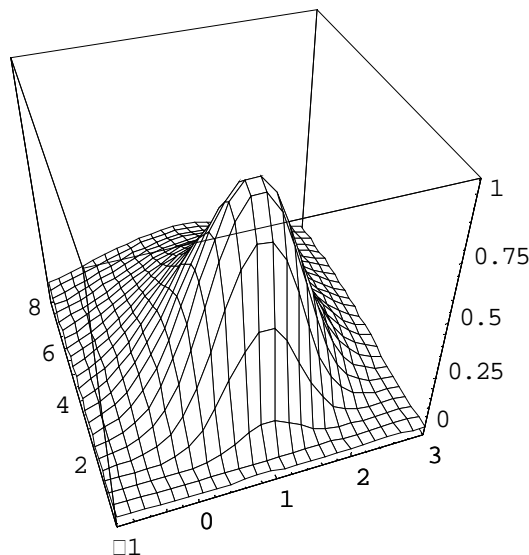


Figura 1.3.2. Verosimiglianza standardizzata per $n = 5$, $\bar{x} = 1$ e $s^2 = 2$.

• **Esempio 1.3.3.** Dato un campione casuale da $X \sim E(0, \sigma)$, si ha la seguente verosimiglianza

$$L(\sigma) = c \prod_{i=1}^n \frac{1}{\sigma} e^{-\frac{x_i}{\sigma}} \mathbf{1}_{[0, \infty)}(x_i) = c \sigma^{-n} e^{-\frac{n\bar{x}}{\sigma}}, \sigma \in \mathbb{R}^+.$$

La verosimiglianza è stata espressa in funzione della realizzazione della statistica \bar{X} . Il grafico della verosimiglianza standardizzata per $n = 5$ e $\bar{x} = 1$ è riportato in Figura 1.3.3. Infine, la log-verosimiglianza risulta

$$l(\sigma) = \log c - n \log \sigma - \frac{n\bar{x}}{\sigma}, \sigma \in \mathbb{R}^+ . \quad \square$$

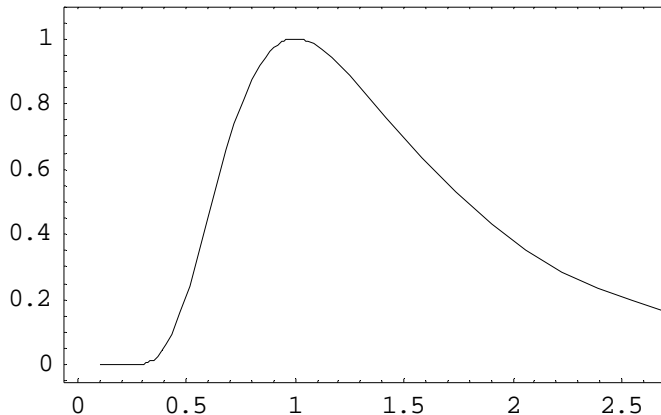


Figura 1.3.3. Verosimiglianza standardizzata per $n = 5$ e $\bar{x} = 1$.

• **Esempio 1.3.4.** Dato un campione casuale da $X \sim U(0, \delta)$, la verosimiglianza risulta

$$L(\delta) = c \prod_{i=1}^n \frac{1}{\delta} \mathbf{1}_{[0,1]} \left(\frac{x_i}{\delta} \right) = c \delta^{-n} \mathbf{1}_{[0,1]} \left(\frac{x_{(n)}}{\delta} \right) = c \delta^{-n} \mathbf{1}_{[x_{(n)}, \infty)}(\delta), \delta \in \mathbb{R}^+,$$

dal momento che $\prod_{i=1}^n \mathbf{1}_{[0,1]}(x_i/\delta) = 1$ quando $\delta \geq x_i$ per ogni i , ovvero quando si ha $\delta \geq x_{(n)}$. La verosimiglianza è stata espressa in funzione della realizzazione della statistica $X_{(n)}$, ovvero la massima statistica ordinata. Il grafico della verosimiglianza standardizzata per $n = 5$ e $x_{(5)} = 1$ è riportato in Figura 1.3.4. Infine, la log-verosimiglianza risulta

$$l(\delta) = \log c - n \log \delta + \log \mathbf{1}_{[x_{(n)}, \infty)}(\delta), \delta \in \mathbb{R}^+ . \quad \square$$

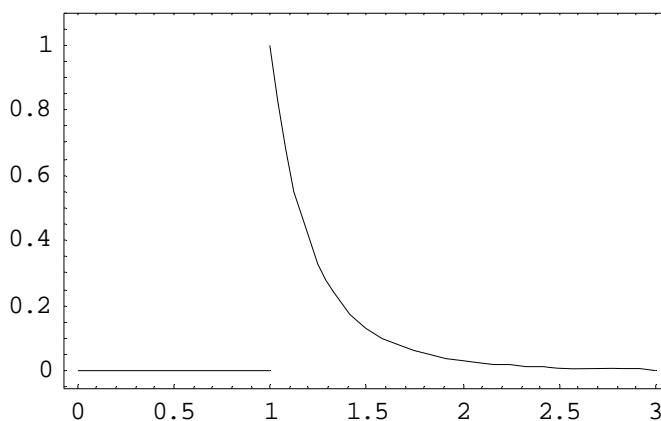


Figura 1.3.4. Verosimiglianza standardizzata per $n = 5$ e $x_{(5)} = 1$.

• **Esempio 1.3.5.** Dato un campione casuale da $X \sim Bi(1, p)$, il modello statistico dipende dal parametro $\theta = p$ con $\Theta = (0, 1)$, ovvero

$$\mathcal{F}_p = \left\{ f_n : f_n(x_1, \dots, x_n; p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \mathbf{1}_{\{0,1\}}(x_i) \right\},$$

da cui si ottiene la seguente verosimiglianza

$$L(p) = c \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \mathbf{1}_{\{0,1\}}(x_i) = c p^{n\bar{x}} (1-p)^{n-n\bar{x}}, \quad p \in (0, 1).$$

La verosimiglianza è stata espressa in funzione della realizzazione della media campionaria \bar{X} . Il grafico della verosimiglianza standardizzata per $n = 5$ e $\bar{x} = 0.2$ è riportato in Figura 1.3.5. Infine, la log-verosimiglianza risulta

$$l(p) = \log c + n\bar{x} \log p + (n - n\bar{x}) \log(1-p), \quad p \in (0, 1). \quad \square$$

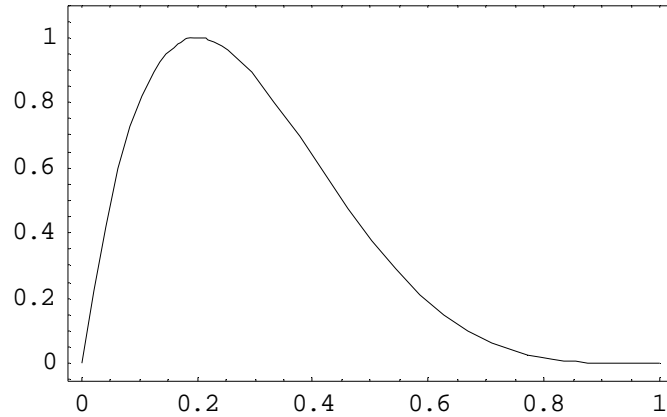


Figura 1.3.5. Verosimiglianza standardizzata per $n = 5$ e $\bar{x} = 0.2$.

• **Esempio 1.3.6.** Dato il modello di regressione lineare dell'Esempio 1.1.8, la verosimiglianza risulta

$$L(a, b, v) = c \prod_{i=1}^n (2\pi v)^{-\frac{1}{2}} e^{-\frac{1}{2v}(y_i - a - bz_i)^2} = c v^{-\frac{n}{2}} e^{-\frac{1}{2v} \sum_{i=1}^n (y_i - a - bz_i)^2}, \quad (a, b, v) \in \mathbb{R}^2 \times \mathbb{R}^+.$$

Siano dunque \bar{y} e s_y^2 la media e la varianza di (y_1, \dots, y_n) , mentre sia

$$s_z^2 = \frac{1}{n} \sum_{i=1}^n z_i^2$$

la varianza di (z_1, \dots, z_n) e

$$s_{zy} = \frac{1}{n} \sum_{i=1}^n z_i y_i$$

la relativa covarianza (tenendo presente dall'Esempio 1.1.8 che $\bar{z} = 0$). Dunque, si ha

$$\begin{aligned} \sum_{i=1}^n (y_i - a - bz_i)^2 &= \sum_{i=1}^n ((y_i - \bar{y}) - bz_i + (\bar{y} - a))^2 \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 + b^2 \sum_{i=1}^n z_i^2 + n(\bar{y} - a)^2 - 2b \sum_{i=1}^n (y_i - \bar{y})z_i \\ &= ns_y^2 + nb^2 s_z^2 + n(\bar{y} - a)^2 - 2nbs_{zy} \\ &= n \left(s_y^2 - \frac{s_{zy}^2}{s_z^2} \right) + n \left(bs_z - \frac{s_{zy}}{s_z} \right)^2 + n(\bar{y} - a)^2, \end{aligned}$$

dove è stato tenuto presente che $\sum_{i=1}^n z_i = 0$ e $\sum_{i=1}^n (y_i - \bar{y}) = 0$. Di conseguenza,

$$L(a, b, v) = cv^{-\frac{n}{2}} e^{-\frac{n}{2v} \left(s_y^2 - \frac{s_{zy}^2}{s_z^2} + \left(bs_z - \frac{s_{zy}}{s_z} \right)^2 + (\bar{y} - a)^2 \right)},$$

con $(a, b, v) \in \mathbb{R}^2 \times \mathbb{R}^+$. La verosimiglianza è stata espressa in funzione delle realizzazioni delle statistiche

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i,$$

e

$$S_y^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

e della statistica

$$S_{zy} = \frac{1}{n} \sum_{i=1}^n z_i Y_i.$$

Infine, la log-verosimiglianza risulta

$$l(a, b, v) = \log c - \frac{n}{2} \log v - \frac{n}{2v} \left(s_y^2 - \frac{s_{zy}^2}{s_z^2} + \left(bs_z - \frac{s_{zy}}{s_z} \right)^2 + (\bar{y} - a)^2 \right),$$

con $(a, b, v) \in \mathbb{R}^2 \times \mathbb{R}^+$. □

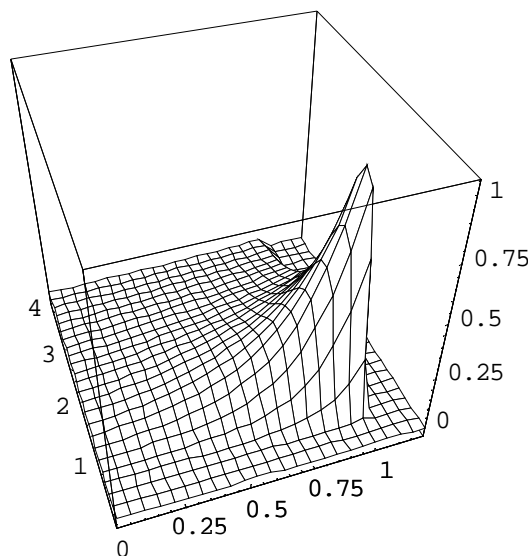


Figura 1.3.6. Verosimiglianza standardizzata per $n = 5$, $\bar{x} = 2$ e $x_{(1)} = 1$.

• **Esempio 1.3.7.** Dato un campione casuale da $X \sim E(\lambda, \sigma)$, il modello statistico dipende dal vettore di parametri $\theta = (\lambda, \sigma)^T$ con $\Theta = \mathbb{R} \times \mathbb{R}^+$, ovvero

$$\mathcal{F}_{\lambda, \sigma} = \left\{ f_n : f_n(x_1, \dots, x_n; \lambda, \sigma) = \prod_{i=1}^n \frac{1}{\sigma} e^{-\frac{x_i - \lambda}{\sigma}} \mathbf{1}_{[0, \infty)}(x_i - \lambda) \right\}$$

e la verosimiglianza risulta

$$\begin{aligned} L(\lambda, \sigma) &= c \prod_{i=1}^n \frac{1}{\sigma} e^{-\frac{x_i - \lambda}{\sigma}} \mathbf{1}_{[0, \infty)}(x_i - \lambda) = c \sigma^{-n} e^{-n \frac{\bar{x} - \lambda}{\sigma}} \mathbf{1}_{[0, \infty)}(x_{(1)} - \lambda) \\ &= c \sigma^{-n} e^{-n \frac{\bar{x} - \lambda}{\sigma}} \mathbf{1}_{(-\infty, x_{(1)}]}(\lambda), \quad (\lambda, \sigma) \in \mathbb{R} \times \mathbb{R}^+, \end{aligned}$$

dove si è tenuto conto che $\prod_{i=1}^n \mathbf{1}_{[0, \infty)}(x_i - \lambda) = 1$ quando $\lambda \leq x_i$ per ogni i , ovvero quando si ha $\lambda \leq x_{(1)}$. La verosimiglianza è stata espressa in funzione delle realizzazioni delle statistiche \bar{X} e $X_{(1)}$. Il grafico della verosimiglianza standardizzata per $n = 5$, $\bar{x} = 2$ e $x_{(1)} = 1$ è riportato in Figura 1.3.6. Infine, la log-verosimiglianza risulta

$$l(\lambda, \sigma) = \log c - n \log \sigma - n \frac{\bar{x} - \lambda}{\sigma} + \log \mathbf{1}_{(-\infty, x_{(1)}]}(\lambda), \quad (\lambda, \sigma) \in \mathbb{R} \times \mathbb{R}^+. \quad \square$$

• **Esempio 1.3.8.** Dato un campione casuale $X \sim U(\lambda, \lambda + \delta)$, il modello statistico dipende dal vettore di parametri $\theta = (\lambda, \delta)^T$ con $\Theta = \mathbb{R}^+ \times \mathbb{R}^+$, ovvero

$$\mathcal{F}_{\lambda, \delta} = \left\{ f_n : f_n(x_1, \dots, x_n; \lambda, \delta) = \prod_{i=1}^n \frac{1}{\delta} \mathbf{1}_{[0, 1]} \left(\frac{x_i - \lambda}{\delta} \right) \right\},$$

da cui

$$\begin{aligned} L(\lambda, \delta) &= c \prod_{i=1}^n \frac{1}{\delta} \mathbf{1}_{[0, 1]} \left(\frac{x_i - \lambda}{\delta} \right) = c \delta^{-n} \mathbf{1}_{[0, 1]} \left(\frac{x_{(1)} - \lambda}{\delta} \right) \mathbf{1}_{[0, 1]} \left(\frac{x_{(n)} - \lambda}{\delta} \right) \\ &= c \delta^{-n} \mathbf{1}_{(-\infty, x_{(1)}]}(\lambda) \mathbf{1}_{[x_{(n)}, \infty)}(\lambda + \delta), \quad (\lambda, \delta) \in \mathbb{R} \times \mathbb{R}^+, \end{aligned}$$

dal momento che $\prod_{i=1}^n \mathbf{1}_{[0, 1]}((x_i - \lambda)/\delta) = 1$ quando $\lambda \leq x_i \leq \lambda + \delta$ per ogni i , ovvero quando si ha $\lambda \leq x_{(1)}$ e $\lambda + \delta \geq x_{(n)}$. La verosimiglianza è stata espressa in funzione delle realizzazioni delle statistiche $X_{(1)}$ e $X_{(n)}$. Il grafico della verosimiglianza standardizzata per $n = 5$, $x_{(1)} = 1$ e $x_{(5)} = 2$ è riportato in Figura 1.3.7. Infine, la log-verosimiglianza risulta

$$l(\lambda, \delta) = \log c - n \log \delta + \log \mathbf{1}_{(-\infty, x_{(1)}]}(\lambda) + \log \mathbf{1}_{[x_{(n)}, \infty)}(\lambda + \delta), \quad (\lambda, \delta) \in \mathbb{R} \times \mathbb{R}^+. \quad \square$$

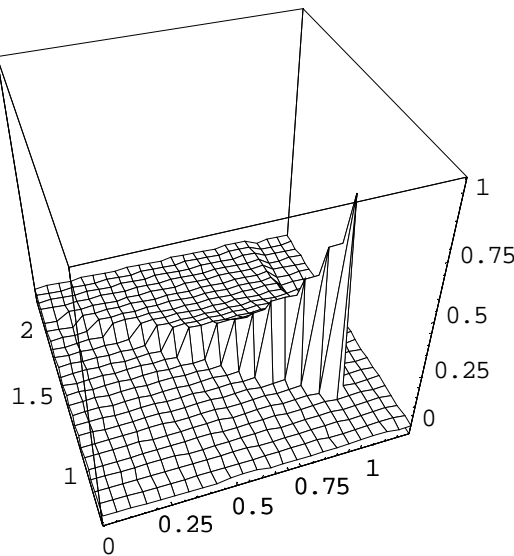


Figura 1.3.7. Verosimiglianza standardizzata per $n = 5$, $x_{(1)} = 1$ e $x_{(5)} = 2$.

• **Esempio 1.3.9.** Dato un campione casuale da $X \sim Po(\mu)$, il modello statistico dipende dal parametro $\theta = \mu$ con $\Theta = \mathbb{R}^+$, ovvero

$$\mathcal{F}_\mu = \left\{ f_n : f_n(x_1, \dots, x_n; \mu) = \prod_{i=1}^n e^{-\mu} \frac{\mu^{x_i}}{x_i!} \mathbf{1}_{\{0,1,\dots\}}(x_i) \right\}.$$

Di conseguenza, la funzione di verosimiglianza risulta

$$L(\mu) = c \prod_{i=1}^n e^{-\mu} \frac{\mu^{x_i}}{x_i!} \mathbf{1}_{\{0,1,\dots\}}(x_i) = c e^{-n\mu} \mu^{n\bar{x}}, \mu \in \mathbb{R}^+.$$

La verosimiglianza è stata espressa in funzione della realizzazione della media campionaria \bar{X} . Il grafico della verosimiglianza standardizzata per $n = 5$ e $\bar{x} = 1$ è riportato in Figura 1.3.8. Infine, la log-verosimiglianza risulta

$$l(\mu) = \log c - n\mu + n\bar{x} \log \mu, \mu \in \mathbb{R}^+. \quad \square$$

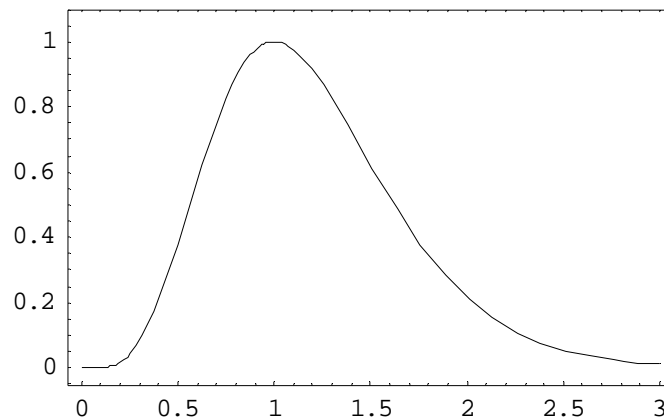


Figura 1.3.8. Verosimiglianza standardizzata per $n = 5$ e $\bar{x} = 1$.

• **Esempio 1.3.10.** Dato un campione casuale da $X \sim L(\mu, 1)$, il modello statistico dipende dal parametro $\theta = \mu$ con $\Theta = \mathbb{R}$, ovvero

$$\mathcal{F}_\mu = \left\{ f_n : f_n(x_1, \dots, x_n; \mu) = \prod_{i=1}^n \frac{1}{2} e^{-|x_i - \mu|} \right\}.$$

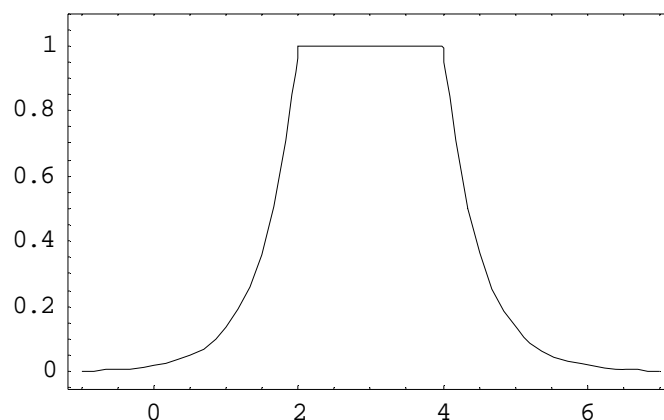


Figura 1.3.9. Verosimiglianza standardizzata per $n = 4$ e $(x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}) = (-1, 2, 4, 9)$.

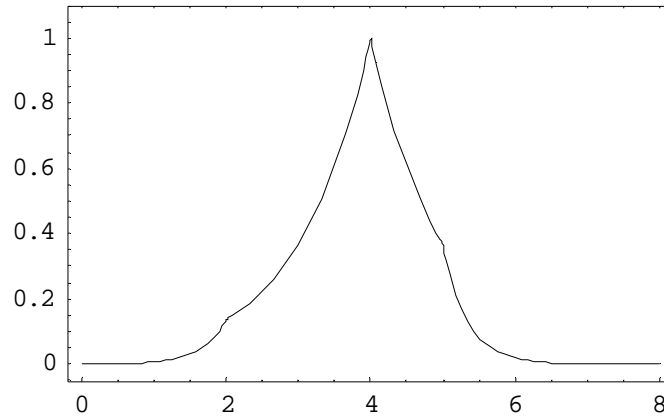


Figura 1.3.10. Verosimiglianza standardizzata per $n = 5$
e $(x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}, x_{(5)}) = (-1, 2, 4, 5, 9)$.

In questo caso si ha la seguente verosimiglianza

$$L(\mu) = c e^{-\sum_{i=1}^n |x_i - \mu|} = c e^{-\sum_{i=1}^n |x_{(i)} - \mu|}, \mu \in \mathbb{R}.$$

La verosimiglianza è stata espressa in funzione della realizzazione della statistica ordinata $(X_{(1)}, \dots, X_{(n)})$. Nelle Figure 1.3.9 e 1.3.10 è riportato il grafico della verosimiglianza standardizzata per una numerosità pari ed una dispari. Infine, la log-verosimiglianza risulta

$$l(\mu) = \log c - \sum_{i=1}^n |x_i - \mu| = \log c - \sum_{i=1}^n |x_{(i)} - \mu|, \mu \in \mathbb{R}. \quad \square$$

• **Esempio 1.3.11.** Si consideri la variabile casuale discreta X con funzione di probabilità

$$f_X(x) = \frac{1}{N} \mathbf{1}_{\{1, \dots, N\}}(x),$$

dove $N \in \mathbb{N}^+$. Questa variabile casuale è la versione discreta di una Uniforme. Dato un campione casuale, il modello statistico dipende dal parametro $\theta = N$ con $\Theta = \mathbb{N}^+$, ovvero

$$\mathcal{F}_N = \left\{ f_n : f_n(x_1, \dots, x_n; N) = \prod_{i=1}^n \frac{1}{N} \mathbf{1}_{\{1, \dots, N\}}(x_i) \right\}.$$

Di conseguenza, la funzione di verosimiglianza risulta

$$L(N) = c \prod_{i=1}^n \frac{1}{N} \mathbf{1}_{\{1, \dots, N\}}(x_i) = c N^{-n} \mathbf{1}_{\{1, \dots, N\}}(x_{(n)}) = c N^{-n} \mathbf{1}_{\{x_{(n)}, x_{(n)+1}, \dots\}}(N),$$

con $N \in \mathbb{N}^+$, e dove si è tenuto conto che $\prod_{i=1}^n \mathbf{1}_{\{1, \dots, N\}}(x_i) = 1$ quando $N \geq x_i$ per ogni i , ovvero quando si ha $N \geq x_{(n)}$. La verosimiglianza è definita solo sugli interi positivi, ovvero lo spazio parametrico. Inoltre, la verosimiglianza è stata espressa in funzione della realizzazione della statistica $X_{(n)}$. Il grafico della verosimiglianza standardizzata per $n = 4$ e $x_{(4)} = 2$ è riportato in Figura 1.3.11. Infine, la log-verosimiglianza risulta

$$l(N) = \log c - n \log N + \log \mathbf{1}_{\{x_{(n)}, x_{(n)+1}, \dots\}}(N), N \in \mathbb{N}^+. \quad \square$$

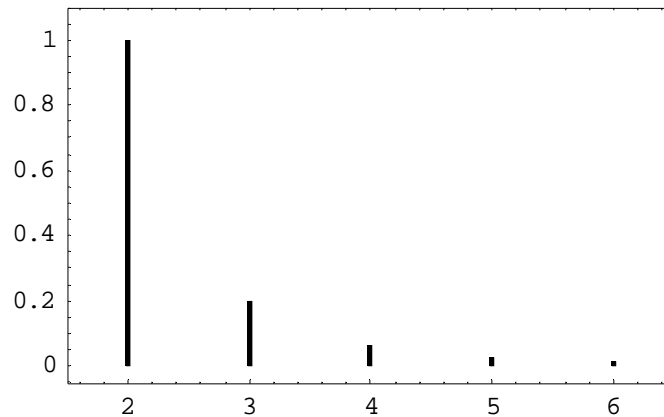


Figura 1.3.11. Verosimiglianza standardizzata per $n = 4$ e $x_{(4)} = 2$.

• **Esempio 1.3.12.** Si consideri la variabile casuale continua X con funzione di densità

$$f_X(x) = \frac{2x}{p} \mathbf{1}_{[0,p]}(x) + \frac{2(1-x)}{1-p} \mathbf{1}_{(p,1]}(x),$$

dove $p \in (0, 1)$. Questa variabile casuale è detta anche Triangolare. In questo caso, dato un campione casuale, il modello statistico dipende dal parametro $\theta = p$ con $\Theta = (0, 1)$, ovvero

$$\mathcal{F}_p = \left\{ f_n : f_n(x_1, \dots, x_n; p) = \prod_{i=1}^n \left(\frac{2x_i}{p} \mathbf{1}_{[0,p]}(x_i) + \frac{2(1-x_i)}{1-p} \mathbf{1}_{(p,1]}(x_i) \right) \right\}.$$

Di conseguenza, la verosimiglianza risulta

$$\begin{aligned} L(p) &= c \prod_{i=1}^n \left(\frac{2x_i}{p} \mathbf{1}_{[0,p]}(x_i) + \frac{2(1-x_i)}{1-p} \mathbf{1}_{(p,1]}(x_i) \right) \\ &= c \prod_{i=1}^n \left(\frac{2x_{(i)}}{p} \mathbf{1}_{[0,p]}(x_{(i)}) + \frac{2(1-x_{(i)})}{1-p} \mathbf{1}_{(p,1]}(x_{(i)}) \right). \end{aligned}$$

La verosimiglianza è stata espressa in funzione della realizzazione della statistica ordinata $(X_{(1)}, \dots, X_{(n)})$. Dalla Figura 1.3.12, si può notare la struttura complessa che può assumere la verosimiglianza per questo modello. \square

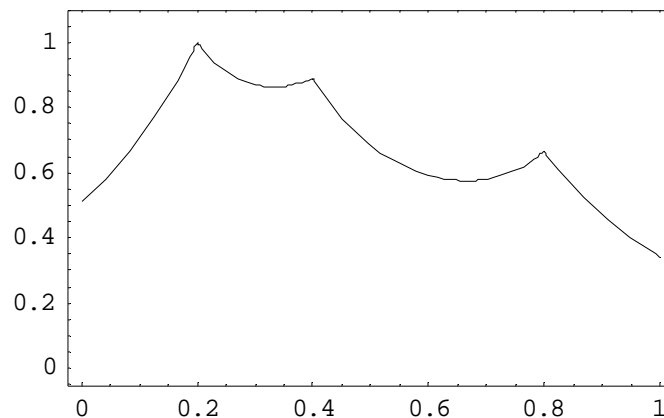


Figura 1.3.12. Verosimiglianza standardizzata per $n = 3$ e $(x_{(1)}, x_{(2)}, x_{(3)}) = (0.2, 0.4, 0.8)$.

Capitolo 2

La stima per punti

2.1. Introduzione

2.1.1. La definizione di stimatore. Una volta scelto il modello statistico \mathcal{F}_θ , un primo obiettivo dell'inferenza è quello di selezionare un valore di θ sulla base del campione osservato. Il procedimento di stima fa corrispondere ad ogni campione $(x_1, \dots, x_n) \in \mathcal{C}_n$ un valore $\theta \in \Theta$, ovvero considera un vettore di funzioni $\tilde{\Theta}$ a k componenti, detto stimatore, tale che

$$\tilde{\Theta} : \mathcal{C}_n \rightarrow \Theta .$$

Uno stimatore è per definizione un vettore di statistiche. La realizzazione campionaria dello stimatore $\tilde{\theta} = \tilde{\Theta}(x_1, \dots, x_n)$ è detta stima. Questo tipo di procedimento di stima è detto per punti perchè ad ogni campione fa corrispondere un punto dello spazio parametrico.

In questo capitolo vengono discusse una serie di proprietà desiderabili per uno stimatore $\tilde{\Theta}$. Tuttavia, anche se lo stimatore gode di proprietà ottimali, si deve sottolineare fin da ora che la stima $\tilde{\theta}$ può essere molto differente dal “vero valore” del parametro θ_0 a causa della variabilità campionaria. Dunque, in un procedimento di stima per punti, la stima $\tilde{\theta}$ deve sempre essere accompagnata da un indice di “precisione” dello stimatore $\tilde{\Theta}$ nello stimare il vero parametro θ_0 . Questo argomento verrà trattato in maggiore dettaglio nel prossimo capitolo.

2.1.2. I problemi regolari di stima. Al fine di ottenere un procedimento di stima con proprietà desiderabili, è opportuno definire alcune condizioni a cui il modello statistico deve preferibilmente sottostare. Dato un modello statistico \mathcal{F}_θ , si dice che il problema di stima è regolare se:

- il modello statistico è identificabile, ovvero le distribuzioni $f_n(x_1, \dots, x_n; \theta)$ specificate dal modello statistico \mathcal{F}_θ sono distinte per valori distinti di $\theta \in \Theta$;
- lo spazio parametrico Θ è un aperto di \mathbb{R}^k ;
- le derivate

$$\frac{\partial}{\partial \theta} f_n(x_1, \dots, x_n; \theta)$$

e

$$\frac{\partial^2}{\partial \theta \partial \theta^\top} f_n(x_1, \dots, x_n; \theta)$$

esistono per ogni $\theta \in \Theta$ e per ogni $(x_1, \dots, x_n) \in \mathcal{C}_n$ (a meno di insiemi di probabilità nulla);

d) risulta

$$\int_{\mathcal{C}_n} \frac{\partial}{\partial \theta} f_n(x_1, \dots, x_n; \theta) d\nu = \frac{\partial}{\partial \theta} \int_{\mathcal{C}_n} f_n(x_1, \dots, x_n; \theta) d\nu$$

e

$$\int_{\mathcal{C}_n} \frac{\partial^2}{\partial \theta \partial \theta^\top} f_n(x_1, \dots, x_n; \theta) d\nu = \frac{\partial^2}{\partial \theta \partial \theta^\top} \int_{\mathcal{C}_n} f_n(x_1, \dots, x_n; \theta) d\nu ,$$

dove l'integrazione è da intendersi opportunamente estesa alle componenti del vettore o della matrice e dove una scrittura del tipo

$$\int_{\mathcal{C}_n} g_n(x_1, \dots, x_n) d\nu$$

va interpretata come integrale di Riemann o come somma su \mathcal{C}_n a secondo che il vettore di variabili casuali (X_1, \dots, X_n) sia continuo o discreto.

Nel caso di un campionamento casuale, la condizione a) si riduce a richiedere che le distribuzioni $f(x; \theta)$ specificate dal modello statistico siano distinte per valori distinti di $\theta \in \Theta$. La condizione c) si riduce all'esistenza delle derivate

$$\frac{\partial}{\partial \theta} f(x; \theta)$$

e

$$\frac{\partial^2}{\partial \theta \partial \theta^T} f(x; \theta)$$

per ogni $\theta \in \Theta$ e per ogni $x \in \mathcal{S}_\theta$ (a meno di insiemi di probabilità nulla), mentre la condizione d) risulta

$$\int_{\mathcal{S}_\theta} \frac{\partial}{\partial \theta} f(x; \theta) d\nu = \frac{\partial}{\partial \theta} \int_{\mathcal{S}_\theta} f(x; \theta) d\nu$$

e

$$\int_{\mathcal{S}_\theta} \frac{\partial^2}{\partial \theta \partial \theta^T} f(x; \theta) d\nu = \frac{\partial^2}{\partial \theta \partial \theta^T} \int_{\mathcal{S}_\theta} f(x; \theta) d\nu .$$

La condizione a) permette di individuare in maniera univoca una specifica distribuzione a partire dal valore del parametro e quindi di ottenere procedimenti di stima coerenti. La condizione b) serve a imporre che il vero valore del parametro θ_0 non sia sulla frontiera di Θ . Infine, le condizioni c) e d) sono condizioni di regolarità matematica, che vengono adoperate nel seguito per ottenere alcune proprietà degli stimatori. Le condizioni a)-d) sono soddisfatte nella gran parte dei modelli statistici di uso comune.

• **Esempio 2.1.1.** Dato un campione casuale da $X \sim N(\mu, 1)$, la condizione a) è verificata in quanto

$$\phi(x - \mu_1) = \phi(x - \mu_2)$$

per ogni x solo se $\mu_1 = \mu_2$. La condizione b) è verificata dal momento che $\Theta = \mathbb{R}$. È immediato verificare la condizione c), dal momento che

$$\frac{\partial}{\partial \mu} \phi(x - \mu) = -\phi'(x - \mu)$$

e

$$\frac{\partial^2}{\partial \mu^2} \phi(x - \mu) = \phi''(x - \mu)$$

esistono per ogni $\mu \in \mathbb{R}$ e $x \in \mathbb{R}$. Inoltre, risulta

$$\frac{\partial}{\partial \mu} \int_{-\infty}^{\infty} \phi(x - \mu) dx = \frac{\partial^2}{\partial \mu^2} \int_{-\infty}^{\infty} \phi(x - \mu) dx = 0 ,$$

essendo

$$\int_{-\infty}^{\infty} \phi(x - \mu) dx = 1 ,$$

mentre

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial \mu} \phi(x - \mu) dx = - \int_{-\infty}^{\infty} \phi'(x - \mu) dx = - [\phi(x - \mu)]_{-\infty}^{\infty} = 0$$

e

$$\int_{-\infty}^{\infty} \frac{\partial^2}{\partial \mu^2} \phi(x - \mu) dx = \int_{-\infty}^{\infty} \phi''(x - \mu) dx = - [\phi'(x - \mu)]_{-\infty}^{\infty} = 0 ,$$

e di conseguenza anche la condizione *d*) è verificata. Si deve dunque concludere il problema di stima è regolare. \square

2.2. Gli stimatori corretti

2.2.1. La definizione di stimatore corretto. Dato un modello statistico \mathcal{F}_θ , se si suppone che il vero valore del parametro è $\theta_0 \in \Theta$, allora si richiede usualmente che la distribuzione dello stimatore sia concentrata intorno a questo valore, ovvero si richiede la proprietà della correttezza. *A priori* dal campionamento, questa proprietà assicura che la determinazione campionaria dello stimatore sia tendenzialmente prossima a θ_0 .

Lo stimatore $\tilde{\Theta}$ è detto corretto per θ se

$$E[\tilde{\Theta}] = \theta ,$$

per ogni $\theta \in \Theta$, dove il valore medio esiste ed è calcolato rispetto alla distribuzione specificata dal modello con il valore del parametro pari a θ . Nel seguito si assumerà sempre questa ultima convenzione per quanto riguarda il calcolo dei valori medi senza ulteriore specificazione.

Uno stimatore non corretto è detto distorto e la quantità

$$B[\tilde{\Theta}] = E[\tilde{\Theta}] - \theta$$

è detta distorsione.

Una proprietà più blanda della correttezza è la cosiddetta correttezza asintotica. Lo stimatore $\tilde{\Theta}$ è detto asintoticamente corretto per θ se per ogni $\theta \in \Theta$

$$\lim_{n \rightarrow \infty} E[\tilde{\Theta}] = \theta .$$

• **Esempio 2.2.1.** Dato un campione casuale da $X \sim N(\mu, \nu)$, tenendo presente l'Esempio 1.2.5, si ha

$$E[(\bar{X}, S_c^2)] = (\mu, \nu) ,$$

e quindi (\bar{X}, S_c^2) è corretto per (μ, ν) . \square

• **Esempio 2.2.2.** Dato un campione casuale da $X \sim N(0, \nu)$, tenendo presente i risultati dell'Esempio 1.2.5, dal momento che

$$E[S^2] = \frac{n-1}{n} \nu ,$$

allora la varianza campionaria S^2 è uno stimatore distorto per ν . La distorsione è pari a

$$B[S^2] = E[S^2] - v = -\frac{v}{n}.$$

Lo stimatore S^2 è asintoticamente corretto per v , dal momento che

$$\lim_{n \rightarrow \infty} E[S^2] = v. \quad \square$$

• **Esempio 2.2.3.** Dato un campione casuale da $X \sim U(0, \delta)$, tenendo presente i risultati dell'Esempio 1.2.6 e le proprietà della Beta (vedi §A.1.5), risulta

$$E[X_{(n)}] = \frac{n\delta}{n+1}.$$

Quindi, $X_{(n)}$ è uno stimatore distorto per δ , con distorsione pari a

$$B[X_{(n)}] = E[X_{(n)}] - \delta = -\frac{\delta}{n+1}.$$

Lo stimatore $X_{(n)}$ è asintoticamente corretto per δ , dal momento che

$$\lim_{n \rightarrow \infty} E[X_{(n)}] = \delta. \quad \square$$

2.3. Gli stimatori coerenti

2.3.1. La definizione di stimatore coerente. Se si suppone che il vero valore del parametro è dato da $\theta_0 \in \Theta$, si richiede usualmente che la distribuzione dello stimatore si concentri sempre di più intorno a θ_0 all'aumentare della numerosità campionaria, ovvero si richiede la cosiddetta proprietà della coerenza.

Se $\mathcal{F}_{\theta, n}$ è un modello statistico per ogni n , uno stimatore $\tilde{\Theta}_n$ si dice coerente (in senso debole) per θ , se la successione di stimatori $(\tilde{\Theta}_n)_{n \geq 1}$ converge in probabilità a θ_0 , ovvero

$$\tilde{\Theta}_n \xrightarrow{p} \theta_0$$

per $n \rightarrow \infty$. Di conseguenza, condizione sufficiente affinché lo stimatore sia coerente per θ è che sia asintoticamente corretto per θ e che

$$\lim_{n \rightarrow \infty} \text{Var}[\tilde{\Theta}_n] = 0.$$

Inoltre, se si ha

$$\tilde{\Theta}_n \xrightarrow{qc} \theta_0$$

per $n \rightarrow \infty$, allora lo stimatore è detto coerente in senso forte per θ . Ovviamente, la coerenza in senso forte implica la coerenza in senso debole.

• **Esempio 2.3.1.** Dato un campione casuale da $X \sim N(\mu, 1)$, tenendo presente i risultati dell'Esempio 1.2.3, dal momento che \bar{X}_n è uno stimatore corretto per μ e che

$$\lim_{n \rightarrow \infty} \text{Var}[\bar{X}_n] = \lim_{n \rightarrow \infty} \frac{1}{n} = 0,$$

allora $\bar{X}_n \xrightarrow{p} \mu_0$, ovvero \bar{X}_n è uno stimatore coerente per μ . Dal momento che per la Legge Forte dei Grandi Numeri di Khintchine si ha $\bar{X}_n \xrightarrow{qc} \mu_0$ (vedi Barabesi, 2020), \bar{X}_n è uno stimatore coerente in senso forte per μ . \square

• **Esempio 2.3.2.** Dato un campione casuale da $X \sim N(\mu, \nu)$, tenendo presente i risultati dell'Esempio 1.2.5, risulta

$$\lim_{n \rightarrow \infty} \text{Var}[(\bar{X}_n, S_{c,n}^2)] = \lim_{n \rightarrow \infty} \begin{pmatrix} \frac{\nu}{n} & 0 \\ 0 & \frac{2\nu^2}{n-1} \end{pmatrix} = \mathbf{0}.$$

Dal momento che $(\bar{X}_n, S_{c,n}^2)$ è corretto per (μ, ν) , allora risulta $(\bar{X}_n, S_{c,n}^2) \xrightarrow{p} (\mu_0, \nu_0)$, ovvero $(\bar{X}_n, S_{c,n}^2)$ è uno stimatore coerente per (μ, ν) . \square

2.4. Gli stimatori efficienti

2.4.1. L'informazione di Fisher. Dato un modello statistico \mathcal{F}_θ , in questa sezione vengono introdotte alcune quantità fondamentali nell'inferenza statistica classica. Data l'importanza degli argomenti trattati, è preferibile lavorare inizialmente con un solo parametro, ovvero si assume $k = 1$.

Se per il modello statistico considerato è valida la condizione c) della §2.1.2, allora si dice funzione punteggio la quantità

$$\begin{aligned} s_n(\theta) = s_n(\theta; x_1, \dots, x_n) &= \frac{d}{d\theta} l(\theta; x_1, \dots, x_n) = \frac{d}{d\theta} \log f_n(x_1, \dots, x_n; \theta) \\ &= \frac{1}{f_n(x_1, \dots, x_n; \theta)} \frac{d}{d\theta} f_n(x_1, \dots, x_n; \theta). \end{aligned}$$

La funzione punteggio $s_n(\theta; x_1, \dots, x_n)$ è la realizzazione campionaria della variabile casuale $s_n(\theta; X_1, \dots, X_n)$, per cui adoperando l'assunzione d) della §2.1.2 si ha

$$\begin{aligned} E[s_n(\theta; X_1, \dots, X_n)] &= \int_{\mathcal{C}_n} s_n(\theta; x_1, \dots, x_n) f_n(x_1, \dots, x_n; \theta) d\nu \\ &= \int_{\mathcal{C}_n} \frac{d}{d\theta} f_n(x_1, \dots, x_n; \theta) d\nu \\ &= \frac{d}{d\theta} \int_{\mathcal{C}_n} f_n(x_1, \dots, x_n; \theta) d\nu = \frac{d}{d\theta} 1 = 0. \end{aligned}$$

Inoltre, tenendo presente la relazione

$$\begin{aligned} \frac{d^2}{d\theta^2} \log f_n(x_1, \dots, x_n; \theta) &= \frac{1}{f_n(x_1, \dots, x_n; \theta)} \frac{d^2}{d\theta^2} f_n(x_1, \dots, x_n; \theta) + \\ &\quad - \left(\frac{1}{f_n(x_1, \dots, x_n; \theta)} \frac{d}{d\theta} f_n(x_1, \dots, x_n; \theta) \right)^2, \end{aligned}$$

e adoperando di nuovo l'assunzione d) della §2.1.2 risulta

$$\begin{aligned}
\text{Var}[s_n(\theta; X_1, \dots, X_n)] &= \text{E}[s_n(\theta; X_1, \dots, X_n)^2] \\
&= \int_{C_n} \left(\frac{1}{f_n(x_1, \dots, x_n; \theta)} \frac{d}{d\theta} f_n(x_1, \dots, x_n; \theta) \right)^2 f_n(x_1, \dots, x_n; \theta) d\nu \\
&= \int_{C_n} \frac{d^2}{d\theta^2} f_n(x_1, \dots, x_n; \theta) d\nu - \int_{C_n} \frac{d^2}{d\theta^2} \log f_n(x_1, \dots, x_n; \theta) f_n(x_1, \dots, x_n; \theta) d\nu \\
&= \frac{d^2}{d\theta^2} \int_{C_n} f_n(x_1, \dots, x_n; \theta) d\nu - \int_{C_n} \frac{d}{d\theta} s_n(\theta; x_1, \dots, x_n) f_n(x_1, \dots, x_n; \theta) d\nu \\
&= -\text{E} \left[\frac{d}{d\theta} s_n(\theta; X_1, \dots, X_n) \right].
\end{aligned}$$

La precedente varianza è detta informazione di Fisher ed è indicata come

$$I_n(\theta) = \text{Var}[s_n(\theta; X_1, \dots, X_n)] = -\text{E} \left[\frac{d}{d\theta} s_n(\theta; X_1, \dots, X_n) \right].$$

Nel caso di un campionamento casuale la funzione punteggio può essere espressa come

$$\begin{aligned}
s_n(\theta; x_1, \dots, x_n) &= \frac{d}{d\theta} \sum_{i=1}^n \log f(x_i; \theta) = \sum_{i=1}^n \frac{d}{d\theta} \log f(x_i; \theta) \\
&= \sum_{i=1}^n \frac{1}{f(x_i; \theta)} \frac{d}{d\theta} f(x_i; \theta) = \sum_{i=1}^n s(\theta; x_i),
\end{aligned}$$

dove

$$s(\theta) = s(\theta; x_i) = \frac{d}{d\theta} \log f(x_i; \theta) = \frac{1}{f(x_i; \theta)} \frac{d}{d\theta} f(x_i; \theta),$$

è la funzione punteggio relativa all' i -esima osservazione campionaria. Le variabili casuali $s(\theta; X_i)$ sono indipendenti in quanto trasformate di variabili casuali indipendenti ed ugualmente distribuite. Dunque, si ha

$$\text{Var}[s_n(\theta; X_1, \dots, X_n)] = \sum_{i=1}^n \text{Var}[s(\theta; X_i)] = n \text{Var}[s(\theta; X)] = -n \text{E} \left[\frac{d}{d\theta} s(\theta; X) \right].$$

Di conseguenza, l'informazione di Fisher può essere espressa come

$$I_n(\theta) = n \text{Var}[s(\theta; X)] = -n \text{E} \left[\frac{d}{d\theta} s(\theta; X) \right] = nI(\theta),$$

dove $I(\theta) = \text{Var}[s(\theta; X)]$ è l'informazione di Fisher relativa ad una sola osservazione campionaria.

• **Esempio 2.4.1.** Dato un campione casuale da $X \sim N(\mu, 1)$, dall'Esempio 2.1.1 è noto che le condizioni c) e d) della §2.1.2 sono verificate. Dunque, tenendo presente l'Esempio 1.3.1, si ha la funzione punteggio

$$s_n(\mu; x_1, \dots, x_n) = \frac{d}{d\mu} \left(\log c - \frac{n}{2} (s^2 + (\bar{x} - \mu)^2) \right) = n(\bar{x} - \mu),$$

da cui si ottiene l'informazione di Fisher

$$I_n(\mu) = -\text{E} \left[\frac{d}{d\mu} n(\bar{X} - \mu) \right] = n.$$

I risultati teorici di questa sezione risultano verificati, dal momento che per via diretta si ottiene

$$E[s_n(\mu; X_1, \dots, X_n)] = E[n(\bar{X} - \mu)] = nE[\bar{X} - \mu] = 0$$

e

$$I_n(\mu) = \text{Var}[s_n(\mu; X_1, \dots, X_n)] = \text{Var}[n(\bar{X} - \mu)] = n^2 \text{Var}[\bar{X}] = n^2 \frac{1}{n} = n.$$

Dal momento che si dispone di un campione casuale, i precedenti risultati possono essere ottenuti più facilmente dalla funzione punteggio riferita ad una sola osservazione campionaria, ovvero

$$s(\mu; x_i) = \frac{d}{d\mu} \left(\log c - \frac{1}{2} (x_i - \mu)^2 \right) = x_i - \mu,$$

da cui l'informazione di Fisher relativa ad una sola osservazione campionaria risulta

$$I(\mu) = -E \left[\frac{d}{d\mu} (X - \mu) \right] = 1.$$

Dalla precedente espressione si ha di nuovo

$$I_n(\mu) = nI(\mu) = n. \quad \square$$

• **Esempio 2.4.2.** Dato un campione casuale da $X \sim E(0, \sigma)$, è facile verificare le condizioni *c*) e *d*) della §2.1.2. Dunque, tenendo presente l'Esempio 1.3.3, si ha la seguente funzione punteggio

$$s_n(\sigma; x_1, \dots, x_n) = \frac{d}{d\sigma} \left(\log c - n \log \sigma - \frac{n\bar{x}}{\sigma} \right) = -\frac{n}{\sigma} + \frac{n\bar{x}}{\sigma^2},$$

da cui, in base ai risultati dell'Esempio 1.2.4, si ottiene l'informazione di Fisher

$$I_n(\sigma) = -E \left[\frac{d}{d\sigma} \left(-\frac{n}{\sigma} + \frac{n\bar{X}}{\sigma^2} \right) \right] = -E \left[\frac{n}{\sigma^2} - \frac{2n\bar{X}}{\sigma^3} \right] = -\frac{n}{\sigma^2} + \frac{2n}{\sigma^3} E[\bar{X}] = \frac{n}{\sigma^2}.$$

Dal momento che si dispone di un campione casuale, i precedenti risultati possono essere ottenuti più facilmente dalla funzione punteggio riferita ad una sola osservazione campionaria, ovvero

$$s(\sigma; x_i) = \frac{d}{d\sigma} \left(-\log \sigma - \frac{x_i}{\sigma} \right) = -\frac{1}{\sigma} + \frac{x_i}{\sigma^2},$$

da cui, tenendo presente le proprietà della Esponenziale (vedi §A.1.4), l'informazione di Fisher relativa ad una sola osservazione campionaria risulta

$$I(\sigma) = -E \left[\frac{d}{d\sigma} \left(-\frac{1}{\sigma} + \frac{X}{\sigma^2} \right) \right] = -\frac{1}{\sigma^2} + \frac{2}{\sigma^3} E[X] = \frac{1}{\sigma^2}.$$

Dalla precedente espressione si ottiene di nuovo

$$I_n(\sigma) = nI(\sigma) = \frac{n}{\sigma^2}. \quad \square$$

• **Esempio 2.4.3.** Dato un campione casuale da $X \sim U(0, \delta)$, la condizione *d*) della §2.1.2 non è verificata. Infatti risulta

$$\frac{d}{d\delta} \frac{1}{\delta} = -\frac{1}{\delta^2}$$

per ogni $\delta \in \mathbb{R}^+$ e $x \in [0, \delta]$, da cui

$$\int_0^\delta \frac{d}{d\delta} \frac{1}{\delta} dx = -\frac{1}{\delta}.$$

Tuttavia, si ha

$$\frac{d}{d\delta} \int_0^\delta \frac{1}{\delta} dx = 0 \neq -\frac{1}{\delta}.$$

Anche se la funzione punteggio esiste, non ha tuttavia le proprietà ottenute in questa sezione. \square

I precedenti risultati possono essere generalizzati al caso di un vettore di parametri $\boldsymbol{\theta}$. Se è valida la condizione c) della §2.1.2, si dice vettore delle funzioni punteggio la quantità $\mathbf{s}_n(\boldsymbol{\theta}; x_1, \dots, x_n)$ data da

$$\mathbf{s}_n(\boldsymbol{\theta}) = \mathbf{s}_n(\boldsymbol{\theta}; x_1, \dots, x_n) = \frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}; x_1, \dots, x_n) = \frac{1}{f_n(x_1, \dots, x_n; \boldsymbol{\theta})} \frac{\partial}{\partial \boldsymbol{\theta}} f_n(x_1, \dots, x_n; \boldsymbol{\theta}).$$

Il vettore $\mathbf{s}_n(\boldsymbol{\theta}; x_1, \dots, x_n)$ è la determinazione campionaria del vettore di variabili casuali $\mathbf{s}_n(\boldsymbol{\theta}; X_1, \dots, X_n)$ e quindi generalizzando in modo ovvio le relazioni ottenute in precedenza, se la condizione d) della §2.1.2 è verificata, si può ottenere che

$$E[\mathbf{s}_n(\boldsymbol{\theta}; X_1, \dots, X_n)] = \mathbf{0}$$

e

$$\text{Var}[\mathbf{s}_n(\boldsymbol{\theta}; X_1, \dots, X_n)] = -E\left[\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{s}_n(\boldsymbol{\theta}; X_1, \dots, X_n)\right].$$

Questa matrice di varianza-covarianza è detta matrice di informazione di Fisher ed è denotata con

$$\mathbf{I}_n(\boldsymbol{\theta}) = \text{Var}[\mathbf{s}_n(\boldsymbol{\theta}; X_1, \dots, X_n)] = -E\left[\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{s}_n(\boldsymbol{\theta}; X_1, \dots, X_n)\right].$$

Nel caso di un campionamento casuale, il vettore delle funzioni punteggio risulta

$$\mathbf{s}_n(\boldsymbol{\theta}; x_1, \dots, x_n) = \sum_{i=1}^n \mathbf{s}(\boldsymbol{\theta}; x_i),$$

dove

$$\mathbf{s}(\boldsymbol{\theta}) = \mathbf{s}(\boldsymbol{\theta}; x_i) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f(x_i; \boldsymbol{\theta}) = \frac{1}{f(x_i; \boldsymbol{\theta})} \frac{\partial}{\partial \boldsymbol{\theta}} f(x_i; \boldsymbol{\theta}),$$

è il vettore delle funzioni punteggio relative all' i -esima osservazione campionaria. Anche in questo caso la matrice di informazione di Fisher può essere espressa come

$$\mathbf{I}_n(\boldsymbol{\theta}) = n\text{Var}[\mathbf{s}(\boldsymbol{\theta}; X)] = -nE\left[\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{s}(\boldsymbol{\theta}; X)\right] = n\mathbf{I}(\boldsymbol{\theta}),$$

dove $\mathbf{I}(\boldsymbol{\theta}) = \text{Var}[\mathbf{s}(\boldsymbol{\theta}; X)]$ è la matrice di informazione di Fisher relativa ad una sola osservazione campionaria.

• **Esempio 2.4.4.** Dato un campione casuale da $X \sim N(\mu, \nu)$, non è difficile verificare (anche se laborioso) le condizioni c) e d) della §2.1.2. Tenendo presente l'Esempio 1.3.2, si ha il seguente vettore di funzioni punteggio

$$\begin{aligned} \mathbf{s}_n(\mu, \nu; x_1, \dots, x_n) &= \begin{pmatrix} \frac{\partial}{\partial \mu} (\log c - \frac{n}{2} \log \nu - \frac{n}{2\nu} (s^2 + (\bar{x} - \mu)^2)) \\ \frac{\partial}{\partial \nu} (\log c - \frac{n}{2} \log \nu - \frac{n}{2\nu} (s^2 + (\bar{x} - \mu)^2)) \end{pmatrix} \\ &= \begin{pmatrix} \frac{n}{\nu} (\bar{x} - \mu) \\ -\frac{n}{2\nu} + \frac{n}{2\nu^2} (s^2 + (\bar{x} - \mu)^2) \end{pmatrix}, \end{aligned}$$

da cui, considerando anche i risultati dell'Esempio 1.2.5, si ha la matrice d'informazione di Fisher

$$\begin{aligned} \mathbf{I}_n(\mu, \nu) &= -\mathbb{E} \begin{pmatrix} \frac{\partial}{\partial \mu} (\frac{n}{\nu} (\bar{X} - \mu)) & \frac{\partial}{\partial \nu} (\frac{n}{\nu} (\bar{X} - \mu)) \\ \frac{\partial}{\partial \mu} (-\frac{n}{2\nu} + \frac{n}{2\nu^2} (S^2 + (\bar{X} - \mu)^2)) & \frac{\partial}{\partial \nu} (-\frac{n}{2\nu} + \frac{n}{2\nu^2} (S^2 + (\bar{X} - \mu)^2)) \end{pmatrix} \quad \square \\ &= \begin{pmatrix} \frac{n}{\nu} & 0 \\ \frac{n}{\nu^2} \mathbb{E}[\bar{X} - \mu] & -\frac{n}{2\nu^2} + \frac{n}{\nu^3} (\mathbb{E}[S^2] + \mathbb{E}[(\bar{X} - \mu)^2]) \end{pmatrix} = \begin{pmatrix} \frac{n}{\nu} & 0 \\ 0 & \frac{n}{2\nu^2} \end{pmatrix}. \end{aligned}$$

2.4.2. Il limite inferiore di Rao-Cramér. Dato un modello statistico \mathcal{F}_θ , in questa sezione viene ottenuto un limite inferiore alla varianza di uno stimatore corretto per una data trasformata di θ . Di nuovo, è opportuno lavorare inizialmente con un solo parametro. Si consideri lo stimatore $\tilde{\Theta}$ tale che $\mathbb{E}[\tilde{\Theta}] = g(\theta)$. Si assuma che siano valide le condizioni *c*) e *d*) della §2.1.2 e che sia valida per lo stimatore $\tilde{\Theta}$ la seguente condizione:

e) la funzione $g(\theta)$ esiste ed è derivabile con derivata

$$\begin{aligned} \frac{d}{d\theta} g(\theta) &= \frac{d}{d\theta} \int_{\mathcal{C}_n} \tilde{\Theta}(x_1, \dots, x_n) f_n(x_1, \dots, x_n; \theta) d\nu \\ &= \int_{\mathcal{C}_n} \frac{d}{d\theta} \tilde{\Theta}(x_1, \dots, x_n) f_n(x_1, \dots, x_n; \theta) d\nu. \end{aligned}$$

In base alle assunzioni *c*) della §2.1.2 ed *e*), si ottiene

$$\begin{aligned} \text{Cov}[\tilde{\Theta}, s_n(\theta; X_1, \dots, X_n)] &= \mathbb{E}[\tilde{\Theta} s_n(\theta; X_1, \dots, X_n)] \\ &= \int_{\mathcal{C}_n} \tilde{\Theta}(x_1, \dots, x_n) s_n(\theta; x_1, \dots, x_n) f_n(x_1, \dots, x_n; \theta) d\nu \\ &= \int_{\mathcal{C}_n} \frac{d}{d\theta} \tilde{\Theta}(x_1, \dots, x_n) f_n(x_1, \dots, x_n; \theta) d\nu = \frac{d}{d\theta} g(\theta). \end{aligned}$$

Inoltre, per la disuguaglianza di Schwarz (vedi Barabesi, 2020), se è verificata la condizione *d*) della §2.1.2, si ha

$$\text{Var}[\tilde{\Theta}] \text{Var}[s_n(\theta; X_1, \dots, X_n)] \geq \text{Cov}[\tilde{\Theta} s_n(\theta; X_1, \dots, X_n)]^2,$$

da cui, supponendo $0 < I_n(\theta) < \infty$, con opportune sostituzioni, si ottiene infine la cosiddetta disuguaglianza di Rao-Cramér, ovvero

$$\text{Var}[\tilde{\Theta}] \geq \frac{(\frac{d}{d\theta} g(\theta))^2}{I_n(\theta)}.$$

Se $\tilde{\Theta}$ è uno stimatore corretto per θ , allora $\frac{d}{d\theta} g(\theta) = 1$, e la precedente espressione risulta

$$\text{Var}[\tilde{\Theta}] \geq \frac{1}{I_n(\theta)}.$$

Dalla precedente espressione è chiaro il ruolo dell'informazione di Fisher, nel senso che essa rappresenta un indice della precisione media ottenibile se si vuole stimare θ in modo non distorto.

Nel caso di un campione casuale la disuguaglianza di Rao-Cramér si riduce a

$$\text{Var}[\tilde{\Theta}] \geq \frac{\left(\frac{d}{d\theta} g(\theta)\right)^2}{nI(\theta)},$$

e, nel caso di uno stimatore corretto per θ , a

$$\text{Var}[\tilde{\Theta}] \geq \frac{1}{nI(\theta)}.$$

• **Esempio 2.4.5.** Dato un campione casuale da $X \sim N(\mu, 1)$, dall'Esempio 2.1.1 è noto che le condizioni *c*) e *d*) della §2.1.2 sono verificate. Tenendo presente l'Esempio 2.4.1, dato un qualsiasi stimatore $\tilde{\Theta}$ corretto per μ (e per cui è valida la condizione *e*) di questa sezione) si ha dunque

$$\text{Var}[\tilde{\Theta}] \geq \frac{1}{n}. \quad \square$$

• **Esempio 2.4.6.** Dato un campione casuale da $X \sim E(0, \sigma)$, dall'Esempio 2.4.2 è noto che le condizioni *c*) e *d*) della §2.1.2 sono verificate. Tenendo presente l'Esempio 2.4.2, dato un qualsiasi stimatore $\tilde{\Theta}$ corretto per σ (e per cui è valida la condizione *e*) di questa sezione) si ha

$$\text{Var}[\tilde{\Theta}] \geq \frac{\sigma^2}{n}.$$

Se invece $\tilde{\Theta}$ è uno stimatore corretto per $g(\sigma) = \sigma^2$ allora, dal momento che

$$\frac{d}{d\theta} g(\sigma) = 2\sigma,$$

dalla forma generale della disuguaglianza di Rao-Cramér si ottiene

$$\text{Var}[\tilde{\Theta}] \geq \frac{4\sigma^4}{n}. \quad \square$$

• **Esempio 2.4.7.** Dato un campione casuale da $X \sim U(0, \delta)$, dall'Esempio 2.4.3 è noto che la condizione *d*) della §2.1.2 non è verificata. In questo caso non è possibile ottenere un limite inferiore alla varianza di uno stimatore $\tilde{\delta}$ corretto per δ . \square

I precedenti risultati possono essere generalizzati al caso di un vettore di parametri $\boldsymbol{\theta}$. Si consideri dunque lo stimatore $\tilde{\boldsymbol{\Theta}}$ tale che $E[\tilde{\boldsymbol{\Theta}}] = \mathbf{g}(\boldsymbol{\theta})$. Si assuma che siano valide le condizioni *c*) e *d*) della §2.1.2 e che sia valida per lo stimatore $\tilde{\boldsymbol{\Theta}}$ la generalizzazione della condizione *e*), ovvero: *e*) il vettore di funzioni $\mathbf{g}(\boldsymbol{\theta})$ esiste ed è derivabile con matrice delle derivate

$$\begin{aligned} \mathbf{D}(\boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{g}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathcal{C}_n} \tilde{\boldsymbol{\Theta}}(x_1, \dots, x_n) f_n(x_1, \dots, x_n; \boldsymbol{\theta}) d\nu \\ &= \int_{\mathcal{C}_n} \frac{\partial}{\partial \boldsymbol{\theta}} \tilde{\boldsymbol{\Theta}}(x_1, \dots, x_n) f_n(x_1, \dots, x_n; \boldsymbol{\theta}) d\nu. \end{aligned}$$

Se $\mathbf{I}_n(\boldsymbol{\theta})$ è una matrice definita positiva, si può dimostrare che la generalizzazione della disuguaglianza di Rao-Cramér (vedi Wilks, 1962) è data da

$$\text{Var}[\tilde{\boldsymbol{\Theta}}] \geq \mathbf{D}(\boldsymbol{\theta})^T \mathbf{I}_n(\boldsymbol{\theta})^{-1} \mathbf{D}(\boldsymbol{\theta}),$$

dove la precedente disuguaglianza si deve intendere nel senso che la matrice differenza $\text{Var}[\tilde{\boldsymbol{\Theta}}] - \mathbf{D}(\boldsymbol{\theta})^T \mathbf{I}_n(\boldsymbol{\theta})^{-1} \mathbf{D}(\boldsymbol{\theta})$ è semidefinita positiva. Nel caso di un campione casuale si ha inoltre

$$\text{Var}[\tilde{\Theta}] \geq \frac{1}{n} \mathbf{D}(\boldsymbol{\theta})^T \mathbf{I}(\boldsymbol{\theta})^{-1} \mathbf{D}(\boldsymbol{\theta}).$$

Infine, se $\tilde{\Theta}$ è uno stimatore corretto per $\boldsymbol{\theta}$, allora $\mathbf{D}(\boldsymbol{\theta})$ è la matrice identità di ordine k , da cui

$$\text{Var}[\tilde{\Theta}] \geq \mathbf{I}_n(\boldsymbol{\theta})^{-1},$$

che, nel caso di un campione casuale, si riduce a

$$\text{Var}[\tilde{\Theta}] \geq \frac{1}{n} \mathbf{I}(\boldsymbol{\theta})^{-1}.$$

• **Esempio 2.4.8.** Dato un campione casuale da $X \sim N(\mu, \nu)$, dall'Esempio 2.4.4 è noto che le condizioni *c*) e *d*) della §2.1.2 sono verificate. Dai risultati dell'Esempio 2.4.4, dato un qualsiasi stimatore $\tilde{\Theta}$ corretto per (μ, ν) (e per cui è valida la condizione *e*) di questa sezione) si ha

$$\text{Var}[\tilde{\Theta}] \geq \begin{pmatrix} \frac{\nu}{n} & 0 \\ 0 & \frac{2\nu^2}{n} \end{pmatrix}.$$

Se invece $\tilde{\Theta}$ è un qualsiasi stimatore corretto per $\mathbf{g}(\mu, \nu) = (\mu, \sqrt{\nu})^T$, allora dal momento che

$$\mathbf{D}(\mu, \nu) = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2\sqrt{\nu}} \end{pmatrix},$$

dalla forma più generale della disuguaglianza di Rao-Cramér si ha

$$\text{Var}[\tilde{\Theta}] \geq \begin{pmatrix} \frac{\nu}{n} & 0 \\ 0 & \frac{\nu}{2n} \end{pmatrix}. \quad \square$$

2.4.3. La definizione di stimatore efficiente. Dato un modello statistico \mathcal{F}_θ , quando si dispone di due stimatori corretti per $\boldsymbol{\theta}$, la nostra preferenza va ovviamente a quello con varianza minore, in quanto la sua distribuzione risulta più concentrata intorno al vero valore del parametro. Dal momento che nella precedente sezione è stato stabilito un limite inferiore alla varianza di uno stimatore corretto, è logico introdurre il seguente indice nel caso di un parametro per valutare la bontà di uno stimatore $\tilde{\Theta}$ corretto per $\boldsymbol{\theta}$

$$\text{eff}(\tilde{\Theta}) = \frac{1}{\text{Var}[\tilde{\Theta}] I_n(\boldsymbol{\theta})}.$$

La quantità $\text{eff}(\tilde{\Theta})$ è detta efficienza ed evidentemente si ha $0 \leq \text{eff}(\tilde{\Theta}) \leq 1$ quando sono valide le condizioni *c*) e *d*) della §2.1.2 e la condizione *e*) della §2.4.2. In questo caso, uno stimatore $\tilde{\Theta}$ corretto per $\boldsymbol{\theta}$ è quindi detto efficiente se $\text{eff}(\tilde{\Theta})$ è massima, ovvero se $\text{eff}(\tilde{\Theta}) = 1$. Niente assicura l'esistenza e l'unicità di uno stimatore efficiente per un dato modello statistico.

• **Esempio 2.4.9.** Dato un campione casuale da $X \sim N(\mu, 1)$, tenendo presente i risultati dell'Esempio 1.2.3 e dell'Esempio 2.4.5, \bar{X} è uno stimatore efficiente in quanto è corretto per μ e

$$\text{eff}(\bar{X}) = \frac{1}{\text{Var}[\bar{X}] I_n(\mu)} = \frac{1}{(1/n)n} = 1. \quad \square$$

• **Esempio 2.4.10.** Dato un campione casuale da $X \sim E(0, \sigma)$, tenendo presente i risultati dell'Esempio 1.2.4 e dell'Esempio 2.4.6, \bar{X} è uno stimatore efficiente in quanto è corretto per σ e

$$\text{eff}(\bar{X}) = \frac{1}{\text{Var}[\bar{X}]I_n(\sigma)} = \frac{1}{(\sigma^2/n)(n/\sigma^2)} = 1. \quad \square$$

Nel caso di un vettore di parametri θ , un indice per valutare il grado di efficienza di uno stimatore $\tilde{\Theta}$ corretto per θ è data da

$$\text{eff}(\tilde{\Theta}) = \det(\text{Var}[\tilde{\Theta}]\mathbf{I}_n(\theta))^{-1}.$$

Anche in questo caso si ha $0 \leq \text{eff}(\tilde{\Theta}) \leq 1$ quando sono valide le condizioni *c*) e *d*) della §2.1.2 e la condizione *e*) della §2.4.2. In questo ultimo caso, uno stimatore $\tilde{\Theta}$ corretto per θ è quindi detto efficiente se $\text{eff}(\tilde{\Theta}) = 1$.

• **Esempio 2.4.11.** Dato un campione casuale da $X \sim N(\mu, v)$, tenendo presente i risultati dell'Esempio 2.3.2 e dell'Esempio 2.4.8, se si considera lo stimatore (\bar{X}, S_c^2) corretto per (μ, v) , si ha

$$\text{eff}[(\bar{X}, S_c^2)] = \det\left(\begin{pmatrix} \frac{v}{n} & 0 \\ 0 & \frac{2v^2}{n-1} \end{pmatrix} \begin{pmatrix} \frac{n}{v} & 0 \\ 0 & \frac{n}{2v^2} \end{pmatrix}\right)^{-1} = \frac{n-1}{n},$$

ovvero lo stimatore (\bar{X}, S_c^2) non è efficiente, anche se per n elevato risulta approssimativamente efficiente. \square

Quando si deve confrontare due stimatori, uno dei quali corretto e l'altro distorto, non si può adoperare l'efficienza come criterio di confronto. Nel caso di un parametro, dato uno stimatore distorto $\tilde{\Theta}$ di θ , un criterio che tiene conto anche della distorsione è il cosiddetto errore quadratico medio, ovvero

$$\text{EQM}[\tilde{\Theta}] = \text{B}[\tilde{\Theta}]^2 + \text{Var}[\tilde{\Theta}].$$

Il precedente indice può essere generalizzato al caso di un vettore di parametri θ , ovvero

$$\text{EQM}[\tilde{\Theta}] = \det(\text{B}[\tilde{\Theta}]\text{B}[\tilde{\Theta}]^T + \text{Var}[\tilde{\Theta}]).$$

Se ci si basa sui precedenti indici, uno stimatore leggermente distorto, ma con bassa varianza, può essere preferibile ad uno stimatore efficiente.

• **Esempio 2.4.12.** Dato un campione casuale da $X \sim N(\mu, v)$, supponiamo che il parametro μ sia noto. Per un qualsiasi stimatore $\tilde{\Theta}$ corretto per v si ha

$$\text{Var}[\tilde{\Theta}] \geq \frac{2v^2}{n}.$$

Se si considera lo stimatore

$$\tilde{\Theta}_1 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2,$$

allora tenendo presente i risultati dell'Esempio 1.2.5 e le proprietà della Chi-quadrato (vedi §A.3.2), si ottiene

$$\text{E}[\tilde{\Theta}_1] = v$$

e

$$\text{Var}[\tilde{\Theta}_1] = \frac{2v^2}{n},$$

ovvero $\tilde{\Theta}_1$ è efficiente e inoltre

$$\text{EQM}[\tilde{\Theta}_1] = \frac{2v^2}{n}.$$

Alternativamente, se si considera ora lo stimatore

$$\tilde{\Theta}_2 = \frac{n}{n+2} \tilde{\Theta}_1,$$

si ha

$$\text{E}[\tilde{\Theta}_2] = \frac{nv}{n+2}$$

e

$$\text{Var}[\tilde{\Theta}_2] = \frac{2nv^2}{(n+2)^2}$$

e quindi

$$\text{EQM}[\tilde{\Theta}_2] = \left(\frac{nv}{n+2} - v \right)^2 + \frac{2nv^2}{(n+2)^2} = \frac{2v^2}{n+2}.$$

Dal momento che $\text{EQM}[\tilde{\Theta}_2] < \text{EQM}[\tilde{\Theta}_1]$, se ci si basa sul criterio dell'errore quadratico medio, lo stimatore distorto è preferibile a quello corretto ed efficiente. \square

2.4.4. La famiglia esponenziale di distribuzioni. In questa sezione si vuole determinare la famiglia di distribuzioni per cui è possibile ottenere stimatori efficienti, dato che le condizioni *c*) e *d*) della §2.1.2 sono verificate. Nel caso di uno stimatore $\tilde{\Theta}$ corretto per θ (e per cui è valida la condizione *e*) della §2.4.2), per le proprietà della disuguaglianza di Schwarz (vedi Barabesi, 2020), si ottiene il limite inferiore di Rao-Cramér quando

$$s_n(\theta; X_1, \dots, X_n) = b(\tilde{\Theta} - \theta),$$

per b costante, ovvero quando $s_n(\theta; X_1, \dots, X_n)$ è una variabile casuale linearmente dipendente da $\tilde{\Theta}$. Dunque, se una statistica efficiente esiste, deve soddisfare la precedente relazione. Inoltre, si deve avere

$$\text{Var}[s_n(\theta; X_1, \dots, X_n)] = b^2 \text{Var}[\tilde{\Theta}],$$

ovvero, dal momento che

$$\text{Var}[\tilde{\Theta}] = \frac{1}{I_n(\theta)}$$

(essendo $\tilde{\Theta}$ uno stimatore efficiente) e

$$\text{Var}[s_n(\theta; X_1, \dots, X_n)] = I_n(\theta),$$

allora $b = I_n(\theta)$. Tenendo presente la definizione di funzione punteggio, si deve dunque avere

$$\frac{d}{d\theta} \log f_n(x_1, \dots, x_n; \theta) = I_n(\theta)(\tilde{\Theta}(x_1, \dots, x_n) - \theta),$$

per ogni $(x_1, \dots, x_n) \in \mathcal{C}_n$ (a meno di insiemi di probabilità nulla). La precedente relazione costituisce un'equazione differenziale in θ , per cui risolvendo si ha

$$\log f_n(x_1, \dots, x_n; \theta) = \tilde{\Theta}(x_1, \dots, x_n) \int I_n(\theta) d\theta - \int \theta I_n(\theta) d\theta + c(x_1, \dots, x_n),$$

ovvero si ha la seguente famiglia di distribuzioni

$$f_n(x_1, \dots, x_n; \theta) = q(x_1, \dots, x_n) e^{\tilde{\Theta}(x_1, \dots, x_n) \varphi(\theta) - \eta(\theta)},$$

dove si è posto

$$\varphi(\theta) = \int I_n(\theta) d\theta,$$

e

$$\eta(\theta) = \int \theta I_n(\theta) d\theta,$$

mentre

$$q(x_1, \dots, x_n) = e^{c(x_1, \dots, x_n)}.$$

Se esiste uno stimatore efficiente, la distribuzione specificata dal modello statistico deve essere della precedente forma. Questa famiglia di distribuzioni è detta famiglia esponenziale. È immediato verificare che tutti gli elementi di una stessa famiglia esponenziale devono avere necessariamente lo stesso supporto per ogni θ .

• **Esempio 2.4.13.** Dato un campione casuale da $X \sim N(\mu, 1)$, la classe di funzioni di densità definita in questo modello fa parte della famiglia esponenziale con $\tilde{\Theta}(x_1, \dots, x_n) = \bar{x}$,

$$\varphi(\mu) = n\mu$$

e

$$\eta(\mu) = \frac{n\mu^2}{2},$$

mentre

$$q(x_1, \dots, x_n) = (2\pi)^{-\frac{n}{2}} e^{-\frac{n}{2}(s^2 + \bar{x}^2)}.$$

In effetti, in base ai risultati dell'Esempio 2.4.9, la media campionaria \bar{X} è uno stimatore efficiente. \square

• **Esempio 2.4.14.** Dato un campione casuale da $X \sim E(\lambda, 1)$, la classe di funzioni di densità definita in questo modello non fa parte della famiglia esponenziale, dal momento che il relativo supporto dipende dal parametro λ . \square

In generale, se si considera un vettore di parametri θ , la famiglia esponenziale è definita come

$$f_n(x_1, \dots, x_n; \theta) = q(x_1, \dots, x_n) e^{\tilde{\Theta}(x_1, \dots, x_n)^T \varphi(\theta) - \eta(\theta)},$$

dove senza perdita di generalità si può assumere che le funzioni definite dalle componenti del vettore $\varphi(\theta)$ siano linearmente indipendenti in Θ . Per questo motivo, questa famiglia esponenziale è detta in forma ridotta. Per una trattazione generale delle famiglie esponenziali si può consultare Brown (1986).

• **Esempio 2.4.15.** Dato un campione casuale da $X \sim N(\mu, \nu)$, la classe di funzioni di densità specificata da questo modello fa parte della famiglia esponenziale dal momento che $\tilde{\Theta}(x_1, \dots, x_n) = (\bar{x}, s^2 + \bar{x}^2)^\top$,

$$\varphi(\mu, \nu) = \left(\frac{n\mu}{\nu}, -\frac{n}{2\nu} \right)^\top$$

e

$$\eta(\mu, \nu) = \frac{n\mu^2}{2\nu} + \frac{n}{2} \log \nu,$$

mentre

$$q(x_1, \dots, x_n) = (2\pi)^{-\frac{n}{2}}. \quad \square$$

2.5. Gli stimatori sufficienti

2.5.1. La definizione di stimatore sufficiente. Come è stato evidenziato negli esempi della §1.3, la funzione di verosimiglianza può dipendere solo dalle realizzazioni di alcune statistiche e quindi non è necessaria la conoscenza del campione originale per determinarla. L'informazione contenuta nel campione originale viene mantenuta dalle statistiche, con il vantaggio di avere l'informazione convenientemente sintetizzata.

Dato un modello statistico \mathcal{F}_θ , uno stimatore $\tilde{\Theta}$ è detto sufficiente per θ se per ogni $(x_1, \dots, x_n) \in \mathcal{C}_n$ e $(y_1, \dots, y_n) \in \mathcal{C}_n$ si ha

$$\tilde{\Theta}(x_1, \dots, x_n) = \tilde{\Theta}(y_1, \dots, y_n) \Rightarrow L(\theta; x_1, \dots, x_n) \propto L(\theta; y_1, \dots, y_n),$$

per ogni $\theta \in \Theta$. Dunque, $\tilde{\Theta}$ è sufficiente per θ quando assume lo stesso valore in due punti dello spazio campionario solo se i due punti hanno verosimiglianze proporzionali, ovvero i due punti forniscono la stessa informazione.

Anche se può risultare sorprendente che una trasformazione non biunivoca del campione originale riesca ad estrarre tutta l'informazione relativa al campione stesso, si tenga presente che la proprietà della sufficienza è legata alla scelta del modello statistico, ovvero uno stimatore sufficiente per un modello può non esserlo per un altro.

2.5.2. Alcune proprietà degli stimatori sufficienti. Un primo risultato relativo agli stimatori sufficienti è il cosiddetto criterio di fattorizzazione di Neyman. Se $\tilde{\Theta}$ è sufficiente per θ , allora $L(\theta)$ dipende da (x_1, \dots, x_n) solo attraverso $\tilde{\theta} = \tilde{\Theta}(x_1, \dots, x_n)$, ovvero $L(\theta) \propto g(\tilde{\theta}; \theta)$. Inoltre, dalla definizione di funzione di verosimiglianza, $L(\theta) \propto f_n(x_1, \dots, x_n; \theta)$ e si deve concludere che la funzione

$$h_n(x_1, \dots, x_n) = \frac{f_n(x_1, \dots, x_n; \theta)}{g(\tilde{\theta}; \theta)}$$

non dipende da θ . Dunque, se $\tilde{\Theta}$ è sufficiente per θ , si deve avere la relazione

$$f_n(x_1, \dots, x_n; \theta) = h_n(x_1, \dots, x_n)g(\tilde{\theta}; \theta).$$

Inversamente, se la precedente relazione è vera, allora $L(\theta)$ è funzione solo di $\tilde{\theta}$ e quindi $\tilde{\Theta}$ è sufficiente per θ . Si è dimostrato di conseguenza che $\tilde{\Theta}$ è sufficiente per θ se e solo se $f_n(x_1, \dots, x_n; \theta)$ può essere fattorizzata nella forma precedente per opportune funzioni h_n e g . Questo risultato è detto appunto criterio di fattorizzazione di Neyman.

• **Esempio 2.5.1.** Dato un campione casuale da $X \sim N(\mu, 1)$, tenendo presente l'Esempio 1.3.1, si ha

$$f_n(x_1, \dots, x_n; \mu) = \prod_{i=1}^n (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}(x_i - \mu)^2} = (2\pi)^{-\frac{n}{2}} e^{-\frac{n}{2}(s^2 + (\bar{x} - \mu)^2)} = (2\pi)^{-\frac{n}{2}} e^{-\frac{ns^2}{2}} e^{-\frac{n}{2}(\bar{x} - \mu)^2},$$

da cui si ottiene che

$$h_n(x_1, \dots, x_n) = (2\pi)^{-\frac{n}{2}} e^{-\frac{ns^2}{2}}$$

e

$$g(\bar{x}; \mu) = e^{-\frac{n}{2}(\bar{x} - \mu)^2}.$$

Dunque la media campionaria \bar{X} è stimatore sufficiente per μ . □

• **Esempio 2.5.2.** Dato un campione casuale da $X \sim E(0, \sigma)$, si ha

$$f_n(x_1, \dots, x_n; \sigma) = \prod_{i=1}^n \frac{1}{\sigma} e^{-\frac{x_i}{\sigma}} \mathbf{1}_{[0, \infty)}(x_i) = \sigma^{-n} e^{-\frac{n\bar{x}}{\sigma}} \prod_{i=1}^n \mathbf{1}_{[0, \infty)}(x_i),$$

da cui

$$h_n(x_1, \dots, x_n) = \prod_{i=1}^n \mathbf{1}_{[0, \infty)}(x_i)$$

e

$$g(\bar{x}; \sigma) = \sigma^{-n} e^{-\frac{n\bar{x}}{\sigma}}.$$

Si deve dunque concludere che la media campionaria \bar{X} è stimatore sufficiente per σ . □

• **Esempio 2.5.3.** Dato un campione casuale da $X \sim N(\mu, \nu)$, tenendo presente anche l'Esempio 1.3.2, si ha

$$f_n(x_1, \dots, x_n; \mu, \nu) = (2\pi)^{-\frac{n}{2}} \nu^{-\frac{n}{2}} e^{-\frac{n}{2\nu}(s^2 + (\bar{x} - \mu)^2)},$$

da cui

$$h_n(x_1, \dots, x_n) = (2\pi)^{-\frac{n}{2}}$$

e

$$g(\bar{x}, s^2; \mu, \nu) = \nu^{-\frac{n}{2}} e^{-\frac{n}{2\nu}(s^2 + (\bar{x} - \mu)^2)}.$$

Dunque lo stimatore (\bar{X}, S^2) è sufficiente per (μ, ν) . □

• **Esempio 2.5.4.** Dato il modello di regressione lineare dell'Esempio 1.1.8, tenendo presente l'Esempio 1.3.6, si ha

$$f_n(y_1, \dots, y_n; a, b, \nu) = (2\pi)^{-\frac{n}{2}} \nu^{-\frac{n}{2}} e^{-\frac{n}{2\nu}(s_y^2 - \frac{s_{zy}^2}{s_z^2} + (bs_z - \frac{s_{zy}^2}{s_z^2}) + (\bar{y} - a)^2)},$$

da cui si ottiene che

$$h_n(y_1, \dots, y_n) = (2\pi)^{-\frac{n}{2}}$$

e

$$g(\bar{y}, s_y^2, s_{zy}; a, b, \nu) = \nu^{-\frac{n}{2}} e^{-\frac{n}{2\nu}(s_y^2 - \frac{s_{zy}^2}{s_z^2} + (bs_z - \frac{s_{zy}}{s_z})^2 + (\bar{y} - a)^2)}.$$

Dunque lo stimatore $(\bar{Y}, S_y^2, S_{zy}^2)$ è sufficiente per (a, b, ν) . \square

Dalla fattorizzazione di Neyman, si può ottenere la distribuzione di $\tilde{\Theta}$ integrando come segue

$$\begin{aligned} f_{\tilde{\Theta}}(\tilde{\theta}; \theta) &= \int_{A(\tilde{\theta})} f_n(x_1, \dots, x_n; \theta) d\nu \\ &= g(\tilde{\theta}; \theta) \int_{A(\tilde{\theta})} h_n(x_1, \dots, x_n) d\nu = h^*(\tilde{\theta})g(\tilde{\theta}; \theta), \end{aligned}$$

dove la funzione

$$h^*(\tilde{\theta}) = \int_{A(\tilde{\theta})} h_n(x_1, \dots, x_n) d\nu$$

non dipende da θ , mentre l'insieme $A(\tilde{\theta})$ è stato definito nella §1.2.1. Di conseguenza, tenendo presente la precedente relazione e la fattorizzazione di Neyman, risulta

$$f_n(x_1, \dots, x_n | \tilde{\theta}) = \frac{f_n(x_1, \dots, x_n; \theta)}{f_{\tilde{\Theta}}(\tilde{\theta}; \theta)} = \frac{h_n(x_1, \dots, x_n)}{h^*(\tilde{\theta})},$$

ovvero se $\tilde{\Theta}$ è sufficiente per θ , la distribuzione del vettore di variabili casuali condizionato $(X_1, \dots, X_n | \tilde{\theta})$ non dipende θ . Inversamente, se $f_n(x_1, \dots, x_n | \tilde{\theta})$ non dipende da θ , allora

$$\begin{aligned} f_n(x_1, \dots, x_n; \theta) &= f_n(x_1, \dots, x_n | \tilde{\theta})f_{\tilde{\Theta}}(\tilde{\theta}; \theta) \\ &= f_n(x_1, \dots, x_n | \tilde{\theta})h^*(\tilde{\theta})g(\tilde{\theta}; \theta) = h_n(x_1, \dots, x_n)g(\tilde{\theta}; \theta), \end{aligned}$$

dove

$$h_n(x_1, \dots, x_n) = f_n(x_1, \dots, x_n | \tilde{\theta})h^*(\tilde{\theta})$$

e quindi $\tilde{\Theta}$ risulta sufficiente per θ grazie al criterio di Neyman. Si deve quindi concludere che $\tilde{\Theta}$ è sufficiente per θ se e solo se la distribuzione del vettore di variabili casuali condizionato $(X_1, \dots, X_n | \tilde{\theta})$ non dipende θ .

Questo fatto evidenzia come il campione (x_1, \dots, x_n) venga ottenuto in due fasi: in una prima fase l'esperimento casuale genera il valore $\tilde{\theta}$, ovvero seleziona l'insieme $A(\tilde{\theta})$ della partizione di \mathcal{C}_n secondo una legge di probabilità che dipende da θ , e in una seconda fase sceglie un elemento (x_1, \dots, x_n) di $A(\tilde{\theta})$ secondo una legge di probabilità che non dipende da θ . Dunque, si è ancora ottenuto che se $\tilde{\Theta}$ è sufficiente per θ la conoscenza del campione originale non aggiunge niente rispetto all'informazione che si ha su θ basandosi sul solo stimatore $\tilde{\Theta}$.

• **Esempio 2.5.5.** Dato un campione casuale da $X \sim N(\mu, \nu)$, tenendo presente l'Esempio 1.2.3 e l'Esempio 1.2.5, si ha che $\bar{X} \sim N(\mu, \nu/n)$ e $nS^2/\nu \sim \chi_{n-1}^2$ sono stimatori indipendenti. Dunque, tenendo presente anche l'Esempio 2.5.2, si ha

$$\begin{aligned} f_n(x_1, \dots, x_n | \bar{x}, s^2) &= \frac{f_n(x_1, \dots, x_n; \mu, \nu)}{f_{\bar{X}}(\bar{x}; \mu, \nu)f_{S^2}(s^2; \nu)} \\ &= \frac{(2\pi)^{-\frac{n}{2}} \nu^{-\frac{n}{2}} e^{-\frac{n}{2\nu}(s^2 + (\bar{x} - \mu)^2)}}{\left(\frac{2\pi\nu}{n}\right)^{-\frac{1}{2}} e^{-\frac{n}{2\nu}(\bar{x} - \mu)^2} \Gamma\left(\frac{n-1}{2}\right)^{-1} \frac{n}{2\nu} \left(\frac{ns^2}{2\nu}\right)^{\frac{1}{2}(n-1)-1} e^{-\frac{ns^2}{2\nu}}} \\ &= \pi^{-\frac{1}{2}(n-1)} n^{-\frac{n}{2}} \Gamma\left(\frac{n-1}{2}\right) (s^2)^{-\frac{1}{2}(n-3)}, \end{aligned}$$

e quindi la distribuzione del vettore di variabili casuali condizionato $(X_1, \dots, X_n \mid \bar{x}, s^2)$ non dipende da (μ, ν) . La precedente è in effetti una funzione di densità congiunta, come si potrebbe dimostrare integrando (in maniera laboriosa) per (x_1, \dots, x_n) con i vincoli $\sum_{i=1}^n x_i = n\bar{x}$ e $\sum_{i=1}^n (x_i - \bar{x})^2 = ns^2$. Dunque, si è di nuovo ottenuto il risultato dell'Esempio 2.5.3, ovvero lo stimatore (\bar{X}, S^2) è sufficiente per (μ, ν) . \square

• **Esempio 2.5.6.** Dato un campione casuale da $X \sim U(0, \delta)$, tenendo presente i risultati ottenuti nell'Esempio 1.3.4, si ha

$$f_n(x_1, \dots, x_n \mid x_{(n)}) = \frac{\delta^{-n} \mathbf{1}_{[0,1]}(x_{(n)}/\delta)}{(n/\delta)(x_{(n)}/\delta)^{n-1} \mathbf{1}_{[0,1]}(x_{(n)}/\delta)} = \frac{1}{nx_{(n)}^{n-1}},$$

e quindi la distribuzione del vettore condizionato di variabili casuali $(X_1, \dots, X_n \mid x_{(n)})$ non dipende dal parametro δ . Dunque, lo stimatore $X_{(n)}$ è sufficiente per δ . \square

Se $\tilde{\Theta}$ è uno stimatore sufficiente per θ , allora dati due campioni $(x_1, \dots, x_n) \in \mathcal{C}_n$ e $(y_1, \dots, y_n) \in \mathcal{C}_n$ si ha

$$(x_1, \dots, x_n) \in A(\tilde{\Theta}), (y_1, \dots, y_n) \in A(\tilde{\Theta}) \Rightarrow L(\theta; x_1, \dots, x_n) \propto L(\theta; y_1, \dots, y_n)$$

ovvero elementi appartenenti ad un dato $A(\tilde{\Theta})$ hanno verosimiglianze proporzionali. Quindi, una volta noto $A(\tilde{\Theta})$, siamo in grado di specificare la funzione di verosimiglianza associata a un qualunque $(x_1, \dots, x_n) \in \mathcal{C}_n$. Questo fatto implica che se si considera la trasformata biunivoca $\mathbf{U}(\tilde{\Theta})$, dal momento che la funzione $\mathbf{U}(\tilde{\Theta}(x_1, \dots, x_n))$ individua la medesima partizione dello spazio campionario, allora $\mathbf{U}(\tilde{\Theta})$ è ancora uno stimatore sufficiente per θ .

• **Esempio 2.5.7.** Dato un campione casuale da $X \sim N(\mu, \nu)$, nell'Esempio 2.5.4 è stato provato che lo stimatore (\bar{X}, S^2) è sufficiente per (μ, ν) . Di conseguenza, lo stimatore trasformato

$$(U_1, U_2) = (\bar{X}, \sqrt{S^2}) = (\bar{X}, S)$$

è sufficiente per (μ, ν) , dal momento che la trasformazione è biunivoca. \square

2.5.3. Stimatori sufficienti minimali. Al fine di ottenere la massima sintesi dell'informazione è importante considerare stimatori sufficienti che sono in grado di ottenere il massimo grado di aggregazione dei dati. Dato un modello statistico \mathcal{F}_θ , uno stimatore $\tilde{\Theta}$ è detto sufficiente minimale per θ se è sufficiente per θ e se per ogni $(x_1, \dots, x_n) \in \mathcal{C}_n$ e $(y_1, \dots, y_n) \in \mathcal{C}_n$ si ha

$$\tilde{\Theta}(x_1, \dots, x_n) = \tilde{\Theta}(y_1, \dots, y_n) \Leftrightarrow L(\theta; x_1, \dots, x_n) \propto L(\theta; y_1, \dots, y_n),$$

per ogni $\theta \in \Theta$. Dunque, $\tilde{\Theta}$ è sufficiente minimale per θ se assume lo stesso valore per due campioni se e solo se i due campioni hanno verosimiglianze proporzionali. Se uno stimatore è sufficiente minimale per θ , il rapporto $L(\theta; x_1, \dots, x_n)/L(\theta; y_1, \dots, y_n)$ deve essere una funzione che non dipende da θ se e solo se $\tilde{\Theta}(x_1, \dots, x_n) = \tilde{\Theta}(y_1, \dots, y_n)$.

Se si considera la trasformata biunivoca $\mathbf{U}(\tilde{\Theta})$, dal momento che la funzione $\mathbf{U}(\tilde{\Theta}(x_1, \dots, x_n))$ individua la medesima partizione dello spazio campionario, allora $\mathbf{U}(\tilde{\Theta})$ è ancora uno stimatore sufficiente minimale per θ .

• **Esempio 2.5.7.** Dato un campione casuale da $X \sim N(\mu, \nu)$, si ha

$$\frac{L(\mu, \nu; x_1, \dots, x_n)}{L(\mu, \nu; y_1, \dots, y_n)} = \frac{c \nu^{-\frac{n}{2}} e^{-\frac{n}{2\nu}(s_1^2 + (\bar{x}_1 - \mu)^2)}}{c \nu^{-\frac{n}{2}} e^{-\frac{n}{2\nu}(s_2^2 + (\bar{x}_2 - \mu)^2)}} = c e^{-\frac{n}{2\nu}(s_1^2 - s_2^2 + (\bar{x}_1 - \mu)^2 - (\bar{x}_2 - \mu)^2)},$$

dove \bar{x}_1, s_1^2 e \bar{x}_2, s_2^2 sono rispettivamente la media e la varianza campionaria relative a (x_1, \dots, x_n) e (y_1, \dots, y_n) . La precedente espressione non dipende da (μ, ν) se e solo se l'argomento della funzione esponenziale si annulla, ovvero quando $\bar{x}_1 = \bar{x}_2$ e $s_1^2 = s_2^2$. Si deve dunque concludere che lo stimatore (\bar{X}, S^2) è sufficiente minimale per (μ, ν) . \square

• **Esempio 2.5.8.** Dato un campione casuale da $X \sim E(0, \sigma)$, allora si ha

$$\frac{L(\sigma; x_1, \dots, x_n)}{L(\sigma; y_1, \dots, y_n)} = \frac{c \sigma^{-n} e^{-n\bar{x}_1/\sigma}}{c \sigma^{-n} e^{-n\bar{x}_2/\sigma}} = c e^{-\frac{n}{\sigma}(\bar{x}_1 - \bar{x}_2)},$$

dove \bar{x}_1 e \bar{x}_2 sono le medie campionarie relative a (x_1, \dots, x_n) e (y_1, \dots, y_n) . La precedente espressione non dipende dal parametro σ se e solo se l'argomento della funzione esponenziale si annulla, ovvero quando $\bar{x}_1 = \bar{x}_2$. Si deve dunque concludere che lo stimatore \bar{X} è sufficiente minimale per σ . \square

Se la distribuzione $f_n(x_1, \dots, x_n; \theta)$ specificata dal modello statistico \mathcal{F}_θ appartiene alla famiglia esponenziale in forma ridotta, allora $\tilde{\Theta}$ è ovviamente uno stimatore sufficiente per θ . Inoltre, dati due campioni $(x_1, \dots, x_n) \in \mathcal{C}_n$ e $(y_1, \dots, y_n) \in \mathcal{C}_n$, per questa famiglia si ha

$$\frac{L(\theta; x_1, \dots, x_n)}{L(\theta; y_1, \dots, y_n)} = \frac{c e^{\tilde{\Theta}(x_1, \dots, x_n)^\top \varphi(\theta) - \eta(\theta)}}{c e^{\tilde{\Theta}(y_1, \dots, y_n)^\top \varphi(\theta) - \eta(\theta)}} = c e^{(\tilde{\Theta}(x_1, \dots, x_n) - \tilde{\Theta}(y_1, \dots, y_n))^\top \varphi(\theta)}.$$

La precedente espressione non dipende dai parametri se e solo se $\tilde{\Theta}(x_1, \dots, x_n) = \tilde{\Theta}(y_1, \dots, y_n)$, ovvero $\tilde{\Theta}$ è sufficiente minimale per θ . Dunque, se la distribuzione specificata dal modello statistico appartiene alla famiglia esponenziale in forma ridotta, si può ottenere immediatamente lo stimatore sufficiente minimale per θ .

• **Esempio 2.5.9.** Dato un campione casuale da $X \sim N(\mu, \nu)$, tenendo presente i risultati dell'Esempio 2.4.13, lo stimatore sufficiente minimale per (μ, ν) è dato da $(\bar{X}, S^2 + \bar{X}^2)$. Questa è una trasformata biunivoca dello stimatore (\bar{X}, S^2) , che dunque a sua volta risulta sufficiente minimale per (μ, ν) , ovvero si sono verificati i risultati dell'Esempio 2.5.7. \square

2.5.4. Il metodo di Blackwell-Rao. Avendo a disposizione uno stimatore corretto e uno sufficiente, a partire da questi stimatori si può costruire un nuovo stimatore corretto e sufficiente che ha varianza minore di quello corretto iniziale. Per semplicità si analizza inizialmente il caso di un parametro.

Se $\tilde{\Theta}$ è sufficiente per θ e $\bar{\Theta}$ è corretto per θ , allora si consideri lo stimatore $U(\tilde{\Theta})$ tale che

$$U(\tilde{\theta}) = E[\bar{\Theta} | \tilde{\theta}].$$

Lo stimatore $U(\tilde{\Theta})$ è corretto per θ dal momento che

$$E[U(\tilde{\Theta})] = E_{\tilde{\theta}}[E[\bar{\Theta} | \tilde{\theta}]] = E[\bar{\Theta}] = \theta.$$

Inoltre, risulta

$$\text{Var}[\bar{\Theta}] = \text{Var}_{\tilde{\theta}}[E[\bar{\Theta} | \tilde{\theta}]] + E_{\tilde{\theta}}[\text{Var}[\bar{\Theta} | \tilde{\theta}]] = \text{Var}[U(\tilde{\Theta})] + E_{\tilde{\theta}}[\text{Var}[\bar{\Theta} | \tilde{\theta}]] \geq \text{Var}[U(\tilde{\Theta})],$$

essendo $E_{\tilde{\theta}}[\text{Var}[\bar{\Theta} | \tilde{\theta}]] \geq 0$. Dunque, lo stimatore $U(\tilde{\Theta})$ è corretto per θ con varianza minore di quella dello stimatore $\bar{\Theta}$. La procedura per la costruzione dello stimatore $U(\tilde{\Theta})$ è detto metodo di Blackwell-Rao.

Si consideri una ulteriore caratteristica di uno stimatore. Uno stimatore $\tilde{\Theta}$ è detto completo se per una qualsiasi trasformata U si ha $E[U(\tilde{\Theta})] = 0$ se e solo se $U(\tilde{\Theta}) = 0$ con probabilità 1. Questa proprietà implica l'unicità di una trasformata di $\tilde{\Theta}$ corretta per θ . Se quindi il metodo di Blackwell-Rao è applicato ad uno stimatore $\bar{\Theta}$ corretto per θ e a uno stimatore $\tilde{\Theta}$ sufficiente e completo per θ , allora

lo stimatore risultante $U(\tilde{\Theta})$ è unico. Questa procedura assicura il raggiungimento dello stimatore corretto per θ avente varianza minima.

• **Esempio 2.5.10.** Dato un campione casuale da $X \sim U(0, \delta)$, dall'Esempio 2.4.7 è noto che il problema di stima non è regolare e che quindi non è possibile ottenere un limite inferiore alla varianza di uno stimatore corretto per δ . Tuttavia, si può costruire lo stimatore corretto per δ con varianza minima attraverso il metodo di Blackwell-Rao. Tenendo presente i risultati ottenuti nell'Esempio 2.5.6, $X_{(n)}$ è stimatore sufficiente per δ . Se $U(X_{(n)})$ è una qualsiasi trasformata positiva è facile dimostrare che $X_{(n)}$ è anche stimatore completo. Inoltre lo stimatore $2X_1$ è corretto per δ , essendo $E[2X_1] = \delta$ come risulta dalle proprietà della Uniforme (vedi A.1.3). La variabile casuale condizionata $(X_1 | x_{(n)})$ è una variabile casuale con distribuzione mista, ovvero con funzione di ripartizione data da

$$F_{(X_1|x_{(n)})}(x_1) = \frac{n-1}{n} \frac{x_1}{x_{(n)}} \mathbf{1}_{[0, x_{(n)})}(x_1) + \frac{1}{n} \mathbf{1}_{[x_{(n)}, \infty)}(x_1),$$

da cui

$$E[2X_1 | x_{(n)}] = \int_0^{x_{(n)}} 2x_1 \frac{n-1}{n} \frac{1}{x_{(n)}} dx_1 + \frac{1}{n} 2x_{(n)} = \frac{n+1}{n} x_{(n)}.$$

Si deve dunque concludere che lo stimatore $(n+1)X_{(n)}/n$ è lo stimatore corretto per δ con varianza minima. \square

Il metodo di Blackwell-Rao può essere generalizzato al caso di un vettore di parametri θ . Se $\tilde{\Theta}$ è uno stimatore sufficiente per θ e $\bar{\Theta}$ è uno stimatore corretto per θ , allora si può dimostrare che lo stimatore

$$U(\tilde{\Theta}) = E[\bar{\Theta} | \tilde{\Theta}],$$

è corretto e sufficiente per θ con matrice di varianza-covarianza minore (nel senso spiegato nella §2.4.2) di quella di $\bar{\Theta}$. Inoltre, uno stimatore $\tilde{\Theta}$ è completo se per un qualsiasi trasformata U si ha $E[U(\tilde{\Theta})] = \mathbf{0}$ se e solo se $U(\tilde{\Theta}) = \mathbf{0}$ con probabilità 1. Anche in questo caso, se il metodo di Blackwell-Rao è applicato allo stimatore $\bar{\Theta}$ corretto per θ e allo stimatore $\tilde{\Theta}$ sufficiente e completo per θ , allora lo stimatore risultante $U(\tilde{\Theta})$ è corretto per θ con matrice di varianza-covarianza minima. Questa affermazione è da intendersi nel senso che la matrice $\text{Var}[U(\tilde{\Theta})] - \text{Var}[\bar{\Theta}]$ è definita negativa per un qualsiasi stimatore corretto $\bar{\Theta}$ per θ .

• **Esempio 2.5.11.** Dato un campione casuale da $X \sim U(\lambda, \lambda + \delta)$, il problema di stima non è regolare e quindi non è possibile ottenere un limite inferiore alla matrice di varianza-covarianza di uno stimatore corretto per (λ, δ) . Tuttavia, si può costruire lo stimatore corretto per (λ, δ) con matrice di varianza-covarianza minima con il metodo di Blackwell-Rao. Tenendo presente i risultati ottenuti nell'Esempio 1.3.8, è facile verificare attraverso il criterio di fattorizzazione di Neyman che $(X_{(1)}, X_{(n)})$ è uno stimatore sufficiente per (λ, δ) . Se $U(X_{(1)}, X_{(n)})$ è una qualsiasi trasformata si può dimostrare che $(X_{(1)}, X_{(n)})$ è anche uno stimatore completo. Inoltre, tenendo presente l'Esempio 1.2.6 e le proprietà della Beta (vedi §A.1.5), allora si ha

$$E[X_{(1)}] = \lambda + \frac{\delta}{n+1}$$

e

$$E[X_{(n)}] = \lambda + \frac{n\delta}{n+1},$$

mentre per le proprietà della Uniforme si ha

$$E[X_1] = \lambda + \frac{\delta}{2},$$

per cui risulta che

$$\left(\frac{n+1}{n-1} X_{(1)} - \frac{2}{n-1} X_1, \frac{n+1}{n-1} (X_{(n)} - X_{(1)}) \right)$$

è uno stimatore corretto per (λ, δ) . La variabile casuale condizionata $(X_1 | x_{(1)}, x_{(n)})$ è una variabile casuale con distribuzione mista, ovvero con funzione di ripartizione

$$F_{(X_1|x_{(1)},x_{(n)})}(x_1) = \frac{1}{n} \mathbf{1}_{[x_{(1)},\infty)}(x_1) + \frac{n-2}{n} \frac{x_1}{x_{(n)} - x_{(1)}} \mathbf{1}_{(x_{(1)},x_{(n)})}(x_1) + \frac{1}{n} \mathbf{1}_{[x_{(n)},\infty)}(x_1),$$

da cui

$$E[X_1 | x_{(1)}, x_{(n)}] = \frac{1}{n} x_{(1)} + \int_{x_{(1)}}^{x_{(n)}} x_1 \frac{n-2}{n} \frac{1}{x_{(n)} - x_{(1)}} dx_1 + \frac{1}{n} x_{(n)} = \frac{x_{(1)} + x_{(n)}}{2}.$$

Dunque, dal momento che

$$\begin{aligned} E\left[\frac{n+1}{n-1} X_{(1)} - \frac{2}{n-1} X_1 \mid x_{(1)}, x_{(n)} \right] &= \frac{n+1}{n-1} x_{(1)} - \frac{2}{n-1} E[X_1 \mid x_{(1)}, x_{(n)}] \\ &= \frac{1}{n-1} (nx_{(1)} - x_{(n)}) \end{aligned}$$

e

$$E\left[\frac{n+1}{n-1} (X_{(n)} - X_{(1)}) \mid x_{(1)}, x_{(n)} \right] = \frac{n+1}{n-1} (x_{(n)} - x_{(1)}),$$

allora si deve dunque concludere che lo stimatore

$$\left(\frac{1}{n-1} (nX_{(1)} - X_{(n)}), \frac{n+1}{n-1} (X_{(n)} - X_{(1)}) \right)$$

è corretto per (λ, δ) con matrice di varianza-covarianza minima. □

Capitolo 3

Le stime di massima verosimiglianza

3.1. Il metodo della massima verosimiglianza

3.1.1. La definizione di stima di massima verosimiglianza. Anche se esistono in letteratura statistica vari metodi per ottenere stime per punti, nel seguito viene considerato solo il cosiddetto metodo della massima verosimiglianza dal momento che è di facile utilizzo e gode di proprietà ottime.

Questo metodo di stima consiste nello scegliere il valore del parametro che massimizza la probabilità di ottenere proprio il campione che è stato estratto. Se si considera un modello statistico \mathcal{F}_θ , dato il campione (x_1, \dots, x_n) si dice stima di massima verosimiglianza di θ quel valore $\hat{\theta} \in \Theta$ tale che

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta).$$

Dal momento che la funzione logaritmo è monotona crescente, la precedente condizione è equivalente a

$$l(\hat{\theta}) = \max_{\theta \in \Theta} l(\theta),$$

che spesso risulta più conveniente dal punto di vista del calcolo. Tenendo presente la precedente definizione, niente assicura l'esistenza e l'unicità della stima di verosimiglianza. Infine, la stima di massima verosimiglianza è la realizzazione campionaria $\hat{\theta} = \hat{\Theta}(x_1, \dots, x_n)$ di uno stimatore $\hat{\Theta}$, detto appunto stimatore di massima verosimiglianza.

Talvolta si è interessati solo ad alcuni componenti del vettore di parametri θ . In questo caso, θ può essere ripartito in due componenti $\theta = (\theta_A^\top, \theta_B^\top)^\top$, tali che $\theta_A \in \Theta_A$ e $\theta_B \in \Theta_B$ e dove $\Theta = \Theta_A \times \Theta_B$. La cosiddetta verosimiglianza profilo relativa a θ_A è definita come

$$L_p(\theta_A) = L_p(\theta_A; x_1, \dots, x_n) = \max_{\theta_B \in \Theta_B} L(\theta_A, \theta_B).$$

La verosimiglianza profilo permette di analizzare la verosimiglianza rispetto a θ_A eliminando l'effetto dovuto a θ_B . Il valore $\hat{\theta}_B = \hat{\theta}_B(\theta_A)$, tale che $L_p(\theta_A) = L(\theta_A, \hat{\theta}_B)$, è in effetti la stima di massima verosimiglianza di θ_B noto θ_A . Infine, la log-verosimiglianza profilo relativa a θ_A è data da

$$l_p(\theta_A) = l_p(\theta_A; x_1, \dots, x_n) = \max_{\theta_B \in \Theta_B} l(\theta_A, \theta_B).$$

• **Esempio 3.1.1.** Dato un campione casuale da $X \sim N(\mu, 1)$, tenendo presente l'Esempio 1.3.1 si ha

$$l(\mu) = \log c - \frac{n}{2} (s^2 + (\bar{x} - \mu)^2) \leq \log c - \frac{ns^2}{2} = l(\bar{x}) = \max_{\mu \in \mathbb{R}} l(\mu),$$

dal momento che $(\bar{x} - \mu)^2 \geq 0$. Dunque $\hat{\mu} = \bar{x}$ è la stima di massima verosimiglianza di μ . Lo stimatore di massima verosimiglianza risulta \bar{X} . Dall'Esempio 1.2.3 si ha che $\bar{X} \sim N(\mu, 1/n)$. Di conseguenza, lo stimatore di massima verosimiglianza risulta corretto, coerente e sufficiente minimale per μ , ed è inoltre efficiente.

Se si suppone che nel medesimo modello lo spazio parametrico sia dato da $\Theta = [0, \infty)$, allora

$$\begin{aligned}
l(\mu) &= \log c - \frac{n}{2} (s^2 + (\bar{x} - \mu)^2) + \log \mathbf{1}_{[0, \infty)}(\mu) \\
&\leq \max \left(\log c - \frac{ns^2}{2} + \log \mathbf{1}_{[0, \infty)}(\bar{x}), \log c - \frac{n}{2} (s^2 + \bar{x}^2) \right) \\
&= \max(l(\bar{x}), l(0)) = \max_{\mu \in [0, \infty)} l(\mu),
\end{aligned}$$

e dunque la stima di massima verosimiglianza è data da

$$\hat{\mu} = \begin{cases} \bar{x} & \bar{x} > 0 \\ 0 & \bar{x} \leq 0 \end{cases},$$

ovvero $\hat{\mu} = \max(\bar{x}, 0)$. Lo stimatore di massima verosimiglianza risulta $\max(\bar{X}, 0)$, che ovviamente non possiede una distribuzione Normale. In questo caso è più complesso determinare le proprietà dello stimatore di massima verosimiglianza.

Se infine si suppone che nel medesimo modello lo spazio parametrico sia dato da $\Theta = (0, \infty)$, allora la stima di massima verosimiglianza risulta $\hat{\mu} = \bar{x}$ se $\bar{x} > 0$, mentre non esiste se $\bar{x} \leq 0$ in quanto il valore $\mu = 0$ non fa parte dello spazio parametrico. In una situazione di questo genere, dal momento che $\mu = 0$ è un valore per il quale la verosimiglianza ha senso matematico, si usa porre $\hat{\mu} = 0$ e nel seguito sarà sempre fatto uso di questa convenzione. Sulla base di questa osservazione, lo stimatore di massima verosimiglianza risulta dunque dato da $\max(\bar{X}, 0)$. \square

• **Esempio 3.1.2.** Dato un campione casuale da $X \sim N(\mu, \nu)$, tenendo presente l'Esempio 1.3.2 si ha

$$\begin{aligned}
l(\mu, \nu) &= \log c - \frac{n}{2} \log \nu - \frac{n}{2\nu} (s^2 + (\bar{x} - \mu)^2) \\
&\leq \log c - \frac{n}{2} \log \nu - \frac{ns^2}{2\nu} = l(\bar{x}, \nu) = \max_{\mu \in \mathbb{R}} l(\mu, \nu),
\end{aligned}$$

essendo $(\bar{x} - \mu)^2 \geq 0$. Inoltre, dal momento che in generale è valida la relazione $\log x \leq x - 1$ essendo la funzione logaritmo concava, allora di conseguenza risulta

$$\log \left(\frac{d}{v} \right) \leq \frac{d}{v} - 1,$$

ovvero

$$-\log v - \frac{d}{v} \leq -\log d - 1$$

per d costante positiva. Adoperando quest'ultima disuguaglianza con $d = s^2$ nella relazione relativa alla log-verosimiglianza si ha

$$l(\bar{x}, \nu) = \log c - \frac{n}{2} \log \nu - \frac{ns^2}{2\nu} \leq \log c - \frac{n}{2} \log s^2 - \frac{n}{2} = l(\bar{x}, s^2) = \max_{\mu \in \mathbb{R}, \nu \in \mathbb{R}^+} l(\mu, \nu).$$

Dunque, $(\hat{\mu}, \hat{\nu}) = (\bar{x}, s^2)$ è la stima di massima verosimiglianza di (μ, ν) . Di conseguenza, lo stimatore di massima verosimiglianza risulta (\bar{X}, S^2) . Tenendo presente i risultati dell'Esempio 1.2.5, si ha che $\bar{X} \sim N(\mu, \nu/n)$ e $nS^2/\nu \sim \chi_{n-1}^2$ e questi due stimatori sono indipendenti. Si può verificare facilmente che lo stimatore di massima verosimiglianza è asintoticamente corretto, coerente e sufficiente minimale per (μ, ν) , ma non è efficiente non essendo corretto.

Tenendo presente i risultati precedenti la log-verosimiglianza profilo relativa a μ è data da

$$l_p(\mu) = \log c - \frac{n}{2} \log(s^2 + (\bar{x} - \mu)^2) - \frac{n}{2} = l(\mu, s^2 + (\bar{x} - \mu)^2), \mu \in \mathbb{R},$$

mentre la log-verosimiglianza profilo relativa a ν è data da

$$l_p(v) = \log c - \frac{n}{2} \log v - \frac{ns^2}{2v} = l(\bar{x}, v), v \in \mathbb{R}^+.$$

I grafici delle verosimiglianze profilo standardizzate per $n = 5$, $\bar{x} = 1$ e $s^2 = 2$ sono riportati nelle Figure 3.1.1 e 3.1.2. \square

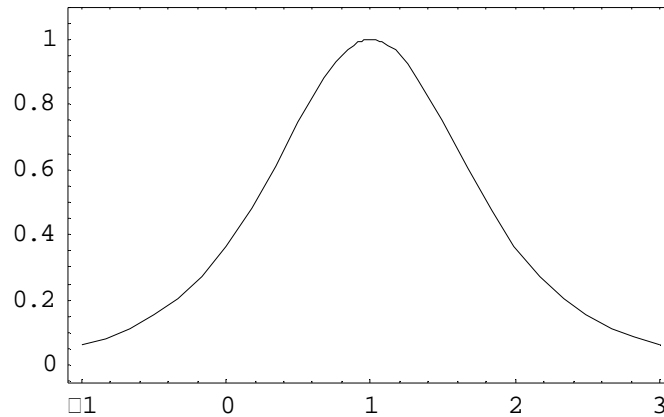


Figura 3.1.1. Verosimiglianza profilo standardizzata relativa a μ per $n = 5$, $\bar{x} = 1$ e $s^2 = 2$.

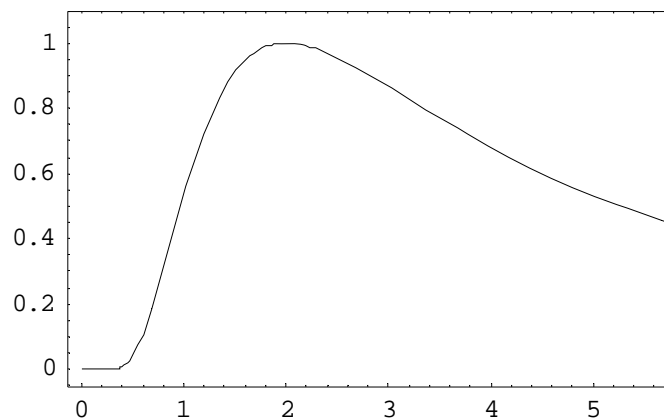


Figura 3.1.2. Verosimiglianza profilo standardizzata relativa a v per $n = 5$, $\bar{x} = 1$ e $s^2 = 2$.

• **Esempio 3.1.3.** Dato un campione casuale da $X \sim E(0, \sigma)$, tenendo presente l'Esempio 1.3.3 e la disuguaglianza dell'Esempio 3.1.2 con $d = \bar{x}$, si ha

$$l(\sigma) = \log c - n \log \sigma - \frac{n\bar{x}}{\sigma} \leq \log c - n \log \bar{x} - n = l(\bar{x}) = \max_{\sigma \in \mathbb{R}^+} l(\sigma),$$

ovvero $\hat{\sigma} = \bar{x}$ è la stima di massima verosimiglianza di σ . Lo stimatore di massima verosimiglianza risulta \bar{X} e sulla base dei risultati dell'Esempio 1.2.4 si ha $\bar{X} \sim G(0, \sigma/n; n)$. Lo stimatore di massima verosimiglianza risulta corretto, coerente e sufficiente minimale per σ , ed inoltre è anche efficiente. \square

• **Esempio 3.1.4.** Dato un campione casuale da $X \sim U(0, \delta)$, sulla base dei risultati ottenuti nell'Esempio 1.3.4, si ha

$$l(\delta) = \log c - n \log \delta + \log \mathbf{1}_{[x_{(n)}, \infty)}(\delta) \leq \log c - n \log x_{(n)} = l(x_{(n)}) = \max_{\delta \in \mathbb{R}^+} l(\delta),$$

dove si è tenuto presente che la funzione logaritmo è crescente e che la log-verosimiglianza può essere definita per $\delta \geq x_{(n)}$ (vedi nota finale all'Esempio 3.1.1). Dunque, $\hat{\delta} = x_{(n)}$ è la stima di massima verosimiglianza di δ . Lo stimatore di massima verosimiglianza risulta $X_{(n)}$ e, tenendo presente i risultati dell'Esempio 1.2.6, si ha $X_{(n)} \sim Be(0, \delta; n, 1)$. Lo stimatore di massima verosimiglianza risulta asintoticamente corretto, coerente e sufficiente per δ . Dal momento che non si è in un problema

di stima regolare non ha senso considerare l'efficienza dello stimatore. Lo stimatore di massima verosimiglianza è comunque una trasformata lineare dello stimatore corretto con varianza minima e per n elevato è approssimativamente equivalente a quest'ultimo stimatore. \square

• **Esempio 3.1.5.** Dato il modello di regressione lineare dell'Esempio 1.1.8, tenendo presente l'Esempio 1.3.6, si ha

$$\begin{aligned} l(a, b, v) &= \log c - \frac{n}{2} \log v - \frac{n}{2v} \left(s_y^2 - \frac{s_{zy}^2}{s_z^2} + \left(bs_z - \frac{s_{zy}}{s_z} \right)^2 + (\bar{y} - a)^2 \right) \\ &\leq \log c - \frac{n}{2} \log v - \frac{n}{2v} \left(s_y^2 - \frac{s_{zy}^2}{s_z^2} \right) = l(\hat{a}, \hat{b}, v) = \max_{a \in \mathbb{R}, b \in \mathbb{R}} l(a, b, v), \end{aligned}$$

dal momento che $(bs_z - s_{zy}/s_z)^2 \geq 0$ e $(\bar{y} - a)^2 \geq 0$, e dove $\hat{a} = \bar{y}$ e $\hat{b} = s_{zy}/s_z^2$. Tenendo in considerazione inoltre la disuguaglianza dell'Esempio 3.1.2 con $d = s_y^2 - s_{zy}^2/s_z^2$, si ha

$$\begin{aligned} l(\hat{a}, \hat{b}, v) &= \log c - \frac{n}{2} \log v - \frac{n}{2v} \left(s_y^2 - \frac{s_{zy}^2}{s_z^2} \right) \\ &\leq \log c - \frac{n}{2} \log \left(s_y^2 - \frac{s_{zy}^2}{s_z^2} \right) - \frac{n}{2} = l(\hat{a}, \hat{b}, s_r^2) = \max_{a \in \mathbb{R}, b \in \mathbb{R}, v \in \mathbb{R}^+} l(a, b, v), \end{aligned}$$

dove $s_r^2 = s_y^2 - s_{zy}^2/s_z^2$. Sulla base dei risultati dell'Esempio 1.3.6, si ottiene che

$$s_r^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{a} - \hat{b}z_i)^2.$$

Dunque, s_r^2 è in effetti la media degli scarti al quadrato dei valori osservati y_i dai valori stimati $\hat{a} + \hat{b}z_i$. Si deve dunque concludere che $(\hat{a}, \hat{b}, s_r^2)$ è la stima di massima verosimiglianza di (a, b, v) . Lo stimatore di massima verosimiglianza è $(\hat{A}, \hat{B}, S_r^2) = (\bar{Y}, S_{zy}/s_z^2, S_y^2 - S_{zy}^2/s_z^2)$.

Gli stimatori \hat{A} e \hat{B} sono combinazioni lineari delle Y_i con i coefficienti delle combinazioni rispettivamente pari a $c_i = 1/n$ e $c_i = z_i/(ns_z^2)$. Per le proprietà della Normale (vedi §A.1.2), \hat{A} e \hat{B} sono statistiche indipendenti distribuite come $\hat{A} \sim N(a, v/n)$ e $\hat{B} \sim N(b, v/ns_z^2)$. Inoltre, S_r^2 dipende dalle variabili casuali $Y_i - \hat{A} - \hat{B}z_i$, che a loro volta sono combinazioni lineari delle Y_i con i coefficienti della combinazione pari rispettivamente a

$$c_{ij} = \begin{cases} -\frac{1}{n} - \frac{z_i z_j}{ns_z^2} & j \neq i \\ 1 - \frac{1}{n} - \frac{z_i^2}{ns_z^2} & j = i \end{cases}.$$

Dunque, ancora per le proprietà della Normale (vedi §A.1.2), con una dimostrazione simile a quella dell'Esempio 1.2.5, si può verificare che \hat{A} , \hat{B} e S_r^2 sono statistiche indipendenti. Inoltre, tenendo presente la relazione dell'Esempio 1.3.6, si ha

$$\begin{aligned} \sum_{i=1}^n \frac{(Y_i - a - bz_i)^2}{v} &= \frac{n}{v} \left(S_y^2 - \frac{S_{zy}^2}{s_z^2} \right) + \frac{n}{v} (bs_z - \hat{B}s_z)^2 + \frac{n}{v} (\bar{Y} - a)^2 \\ &= \frac{nS_r^2}{v} + \frac{(\hat{B} - b)^2}{v/(ns_z^2)} + \frac{(\bar{Y} - a)^2}{v/n}. \end{aligned}$$

Dal momento che

$$\sum_{i=1}^n \frac{(Y_i - a - bz_i)^2}{v} \sim \chi_n^2$$

e che \widehat{A} , \widehat{B} e S_r^2 sono indipendenti, allora per la proprietà *i*) della §A.3.1 deve essere necessariamente

$$\frac{nS_r^2}{v} \sim \chi_{n-2}^2.$$

Lo stimatore di massima verosimiglianza risulta asintoticamente corretto, coerente e sufficiente minimale per (a, b, v) . Tenendo presente i risultati precedenti, la log-verosimiglianza profilo relativa a (a, b) è data da

$$\begin{aligned} l_p(a, b) &= \log c - \frac{n}{2} \log(s_r^2 + s_z^2(\widehat{b} - b)^2 + (\widehat{a} - a)^2) - \frac{n}{2} \\ &= l(a, b, s_r^2 + s_z^2(\widehat{b} - b)^2 + (\widehat{a} - a)^2), (a, b) \in \mathbb{R}^2, \end{aligned}$$

mentre la log-verosimiglianza profilo relativa a a è

$$l_p(a) = \log c - \frac{n}{2} \log(s_r^2 + (\widehat{a} - a)^2) - \frac{n}{2} = l(a, \widehat{b}, s_r^2 + (\widehat{a} - a)^2), a \in \mathbb{R},$$

e la log-verosimiglianza profilo relativa a b è

$$l_p(b) = \log c - \frac{n}{2} \log(s_r^2 + s_z^2(\widehat{b} - b)^2) - \frac{n}{2} = l(\widehat{a}, b, s_r^2 + s_z^2(\widehat{b} - b)^2), b \in \mathbb{R}.$$

I grafici delle verosimiglianze profilo standardizzate per $n = 5$, $s_z^2 = 2$, $\widehat{a} = 0$, $\widehat{b} = 2$ e $s_r^2 = 1$ sono riportati nelle Figure 3.1.3, 3.1.4 e 3.1.5. □

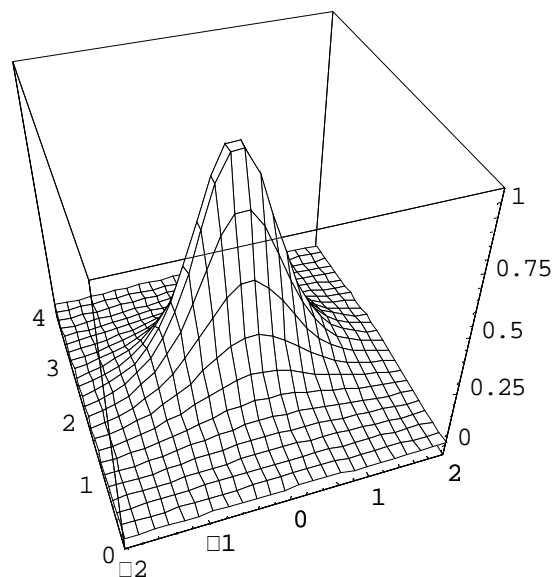


Figura 3.1.3. Verosimiglianza profilo standardizzata relativa a (a, b) per $n = 5$, $s_z^2 = 2$, $\widehat{a} = 0$, $\widehat{b} = 2$ e $s_r^2 = 1$.

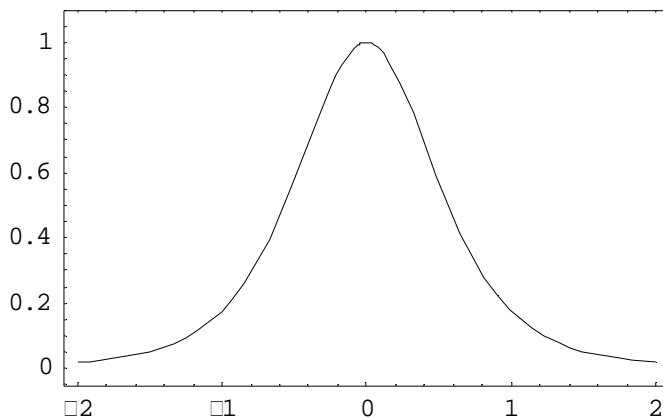


Figura 3.1.4. Verosimiglianza profilo standardizzata relativa a a per $n = 5$, $s_z^2 = 2$, $\hat{a} = 0$, $\hat{b} = 2$ e $s_r^2 = 1$.

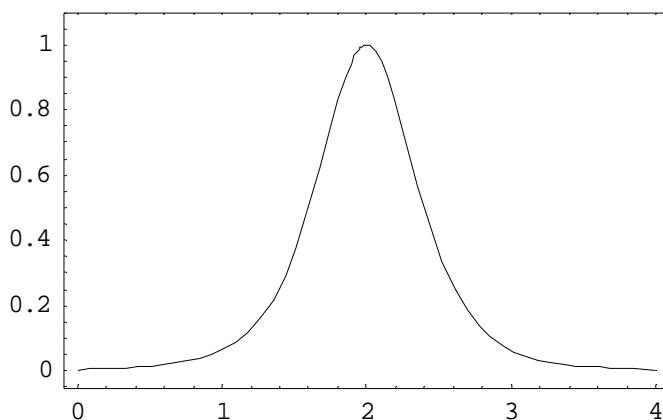


Figura 3.1.5. Verosimiglianza profilo standardizzata relativa a b per $n = 5$, $s_z^2 = 2$, $\hat{a} = 0$, $\hat{b} = 2$ e $s_r^2 = 1$.

• **Esempio 3.1.6.** Dato un campione casuale da $X \sim E(\lambda, \sigma)$, sulla base dei risultati ottenuti nell'Esempio 1.3.7, si ha

$$\begin{aligned} l(\lambda, \sigma) &= \log c - n \log \sigma - n \frac{\bar{x} - \lambda}{\sigma} + \log \mathbf{1}_{(-\infty, x_{(1)}]}(\lambda) \\ &\leq \log c - n \log \sigma - n \frac{\bar{x} - x_{(1)}}{\sigma} = l(x_{(1)}, \sigma) = \max_{\lambda \in \mathbb{R}} l(\lambda, \sigma), \end{aligned}$$

dove si è tenuto presente che la funzione $(\bar{x} - \lambda)$ è decrescente in λ e la log-verosimiglianza può essere definita per $x_{(1)} \leq \lambda$ (vedi nota finale all'Esempio 3.1.1). Adoperando la disuguaglianza dell'Esempio 3.1.2 con $d = \bar{x} - x_{(1)}$ nella precedente relazione, si ha inoltre

$$\begin{aligned} l(x_{(1)}, \sigma) &= \log c - n \log \sigma - n \frac{\bar{x} - x_{(1)}}{\sigma} \leq \log c - n \log(\bar{x} - x_{(1)}) - n \\ &= l(x_{(1)}, \bar{x} - x_{(1)}) = \max_{\lambda \in \mathbb{R}, \sigma \in \mathbb{R}^+} l(\lambda, \sigma). \end{aligned}$$

Dunque, $(\hat{\lambda}, \hat{\sigma}) = (x_{(1)}, \bar{x} - x_{(1)})$ è la stima di massima verosimiglianza di (λ, σ) . Lo stimatore di massima verosimiglianza risulta $(X_{(1)}, \bar{X} - X_{(1)})$. Non è agevole determinare la funzione di densità congiunta di questo stimatore e dunque non è immediato ottenere le relative proprietà. Inoltre, la log-verosimiglianza profilo relativa a λ risulta

$$l_p(\lambda) = \log c - n \log(\bar{x} - \lambda) - n + \log \mathbf{1}_{(-\infty, x_{(1)}]}(\lambda) = l(\lambda, \bar{x} - \lambda), \lambda \in \mathbb{R},$$

mentre la log-verosimiglianza profilo relativa a σ è data da

$$l_p(\sigma) = \log c - n \log \sigma - n \frac{\bar{x} - x_{(1)}}{\sigma} = l(x_{(1)}, \sigma), \sigma \in \mathbb{R}^+.$$

I grafici delle verosimiglianze profilo standardizzate per $n = 5$, $\bar{x} = 2$ e $x_{(1)} = 1$ sono riportati nelle Figure 3.1.6 e 3.1.7. □

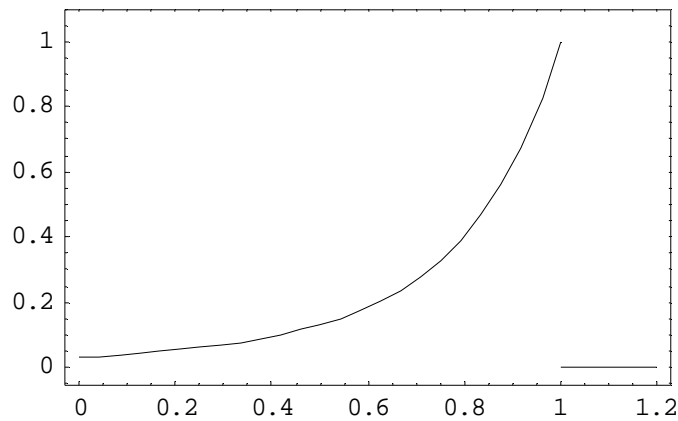


Figura 3.1.6. Verosimiglianza profilo standardizzata relativa a λ per $n = 5$, $\bar{x} = 2$ e $x_{(1)} = 1$.

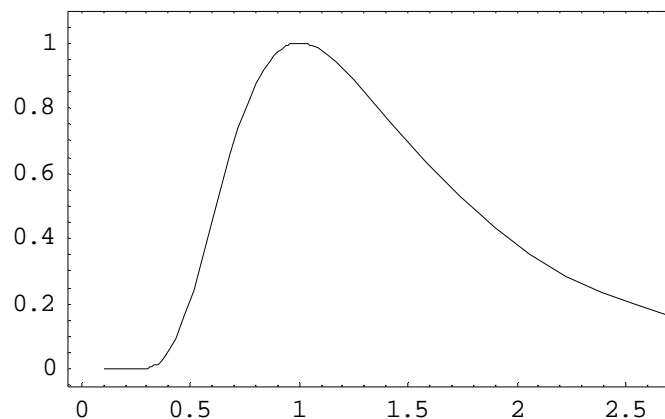


Figura 3.1.7. Verosimiglianza profilo standardizzata relativa a σ per $n = 5$, $\bar{x} = 2$ e $x_{(1)} = 1$.

• **Esempio 3.1.7.** Dato un campione casuale da $X \sim L(\mu, 1)$, sulla base dei risultati dell'Esempio 1.3.10, $l(\mu)$ raggiunge il massimo quando la funzione

$$h(\mu) = \sum_{i=1}^n |x_{(i)} - \mu|$$

raggiunge il minimo. Per n dispari, ovvero $n = 2l + 1$, il minimo di $h(\mu)$ è raggiunto per $x_{(l+1)}$, mentre per n pari, ovvero $n = 2l$, il minimo di $h(\mu)$ è raggiunto per qualsiasi valore compreso fra $x_{(l)}$ e $x_{(l+1)}$. Dunque, in questo ultimo caso esistono infinite stime di verosimiglianza. Tenendo presente la definizione data nella §1.2.4, si deve concludere che il minimo di $h(\mu)$ è raggiunto per la mediana campionaria \tilde{x} se n è dispari e quindi $\hat{\mu} = \tilde{x}$ è la stima di massima verosimiglianza di μ . Di conseguenza, lo stimatore di massima verosimiglianza risulta \tilde{X} se n è dispari, mentre non è univocamente definito se n è pari. □

3.1.2. Alcuni aspetti relativi al calcolo delle stime di massima verosimiglianza. Negli esempi della §3.1.1 la massimizzazione della verosimiglianza è stata volutamente ottenuta senza ricorrere all'uso delle derivate. I motivi di questa scelta sono vari: *i*) innanzitutto nel punto di massimo può non esistere la derivata (vedi Esempio 3.1.7); *ii*) il punto di massimo potrebbe essere sulla frontiera dello

spazio parametrico (vedi Esempio 3.1.4); *iii*) la massimizzazione con l'uso delle derivate comporta il calcolo del vettore delle derivate prime e della matrice hessiana che può essere oneroso quando il numero dei parametri è elevato (vedi Esempio 3.1.6).

Quando tuttavia la condizione *c*) della §2.1.2 è verificata e se risulta $l(\boldsymbol{\theta}) \rightarrow -\infty$ quando $\boldsymbol{\theta}$ si avvicina alla frontiera di Θ , la stima di massima verosimiglianza $\hat{\boldsymbol{\theta}}$ è una delle soluzioni dell'equazione vettoriale, detta equazione di verosimiglianza, data da

$$\mathbf{s}_n(\boldsymbol{\theta}; x_1, \dots, x_n) = \frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}; x_1, \dots, x_n) = \mathbf{0}.$$

La stima di massima verosimiglianza si ottiene per quella soluzione per cui si ha il massimo assoluto della verosimiglianza. Inoltre, se la matrice hessiana cambiata di segno

$$\hat{\mathbf{I}}_n(\boldsymbol{\theta}) = -\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{s}_n(\boldsymbol{\theta}; x_1, \dots, x_n) = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} l(\boldsymbol{\theta}; x_1, \dots, x_n)$$

è definita positiva, allora la log-verosiglianza risulta concava. In questo caso esiste un sola soluzione all'equazione di verosimiglianza che corrisponde appunto alla stima di massima verosimiglianza. La quantità $\hat{\mathbf{I}}_n(\hat{\boldsymbol{\theta}})$ è detta informazione osservata di Fisher.

• **Esempio 3.1.8.** Dato un campione casuale da $X \sim Bi(1, p)$, tenendo presente i risultati ottenuti nell'Esempio 1.3.5, si ha che $l(p)$ è derivabile su tutto lo spazio parametrico $(0, 1)$ e

$$\lim_{p \rightarrow 0} l(p) = \lim_{p \rightarrow 1} l(p) = -\infty.$$

Inoltre, si ha la seguente equazione di verosimiglianza

$$s_n(p; x_1, \dots, x_n) = \frac{d}{dp} l(p) = \frac{n\bar{x}}{p} - \frac{n - n\bar{x}}{1 - p} = 0,$$

che ha un'unica soluzione $\hat{p} = \bar{x}$. Infine, essendo

$$\hat{I}_n(p) = -\frac{d^2}{dp^2} l(p) = \frac{n\bar{x}}{p^2} + \frac{n - n\bar{x}}{(1 - p)^2} > 0,$$

allora $\hat{p} = \bar{x}$ è la stima di massima verosimiglianza. Lo stimatore di verosimiglianza risulta \bar{X} e per le proprietà della Binomiale (vedi §A.2.2) si ha $n\bar{X} \sim Bi(n, p)$. Lo stimatore di massima verosimiglianza risulta corretto, coerente e sufficiente minimale per p , ed inoltre è anche efficiente. \square

• **Esempio 3.1.9.** Dato un campione casuale da $X \sim N(\mu, \nu)$, la condizione *c*) della §2.1.2 è verificata e $l(\mu, \nu) \rightarrow -\infty$ quando ci si avvicina alla frontiera dello spazio parametrico. Si ha la seguente equazione di verosimiglianza

$$\mathbf{s}_n(\mu, \nu; x_1, \dots, x_n) = \begin{pmatrix} \frac{\partial}{\partial \mu} l(\mu, \nu) \\ \frac{\partial}{\partial \nu} l(\mu, \nu) \end{pmatrix} = \begin{pmatrix} \frac{n}{\nu} (\bar{x} - \mu) \\ -\frac{n}{2\nu} + \frac{n}{2\nu^2} (s^2 + (\bar{x} - \mu)^2) \end{pmatrix} = \mathbf{0},$$

che ha un'unica soluzione $(\hat{\mu}, \hat{\nu}) = (\bar{x}, s^2)$. Inoltre, essendo

$$\begin{aligned} \hat{\mathbf{I}}_n(\mu, \nu) &= -\begin{pmatrix} \frac{\partial^2}{\partial \mu^2} l(\mu, \nu) & \frac{\partial^2}{\partial \mu \partial \nu} l(\mu, \nu) \\ \frac{\partial^2}{\partial \mu \partial \nu} l(\mu, \nu) & \frac{\partial^2}{\partial \nu^2} l(\mu, \nu) \end{pmatrix} \\ &= \begin{pmatrix} \frac{n}{\nu} & \frac{n}{\nu^2} (\bar{x} - \mu) \\ \frac{n}{\nu^2} (\bar{x} - \mu) & -\frac{n}{2\nu^2} + \frac{n}{\nu^3} (s^2 + (\bar{x} - \mu)^2) \end{pmatrix}, \end{aligned}$$

si ottiene

$$\widehat{\mathbf{I}}_n(\bar{x}, s^2) = \begin{pmatrix} \frac{n}{s^2} & 0 \\ 0 & \frac{n}{2s^4} \end{pmatrix},$$

che è una matrice definita positiva. Dal momento che l'equazione di verosimiglianza ha un'unica soluzione e che il massimo non può essere sulla frontiera dello spazio parametrico, si deve concludere che $(\widehat{\mu}, \widehat{v}) = (\bar{x}, s^2)$ è la stima di massima verosimiglianza di (μ, v) . Dunque, i risultati dell'Esempio 3.1.2 sono stati ottenuti di nuovo. \square

Anche se la condizione c) della §2.1.2 è verificata, può accadere che l'equazione di verosimiglianza non abbia soluzioni esplicite e quindi in questo caso si deve ricorrere a tecniche numeriche. Un metodo iterativo per ottenere un'approssimazione della stima di massima verosimiglianza è quello di Newton-Raphson. Partendo da una stima iniziale $\widehat{\theta}_0$ (anche rozza e ottenuta eventualmente da un metodo alternativo alla massima verosimiglianza), il procedimento è basato sull'equazione ricorsiva

$$\widehat{\theta}_{j+1} = \widehat{\theta}_j + \widehat{\mathbf{I}}_n(\widehat{\theta}_j)^{-1} \mathbf{s}_n(\widehat{\theta}_j), \quad j = 0, 1, \dots$$

L'iterazione ha termine per quel j tale che $\|\widehat{\theta}_j - \widehat{\theta}_{j-1}\| < \epsilon$ con $\epsilon > 0$ errore prefissato e in questo caso l'approssimazione della stima di massima verosimiglianza è data da $\widehat{\theta}_j$. Se la log-verosimiglianza non è concava, niente assicura che la procedura di Newton-Raphson abbia effettivamente determinato un massimo assoluto. In questo caso, è opportuna una rappresentazione grafica preliminare della log-verosimiglianza e la scelta di vari punti iniziali per l'iterazione.

• **Esempio 3.1.10.** Dato un campione casuale da $X \sim W(0, 1; p)$, il modello statistico dipende dal parametro $\theta = p$ con $\Theta = \mathbb{R}^+$, ovvero

$$\mathcal{F}_p = \left\{ f_n : f_n(x_1, \dots, x_n; p) = \prod_{i=1}^n p x_i^{p-1} e^{-x_i} \mathbf{1}_{[0, \infty)}(x_i) \right\},$$

da cui si ottiene la seguente log-verosimiglianza

$$l(p) = \log c + n \log p + (p - 1) \sum_{i=1}^n \log x_i - \sum_{i=1}^n x_i^p, \quad p \in \mathbb{R}^+.$$

Dal momento che

$$s_n(p; x_1, \dots, x_n) = \frac{d}{dp} l(p) = \frac{n}{p} + \sum_{i=1}^n \log x_i - \sum_{i=1}^n x_i^p \log x_i,$$

allora l'equazione di verosimiglianza non ammette soluzioni esplicite. Inoltre risulta

$$\widehat{\mathbf{I}}_n(p) = - \frac{d^2}{dp^2} l(p) = \frac{n}{p^2} + \sum_{i=1}^n x_i^p (\log x_i)^2 > 0.$$

Dunque, dal momento che se esiste almeno un valore $x_i > 1$ risulta

$$\lim_{p \rightarrow 0} l(p) = \lim_{p \rightarrow \infty} l(p) = -\infty,$$

la stima di massima verosimiglianza coincide con la soluzione dell'equazione di verosimiglianza.

Per esemplificare il metodo di Newton-Raphson si è generato artificialmente un campione casuale da questo modello con $p = 3$ e $n = 10$, ottenendo il campione

$$(0.70401, 0.75416, 0.89619, 0.97485, 1.25193, 1.70781, 0.80169, 1.12120, 0.95200, 1.49632).$$

Scegliendo come stima iniziale $\widehat{p}_0 = n^{-1} \sum_{i=1}^n x_i^2$ (una stima piuttosto grezza) e un errore pari a $\epsilon = 0.00001$, il metodo di Newton-Raphson fornisce le iterazioni della Tavola 3.1.1.

Tavola 3.1.1. Iterazioni per il metodo di Newton-Raphson.

| j | \hat{p}_j | $s_n(\hat{p}_j)$ | $l(\hat{p}_j)$ |
|-----|-------------|------------------|----------------|
| 0 | 1.23496 | 6.97228 | -8.78720 |
| 1 | 2.14611 | 2.35514 | -4.79208 |
| 2 | 2.78213 | 0.17488 | -4.00095 |
| 3 | 2.83523 | 0.00005 | -3.99631 |
| 4 | 2.83525 | 0.00000 | -3.99631 |
| 5 | 2.83525 | 0.00000 | -3.99631 |

Dunque la stima di massima verosimiglianza è data da $\hat{p} \simeq 2.83525$ a meno di un errore pari a $\epsilon = 0.00001$. In questo caso, è stata ottenuta una soluzione specifica, nel senso che cambiando campione si deve ripercorrere tutto il procedimento di calcolo. Infine, si deve sottolineare che non si dispone della forma analitica dello stimatore di massima verosimiglianza e quindi non si può affermare nulla circa la sua distribuzione e le relative proprietà. \square

• **Esempio 3.1.11.** Se si assume che $X \sim C(\lambda, 1)$, allora dato un campione casuale il modello statistico dipende dal parametro $\theta = \lambda$ con $\Theta = \mathbb{R}$, ovvero

$$\mathcal{F}_\lambda = \left\{ f_n : f_n(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n \frac{1}{\pi(1 + (x_i - \lambda)^2)} \right\},$$

da cui si ottiene la seguente log-verosimiglianza

$$l(\lambda) = \log c - \sum_{i=1}^n \log(1 + (x_i - \lambda)^2), \lambda \in \mathbb{R}.$$

Dal momento che

$$s_n(\lambda; x_1, \dots, x_n) = \frac{d}{d\lambda} l(\lambda) = \sum_{i=1}^n \frac{2(x_i - \lambda)}{1 + (x_i - \lambda)^2} = 0,$$

allora è evidente che l'equazione di verosimiglianza non ammette soluzioni esplicite. Inoltre risulta

$$\hat{I}_n(\lambda) = - \frac{d^2}{d\lambda^2} l(\lambda) = \sum_{i=1}^n \frac{2(1 - (x_i - \lambda))}{(1 + (x_i - \lambda)^2)^2}.$$

Il segno della precedente quantità dipende da λ e quindi la log-verosimiglianza non è concava.

Per esemplificare il metodo di Newton-Raphson si è generato artificialmente un campione casuale da questo modello con $\lambda = 1$ e $n = 10$, ottenendo

$$(0.27577, 41.37492, 0.50046, 3.98571, 0.52963, 1.94228, -1.13702, 6.32890, 0.97387, 1.90390).$$

Scegliendo come stima iniziale $\hat{\lambda}_0 = (x_{(5)} + x_{(6)})/2$ (ovvero la mediana campionaria) e un errore pari a $\epsilon = 0.00001$, il metodo di Newton-Raphson fornisce le iterazioni della Tavola 3.1.2.

Tavola 3.1.2. Iterazioni per il metodo di Newton-Raphson.

| j | \hat{p}_j | $s_n(\hat{p}_j)$ | $l(\hat{p}_j)$ |
|-----|-------------|------------------|----------------|
| 0 | 1.43889 | 1.72933 | -17.35345 |
| 1 | 0.78475 | -0.67453 | -16.90786 |
| 2 | 0.91315 | -0.04125 | -16.86283 |
| 3 | 0.92214 | -0.00022 | -16.86264 |
| 4 | 0.92219 | 0.00000 | -16.86264 |
| 5 | 0.92219 | 0.00000 | -16.86264 |

Si può inoltre verificare graficamente che per questo campione la log-verosimiglianza presenta un solo massimo e dunque la stima di massima verosimiglianza è data da $\hat{\lambda} \simeq 0.92219$ a meno di un errore pari a $\epsilon = 0.00001$. Per un ulteriore campione la log-verosimiglianza potrebbe presentare più massimi relativi e quindi il procedimento dovrebbe essere eventualmente ripetuto con più stime iniziali. \square

3.2. Le proprietà degli stimatori di massima verosimiglianza

3.2.1. La proprietà di equivarianza. La proprietà di equivarianza assicura la congruenza della stima di massima verosimiglianza quando si riparametrizza il modello originale mediante funzioni biunivoche dei parametri.

Dato un modello statistico \mathcal{F}_θ , se $\mathbf{g} : \Theta \rightarrow \Gamma$ è una funzione biunivoca e $\gamma = \mathbf{g}(\theta)$, allora sia $L_\gamma(\gamma)$ la verosimiglianza relativa al modello statistico riparametrizzato \mathcal{F}_γ . Si ha

$$L_\gamma(\gamma) = L(\mathbf{g}^{-1}(\gamma)) = L(\theta),$$

per cui, dal momento che il massimo di $L(\theta)$ si ha per $\hat{\theta}$, il massimo di $L_\gamma(\gamma)$ si ottiene per $\hat{\gamma} = \mathbf{g}(\hat{\theta})$, che dunque risulta la stima di massima verosimiglianza di γ . Questa proprietà, detta equivarianza, permette di calcolare facilmente la stima di massima verosimiglianza di funzioni biunivoche di θ quando si dispone della stima di massima verosimiglianza $\hat{\theta}$. La proprietà dell'equivarianza evita inoltre incongruenze passando fra le varie parametrizzazioni di uno stesso modello statistico. La stima di massima verosimiglianza $\hat{\gamma}$ è la realizzazione campionaria dello stimatore di verosimiglianza $\mathbf{g}(\hat{\Theta})$, la cui distribuzione può essere trovata agevolmente a partire da quella di $\hat{\Theta}$, dal momento che la trasformazione è biunivoca.

• **Esempio 3.2.1.** Dato un campione casuale da $X \sim E(0, \sigma)$, può risultare interessante stimare $\text{Var}[X] = \sigma^2$. La trasformazione $v = \sigma^2$ per $\sigma > 0$ è biunivoca. Inoltre, dall'Esempio 3.1.3 risulta $\hat{\sigma} = \bar{x}$ e quindi $\hat{v} = \bar{x}^2$. Lo stimatore di massima verosimiglianza risulta \bar{X}^2 , la cui distribuzione può essere determinata facilmente da quella di \bar{X} . \square

3.2.2. La sufficienza e gli stimatori di massima verosimiglianza. Gli stimatori di massima verosimiglianza sono trasformate di stimatori sufficienti. Infatti, dato un modello statistico \mathcal{F}_θ , se esiste uno stimatore $\tilde{\Theta}$ sufficiente per θ , sulla base del criterio di fattorizzazione di Neyman risulta

$$L(\theta) = L(\theta; x_1, \dots, x_n) = c g(\tilde{\Theta}; \theta), \theta \in \Theta.$$

Di conseguenza la stima di verosimiglianza può essere ottenuta semplicemente massimizzando $g(\tilde{\Theta}; \theta)$. Dal momento che quest'ultima funzione dipende da (x_1, \dots, x_n) solo attraverso $\tilde{\theta}$, allora la stima di massima verosimiglianza $\hat{\theta}$ è funzione di $\tilde{\theta}$. Quindi, anche lo stimatore di massima verosimiglianza $\hat{\Theta}$ è una trasformata dello stimatore $\tilde{\Theta}$ sufficiente per θ .

• **Esempio 3.2.2.** Dato un campione casuale da $X \sim N(\mu, v)$, dall'Esempio 3.1.2 è noto che lo stimatore di massima verosimiglianza risulta (\bar{X}, S^2) . Nell'Esempio 2.5.3 si è ottenuto che (\bar{X}, S^2) è sufficiente per (μ, v) e dunque i risultati di questa sezione risultano verificati. \square

• **Esempio 3.2.3.** Dato un campione casuale da $X \sim E(0, \sigma)$, dall'Esempio 3.1.3 è noto che lo stimatore di massima verosimiglianza risulta \bar{X} . Nell'Esempio 2.5.2 si è ottenuto che \bar{X} è sufficiente e quindi i risultati di questa sezione risultano verificati. \square

3.2.3. La famiglia esponenziale e gli stimatori di massima verosimiglianza. Nel caso particolare in cui la distribuzione specificata dal modello statistico \mathcal{F}_θ appartenga alla famiglia esponenziale, gli

stimatori di massima verosimiglianza godono di interessanti proprietà. Per semplicità si analizza inizialmente il caso di un parametro. La verosimiglianza risulta

$$L(\theta) = c e^{\tilde{\Theta}(x_1, \dots, x_n)\varphi(\theta) - \eta(\theta)},$$

e quindi la log-verosimiglianza è data da

$$l(\theta) = \log c + \tilde{\Theta}(x_1, \dots, x_n)\varphi(\theta) - \eta(\theta).$$

Di conseguenza, l'equazione di verosimiglianza risulta

$$s_n(\theta; x_1, \dots, x_n) = \frac{d}{d\theta} l(\theta) = \tilde{\Theta}(x_1, \dots, x_n) \frac{d}{d\theta} \varphi(\theta) - \frac{d}{d\theta} \eta(\theta) = 0.$$

Tenendo presente che la media della funzione punteggio è nulla, dalla precedente relazione si ha inoltre

$$E[s_n(\theta; X_1, \dots, X_n)] = E[\tilde{\Theta}(X_1, \dots, X_n)] \frac{d}{d\theta} \varphi(\theta) - \frac{d}{d\theta} \eta(\theta) = 0,$$

ovvero, combinando le precedenti equazioni, si deve avere

$$E[\tilde{\Theta}(X_1, \dots, X_n)] = \tilde{\Theta}(x_1, \dots, x_n).$$

Si può dimostrare che questa equazione ammette soluzione se $\tilde{\Theta}(x_1, \dots, x_n)$ è un valore interno al supporto dello stimatore $\tilde{\Theta}(X_1, \dots, X_n)$ (vedi Brown, 1986). Se la stima di massima verosimiglianza esiste, dunque deve essere quel valore $\hat{\theta}$ per cui la media dello stimatore $\tilde{\Theta}$ è uguale alla sua determinazione campionaria. Di conseguenza, lo stimatore di verosimiglianza $\hat{\Theta}$ è funzione dello stimatore $\tilde{\Theta}$ che per i risultati della §2.5.3 è sufficiente minimale per θ . Inoltre, l'informazione osservata di Fisher è data da

$$\begin{aligned} \hat{I}_n(\hat{\theta}) &= - \left. \frac{d}{d\theta} s_n(\theta; x_1, \dots, x_n) \right|_{\theta=\hat{\theta}} = - \tilde{\Theta}(x_1, \dots, x_n) \left. \frac{d^2}{d\theta^2} \varphi(\theta) \right|_{\theta=\hat{\theta}} + \left. \frac{d^2}{d\theta^2} \eta(\theta) \right|_{\theta=\hat{\theta}} \\ &= - E[\tilde{\Theta}(X_1, \dots, X_n)] \left. \frac{d^2}{d\theta^2} \varphi(\theta) \right|_{\theta=\hat{\theta}} + \left. \frac{d^2}{d\theta^2} \eta(\theta) \right|_{\theta=\hat{\theta}}, \end{aligned}$$

mentre l'informazione di Fisher è data da

$$I_n(\theta) = - E \left[\frac{d}{d\theta} s_n(\theta; x_1, \dots, x_n) \right] = - E[\tilde{\Theta}(X_1, \dots, X_n)] \frac{d^2}{d\theta^2} \varphi(\theta) + \frac{d^2}{d\theta^2} \eta(\theta).$$

Si deve concludere dunque che $I_n(\hat{\theta}) = \hat{I}_n(\hat{\theta})$. Dal momento che $I_n(\theta) > 0$ per ogni θ , allora si ha anche $\hat{I}_n(\hat{\theta}) > 0$. Di conseguenza, ogni soluzione dell'equazione di verosimiglianza è un punto di massimo e quindi la soluzione è unica. Per inciso si noti che se la distribuzione specificata dal modello statistico appartiene alla famiglia esponenziale, allora la procedura di Newton-Raphson determina sicuramente un massimo assoluto. Infine, per i risultati ottenuti nella §2.4.4, se esiste uno stimatore efficiente, questo è dato da $\tilde{\Theta}$ e lo stimatore di massima verosimiglianza $\hat{\Theta}$ coincide con $\tilde{\Theta}$.

Questi risultati possono essere generalizzati al caso di un vettore di parametri θ . In questo caso la stima di massima verosimiglianza è data dalla soluzione dell'equazione

$$E[\tilde{\Theta}(X_1, \dots, X_n)] = \tilde{\Theta}(x_1, \dots, x_n),$$

ed è unica. Inoltre lo stimatore di verosimiglianza $\hat{\Theta}$ è funzione dello stimatore $\tilde{\Theta}$ sufficiente minimale per θ . Nel caso infine che esista uno stimatore efficiente allora questo è dato dallo stimatore di massima verosimiglianza e risulta $\hat{\Theta} = \tilde{\Theta}$.

• **Esempio 3.2.4.** Dato un campione casuale da $X \sim N(\mu, 1)$, è noto dall'Esempio 2.4.13 che la classe di funzioni di densità definita in questo modello fa parte della famiglia esponenziale. La stima di massima verosimiglianza è data dalla soluzione dell'equazione $E[\bar{X}] = \bar{x}$. Dal momento che $E[\bar{X}] = \mu$ (vedi Esempio 1.2.3), allora la stima di massima verosimiglianza risulta $\hat{\mu} = \bar{x}$, risultato ottenuto in precedenza nell'Esempio 3.1.1. Lo stimatore di verosimiglianza \bar{X} è efficiente (vedi Esempio 2.4.9) e coincide infatti con lo stimatore sufficiente minimale per μ . \square

• **Esempio 3.2.5.** Dato un campione casuale da $X \sim N(\mu, v)$, è noto dall'Esempio 2.4.15 che la classe di funzioni di densità definita in questo modello fa parte della famiglia esponenziale. La stima di massima verosimiglianza è data dall'equazione vettoriale

$$E[(\bar{X}, S^2 + \bar{X}^2)] = (\bar{x}, s^2 + \bar{x}^2).$$

Dal momento che $E[\bar{X}] = \mu$, $E[S^2] = (n - 1)v/n$ e $E[\bar{X}^2] = v/n + \mu^2$ (vedi Esempio 1.2.3 e §1.2.3), la precedente equazione si riduce a

$$(\mu, v + \mu^2) = (\bar{x}, s^2 + \bar{x}^2),$$

per cui la stima di massima verosimiglianza risulta (\bar{x}, s^2) , un risultato ottenuto in precedenza nell'Esempio 3.1.2. Lo stimatore di verosimiglianza (\bar{X}, S^2) non è efficiente dal momento che non è corretto per (μ, v) e quindi, dati i risultati di questa sezione, non esiste in generale uno stimatore efficiente. In ogni caso lo stimatore (\bar{X}, S^2) è sufficiente minimale per (μ, v) . \square

3.3. Le proprietà degli stimatori di massima verosimiglianza per grandi campioni

3.3.1. La coerenza degli stimatori di massima verosimiglianza. Supponiamo che $\mathcal{F}_{\theta, n}$ sia un modello statistico relativo ad un campionamento casuale per ogni n . Siano inoltre valide le seguenti condizioni:

- a) il modello statistico è identificabile secondo la definizione data nella §2.1.2.
- b) lo spazio parametrico Θ è un aperto di \mathbb{R}^k ;
- c) la funzione punteggio $s(\theta; x)$ esiste ed è continua per ogni $\theta \in \Theta$ e per ogni $x \in \mathcal{S}_\theta$ (a meno di insiemi di probabilità nulla);
- d) risulta

$$\int_{\mathcal{S}_\theta} \frac{\partial}{\partial \theta} f(x; \theta) d\nu = \frac{\partial}{\partial \theta} \int_{\mathcal{S}_\theta} f(x; \theta) d\nu.$$

Le condizioni a)-d) date in questa sezione sono leggermente più blande di quelle per un problema di stima regolare. In effetti, anche se per provare i risultati che seguono sono sufficienti solo le precedenti condizioni, molti autori assumono comunque di trovarsi in un problema di stima regolare.

Per semplicità si considera inizialmente un solo parametro θ . Se $\theta_0 \in \Theta$ è il vero valore del parametro, allora innanzitutto si vuole dimostrare che l'equazione di verosimiglianza fornisce una successione di insiemi di soluzioni (ognuno dei quali non vuoto) dalla quale può essere scelta una successione di soluzioni che converge a θ_0 quasi certamente, fatto che equivale dunque ad ottenere uno stimatore $\tilde{\Theta}_n$ fortemente coerente per θ_0 . Per la Legge Forte dei Grandi Numeri di Khintchine (vedi Barabesi, 2020) si ha

$$\frac{1}{n} s_n(\theta; X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n s(\theta; X_i) \xrightarrow{qc} a(\theta),$$

dove

$$a(\theta) = \int_{\mathcal{S}_{\theta_0}} s(\theta; x) f(x; \theta_0) d\nu, \theta \in \Theta,$$

essendo $n^{-1} \sum_{i=1}^n s(\theta; X_i)$ una media campionaria di variabili casuali indipendenti ed ugualmente distribuite con media $a(\theta)$. Per l'assunzione *d*) la funzione $a(\theta)$ è continua e strettamente decrescente in un intorno di θ_0 del tipo $(\theta_0 - \epsilon, \theta_0 + \epsilon)$ con $\epsilon > 0$. Dal momento che $a(\theta_0) = 0$ per le proprietà della funzione punteggio, risulta $a(\theta_0 - \epsilon) > 0$ e $a(\theta_0 + \epsilon) < 0$. Allora si ha

$$\frac{1}{n} s_n(\theta_0 - \epsilon; X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n s(\theta_0 - \epsilon; X_i) \xrightarrow{qc} a(\theta_0 - \epsilon) > 0$$

e

$$\frac{1}{n} s_n(\theta_0 + \epsilon; X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n s(\theta_0 + \epsilon; X_i) \xrightarrow{qc} a(\theta_0 + \epsilon) < 0.$$

Di conseguenza, tenendo presente la condizione *c*), deve esistere un valore $\theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon)$ per cui si ha

$$\frac{1}{n} s_n(\theta; X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n s(\theta; X_i) \xrightarrow{qc} 0.$$

Dal momento che ϵ può essere scelto piccolo a piacere, questo risultato è equivalente ad affermare che esiste una successione di soluzioni dell'equazione di verosimiglianza che converge a θ_0 quasi certamente, ovvero si è ottenuto uno stimatore $\tilde{\Theta}_n$ fortemente coerente per θ_0 .

Si dovrebbero sottolineare alcuni aspetti importanti di questo risultato. In primo luogo, si è visto che esiste una successione di soluzioni dell'equazione di verosimiglianza che permette di costruire uno stimatore $\tilde{\Theta}_n$ fortemente coerente per θ_0 , ma niente viene detto riguardo al modo in cui scegliere questa successione dall'insieme delle soluzioni dell'equazione di verosimiglianza. In secondo luogo, lo stimatore $\tilde{\Theta}_n$ non coincide necessariamente con lo stimatore di massima verosimiglianza $\hat{\Theta}_n$. I due stimatori coincidono nel caso che l'equazione di verosimiglianza abbia un'unica soluzione che massimizza la verosimiglianza. Il risultato visto è comunque importante in quanto assicura la coerenza forte dello stimatore ottenuto attraverso l'equazione di verosimiglianza, e questo fatto giustifica fra l'altro il procedimento di Newton-Raphson per grandi campioni.

I precedenti risultati possono essere estesi al caso di un vettore di parametri θ . Se $\theta_0 \in \Theta$ è il vero valore del parametro, l'equazione di verosimiglianza fornisce una successione di insiemi di soluzioni, dalla quale può essere scelta una successione di soluzioni che converge a θ_0 quasi certamente, ovvero esiste uno stimatore $\tilde{\Theta}_n \xrightarrow{qc} \theta_0$ (vedi Wilks, 1962).

• **Esempio 3.3.1.** Dato un campione casuale da $X \sim Bi(1, p)$, le condizioni *a*)-*d*) di questa sezione possono essere facilmente verificate. Inoltre, dall'Esempio 3.1.8 si ha che per ogni n l'equazione di verosimiglianza ha una soluzione unica che massimizza la log-verosimiglianza. Dunque, la soluzione dell'equazione di verosimiglianza coincide con lo stimatore di massima verosimiglianza e si ha $\bar{X}_n \xrightarrow{qc} p_0$. \square

• **Esempio 3.3.2.** Dato un campione casuale da $X \sim N(\mu, \nu)$, le condizioni sono *a*)-*d*) di questa sezione sono verificate, in quanto ci si trova in un problema di stima regolare. Inoltre, dall'Esempio 3.1.9 risulta che per ogni n l'equazione di verosimiglianza ha una soluzione unica che massimizza la log-verosimiglianza. La soluzione dell'equazione di verosimiglianza coincide con lo stimatore di massima verosimiglianza e si ha $(\bar{X}_n, S_n^2) \xrightarrow{qc} (\mu_0, \nu_0)$. \square

Con una dimostrazione più complessa della precedente, si può ottenere anche la coerenza forte dello stimatore di massima verosimiglianza $\hat{\Theta}$. Supponiamo dunque che $\mathcal{F}_{\theta,n}$ sia un modello statistico relativo ad un campionamento casuale per ogni n e siano verificate le seguenti condizioni:

- a') il modello statistico è identificabile secondo la definizione data nella §2.1.2.
- b') lo spazio parametrico Θ è un compatto di \mathbb{R}^k ;
- c') la funzione $f(x; \theta)$ è semicontinua superiormente in θ per ogni $\theta \in \Theta$ e per ogni $x \in \mathcal{S}_\theta$ (a meno di insiemi di probabilità nulla);
- d') esiste una funzione $k(x)$ tale che

$$\int_{\mathcal{S}_{\theta_0}} |k(x)| f(x; \theta_0) d\nu < \infty ,$$

dove θ_0 è il vero valore del parametro, e

$$\log \left(\frac{f(x; \theta)}{f(x; \theta_0)} \right) < k(x)$$

per ogni $\theta \in \Theta$ e per ogni $x \in \mathcal{S}_\theta$ (a meno di insiemi di probabilità nulla).

Se le condizioni a')-d') sono verificate, si può dimostrare che lo stimatore di massima verosimiglianza è fortemente coerente, ovvero $\hat{\Theta}_n \xrightarrow{qc} \theta_0$ per $n \rightarrow \infty$ (vedi Ferguson, 1996).

3.3.2. La distribuzione per grandi campioni degli stimatori di massima verosimiglianza.

Supponiamo che $\mathcal{F}_{\theta,n}$ sia un modello statistico relativo ad un campionamento casuale per ogni n e siano inoltre valide le seguenti condizioni:

- a) il modello statistico è identificabile secondo la definizione data nella §2.1.2.
- b) lo spazio parametrico Θ è un aperto di \mathbb{R}^k ;
- c) le derivate

$$\frac{\partial}{\partial \theta} f(x; \theta)$$

e

$$\frac{\partial^2}{\partial \theta \partial \theta^T} f(x; \theta)$$

esistono e sono continue per ogni $\theta \in \Theta$ e per ogni $x \in \mathcal{S}_\theta$ (a meno di insiemi di probabilità nulla);

d) risulta

$$\int_{\mathcal{S}_\theta} \frac{\partial}{\partial \theta} f(x; \theta) d\nu = \frac{\partial}{\partial \theta} \int_{\mathcal{S}_\theta} f(x; \theta) d\nu$$

e

$$\int_{\mathcal{S}_\theta} \frac{\partial^2}{\partial \theta \partial \theta^T} f(x; \theta) d\nu = \frac{\partial^2}{\partial \theta \partial \theta^T} \int_{\mathcal{S}_\theta} f(x; \theta) d\nu .$$

e) lo stimatore di massima verosimiglianza $\hat{\Theta}_n$ coincide con lo stimatore ottenuto dall'unica soluzione dell'equazione di verosimiglianza;

f) esiste una funzione $k(x)$ tale che

$$\int_{\mathcal{S}_{\theta_0}} |k(x)| f(x; \theta_0) d\nu < \infty ,$$

dove θ_0 è il vero valore del parametro, e ogni componente di $\frac{\partial}{\partial \theta} \mathbf{s}(\theta; x)$ è limitato uniformemente in valore assoluto da $k(x)$ in un intorno di θ_0 ;

g) l'informazione di Fisher $\mathbf{I}(\theta)$ relativa ad una sola osservazione è definita positiva.

Le condizioni a)-d) sono leggermente più forti di quelle per un problema di stima regolare, nel senso che nella condizione c) si richiede che la derivata $\frac{\partial^2}{\partial\theta\partial\theta^T} f(x; \theta)$ sia anche continua.

Si vuole ora determinare la distribuzione per grandi campioni dello stimatore di massima verosimiglianza. Per semplicità si analizza inizialmente il caso di un parametro. Se $\theta_0 \in \Theta$ è il vero valore del parametro, per un dato campione $(x_1, \dots, x_n) \in \mathcal{C}_n$, consideriamo l'espansione in serie di Taylor della funzione punteggio $s_n(\theta; x_1, \dots, x_n) = \sum_{i=1}^n s(\theta; x_i)$ in θ_0 , ovvero

$$s_n(\theta_0; x_1, \dots, x_n) = s_n(\hat{\theta}_n; x_1, \dots, x_n) + \frac{d}{d\theta} s_n(\theta; x_1, \dots, x_n) \Big|_{\theta=\theta^*} (\theta_0 - \hat{\theta}_n),$$

dove $\hat{\theta}_n = \hat{\Theta}_n(x_1, \dots, x_n)$ e $|\theta_0 - \theta^*| < |\theta_0 - \hat{\theta}_n|$. Dal momento che per l'assunzione e) risulta $s_n(\hat{\theta}_n; x_1, \dots, x_n) = 0$, dalla precedente relazione si ha

$$\frac{1}{\sqrt{n}} s_n(\theta_0; x_1, \dots, x_n) = - \frac{1}{n} \frac{d}{d\theta} s_n(\theta; x_1, \dots, x_n) \Big|_{\theta=\theta^*} \sqrt{n}(\hat{\theta}_n - \theta_0).$$

La variabile casuale

$$\frac{1}{\sqrt{n}} s_n(\theta_0; X_1, \dots, X_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n s(\theta_0; X_i)$$

è data dalla somma di n variabili casuali indipendenti ed ugualmente distribuite. Dalla §2.4.1 risulta inoltre $E[s(\theta_0; X)] = 0$ e $\text{Var}[s(\theta_0; X)] = I(\theta_0)$ e quindi per il Teorema Fondamentale del Limite (se l'assunzione g) è verificata) si ha

$$\frac{1}{\sqrt{n}} s_n(\theta_0; X_1, \dots, X_n) \xrightarrow{d} N(0, I(\theta_0)).$$

Si osservi ora che la variabile casuale

$$- \frac{1}{n} \frac{d}{d\theta} s_n(\theta; X_1, \dots, X_n) \Big|_{\theta=\theta_0} = - \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} s(\theta; X_i) \Big|_{\theta=\theta_0}$$

è data dalla media di n variabili casuali indipendenti ed ugualmente distribuite. Dunque, dalla §2.4.1 risulta

$$E \left[- \frac{d}{d\theta} s(\theta; X_i) \Big|_{\theta=\theta_0} \right] = I(\theta_0)$$

e quindi per la Legge Forte dei Grandi Numeri di Khintchine (se l'assunzione g) è verificata) si ottiene

$$- \frac{1}{n} \frac{d}{d\theta} s_n(\theta; X_1, \dots, X_n) \Big|_{\theta=\theta_0} \xrightarrow{qc} I(\theta_0).$$

Infine, per le assunzioni a)-e) si ha $\hat{\Theta}_n \xrightarrow{qc} \theta_0$ da cui $\theta^* \rightarrow \theta_0$ e per la Legge Forte Uniforme dei Grandi Numeri (se l'assunzione f) è verificata) si ottiene (vedi Wilks, 1962)

$$- \frac{1}{n} \frac{d}{d\theta} s_n(\theta; X_1, \dots, X_n) \Big|_{\theta=\theta^*} \xrightarrow{qc} I(\theta_0).$$

Tenendo presente la relazione iniziale, si noti che le variabili casuali $n^{-1/2} s_n(\theta_0; X_1, \dots, X_n)$ e $- n^{-1} \frac{d}{d\theta} s_n(\theta; X_1, \dots, X_n) \Big|_{\theta=\theta^*} \sqrt{n}(\hat{\Theta}_n - \theta_0)$ devono condividere la medesima distribuzione, e quindi applicando il Teorema di Slutsky (vedi Barabesi, 2020) si ottiene infine la distribuzione per grandi campioni dello stimatore di massima verosimiglianza, ovvero

$$\sqrt{n}(\hat{\Theta}_n - \theta_0) \xrightarrow{d} N\left(0, \frac{1}{I(\theta_0)}\right).$$

Il precedente risultato può essere generalizzato al caso di un vettore di parametri, nel senso che (vedi Wilks, 1962)

$$\sqrt{n}(\hat{\Theta}_n - \theta_0) \xrightarrow{d} N_k(\mathbf{0}, \mathbf{I}(\theta_0)^{-1}).$$

La distribuzione per grandi campioni dello stimatore di massima verosimiglianza è stata ottenuta quando questo coincide con lo stimatore ottenuto dall'unica soluzione dell'equazione di verosimiglianza. Risultati più complessi possono essere ottenuti, anche se in questo ambito non saranno ulteriormente considerati (per ulteriori analisi vedi Lehmann, 1983).

È importante infine osservare che, se sono verificate le assunzioni a)-f) di questa sezione, dai risultati ottenuti lo stimatore di massima verosimiglianza risulta asintoticamente efficiente, nel senso che per grandi campioni la sua varianza tende al limite inferiore di Rao-Cramér.

• **Esempio 3.3.3.** Dato un campione casuale da $X \sim N(\mu, 1)$, allora le condizioni a)-e) di questa sezione sono verificate (vedi Esempio 2.1.1 e Esempio 3.3.1). È laborioso, anche se semplice verificare la condizione f). Infine, la condizione g) è verificata dal momento che $I(\mu) = 1$ (vedi Esempio 2.4.1). Inoltre, dall'Esempio 3.1.1 è noto che lo stimatore di massima verosimiglianza è dato da \bar{X}_n e inoltre risulta $1/I(\mu_0) = 1$. Di conseguenza, si ottiene che

$$\sqrt{n}(\bar{X}_n - \mu_0) \xrightarrow{d} N(0, 1),$$

ovvero la distribuzione per grandi campioni coincide in questo caso con quella per campioni finiti. □

• **Esempio 3.3.4.** Dato un campione casuale da $X \sim E(0, \sigma)$, si può dimostrare che le condizioni di questa sezione sono verificate. Inoltre, dall'Esempio 3.1.3 è noto che lo stimatore di massima verosimiglianza è \bar{X}_n , mentre dall'Esempio 2.4.2 si ha $1/I(\sigma_0) = \sigma_0^2$. Di conseguenza, si ha

$$\sqrt{n}(\bar{X}_n - \sigma_0) \xrightarrow{d} N(0, \sigma_0^2). \quad \square$$

• **Esempio 3.3.5.** Dato un campione casuale da $X \sim U(0, \delta)$, le condizioni di questa sezione non sono verificate (vedi Esempio 2.4.3). Dall'Esempio 3.1.4 lo stimatore di massima verosimiglianza è dato da $X_{(n)}$ e dalla teoria delle statistiche estremali è noto che (Ferguson, 1996)

$$-\frac{n}{\delta_0} (X_{(n)} - \delta_0) \xrightarrow{d} E(0, 1).$$

Dunque, la distribuzione per grandi campioni dello stimatore $X_{(n)}$ non è Normale. □

• **Esempio 3.3.6.** Dato un campione casuale da $X \sim W(0, 1; p)$, si può dimostrare che le condizioni di questa sezione sono verificate per $p > 2$. Dall'Esempio 3.1.10 è noto che lo stimatore di massima verosimiglianza non ha una forma esplicita e non è possibile ottenere la relativa distribuzione per campioni finiti. Tuttavia, dal momento che si ha $1/I(p_0) = 6p_0^2/\pi^2$, per $p_0 > 2$ la distribuzione per grandi campioni dello stimatore di massima verosimiglianza risulta

$$\sqrt{n}(\hat{P}_n - p_0) \xrightarrow{d} N\left(0, \frac{6p_0^2}{\pi^2}\right). \quad \square$$

• **Esempio 3.3.7.** Dato un campione casuale da $X \sim N(\mu, \nu)$, si può dimostrare che le condizioni di questa sezione sono verificate. Inoltre, dall'Esempio 3.1.2 è noto che lo stimatore di massima verosimiglianza è dato da (\bar{X}_n, S_n^2) , mentre dall'Esempio 2.4.4 risulta

$$\mathbf{I}(\mu_0, \nu_0)^{-1} = \begin{pmatrix} \nu_0 & 0 \\ 0 & 2\nu_0^2 \end{pmatrix}.$$

Dunque si ha

$$\sqrt{n}(\bar{X}_n - \mu_0, S_n^2 - \nu_0)^T \xrightarrow{d} N_2(\mathbf{0}, \mathbf{I}(\mu_0, \nu_0)^{-1}). \quad \square$$

Supponendo vere le condizioni date in questa sezione, nel caso in cui $g: \Theta \rightarrow \Gamma$ sia una funzione biunivoca e $\gamma = g(\theta)$, tenendo presente la proprietà di equivarianza degli stimatori di massima verosimiglianza e i risultati della §2.4.2, si ha immediatamente che

$$\sqrt{n}(g(\hat{\Theta}_n) - \gamma_0) \xrightarrow{d} N\left(0, \frac{d(\theta_0)^2}{I(\theta_0)}\right),$$

dove $d(\theta) = \frac{d}{d\theta} g(\theta)$ e $\gamma_0 = g(\theta_0)$, supponendo che $d(\theta_0)$ esista e $d(\theta_0) \neq 0$.

Sulla base del precedente risultato è interessante determinare la trasformazione g per cui la varianza per grandi campioni di $\sqrt{n}(g(\hat{\Theta}_n) - \gamma_0)$ risulta pari a 1. In questo caso, si deve avere

$$\frac{d(\theta)^2}{I(\theta)} = 1$$

per ogni θ , ovvero

$$\frac{d}{d\theta} g(\theta) = I(\theta)^{1/2}.$$

Questa equazione differenziale ha per soluzione la funzione

$$g(\theta) = \int I(\theta)^{1/2} d\theta,$$

detta trasformazione stabilizzatrice della varianza.

Questi risultati possono essere estesi al caso di un vettore di parametri. Se dunque $\mathbf{g}: \Theta \rightarrow \Gamma$ è una funzione biunivoca e $\boldsymbol{\gamma} = \mathbf{g}(\boldsymbol{\theta})$, allora si ha

$$\sqrt{n}(\mathbf{g}(\hat{\Theta}_n) - \boldsymbol{\gamma}_0) \xrightarrow{d} N_k(\mathbf{0}, \mathbf{D}(\boldsymbol{\theta}_0)^T \mathbf{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{D}(\boldsymbol{\theta}_0)),$$

dove $\mathbf{D}(\boldsymbol{\theta})$ è definita nella §2.4.2 e $\boldsymbol{\gamma}_0 = \mathbf{g}(\boldsymbol{\theta}_0)$.

• **Esempio 3.3.8.** Dato un campione casuale da $X \sim E(0, \sigma)$, tenendo presente i risultati dell'Esempio 3.3.4, si consideri la trasformata biunivoca $g(\sigma) = \sigma^2$ per $\sigma > 0$. Dal momento che risulta

$$d(\sigma) = \frac{d}{d\sigma} g(\sigma) = 2\sigma,$$

allora si ha

$$\sqrt{n}(\bar{X}_n^2 - \gamma_0) \xrightarrow{d} N(0, 4\gamma_0^2),$$

dove $\gamma_0 = \sigma_0^2$. La trasformazione stabilizzatrice della varianza è data da

$$g(\sigma) = \int \frac{1}{\sigma} d\sigma = \log \sigma$$

e dunque

$$\sqrt{n}(\log \bar{X}_n - \gamma_0) \xrightarrow{d} N(0, 1),$$

dove $\gamma_0 = \log \sigma_0$. □

3.3.3. Una nota sull'informazione osservata di Fisher. Nella §2.1.1 è stato affermato che, in un procedimento di stima per punti, una stima deve essere accompagnata da una misura della sua attendibilità nello stimare il vero parametro θ_0 .

Dato un modello statistico $\mathcal{F}_{\theta,n}$, se il relativo stimatore di massima verosimiglianza $\hat{\Theta}$ è corretto per θ (o almeno asintoticamente corretto), una naturale misura della precisione della stima è data da $\text{Var}[\hat{\Theta}]$. Questa quantità dipende da θ e va quindi stimata sulla base del campione. Si indichi con $\hat{\text{Var}}[\hat{\Theta}]$ la stima di $\text{Var}[\hat{\Theta}]$. Nel caso che $\text{Var}[\hat{\Theta}]$ sia una trasformata biunivoca di θ , si può sfruttare la proprietà dell'equivarianza, ovvero $\hat{\text{Var}}[\hat{\Theta}]$ può essere ottenuta semplicemente sostituendo $\hat{\theta}$ al posto di θ nell'espressione di $\text{Var}[\hat{\Theta}]$ (vedi Esempio 3.2.1). Se la trasformata non è biunivoca, si tenga presente che la stima che si ottiene in questo modo non è la stima di massima verosimiglianza di $\text{Var}[\hat{\Theta}]$.

Se $\hat{\Theta}_n$ è coerente per θ , un indice di precisione per $\hat{\theta}_n$ è dato dalla matrice inversa dell'informazione osservata di Fisher, ovvero da $\hat{\mathbf{I}}_n(\hat{\theta}_n)^{-1}$. Questo indice misura la curvatura locale della log-verosimiglianza in un intorno di $\hat{\theta}_n$ e quindi misura anche il grado relativo di preferenza che la verosimiglianza assegna a $\hat{\theta}_n$ rispetto ad altri valori di θ .

Supponendo verificate le condizioni della §3.3.2 e i risultati conseguiti nella medesima sezione, una ulteriore giustificazione per la scelta del precedente indice è data dal fatto che

$$\frac{1}{n} \hat{\mathbf{I}}_n(\hat{\theta}_n) \xrightarrow{qc} \mathbf{I}(\theta_0),$$

da cui segue che in effetti $\hat{\mathbf{I}}_n(\hat{\theta}_n)^{-1}$ è una stima della matrice di varianza-covarianza per grandi campioni di $\hat{\Theta}_n$.

• **Esempio 3.3.9.** Dato un campione casuale da $X \sim E(0, \sigma)$, si ha

$$\hat{I}_n(\sigma) = -\frac{n}{\sigma^2} + \frac{2n\bar{x}}{\sigma^3},$$

e, dal momento che la stima di massima verosimiglianza di σ è data \bar{x} , si ha $\hat{I}_n(\bar{x})^{-1} = \bar{x}^2/n$. Tenendo presente l'Esempio 3.2.1, la stima di massima verosimiglianza di $\text{Var}(\bar{X}) = \sigma^2/n$ è data da $\hat{\text{Var}}(\bar{X}) = \bar{x}^2/n$, che quindi coincide con $\hat{I}_n(\bar{x})^{-1}$. □

• **Esempio 3.3.10.** Dato un campione casuale da $X \sim N(\mu, \nu)$, tenendo presente l'Esempio 3.1.9, dal momento che la stima di massima verosimiglianza di (μ, σ) è data da (\bar{x}, s^2) , si ha

$$\hat{\mathbf{I}}_n(\bar{x}, s^2)^{-1} = \begin{pmatrix} \frac{s^2}{n} & 0 \\ 0 & \frac{2s^4}{n} \end{pmatrix}.$$

Dunque, la stima di massima verosimiglianza $\hat{\text{Var}}[(\bar{X}, S^2)]$ coincide con $\hat{\mathbf{I}}_n(\bar{x}, s^2)^{-1}$. □

3.4. Alcune esemplificazioni del metodo della massima verosimiglianza

3.4.1. La stima di massima verosimiglianza e la distribuzione di Poisson. Si consideri un campionamento casuale da $X \sim Po(\mu)$. La classe di funzioni di probabilità definita da questo modello fa parte della famiglia esponenziale con $\bar{\Theta}(x_1, \dots, x_n) = \bar{x}$,

$$\varphi(\mu) = n \log \mu$$

e

$$\eta(\mu) = n\mu,$$

mentre

$$q(x_1, \dots, x_n) = \prod_{i=1}^n \frac{\mathbf{1}_{\{0,1,\dots\}}(x_i)}{x_i!}.$$

Dunque, la stima di massima verosimiglianza è data dalla soluzione dell'equazione $E[\bar{X}] = \bar{x}$. Dal momento che $E[\bar{X}] = \mu$ (vedi §A.2.2), la stima di massima verosimiglianza risulta $\hat{\mu} = \bar{x}$ e lo stimatore di massima verosimiglianza \bar{X} è corretto per μ .

Dalle proprietà della famiglia esponenziale (vedi §3.2.3) si ricava che $I_n(\mu) = n/\mu$. Inoltre, per le proprietà della media campionaria (vedi §1.2.2) e della Poisson (vedi §A.2.2), si ha $\text{Var}[\bar{X}] = \mu/n$ e si deve concludere che \bar{X} è anche uno stimatore efficiente. Tenendo presente i risultati della §3.2.3, \bar{X} è anche sufficiente minimale per μ .

Per quanto riguarda le proprietà per grandi campioni, le condizioni a)-g) della §3.3.2 sono verificate e quindi \bar{X}_n è fortemente coerente per μ , mentre

$$\sqrt{n}(\bar{X}_n - \mu_0) \xrightarrow{d} N(0, \mu_0).$$

Dal momento che $\text{Var}[\bar{X}]$ è una funzione biunivoca di μ , allora una misura della precisione di \bar{X} è data da $\widehat{\text{Var}}[\bar{X}] = \bar{x}/n$. Questa stima coincide con $\widehat{I}_n(\bar{x})^{-1}$.

Si considera ora un'applicazione con dati reali. I dati della Tavola 3.4.1 si riferiscono al numero di taxi arrivati in intervalli di un minuto alla stazione di Euston a Londra fra le 9.00 e le 10.00 in una mattina del 1950. Se gli arrivi sono casuali allora è noto dalla teoria dei processi stocastici puntuali che i dati provengono da una variabile casuale di Poisson $Po(\mu)$, dove μ rappresenta l'intensità del processo. È immediato verificare che la stima di massima verosimiglianza risulta $\bar{x} = 79/60 \simeq 1.3167$, che è l'intensità stimata di taxi per minuto. Inoltre, si ha $\widehat{\text{Var}}[\bar{X}] = 0.0219$. Il coefficiente di variazione stimato risulta $\widehat{\text{Var}}[\bar{X}]^{1/2}/\bar{x} = 0.1124$, un valore piuttosto basso che suggerisce una certa precisione della stima.

Tavola 3.4.1. Numero di taxi in intervalli di un minuto.

| numero di taxi | frequenza |
|----------------|-----------|
| 0 | 18 |
| 1 | 18 |
| 2 | 14 |
| 3 | 7 |
| 4 | 3 |
| più di 5 | 0 |

Fonte: Kendall (1951)

3.4.2. La stima di massima verosimiglianza e la regressione di Poisson. Supponiamo che il campione (y_1, \dots, y_n) sia la determinazione di un vettore di variabili casuali (Y_1, \dots, Y_n) , le cui componenti sono indipendenti e tali che $Y_i \sim Po(bz_i)$. In questo caso, (z_1, \dots, z_n) rappresenta un vettore di quantità note e fisse. Il campione non è casuale, in quanto le Y_i , sebbene indipendenti, non

sono ugualmente distribuite. Questo è in effetti un modello di regressione e può essere opportuno quando le variabili di risposta sono intere (per ulteriori dettagli si veda Agresti, 1995).

Il modello statistico risultante dipende dal parametro $\theta = b$ con $\Theta = \mathbb{R}^+$ ed è dato da

$$\mathcal{F}_b = \left\{ f_n : f_n(y_1, \dots, y_n; b) = \prod_{i=1}^n e^{-bz_i} \frac{(bz_i)^{y_i}}{y_i!} \mathbf{1}_{\{0,1,\dots\}}(y_i) \right\}.$$

La classe di funzioni di probabilità definita da questo modello fa parte della famiglia esponenziale con $\tilde{\Theta}(y_1, \dots, y_n) = \bar{y}$,

$$\varphi(b) = n \log b$$

e

$$\eta(b) = nb\bar{z},$$

mentre

$$q(y_1, \dots, y_n) = \prod_{i=1}^n \frac{z_i^{y_i}}{y_i!} \mathbf{1}_{\{0,1,\dots\}}(y_i),$$

dove $\bar{z} = n^{-1} \sum_{i=1}^n z_i$. Dunque, la stima di massima verosimiglianza è data dalla soluzione dell'equazione $E[\bar{Y}] = \bar{y}$. Dal momento che $E[\bar{Y}] = b\bar{z}$, la stima di massima verosimiglianza risulta $\hat{b} = \bar{y}/\bar{z}$. Di conseguenza, lo stimatore di massima verosimiglianza è \bar{Y}/\bar{z} ed è corretto per b . Dalle proprietà della famiglia esponenziale (vedi §3.2.3) si può ricavare che $I_n(b) = n\bar{z}/b$. Inoltre, per le proprietà della media campionaria (vedi §1.2.2) e della Poisson (vedi §A.2.2), si ha $\text{Var}[\bar{Y}/\bar{z}] = b/(n\bar{z})$ e quindi si deve concludere che \bar{Y}/\bar{z} è anche uno stimatore efficiente. Tenendo presente i risultati della §3.2.3, si ha inoltre che \bar{Y}/\bar{z} è sufficiente minimale per b . Dal momento che $\text{Var}[\bar{Y}/\bar{z}]$ è una funzione biunivoca di b , allora una misura della precisione di \bar{Y}/\bar{z} è data da $\widehat{\text{Var}}[\bar{Y}/\bar{z}] = \bar{y}/(n\bar{z}^2)$. Questa stima coincide con $\hat{I}_n(\bar{y}/\bar{z})^{-1}$.

Si considera ora un'applicazione con dati reali. I dati della Tavola 3.4.2 si riferiscono ad alcuni pezzi di stoffa prodotti da una industria.

Tavola 3.4.2. Lunghezza (in metri) e numero di difetti nei pezzi di stoffa.

| pezzo | z_i | y_i | pezzo | z_i | y_i |
|-------|-------|-------|-------|-------|-------|
| 1 | 551 | 6 | 17 | 543 | 8 |
| 2 | 651 | 4 | 18 | 842 | 9 |
| 3 | 832 | 17 | 19 | 905 | 23 |
| 4 | 375 | 9 | 20 | 542 | 9 |
| 5 | 715 | 14 | 21 | 522 | 6 |
| 6 | 868 | 8 | 22 | 122 | 1 |
| 7 | 271 | 5 | 23 | 657 | 9 |
| 8 | 630 | 7 | 24 | 170 | 4 |
| 9 | 491 | 7 | 25 | 738 | 9 |
| 10 | 372 | 7 | 26 | 371 | 14 |
| 11 | 645 | 6 | 27 | 735 | 17 |
| 12 | 441 | 8 | 28 | 749 | 10 |
| 13 | 895 | 28 | 29 | 495 | 7 |
| 14 | 458 | 4 | 30 | 716 | 3 |
| 15 | 642 | 10 | 31 | 952 | 9 |
| 16 | 492 | 4 | 32 | 417 | 2 |

Fonte: Bissel (1972)

Ogni pezzo di stoffa è stato misurato e successivamente è stato contato il numero di difetti di lavorazione che presentava. Se i difetti sono situati in modo casuale sui pezzi di stoffa, in base alla

teoria dei processi stocastici puntuali, un modello di regressione di Poisson sembra dunque ragionevole nell'analisi di questi dati.

È immediato verificare che la stima di massima verosimiglianza risulta $\bar{y}/\bar{z} = 8.875/587.65625 \simeq 0.0151$, quantità che rappresenta l'intensità di errori per metro nella produzione di stoffa. Inoltre, è immediato verificare che risulta $\widehat{\text{Var}}[\bar{Y}/\bar{z}]^{1/2} = 0.0009$. Di conseguenza il coefficiente di variazione stimato risulta $\widehat{\text{Var}}[\bar{Y}/\bar{z}]^{1/2}/(\bar{y}/\bar{z}) = 0.0596$, un valore piuttosto basso che suggerisce un certa precisione della stima.

I dati della Tavola 3.4.2 sono infine rappresentati graficamente nella Figura 3.4.1 insieme con la retta della regressione di Poisson.

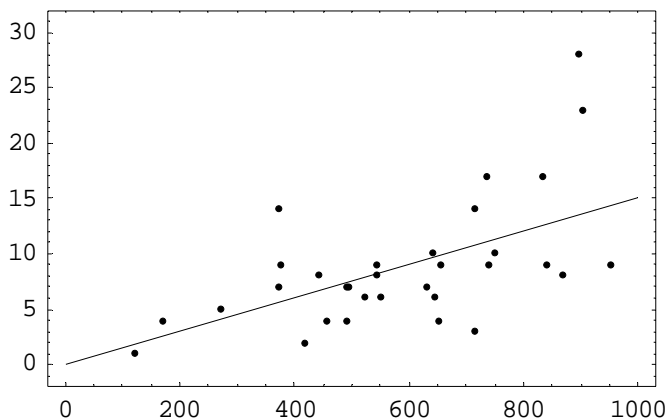


Figura 3.4.1. Diagramma per punti e retta di regressione di Poisson per i dati della Tavola 3.4.2.

3.4.3. La stima di massima verosimiglianza e la regressione logistica. Supponiamo che il campione (y_1, \dots, y_n) sia la determinazione di un vettore di variabili casuali (Y_1, \dots, Y_n) , le cui componenti sono indipendenti e tali che $Y_i \sim Bi(1, p_i)$. Inoltre, si pone

$$p_i = p_i(a, b) = \frac{e^{a+bz_i}}{1 + e^{a+bz_i}},$$

dove (z_1, \dots, z_n) rappresenta un vettore di quantità note e fisse. Questo è in effetti un modello di regressione quando si ha una variabile di risposta dicotomica e le probabilità di successo vengono modellate secondo la precedente funzione, detta curva logistica (per ulteriori dettagli si veda Agresti, 1995). Il campione non è casuale, in quanto le Y_i , sebbene indipendenti, non sono ugualmente distribuite.

Il modello statistico risultante dipende dal vettore di parametri $\theta = (a, b)^T$ con $\Theta = \mathbb{R}^2$ ed è dato da

$$\mathcal{F}_{a,b} = \left\{ f_n : f_n(y_1, \dots, y_n; a, b) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} \mathbf{1}_{\{0,1\}}(y_i) \right\}.$$

La classe di funzioni di probabilità definita dal precedente modello fa parte della famiglia esponenziale una volta che si pone $\tilde{\theta}(y_1, \dots, y_n) = (\sum_{i=1}^n y_i, \sum_{i=1}^n z_i y_i)^T$,

$$\varphi(a, b) = (a, b)^T$$

e

$$\eta(a, b) = - \sum_{i=1}^n \log(1 - p_i),$$

mentre

$$q(y_1, \dots, y_n) = \prod_{i=1}^n \mathbf{1}_{\{0,1\}}(y_i).$$

Dunque, la stima di massima verosimiglianza è data dalla soluzione dell'equazione vettoriale

$$\mathbb{E} \left[\left(\sum_{i=1}^n Y_i, \sum_{i=1}^n z_i Y_i \right) \right] = \left(\sum_{i=1}^n y_i, \sum_{i=1}^n z_i y_i \right).$$

Dal momento che

$$\mathbb{E} \left[\left(\sum_{i=1}^n Y_i, \sum_{i=1}^n z_i Y_i \right) \right] = \left(\sum_{i=1}^n p_i, \sum_{i=1}^n z_i p_i \right),$$

allora la stima di massima verosimiglianza si ottiene dalla soluzione dell'equazione vettoriale

$$\left(\sum_{i=1}^n p_i, \sum_{i=1}^n z_i p_i \right) = \left(\sum_{i=1}^n y_i, \sum_{i=1}^n z_i y_i \right),$$

che non ha soluzioni esplicite. Dunque, dato un certo campione si deve applicare la procedura di Newton-Raphson per ottenere la stima di massima verosimiglianza. Al fine di implementare questa procedura, la log-verosimiglianza è data da

$$l(a, b) = \log c + \sum_{i=1}^n y_i \log p_i + \sum_{i=1}^n (1 - y_i) \log(1 - p_i), \quad (a, b) \in \mathbb{R}^2,$$

da cui, dal momento che si verifica

$$\frac{\partial}{\partial a} p_i(a, b) = p_i(1 - p_i),$$

e

$$\frac{\partial}{\partial b} p_i(a, b) = z_i p_i(1 - p_i),$$

il vettore delle funzioni punteggio risulta

$$\mathbf{s}_n(a, b) = \begin{pmatrix} \frac{\partial}{\partial a} l(a, b) \\ \frac{\partial}{\partial b} l(a, b) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i - \sum_{i=1}^n p_i \\ \sum_{i=1}^n z_i y_i - \sum_{i=1}^n z_i p_i \end{pmatrix},$$

mentre

$$\hat{\mathbf{I}}_n(a, b) = - \begin{pmatrix} \frac{\partial^2}{\partial a^2} l(a, b) & \frac{\partial^2}{\partial a \partial b} l(a, b) \\ \frac{\partial^2}{\partial a \partial b} l(a, b) & \frac{\partial^2}{\partial b^2} l(a, b) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n p_i(1 - p_i) & \sum_{i=1}^n z_i p_i(1 - p_i) \\ \sum_{i=1}^n z_i p_i(1 - p_i) & \sum_{i=1}^n z_i^2 p_i(1 - p_i) \end{pmatrix}.$$

Si considera ora un'applicazione con dati reali. I dati della Tavola 3.4.3 si riferiscono ai voli shuttle prima dell'incidente che ne sospese il programma. In particolare, è stata misurata la temperatura esterna (in F°) al momento del volo ed è stato riportato se vi è stato danneggiamento termico ($y_i = 1$) o meno ($y_i = 0$) nei pannelli di protezione.

Tavola 3.4.3. Temperatura (in F°) e danneggiamento termico nei voli shuttle.

| shuttle | z_i | y_i | shuttle | z_i | y_i |
|---------|-------|-------|---------|-------|-------|
| 1 | 66 | 0 | 13 | 67 | 0 |
| 2 | 70 | 1 | 14 | 53 | 1 |
| 3 | 69 | 0 | 15 | 67 | 0 |
| 4 | 68 | 0 | 16 | 75 | 0 |
| 5 | 67 | 0 | 17 | 70 | 0 |
| 6 | 72 | 0 | 18 | 81 | 0 |
| 7 | 73 | 0 | 19 | 76 | 0 |
| 8 | 70 | 0 | 20 | 79 | 0 |
| 9 | 57 | 1 | 21 | 75 | 1 |
| 10 | 63 | 1 | 22 | 76 | 0 |
| 11 | 70 | 1 | 23 | 58 | 1 |
| 12 | 78 | 0 | | | |

Fonte: Dalal, Fowlkes e Hoadley (1989)

Scegliendo come stima iniziale $(\hat{a}_0, \hat{b}_0) = (0, 0)$ e un errore pari a $\epsilon = 0.00001$, il metodo di Newton-Raphson fornisce le iterazioni della Tavola 3.4.4.

Tavola 3.4.4. Iterazioni per il metodo di Newton-Raphson.

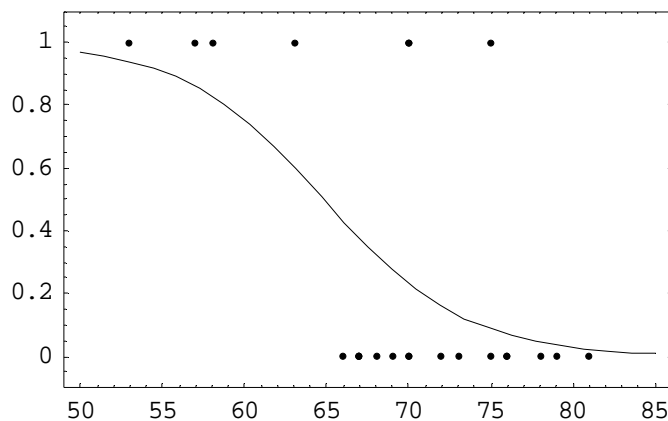
| j | \hat{a}_j | \hat{b}_j | $s_{n1}(\hat{a}_j, \hat{b}_j)$ | $s_{n2}(\hat{a}_j, \hat{b}_j)$ | $l(\hat{a}_j, \hat{b}_j)$ |
|-----|-------------|-------------|--------------------------------|--------------------------------|---------------------------|
| 0 | 0.00000 | 0.00000 | -4.50000 | -354.00000 | -15.94238 |
| 1 | 9.61905 | -0.14952 | -0.78714 | -63.33039 | -10.59252 |
| 2 | 13.65574 | -0.21125 | -0.11878 | -10.00817 | -10.17932 |
| 3 | 14.93829 | -0.23060 | -0.00529 | -0.49352 | -10.15770 |
| 4 | 15.04229 | -0.23215 | -0.00002 | -0.00193 | -10.15760 |
| 5 | 15.04290 | -0.23216 | 0.00000 | 0.00000 | -10.15760 |
| 6 | 15.04290 | -0.23216 | 0.00000 | 0.00000 | -10.15760 |

La stima di massima verosimiglianza risulta $(\hat{a}, \hat{b}) \simeq (15.04290, -0.23216)$ a meno di un errore pari a $\epsilon = 0.00001$. Inoltre, la matrice inversa dell'informazione osservata di Fisher risulta

$$\hat{\mathbf{I}}_n(\hat{a}, \hat{b})^{-1} = \begin{pmatrix} 54.44427 & -0.79639 \\ -0.79639 & 0.01171 \end{pmatrix}.$$

Infine, la curva logistica stimata è data da

$$p(z) = \frac{e^{15.04290 - 0.23216z}}{1 + e^{15.04290 - 0.23216z}}.$$

**Figura 3.4.2.** Diagramma per punti e curva di regressione logistica stimata per i dati della Tavola 3.4.3.

La curva logistica è decrescente, ovvero la probabilità di danneggiamento termico aumenta al diminuire della temperatura. I dati della Tavola 3.4.3 sono stati rappresentati graficamente nella Figura 3.4.2 insieme con la curva di regressione logistica stimata.

3.4.4. La stima di massima verosimiglianza e la distribuzione ipergeometrica. Nell'ambito della statistica ambientale, si vuole stimare spesso la numerosità N di una fissa popolazione biologica. In questo caso si cattura un numero M di animali che vengono “marcati” in modo opportuno. Successivamente, si ricatturano n animali nella popolazione e si determina fra di questi il numero x dei marcati. Questo è appunto il cosiddetto metodo di “cattura-ricattura”. Se la probabilità di essere catturati è la stessa per tutti gli animali della popolazione, non è difficile verificare che il campione consiste in una singola realizzazione x di una $X \sim I(n, M, N)$ (con n e M noti) e dunque il modello statistico risultante dipende dal parametro $\theta = N$ con $\Theta = \mathbb{N}^+$, dove

$$\mathcal{F}_N = \left\{ f_n : f_n(x; N) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \mathbf{1}_{\{\max(0, n-N+M), \dots, \min(n, M)\}}(x) \right\}.$$

Di conseguenza, la verosimiglianza risulta

$$L(N) = c \frac{(N-M)!(N-n)!}{(N-M-n+x)!N!} \mathbf{1}_{\{\max(0, n-N+M), \dots, \min(n, M)\}}(x).$$

Dal momento che

$$\frac{L(N)}{L(N-1)} = \frac{(N-M)(N-n)}{(N-M-n+x)N},$$

allora si ha $L(N)/L(N-1) > 1$ se $N < nM/x$, ovvero la successione $L(N)$ è crescente per $N < nM/x$ e decrescente successivamente. Dunque, la stima di massima verosimiglianza \hat{N} si ottiene per uno dei due valori $\llbracket nM/x \rrbracket$ o $\llbracket nM/x \rrbracket + 1$ che massimizza $L(N)$, dove $\llbracket \cdot \rrbracket$ indica la funzione troncamento. Si può dimostrare inoltre che una stima di $\text{Var}[\hat{N}]$ è data da (vedi Seber, 1973)

$$\tilde{\text{Var}}[\hat{N}] = \frac{(n+1)(M+1)(M-x)(n-x)}{(x+1)^2(x+2)},$$

che tuttavia non corrisponde alla stima di massima verosimiglianza.

In un applicazione con dati reali Odum e Pontin (1961) hanno marcato 600 formiche di una colonia della specie *Lasius Flavus* con una sostanza radiottiva (P^{32}) e ne hanno ricatturate 437 di cui 68 sono risultate marcate. In questo caso, la stima di verosimiglianza è dunque data da $\hat{N} = 3213$, mentre si ha $\tilde{\text{Var}}[\hat{N}]^{1/2} = 323.9760$.

Capitolo 4

La verifica di ipotesi

4.1. Il test statistico

4.1.1. La definizione di test statistico. Sulla base del campione osservato (x_1, \dots, x_n) si è interessati a stabilire se il vero valore del parametro appartiene ad un certo sottoinsieme dello spazio parametrico. Dato un modello statistico \mathcal{F}_θ , se gli insiemi Θ_0 e Θ_1 costituiscono una partizione di Θ , la verifica statistica di ipotesi consiste in un procedimento decisionale di scelta fra l'ipotesi di base $H_0 : \theta \in \Theta_0$ e l'ipotesi alternativa $H_1 : \theta \in \Theta_1$. L'insieme delle ipotesi ammissibili e la sua partizione in H_0 e H_1 è detto sistema di ipotesi. Se Θ_0 (alternativamente Θ_1) contiene un solo punto l'ipotesi H_0 (alternativamente H_1) è detta semplice. In caso contrario l'ipotesi è detta composta.

Lo strumento statistico che sulla base del campione consente di concludere in favore dell'una o dell'altra ipotesi è il test statistico. Dato un modello statistico \mathcal{F}_θ e il sistema di ipotesi $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_1$, si dice test una funzione

$$D : \mathcal{C}_n \rightarrow \{H_0, H_1\}.$$

Il test è in effetti una regola decisionale che suddivide \mathcal{C}_n negli insiemi complementari \mathcal{C}_0 e \mathcal{C}_1 , in modo tale che si accetta H_0 se $(x_1, \dots, x_n) \in \mathcal{C}_0$, mentre si accetta H_1 se $(x_1, \dots, x_n) \in \mathcal{C}_1$. L'insieme \mathcal{C}_1 è anche detto regione critica del test.

Usualmente il procedimento decisionale è basato sulla determinazione di una statistica piuttosto che sul campione osservato. Dunque, se $T = T(X_1, \dots, X_n)$ è una statistica con supporto \mathcal{T} , si dice test basato su T la funzione

$$D : \mathcal{T} \rightarrow \{H_0, H_1\},$$

mentre T è detta statistica test. Il test basato su T è una regola decisionale che suddivide \mathcal{T} negli insiemi complementari \mathcal{T}_0 e \mathcal{T}_1 , in modo tale che si accetta H_0 se per la realizzazione $t = T(x_1, \dots, x_n)$ di T si ha $t \in \mathcal{T}_0$, mentre si accetta H_1 se $t \in \mathcal{T}_1$. L'insieme \mathcal{T}_1 è detto regione critica del test basato su T . L'insieme \mathcal{C}_1 è indotto dall'insieme \mathcal{T}_1 , essendo

$$\mathcal{C}_1 = \{(x_1, \dots, x_n) : t = T(x_1, \dots, x_n) \in \mathcal{T}_1\}.$$

• **Esempio 4.1.1.** Dato un campione casuale da $X \sim N(\mu, 1)$, si consideri il sistema di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$, dove μ_0 è una quantità nota. L'ipotesi H_0 è semplice, mentre l'ipotesi H_1 è composta. Se si suppone che la statistica test sia \bar{X} , risulta $\mathcal{T} = \mathbb{R}$. Una possibile scelta per \mathcal{T}_0 potrebbe essere data da

$$\mathcal{T}_0 = \{\bar{x} : |\bar{x} - \mu_0| < a\},$$

con a costante. Di conseguenza, la regione critica del test basato su \bar{X} risulta

$$\mathcal{T}_1 = \{\bar{x} : |\bar{x} - \mu_0| \geq a\}.$$

Questa scelta di \mathcal{T}_1 appare logica, in quanto più la realizzazione della media campionaria differisce dal valore ipotizzato μ_0 per la media, più si è propensi ad accettare l'ipotesi alternativa. Resta aperto il problema della scelta della costante a . Il problema che verrà analizzato nelle sezioni successive è quello di determinare la statistica test che permette la partizione dello spazio campionario "ottima" secondo alcuni criteri. \square

4.1.2. La funzione potenza. Uno strumento per misurare la capacità discriminatoria del test basato su una statistica è dato dalla cosiddetta funzione potenza. Dato un modello statistico \mathcal{F}_θ e il sistema di ipotesi $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_1$, la funzione potenza del test basato su T è data da

$$P_T(\theta) = \Pr_\theta(T \in \mathcal{T}_1),$$

dove \Pr_θ è da intendersi nel senso che la probabilità è indotta dalla distribuzione specificata dal modello con il valore del parametro pari a θ . Per ogni $\theta \in \Theta_0$ la funzione potenza $P_T(\theta)$ fornisce la probabilità di respingere H_0 quando questa è vera, ovvero la probabilità di commettere il cosiddetto errore di I specie. Analogamente, per ogni $\theta \in \Theta_1$ la quantità $1 - P_T(\theta)$ fornisce la probabilità di accettare H_0 quando è vera H_1 , ovvero la probabilità di commettere il cosiddetto errore di II specie. Per ogni $\theta \in \Theta_1$, la funzione potenza $P_T(\theta)$ fornisce la probabilità di accettare H_1 quando questa è vera.

Si dice che il test basato su T è al livello di significatività α se

$$\sup_{\theta \in \Theta_0} P_T(\theta) = \alpha.$$

Il livello di significatività α rappresenta la massima probabilità di commettere un errore di I specie.

• **Esempio 4.1.2.** Dato un campione casuale da $X \sim N(\mu, 1)$, si consideri il sistema di ipotesi $H_0 : \mu \leq 0$ contro $H_1 : \mu > 0$. Entrambe le ipotesi sono composte. Se si suppone che la statistica test sia \bar{X} , dato che $\mathcal{T} = \mathbb{R}$, si può scegliere

$$\mathcal{T}_0 = \left(-\infty, \frac{z_{1-\alpha}}{\sqrt{n}} \right)$$

e

$$\mathcal{T}_1 = \left[\frac{z_{1-\alpha}}{\sqrt{n}}, \infty \right).$$

Questa selezione di \mathcal{T}_1 appare logica, in quanto più la realizzazione della media campionaria è elevata, più si è propensi ad accettare l'ipotesi alternativa. Dal momento che risulta $\bar{X} \sim N(\mu, 1/n)$ per ogni $\mu \in \mathbb{R}$, allora la funzione potenza è data da

$$\begin{aligned} P_{\bar{X}}(\mu) &= \Pr_\mu \left(\bar{X} \geq \frac{z_{1-\alpha}}{\sqrt{n}} \right) = \Pr_\mu(\sqrt{n}(\bar{X} - \mu) \geq z_{1-\alpha} - \sqrt{n}\mu) \\ &= 1 - \Phi(z_{1-\alpha} - \sqrt{n}\mu) = \Phi(\sqrt{n}\mu - z_{1-\alpha}), \mu \in \mathbb{R}. \end{aligned}$$

Il grafico della precedente funzione potenza per $n = 10$ e $\alpha = 0.05$ è riportato in Figura 4.1.1.

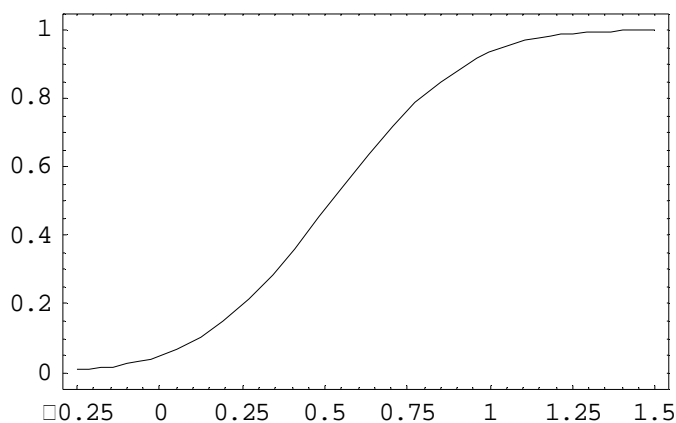


Figura 4.1.1. Funzione potenza $P_{\bar{X}}(\mu)$ per $n = 10$ e $\alpha = 0.05$.

Dal momento che $P_{\bar{X}}(\mu)$ è crescente e che

$$\sup_{\mu \leq 0} P_{\bar{X}}(\mu) = P_{\bar{X}}(0) = \Phi(-z_{1-\alpha}) = 1 - \Phi(z_{1-\alpha}) = \alpha,$$

il test basato su \bar{X} è al livello di significatività α . □

4.1.3. Le proprietà del test statistico. Dalla discussione fatta nella precedente sezione risulta evidente che esiste la possibilità di commettere due tipi di errore decisionale, ovvero può accadere che la realizzazione della statistica test sia in \mathcal{T}_1 quando il vero valore del parametro è in Θ_0 (errore di I specie) e che la realizzazione della statistica test sia in \mathcal{T}_0 quando il vero valore del parametro è in Θ_1 (errore di II specie). Dal momento che non si può sperare di rendere contemporaneamente pari a zero questi tipi di errore, si deve stabilire un insieme di proprietà desiderabili per un test.

Una prima proprietà opportuna per un test è quella della correttezza. Dato un modello statistico \mathcal{F}_θ e il sistema di ipotesi $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_1$, un test basato su T al livello di significatività α con funzione potenza $P_T(\theta)$ è detto corretto al livello di significatività α se

$$P_T(\theta) \geq \alpha, \forall \theta \in \Theta_1.$$

La proprietà della correttezza permette di controllare l'errore di I specie e al tempo stesso assicura che la probabilità di accettare H_1 quando è vera risulta maggiore dell'errore di I specie.

Una seconda proprietà riguarda il comportamento per grandi campioni del test. Se $\mathcal{F}_{\theta,n}$ è un modello statistico per ogni n , dato il sistema di ipotesi $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_1$, si consideri il test al livello di significatività α basato sulla successione di statistiche $(T_n)_{n \geq 1}$. Il test è detto coerente se

$$\lim_{n \rightarrow \infty} P_{T_n}(\theta) = 1, \forall \theta \in \Theta_1.$$

La proprietà della coerenza assicura che la probabilità di commettere un errore di II specie tende a 0 quando si dispone di grandi campioni.

• **Esempio 4.1.3.** Dato un campione casuale da $X \sim N(\mu, 1)$, si consideri il sistema di ipotesi $H_0 : \mu \leq 0$ contro $H_1 : \mu > 0$ (vedi Esempio 4.1.2). Il test basato su \bar{X} è corretto in quanto si ha $P_{\bar{X}}(\mu) \leq \alpha$ per ogni $\mu \leq 0$ e $P_{\bar{X}}(\mu) > \alpha$ per ogni $\mu > 0$. Dal momento che la successione di funzioni $(\Phi(\sqrt{n}\mu - z_{1-\alpha}))_{n \geq 1}$ converge uniformemente ad una funzione costante pari ad 1 per $\mu > 0$, allora

$$\lim_{n \rightarrow \infty} P_{T_n}(\mu) = 1, \forall \mu > 0,$$

ovvero il test basato sulla successione di statistiche $(\bar{X}_n)_{n \geq 1}$ risulta coerente. □

4.1.4. Il test di livello assegnato. Idealmente, si vorrebbe che $P_T(\theta)$ fosse più alta possibile quando $\theta \in \Theta_1$ e la più piccola possibile quando $\theta \in \Theta_0$. Come già evidenziato nella sezione precedente, questi requisiti sono conflittuali tra loro. Un possibile modo di procedere è quello di fissare il livello di significatività α e scegliere quel test che ha più alta potenza per ogni $\theta \in \Theta_1$.

Dato un modello statistico \mathcal{F}_θ e il sistema di ipotesi $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_1$, un test basato su T al livello di significatività α con funzione potenza $P_T(\theta)$ è detto uniformemente più potente al livello di significatività α se

$$P_T(\theta) \geq P_{T^*}(\theta), \forall \theta \in \Theta_1,$$

per ogni altro test basato su una qualsiasi statistica T^* al livello di significatività α .

Sfortunatamente i test uniformemente più potenti esistono solo in varie situazioni speciali. Per campioni finiti, è possibile costruire test uniformemente più potenti quando entrambe le ipotesi sono semplici o quando la statistica test ha una particolare struttura. Sotto certe condizioni, in generale si può costruire test uniformemente più potenti quando si dispone di grandi campioni. Questi argomenti saranno trattati in dettaglio nelle prossime sezioni.

Con la precedente impostazione del problema l'ipotesi di base e l'ipotesi alternativa vengono trattate in modo non simmetrico. Questo è giustificato dal fatto che usualmente H_0 costituisce una affermazione in qualche modo privilegiata e quindi si preferisce controllare il livello di significatività del test (ovvero l'errore di I specie) che comporta l'erroneo rifiuto di questa ipotesi privilegiata. Anche se per sviluppare la teoria è necessario fissare il livello di significatività α , è bene sottolineare che quando si lavora operativamente non esiste nessuna regola ragionevole per stabilirne la scelta. Questa considerazione porta al concetto di livello di significatività osservato o valore-P.

Dato un modello statistico \mathcal{F}_θ e il sistema di ipotesi $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_1$, se la regione critica del test basato su T è data da $\mathcal{T}_1 = \{t : t \geq c\}$, per un determinato valore campionario $t = T(x_1, \dots, x_n)$ si dice livello di significatività osservato la quantità

$$\alpha_{oss} = \sup_{\theta \in \Theta_0} \Pr_\theta(T \geq t),$$

mentre, se la regione critica è data da $\mathcal{T}_1 = \{t : t \leq c\}$, allora si dice livello di significatività osservato la quantità

$$\alpha_{oss} = \sup_{\theta \in \Theta_0} \Pr_\theta(T \leq t).$$

Quando invece la statistica test T ha una distribuzione simmetrica, se la regione critica del test basato su T è data da $\mathcal{T}_1 = \{t : t \leq c_1, t \geq c_2\}$, si dice livello di significatività osservato la quantità

$$\alpha_{oss} = 2 \min \left\{ \sup_{\theta \in \Theta_0} \Pr_\theta(T \leq t), \sup_{\theta \in \Theta_0} \Pr_\theta(T \geq t) \right\}.$$

Il livello di significatività osservato rappresenta la probabilità di ottenere, quando H_0 è vera, un valore campionario t di T estremo (nella appropriata direzione) almeno quanto quello osservato. Dunque, il livello di significatività osservato fornisce una misura su quanto l'ipotesi di base risulta compatibile con i dati campionari. Un livello di significatività osservato basso porta a ritenere poco compatibile con i dati campionari l'ipotesi di base, mentre con un livello di significatività osservato elevato è vera l'affermazione contraria.

In una verifica di ipotesi si può semplicemente riportare il livello di significatività osservato, oppure si può arrivare ad una decisione sull'accettazione di H_0 fissando un livello di significatività α . Se il livello di significatività osservato è minore o uguale ad α , allora si respinge H_0 , altrimenti si accetta H_0 . Il livello di significatività osservato diventa in questo caso il più elevato livello di significatività per cui si accetta H_0 . In questo caso il livello di significatività osservato diventa non solo uno strumento per la decisione nella verifica di ipotesi, ma anche una misura quantitativa di questa decisione.

4.2. Il test del rapporto delle verosimiglianze

4.2.1. Il Lemma di Neyman-Pearson. Dato un modello statistico \mathcal{F}_θ , si supponga che lo spazio parametrico sia composto da due soli punti, ovvero $\Theta = \{\theta_0, \theta_1\}$. Si consideri il sistema di ipotesi $H_0 : \theta = \theta_0$ contro $H_1 : \theta = \theta_1$, dove evidentemente entrambe le ipotesi sono semplici. Si desidera costruire il test corretto al livello di significatività α tale che sia il più potente.

Al fine di verificare questo sistema di ipotesi, si dice test del rapporto delle verosimiglianze quel test basato sulla statistica test R (detta appunto rapporto delle verosimiglianze), la cui realizzazione campionaria è data da

$$r = \frac{L(\theta_0; x_1, \dots, x_n)}{L(\theta_1; x_1, \dots, x_n)}.$$

La statistica test R ha una interpretazione intuitiva, nel senso che se la realizzazione $r = R(x_1, \dots, x_n)$ è tale che $r > 1$, allora il campione attribuisce maggiore verosimiglianza a θ_0 piuttosto che a θ_1 e quindi si è più propensi ad accettare H_0 . Alternativamente, se $r < 1$, allora il campione attribuisce maggiore verosimiglianza a θ_1 piuttosto che a θ_0 e quindi si è più propensi ad accettare H_1 . Di conseguenza, si può scegliere come regione critica

$$T_1 = \{r : r \leq r_\alpha\},$$

dove

$$\Pr_{\theta_0}(R \leq r_\alpha) = \alpha.$$

Ovviamente, T_1 è equivalente alla regione critica indotta da R sullo spazio campionario data da

$$C_1 = \{(x_1, \dots, x_n) : r = R(x_1, \dots, x_n) \leq r_\alpha\}.$$

Se si è osservato il valore campionario r , il livello di significatività osservato del test è dato da

$$\alpha_{oss} = \Pr_{\theta_0}(R \leq r).$$

Supponiamo di considerare un test basato su una statistica T^* corretto al livello di significatività α e sia inoltre C_1^* la regione critica indotta da T^* sullo spazio campionario. Risulta

$$\alpha = \int_{C_1} f_n(x_1, \dots, x_n; \theta_0) d\nu = \int_{C_1^*} f_n(x_1, \dots, x_n; \theta_0) d\nu,$$

da cui

$$\int_{C_1 - C_1^*} f_n(x_1, \dots, x_n; \theta_0) d\nu = \int_{C_1^* - C_1} f_n(x_1, \dots, x_n; \theta_0) d\nu,$$

dal momento che l'integrale sull'insieme $C_1 \cap C_1^*$ è comune alle due regioni. Inoltre, se $(x_1, \dots, x_n) \in C_1 - C_1^*$ e quindi $(x_1, \dots, x_n) \in C_1$, dalla definizione di C_1 risulta

$$f_n(x_1, \dots, x_n; \theta_0) \leq r_\alpha f_n(x_1, \dots, x_n; \theta_1).$$

Alternativamente, se $(x_1, \dots, x_n) \in C_1^* - C_1$ e quindi $(x_1, \dots, x_n) \in C_1^* - C_1$, si ha

$$f_n(x_1, \dots, x_n; \theta_0) > r_\alpha f_n(x_1, \dots, x_n; \theta_1).$$

Dunque, dalle precedenti relazioni si ottiene

$$\begin{aligned} \int_{C_1 - C_1^*} r_\alpha f_n(x_1, \dots, x_n; \theta_1) d\nu &\geq \int_{C_1 - C_1^*} f_n(x_1, \dots, x_n; \theta_0) d\nu = \int_{C_1^* - C_1} f_n(x_1, \dots, x_n; \theta_0) d\nu \\ &> \int_{C_1^* - C_1} r_\alpha f_n(x_1, \dots, x_n; \theta_1) d\nu, \end{aligned}$$

da cui

$$\int_{C_1 - C_1^*} f_n(x_1, \dots, x_n; \theta_1) d\nu > \int_{C_1^* - C_1} f_n(x_1, \dots, x_n; \theta_1) d\nu.$$

Sommando ad ambo i membri l'integrale sull'insieme $C_1 \cap C_1^*$, si ha infine

$$\int_{C_1} f_n(x_1, \dots, x_n; \theta_1) d\nu > \int_{C_1^*} f_n(x_1, \dots, x_n; \theta_1) d\nu,$$

che è equivalente a $P_R(\theta_1) > P_{T^*}(\theta_1)$. Si deve dunque concludere che il test del rapporto delle verosimiglianze è il più potente fra quelli corretti quando entrambe le ipotesi sono semplici. Questo risultato costituisce il cosiddetto Lemma di Neyman-Pearson.

• **Esempio 4.2.1.** Dato un campione casuale da $X \sim N(\mu, 1)$, si supponga che lo spazio parametrico sia $\Theta = \{\mu_0, \mu_1\}$ dove $\mu_1 < \mu_0$ e si consideri il sistema di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu = \mu_1$. Tenendo presente l'Esempio 1.3.1, la determinazione campionaria del rapporto delle verosimiglianze è data da

$$r = \frac{c e^{-\frac{n}{2}(s^2 + (\bar{x} - \mu_0)^2)}}{c e^{-\frac{n}{2}(s^2 + (\bar{x} - \mu_1)^2)}} = e^{n(\mu_0 - \mu_1)\bar{x} - \frac{n}{2}(\mu_0^2 - \mu_1^2)}.$$

Dal momento che r è una funzione biunivoca di \bar{x} , la regione critica \mathcal{C}_1 indotta da R è la stessa indotta da \bar{X} e quindi i test basati sulle due statistiche sono equivalenti. Inoltre, dal momento che se è vera H_0 si ha $\sqrt{n}(\bar{X} - \mu_0) \sim N(0, 1)$, allora

$$\Pr_{\mu_0}(\sqrt{n}(\bar{X} - \mu_0) \leq z_\alpha) = \Pr_{\mu_0}\left(\bar{X} \leq \mu_0 + \frac{z_\alpha}{\sqrt{n}}\right) = \alpha.$$

Di conseguenza, la regione critica del test più potente corretto al livello di significatività α è data da

$$\mathcal{T}_1 = \left\{ \bar{x} : \bar{x} \leq \mu_0 + \frac{z_\alpha}{\sqrt{n}} \right\}.$$

Se si è osservato il valore campionario \bar{x} , il livello di significatività osservato risulta

$$\alpha_{oss} = \Pr_{\mu_0}(\bar{X} \leq \bar{x}) = \Phi(\sqrt{n}(\bar{x} - \mu_0)). \quad \square$$

La precedente procedura può essere generalizzata per determinare test uniformemente più potenti in sistemi di ipotesi più complessi quando il rapporto delle verosimiglianze ha una particolare struttura. Si consideri il sistema di ipotesi $H_0 : \theta = \theta_0$ contro $H_1 : \theta = \theta_1 < \theta_0$, dove la disuguaglianza è da intendersi in senso lessicografico. Questo tipo di ipotesi alternativa è detto direzionale. Supponiamo inoltre che il rapporto delle verosimiglianze R sia una trasformata monotona crescente di una statistica T . I test basati su R e T sono equivalenti, dal momento che inducono la medesima regione critica

$$\begin{aligned} \mathcal{C}_1 &= \{(x_1, \dots, x_n) : r = R(x_1, \dots, x_n) \leq r_\alpha\} \\ &= \{(x_1, \dots, x_n) : t = T(x_1, \dots, x_n) \leq t_\alpha\}, \end{aligned}$$

dove $\Pr_{\theta_0}(T \leq t_\alpha) = \alpha$. Per un dato θ_1 , il Lemma di Neyman-Pearson assicura che il test basato su T è il più potente. Dal momento che t_α non dipende dal valore di θ_1 in quanto T è una statistica, la precedente affermazione è valida per ogni θ specificato in H_1 , ovvero si è ottenuto il test uniformemente più potente per questo sistema di ipotesi.

In maniera analoga, nel caso che si consideri il sistema di ipotesi $H_0 : \theta = \theta_0$ contro $H_1 : \theta = \theta_1 > \theta_0$, se il rapporto delle verosimiglianze R è una trasformata monotona decrescente di una statistica T , allora il test basato su T è uniformemente più potente.

Si consideri infine il sistema di ipotesi $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$. Questo tipo di ipotesi alternativa è detta bilaterale. Dalla precedente discussione risulta chiaro che per il sistema di ipotesi precedente non può esistere un test uniformemente più potente, dal momento che il rapporto delle verosimiglianze non può essere contemporaneamente funzione monotona crescente e decrescente di una statistica T .

• **Esempio 4.2.2.** Dato un campione casuale da $X \sim N(\mu, 1)$, si supponga che lo spazio parametrico sia $\Theta = (-\infty, \mu_0]$ e si consideri il sistema di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu < \mu_0$. Dall'Esempio 4.2.1 è noto che il rapporto delle verosimiglianze è una trasformata monotona crescente di \bar{X} . Quindi il test basato su \bar{X} è quello uniformemente più potente anche in questo sistema di ipotesi. La regione critica risulta identica a quella dell'Esempio 4.2.1. \square

4.2.2. Il test del rapporto delle verosimiglianze. In questa sezione si generalizza la formulazione del rapporto delle verosimiglianze al caso in cui una o entrambe le ipotesi siano composte. Dato un modello statistico \mathcal{F}_θ e il sistema di ipotesi $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_1$, il test del rapporto delle verosimiglianze è basato sulla statistica test R la cui realizzazione campionaria è data da

$$r = \frac{\max_{\theta \in \Theta_0} L(\theta; x_1, \dots, x_n)}{\max_{\theta \in \Theta} L(\theta; x_1, \dots, x_n)}.$$

Questa statistica test ha un'interpretazione intuitiva, nel senso che se si sta confrontando la "plausibilità" di un valore θ rispetto ad un altro sulla base di un campione (x_1, \dots, x_n) , siamo portati a scegliere quel valore che fornisce la verosimiglianza più alta. Estendendo la discussione fatta nella precedente sezione, se non esiste un valore θ che fornisce una verosimiglianza sensibilmente più alta in Θ rispetto alla verosimiglianza massima in Θ_0 , siamo propensi ad accettare H_0 . Ovviamente, si ha $0 \leq r \leq 1$. Se la realizzazione r è prossima ad 1 si è più propensi ad accettare H_0 , mentre se la realizzazione r è prossima ad 0 si è più propensi ad accettare H_1 . Di conseguenza, si può scegliere come regione critica $\mathcal{T}_1 = \{r : r \leq r_\alpha\}$, dove

$$\sup_{\theta \in \Theta_0} \Pr_\theta(R \leq r_\alpha) = \alpha.$$

Se si è osservato il valore campionario r , il livello di significatività osservato del test è dato da

$$\alpha_{oss} = \sup_{\theta \in \Theta_0} \Pr_\theta(R \leq r).$$

Se risulta $\theta = (\theta_A^T, \theta_B^T)^T$ dove θ_A è un vettore a $(k - h)$ componenti e θ_B è un vettore a h componenti, dato il sistema di ipotesi $H_0 : \theta_A = \theta_{0A}$ contro $H_1 : \theta_B \neq \theta_{0B}$, la realizzazione campionaria della statistica test R è data da

$$r = \frac{L_p(\theta_{0A}; x_1, \dots, x_n)}{L(\hat{\theta}; x_1, \dots, x_n)},$$

che evidenzia la dipendenza di R dalla verosimiglianza profilo.

• **Esempio 4.2.3.** Dato un campione casuale da $X \sim N(\mu, 1)$, si consideri il sistema di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$. Tenendo presente l'Esempio 1.3.1 e l'Esempio 3.1.1, la determinazione campionaria del rapporto delle verosimiglianze è data da

$$r = \frac{c e^{-\frac{n}{2}(s^2 + (\bar{x} - \mu_0)^2)}}{c e^{-\frac{n}{2}s^2}} = e^{-\frac{n}{2}(\bar{x} - \mu_0)^2}.$$

Dal momento che r è una funzione biunivoca di $(\bar{x} - \mu_0)^2$, la regione critica \mathcal{C}_1 indotta da R è la stessa indotta da $(\bar{X} - \mu_0)^2$ e quindi i relativi test sono equivalenti. Inoltre, dal momento che se è vera H_0 si ha $\sqrt{n}(\bar{X} - \mu_0) \sim N(0, 1)$, allora $n(\bar{X} - \mu_0)^2 \sim \chi_1^2$. Tenendo presente che r è una funzione monotona decrescente di $(\bar{x} - \mu_0)^2$ e dunque si respinge H_0 per realizzazioni elevate di $(\bar{X} - \mu_0)^2$, allora

$$\Pr_{\mu_0}(n(\bar{X} - \mu_0)^2 \geq \chi_{1,1-\alpha}^2) = \alpha.$$

Di conseguenza, la regione critica del test basato sul rapporto delle verosimiglianze è data da

$$\mathcal{T}_1 = \{\bar{x} : n(\bar{x} - \mu_0)^2 \geq \chi_{1,1-\alpha}^2\}.$$

Essendo $z_{1-\alpha/2} = \sqrt{\chi_{1,1-\alpha}^2}$ per la relazione fra la $N(0, 1)$ e la χ_1^2 , la precedente regione critica può essere anche espressa in modo equivalente come

$$\mathcal{T}_1 = \left\{ \bar{x} : \bar{x} \leq \mu_0 - \frac{z_{1-\alpha/2}}{\sqrt{n}}, \bar{x} \geq \mu_0 + \frac{z_{1-\alpha/2}}{\sqrt{n}} \right\}.$$

Dal momento che la distribuzione di \bar{X} è simmetrica, se si è osservato il valore campionario \bar{x} , il livello di significatività osservato risulta

$$\begin{aligned} \alpha_{oss} &= 2 \min(\Pr_{\mu_0}(\bar{X} \leq \bar{x}), \Pr_{\mu_0}(\bar{X} \geq \bar{x})) \quad \square \\ &= 2 \min(\Phi(\sqrt{n}(\bar{x} - \mu_0)), 1 - \Phi(\sqrt{n}(\bar{x} - \mu_0))) = 2(1 - \Phi(\sqrt{n}|\bar{x} - \mu_0|)). \end{aligned}$$

4.2.3. La distribuzione per grandi campioni del rapporto delle verosimiglianze. Il test del rapporto delle verosimiglianze R_n ha proprietà ottime quando si dispone di grandi campioni. Per semplicità si analizza inizialmente il caso di un solo parametro.

Dato un modello statistico \mathcal{F}_θ relativo ad un campionamento casuale, si consideri il sistema di ipotesi $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$ e si supponga che le condizioni a)-g) della §3.3.2 siano verificate. Se H_0 è vera, per un dato campione $(x_1, \dots, x_n) \in \mathcal{C}_n$, si consideri l'espansione in serie di Taylor di

$$\log f_n(x_1, \dots, x_n; \theta) = \sum_{i=1}^n \log f(x_i; \theta)$$

in θ_0 , ovvero

$$\begin{aligned} \log f_n(x_1, \dots, x_n; \theta_0) &= \log f_n(x_1, \dots, x_n; \hat{\theta}_n) \\ &+ \frac{d}{d\theta} \log f_n(x_1, \dots, x_n; \theta) \Big|_{\theta=\theta^*} (\theta_0 - \hat{\theta}_n) \\ &+ \frac{1}{2} \frac{d^2}{d\theta^2} \log f_n(x_1, \dots, x_n; \theta) \Big|_{\theta=\theta^*} (\theta_0 - \hat{\theta}_n)^2, \end{aligned}$$

dove $\hat{\theta}_n = \hat{\Theta}_n(x_1, \dots, x_n)$ rappresenta la stima dell'equazione di verosimiglianza, mentre $|\theta_0 - \theta^*| < |\theta_0 - \hat{\theta}_n|$. Dal momento che

$$\frac{d}{d\theta} \log f_n(x_1, \dots, x_n; \theta) \Big|_{\theta=\theta^*} \simeq s_n(\hat{\theta}_n; x_1, \dots, x_n) = 0$$

per l'assunzione b), la precedente relazione può essere riscritta come

$$-2 \log \frac{f_n(x_1, \dots, x_n; \theta_0)}{f_n(x_1, \dots, x_n; \hat{\theta}_n)} = -\frac{1}{n} \frac{d}{d\theta} s_n(x_1, \dots, x_n; \theta) \Big|_{\theta=\theta^*} (\sqrt{n}(\hat{\theta}_n - \theta_0))^2.$$

La quantità a primo membro è $-2 \log r_n$, dove $r_n = R_n(x_1, \dots, x_n)$. Inoltre, in base ai risultati della §3.3.2, si ha

$$-\frac{1}{n} \frac{d}{d\theta} s_n(X_1, \dots, X_n; \theta) \Big|_{\theta=\theta^*} (\sqrt{n}(\hat{\Theta}_n - \theta_0))^2 \xrightarrow{d} \chi_1^2,$$

da cui segue che

$$-2 \log R_n \xrightarrow{d} \chi_1^2.$$

Si è determinata la distribuzione per grandi campioni di una trasformata biunivoca del rapporto delle verosimiglianze quando H_0 è vera. Dal momento che la regione critica indotta da $-2 \log R_n$ è identica a quella indotta da R_n , i relativi test sono equivalenti. Si può dimostrare che il test basato

sulla successione di statistiche $(-2 \log R_n)_{n \geq 1}$ è coerente al livello di significatività α e che è uniformemente più potente (in senso asintotico) per una vasta classe di test (Wilks, 1962).

Dal momento che $-2 \log r$ è una funzione monotona decrescente di r , si respinge H_0 per realizzazioni elevate di $-2 \log R_n$. Essendo

$$\lim_{n \rightarrow \infty} \Pr_{\theta_0}(-2 \log R_n \geq \chi_{1,1-\alpha}^2) = \alpha,$$

la regione critica per grandi campioni del test basato sul rapporto delle verosimiglianze è data da

$$\mathcal{T}_{n,1} = \{r : -2 \log r \geq \chi_{1,1-\alpha}^2\}.$$

La regione critica per campioni finiti data da $\mathcal{T}_1 = \{r : r \leq r_\alpha\}$, dove $\Pr_{\theta_0}(R \leq r_\alpha) = \alpha$, e la regione critica $\mathcal{T}_{n,1}$ in generale non coincidono (per un'eccezione si veda l'Esempio 4.2.2). In particolare, il livello reale di significatività del test α_n non coincide con il livello nominale di significatività α (ovvero quello prefissato). Dal momento che si vuole controllare al massimo l'errore di I specie, è opportuno che sia $\alpha_n \leq \alpha$ e se il test soddisfa a questa condizione allora è detto conservatore. Infine, se si è osservato il valore campionario r , il livello di significatività osservato per grandi campioni del test è dato da

$$\alpha_{n,oss} = \Pr(\chi_1^2 \geq -2 \log r),$$

che ovviamente non coincide in generale con il livello di significatività osservato per campioni finiti $\alpha_{oss} = \Pr_{\theta_0}(R \leq r)$.

Per ottenere i risultati di questa sezione, è stato implicitamente dimostrato che

$$I_n(\theta_0)(\hat{\Theta}_n - \theta_0)^2 \xrightarrow{d} \chi_1^2$$

e che

$$\hat{I}_n(\hat{\Theta}_n)(\hat{\Theta}_n - \theta_0)^2 \xrightarrow{d} \chi_1^2.$$

Su queste due statistiche si possono dunque costruire test equivalenti al test del rapporto delle verosimiglianze per grandi campioni.

• **Esempio 4.2.4.** Dato un campione casuale da $X \sim E(0, \sigma)$, si consideri il sistema di ipotesi $H_0 : \sigma = \sigma_0$ contro $H_1 : \sigma \neq \sigma_0$, dove σ_0 è una quantità nota. Tenendo presente l'Esempio 1.3.3 e l'Esempio 3.1.3, la determinazione campionaria del rapporto delle verosimiglianze è data da

$$r = \frac{c e^{-n \log \sigma_0 - n\bar{x}/\sigma_0}}{c e^{-n \log \bar{x} - n}} = e^{n \log(\bar{x}/\sigma_0) - n\bar{x}/\sigma_0 + n},$$

ovvero r è una funzione (non biunivoca) di \bar{x} . È evidentemente laborioso determinare la distribuzione per campioni finiti del rapporto delle verosimiglianze. Tuttavia, dal momento che le condizioni richieste in questa sezione sono soddisfatte, se H_0 è vera risulta

$$-2 \log R_n = -2n \log \left(\frac{\bar{X}_n}{\sigma_0} \right) + \frac{2n\bar{X}_n}{\sigma_0} - 2n \xrightarrow{d} \chi_1^2.$$

Di conseguenza, la regione critica per grandi campioni del test basato sul rapporto delle verosimiglianze è data da

$$\mathcal{T}_{n,1} = \left\{ \bar{x} : -2n \log \left(\frac{\bar{x}}{\sigma_0} \right) + \frac{2n\bar{x}}{\sigma_0} - 2n \geq \chi_{1,1-\alpha}^2 \right\}.$$

Infine, la significatività osservata per grandi campioni risulta

$$\alpha_{n,oss} = \Pr\left(\chi_1^2 \geq -2n \log\left(\frac{\bar{x}}{\sigma_0}\right) + \frac{2n\bar{x}}{\sigma_0} - 2n\right). \quad \square$$

• **Esempio 4.2.5.** Dato un campione casuale da $X \sim Bi(1, p)$, si consideri il sistema di ipotesi $H_0 : p = p_0$ contro $H_1 : p \neq p_0$. Tenendo presente l'Esempio 1.3.5 e l'Esempio 3.1.8, la determinazione campionaria del rapporto delle verosimiglianze è data da

$$r = \frac{c e^{n\bar{x} \log p_0 + (n-n\bar{x}) \log(1-p_0)}}{c e^{n\bar{x} \log \bar{x} + (n-n\bar{x}) \log(1-\bar{x})}} = e^{n\bar{x} \log(p_0/\bar{x}) + (n-n\bar{x}) \log((1-p_0)/(1-\bar{x}))}$$

ovvero r è una funzione (non biunivoca) di \bar{x} . È evidentemente laborioso determinare la distribuzione per campioni finiti del rapporto delle verosimiglianze. Tuttavia, dal momento che le condizioni richieste in questa sezione sono soddisfatte, allora se H_0 è vera risulta

$$-2 \log R_n = -2n\bar{X}_n \log\left(\frac{p_0}{\bar{X}_n}\right) - 2(n - n\bar{X}_n) \log\left(\frac{1-p_0}{1-\bar{X}_n}\right) \xrightarrow{d} \chi_1^2.$$

Di conseguenza, la regione critica per grandi campioni del test basato sul rapporto delle verosimiglianze è data da

$$\mathcal{T}_{n,1} = \left\{ \bar{x} : -2n\bar{x} \log\left(\frac{p_0}{\bar{x}}\right) - 2(n - n\bar{x}) \log\left(\frac{1-p_0}{1-\bar{x}}\right) \geq \chi_{1,1-\alpha}^2 \right\}.$$

Infine, la significatività osservata per grandi campioni risulta

$$\alpha_{n,oss} = \Pr\left(\chi_1^2 \geq -2n\bar{x} \log\left(\frac{p_0}{\bar{x}}\right) - 2(n - n\bar{x}) \log\left(\frac{1-p_0}{1-\bar{x}}\right)\right). \quad \square$$

I precedenti risultati possono essere generalizzati al caso di un vettore di parametri. Se \mathcal{F}_θ è un modello statistico relativo ad un campionamento casuale, si consideri il sistema di ipotesi $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$ e supponiamo che le condizioni a)-g) della §3.3.2 siano verificate. Si può dimostrare che se H_0 è vera risulta (vedi Wilks, 1962)

$$-2 \log R_n \xrightarrow{d} \chi_k^2.$$

Dunque, la regione critica per grandi campioni del test basato sul rapporto delle verosimiglianze è data da

$$\mathcal{T}_{n,1} = \{r : -2 \log r \geq \chi_{k,1-\alpha}^2\}.$$

Inoltre, per un dato valore campionario r , il livello di significatività osservato per grandi campioni risulta

$$\alpha_{n,oss} = \Pr(\chi_k^2 \geq -2 \log r).$$

Anche nel caso di un vettore di parametri si può dimostrare che (vedi Wilks, 1962)

$$(\hat{\Theta}_n - \theta_0)^T \mathbf{I}_n(\theta_0) (\hat{\Theta}_n - \theta_0) \xrightarrow{d} \chi_k^2$$

e che

$$(\hat{\Theta}_n - \theta_0)^T \hat{\mathbf{I}}_n(\hat{\Theta}_n) (\hat{\Theta}_n - \theta_0) \xrightarrow{d} \chi_k^2.$$

Su queste due statistiche si possono costruire test equivalenti a quello del rapporto delle verosimiglianze quando si dispone di grandi campioni.

Si supponga infine che $\theta = (\theta_A^T, \theta_B^T)^T$ dove θ_A è un vettore a $(k - h)$ componenti e θ_B è un vettore a h componenti. Si consideri il sistema di ipotesi $H_0 : \theta_A = \theta_{0A}$ contro $H_1 : \theta_A \neq \theta_{0A}$ e supponiamo che le condizioni a)-g) della §3.3.2 siano verificate. Si può dimostrare che se H_0 è vera risulta (vedi Wilks, 1962)

$$-2 \log R_n \xrightarrow{d} \chi_{k-h}^2,$$

ovvero in questo caso i gradi di libertà della distribuzione di $-2 \log R_n$ per grandi campioni diminuiscono del numero di componenti di θ da stimare sotto H_0 . Dunque, la regione critica per grandi campioni del test basato sul rapporto delle verosimiglianze è data da

$$\mathcal{T}_{n,1} = \{r : -2 \log r \geq \chi_{k-h,1-\alpha}^2\}.$$

Per un dato valore campionario r , il livello di significatività osservato per grandi campioni risulta

$$\alpha_{n,oss} = \Pr(\chi_{k-h}^2 \geq -2 \log r).$$

4.3. Alcune applicazioni del test del rapporto delle verosimiglianze

4.3.1. Il test t di Student ad un campione (bilaterale). Si consideri un campionamento casuale da $X \sim N(\mu, v)$ e si supponga di voler verificare il sistema di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$. Si ha

$$\Theta = \{(\mu, v) : (\mu, v) \in \mathbb{R} \times \mathbb{R}^+\}$$

e

$$\Theta_0 = \{(\mu, v) : \mu = \mu_0, v \in \mathbb{R}^+\}.$$

In base ai risultati dell'Esempio 3.1.2, si ottiene

$$\max_{(\mu, v) \in \Theta} l(\mu, v) = l(\bar{x}, s^2).$$

Inoltre, tenendo presente la disuguaglianza dell'Esempio 3.1.2 con $d = s^2 + (\bar{x} - \mu_0)^2$, si ha

$$\begin{aligned} l(\mu_0, v) &= \log c - \frac{n}{2} \log v - \frac{n}{2v} [s^2 + (\bar{x} - \mu_0)^2] \leq \log c - \frac{n}{2} \log[s^2 + (\bar{x} - \mu_0)^2] - \frac{n}{2} \\ &= l[\mu_0, s^2 + (\bar{x} - \mu_0)^2] = l_p(\mu_0) = \max_{(\mu, v) \in \Theta_0} l(\mu, v). \end{aligned}$$

La determinazione campionaria di R è data da

$$\begin{aligned} r &= e^{l_p(\mu_0) - l(\bar{x}, s^2)} = e^{-\frac{n}{2} \log(s^2 + (\bar{x} - \mu_0)^2) + \frac{n}{2} \log s^2} \\ &= \left(1 + \frac{(\bar{x} - \mu_0)^2}{s^2}\right)^{-\frac{1}{2}n} = \left(1 + \frac{n(\bar{x} - \mu_0)^2}{(n-1)s_c^2}\right)^{-\frac{1}{2}n} = \left(1 + \frac{t^2}{n-1}\right)^{-\frac{1}{2}n}, \end{aligned}$$

dove

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s_c}.$$

Se H_0 è vera, t è la determinazione campionaria di una statistica T distribuita come una t di Student con $(n - 1)$ gradi di libertà, dal momento che

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S_c} = \frac{\sqrt{n}(\bar{X} - \mu_0)/\sqrt{v}}{\sqrt{((n-1)S_c^2/v)/(n-1)}}$$

e che inoltre le statistiche $\sqrt{n}(\bar{X} - \mu_0)/\sqrt{v} \sim N(0, 1)$ e $(n-1)S_c^2/v \sim \chi_{n-1}^2$ sono indipendenti (vedi Esempio 1.2.5). Inoltre r è una funzione monotona decrescente di $|t|$ e quindi i test costruiti su R e $|T|$ sono equivalenti. Tenendo presente la relazione fra le statistiche T e $|T|$ e dal momento che rifiutare H_0 per piccole determinazioni di R è equivalente a rifiutare H_0 per determinazioni elevate di $|T|$, la regione critica del test risulta

$$\mathcal{T}_1 = \{t : t \leq -t_{n-1, 1-\alpha/2}, t \geq t_{n-1, 1-\alpha/2}\}.$$

Dal momento che la distribuzione di T è simmetrica, se si è osservato il valore campionario t , il livello di significatività osservato risulta infine

$$\alpha_{oss} = 2 \Pr(t_{n-1} \geq |t|).$$

Anche se in questo caso la distribuzione della statistica test è nota per campioni finiti, si noti tuttavia che per grandi campioni, quando H_0 è vera, la quantità

$$-2 \log r = n \log \left(1 + \frac{t^2}{n-1} \right),$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_1^2$, in quanto vi sono due parametri, di cui uno da stimare sotto ipotesi di base.

• **Esempio 4.3.1.** Gli antichi greci indicavano come rettangolo aureo un rettangolo con un rapporto fra i lati dato da $\rho = 1 \div (\sqrt{5} + 1)/2 \simeq 0.618$. Questo tipo di rettangolo era spesso adoperato nella loro architettura (per esempio nella struttura del Partenone). I dati della Tavola 4.3.1 riguardano il rapporto fra i lati di un campione casuale di 20 rettangoli usati dai nativi americani Shoshoni per decorare le loro tende.

Tavola 4.3.1. Rapporto dei lati dei rettangoli.

| rettangolo | x_i | rettangolo | x_i |
|------------|-------|------------|-------|
| 1 | 0.693 | 11 | 0.654 |
| 2 | 0.662 | 12 | 0.615 |
| 3 | 0.690 | 13 | 0.668 |
| 4 | 0.606 | 14 | 0.601 |
| 5 | 0.570 | 15 | 0.576 |
| 6 | 0.749 | 16 | 0.670 |
| 7 | 0.672 | 17 | 0.606 |
| 8 | 0.628 | 18 | 0.611 |
| 9 | 0.609 | 19 | 0.553 |
| 10 | 0.844 | 20 | 0.933 |

Fonte: Dubois (1970)

Si vuole verificare se gli Shoshoni avessero conoscenza del rettangolo aureo, ovvero, supponendo valido il modello di questa sezione, si vuole verificare il sistema di ipotesi $H_0 : \mu = \rho$ contro $H_1 : \mu \neq \rho$. Dal momento che per questi dati è immediato verificare che $\bar{x} = 0.6605$ e $s_c = 0.0925$, allora si ha $t = 2.0529$. Di conseguenza, il livello di significatività osservato risulta

$$\alpha_{oss} = 2 \Pr(t_{19} \geq 2.0529) = 0.0541.$$

Dato che la significatività osservata è piuttosto bassa, vi è una moderata indicazione a respingere l'ipotesi di base, ovvero sembra dubbio che gli Shoshoni avessero conoscenza del rettangolo aureo. Infatti, si può respingere H_0 ad ogni livello di significatività $\alpha > 0.0541$. Si noti che $-2 \log r = 4.0066$ e dunque si ha

$$\alpha_{n,oss} = \Pr(\chi_1^2 \geq 4.0066) = 0.0453,$$

un risultato che conferma le conclusioni precedenti. \square

4.3.2. Il test t di Student ad un campione (direzionale). Si consideri ancora il modello della §4.3.1 e si supponga ora di voler verificare il sistema di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu < \mu_0$. Dunque, si ha

$$\Theta = \{(\mu, v) : \mu \leq \mu_0, v \in \mathbb{R}^+\}$$

e

$$\Theta_0 = \{(\mu, v) : \mu = \mu_0, v \in \mathbb{R}^+\}.$$

Per quanto riguarda la massimizzazione su Θ_0 nulla cambia rispetto alla §4.3.1. Inoltre, tenendo presente la disuguaglianza dell'Esempio 3.1.2 con $d = s^2 + (\bar{x} - \mu)^2$, con un procedimento simile a quello dell'Esempio 3.1.1, si ha

$$\begin{aligned} l(\mu, v) &= \log c - \frac{n}{2} \log v - \frac{n}{2v} (s^2 + (\bar{x} - \mu)^2) + \log \mathbf{1}_{(-\infty, \mu_0]}(\mu) \\ &\leq \log c - \frac{n}{2} \log(s^2 + (\bar{x} - \mu)^2) - \frac{n}{2} + \log \mathbf{1}_{(-\infty, \mu_0]}(\mu) = l(\mu, s^2 + (\bar{x} - \mu)^2) \\ &\leq \max\left(\log c - \frac{n}{2} \log s^2 - \frac{n}{2} + \log \mathbf{1}_{(-\infty, \mu_0]}(\bar{x}), \log c - \frac{n}{2} \log[s^2 + (\bar{x} - \mu_0)^2] - \frac{n}{2}\right) \\ &= \max(l(\bar{x}, s^2) + \log \mathbf{1}_{(-\infty, \mu_0]}(\bar{x}), l(\mu_0, s^2 + (\bar{x} - \mu_0)^2)) = \max_{(\mu, v) \in \Theta_0} l(\mu, v). \end{aligned}$$

Di conseguenza la determinazione campionaria di R è data da

$$\begin{aligned} r &= \mathbf{1}_{(-\infty, \mu_0]}(\bar{x}) e^{l_p(\mu_0) - l(\bar{x}, s^2)} + \mathbf{1}_{(\mu_0, \infty)}(\bar{x}) \\ &= \mathbf{1}_{(-\infty, 0]}(t) \left(1 + \frac{t^2}{n-1}\right)^{-n/2} + \mathbf{1}_{(0, \infty)}(t), \end{aligned}$$

dove

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s_c}.$$

Per i motivi esposti nella sezione precedente, quando H_0 è vera, t è la determinazione campionaria di una statistica T distribuita come una t_{n-1} . Dal momento che r è una funzione monotona crescente di t , allora i test costruiti su R e T sono equivalenti. Inoltre, rifiutare H_0 per piccole determinazioni di R è equivalente a rifiutare H_0 per piccole determinazioni di T , e dunque la regione critica del test è data da

$$\mathcal{T}_1 = \{t : t \leq t_{n-1, \alpha}\}.$$

Infine, per un dato valore campionario t , il livello di significatività osservato risulta

$$\alpha_{oss} = \Pr(t_{n-1} \leq t).$$

È immediato estendere questi risultati al sistema di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu > \mu_0$. In questo caso la regione critica del test è data da

$$\mathcal{T}_1 = \{t : t \geq t_{n-1, 1-\alpha}\},$$

mentre il livello di significatività osservato risulta

$$\alpha_{oss} = \Pr(t_{n-1} \geq t).$$

Infine, la statistica $-2 \log R_n$ non converge ad una Chi-quadrato, in quanto l'ipotesi di base non è del tipo specificato nella §4.2.3.

• **Esempio 4.3.2.** Su un campione casuale di 20 soggetti in sovrappeso (con un peso corporeo superiore a 100 chilogrammi) è stato misurato il livello di colesterolo ottenendo i dati della Tavola 4.3.2.

Tavola 4.3.2. Livello di colesterolo (in mg per 100 ml).

| soggetto | x_i | soggetto | x_i |
|----------|-------|----------|-------|
| 1 | 344 | 11 | 226 |
| 2 | 185 | 12 | 175 |
| 3 | 263 | 13 | 242 |
| 4 | 246 | 14 | 252 |
| 5 | 224 | 15 | 153 |
| 6 | 212 | 16 | 183 |
| 7 | 188 | 17 | 137 |
| 8 | 250 | 18 | 202 |
| 9 | 148 | 19 | 194 |
| 10 | 169 | 20 | 213 |

Fonte: Selvin (1991)

Da numerose esperienze cliniche risulta che il livello di colesterolo nel sangue di una persona sana è di circa 190 mg per 100 ml. Si sospetta che i soggetti in forte sovrappeso abbiano un livello del colesterolo più alto della norma e dunque si vuole verificare il sistema di ipotesi $H_0 : \mu = 190$ contro $H_1 : \mu > 190$. Dal momento che per questi dati si ottiene $\bar{x} = 210.3$ e $s_c = 48.3399$, allora si ha $t = 1.8780$. Di conseguenza, il livello di significatività osservato è dato da

$$\alpha_{oss} = \Pr(t_{19} \geq 1.8780) = 0.0379.$$

Dal momento che la significatività osservata è bassa, si è propensi ad affermare che i soggetti in sovrappeso abbiano un livello del colesterolo più alto della norma, in quanto si può respingere H_0 ad ogni livello di significatività $\alpha > 0.0379$. \square

4.3.3. Il test Chi-quadrato per la verifica di ipotesi sulla varianza. Si consideri un campionamento casuale da $X \sim N(\mu, v)$ e si supponga di voler verificare il sistema di ipotesi $H_0 : v = v_0$ contro $H_1 : v \neq v_0$. Anche se questa verifica di ipotesi non si adopera usualmente nella pratica, sarà utilizzata nel prossimo capitolo per ulteriori sviluppi teorici. Si ha

$$\Theta = \{(\mu, v) : (\mu, v) \in \mathbb{R} \times \mathbb{R}^+\}$$

e

$$\Theta_0 = \{(\mu, v) : v = v_0, \mu \in \mathbb{R}\}.$$

In base ai risultati dell'Esempio 3.1.2, risulta

$$\max_{(\mu, v) \in \Theta} l(\mu, v) = l(\bar{x}, s^2).$$

Inoltre, si ha

$$\begin{aligned} l(\mu, v_0) &= \log c - \frac{n}{2} \log v_0 - \frac{n}{2v_0} (s^2 + (\bar{x} - \mu)^2) \\ &\leq \log c - \frac{n}{2} \log v_0 - \frac{ns^2}{2v_0} = l(\bar{x}, v_0) = l_p(v_0) = \max_{(\mu, v) \in \Theta_0} l(\mu, v). \end{aligned}$$

Di conseguenza, la determinazione campionaria di R è data da

$$\begin{aligned} r &= e^{l_p(v_0) - l(\bar{x}, s^2)} = e^{\frac{n}{2} \log\left(\frac{s^2}{v_0}\right) - \frac{ns^2}{2v_0} + \frac{n}{2}} \\ &= \left(\frac{s^2}{v_0} e^{-\frac{s^2}{v_0} + 1}\right)^{\frac{1}{2}n} = \left(\frac{u}{n} e^{-\frac{u}{n} + 1}\right)^{\frac{1}{2}n}, \end{aligned}$$

dove

$$u = \frac{ns^2}{v_0}.$$

Quando H_0 è vera, u è la determinazione campionaria di una statistica U distribuita come una Chi-quadrato con $(n-1)$ gradi di libertà (vedi Esempio 1.2.5). In questo caso, r non è una funzione biunivoca di u e quindi i test costruiti su R e U non sono equivalenti. Tuttavia, dal momento che $r = r(u)$ è una funzione prima crescente e successivamente decrescente, allora la regione critica del test è del tipo $\mathcal{T}_1 = \{u : u \leq c_1, u \geq c_2\}$, dove c_1 e c_2 vanno scelti in modo tale che $r(c_1) = r(c_2)$ e $\Pr(c_1 < \chi_{n-1}^2 < c_2) = 1 - \alpha$. Questa procedura risulta complessa e usualmente si adotta come regione critica

$$\mathcal{T}_1 = \{u : u \leq \chi_{n-1, \alpha/2}^2, u \geq \chi_{n-1, 1-\alpha/2}^2\}.$$

Dato che la distribuzione di U non è simmetrica, il livello di significatività osservato non è definito. Tuttavia, dato che per n elevato la distribuzione di U è approssimativamente simmetrica, se si è osservato il valore campionario u , usualmente si pone

$$\alpha_{oss} = 2 \min(\Pr(\chi_{n-1}^2 \leq u), \Pr(\chi_{n-1}^2 \geq u)).$$

Infine, per grandi campioni la quantità

$$-2 \log r = -n \log\left(\frac{u}{n}\right) + u - n$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_1^2$, in quanto vi sono due parametri, di cui uno da stimare sotto ipotesi di base.

4.3.4. Il test t di Student a due campioni. Si consideri due campioni casuali indipendenti $(x_{11}, \dots, x_{1n_1})$ e $(x_{21}, \dots, x_{2n_2})$ rispettivamente da $X_1 \sim N(\mu_1, v)$ e $X_2 \sim N(\mu_2, v)$, ovvero da due variabili casuali Normali indipendenti con differenti medie e uguale varianza. Quando si assume l'omogeneità delle varianze come in questo caso, le variabili casuali sono dette omoschedastiche. Sia inoltre $n = n_1 + n_2$ la numerosità campionaria globale. Il modello statistico è dato da

$$\mathcal{F}_{\mu_1, \mu_2, v} = \left\{ f_n : f_n(x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{2n_2}; \mu_1, \mu_2, v) = \prod_{j=1}^2 \prod_{i=1}^{n_j} \frac{1}{\sqrt{v}} \phi\left(\frac{x_{ji} - \mu_j}{\sqrt{v}}\right) \right\},$$

dove $(\mu_1, \mu_2, v) \in \mathbb{R}^2 \times \mathbb{R}^+$. Si vuole verificare il sistema di ipotesi $H_0 : \mu_1 = \mu_2 = \mu$ contro $H_1 : \mu_1 \neq \mu_2$. Dunque, si ha

$$\Theta = \{(\mu_1, \mu_2, v) : (\mu_1, \mu_2, v) \in \mathbb{R}^2 \times \mathbb{R}^+\}$$

e

$$\Theta_0 = \{(\mu_1, \mu_2, v) : \mu_1 = \mu_2, v \in \mathbb{R}^+\}.$$

Se \bar{x}_j e s_j^2 rappresentano rispettivamente le realizzazioni campionarie della media campionaria \bar{X}_j e della varianza campionaria S_j^2 per il j -esimo campione, allora estendendo i risultati dell'Esempio 1.3.1 la funzione di log-verosimiglianza è data da

$$\begin{aligned}
l(\mu_1, \mu_2, \nu) &= \log c + \sum_{j=1}^2 \left(-\frac{n_j}{2} \log \nu - \frac{n_j}{2\nu} (s_j^2 + (\bar{x}_j - \mu_j)^2) \right) \\
&= \log c - \frac{n}{2} \log \nu - \frac{ns_w^2}{2\nu} - \frac{1}{2\nu} \sum_{j=1}^2 n_j (\bar{x}_j - \mu_j)^2,
\end{aligned}$$

dove

$$s_w^2 = \frac{1}{n} \sum_{j=1}^2 n_j s_j^2.$$

In maniera analoga all'Esempio 3.1.2, adoperando la disuguaglianza del medesimo esempio con $d = s_w^2$, si ha

$$\begin{aligned}
l(\mu_1, \mu_2, \nu) &\leq \log c - \frac{n}{2} \log \nu - \frac{ns_w^2}{2\nu} = l(\bar{x}_1, \bar{x}_2, \nu) \\
&\leq \log c - \frac{n}{2} \log s_w^2 - \frac{n}{2} = l(\bar{x}_1, \bar{x}_2, s_w^2) = \max_{(\mu_1, \mu_2, \nu) \in \Theta} l(\mu_1, \mu_2, \nu).
\end{aligned}$$

Inoltre, tenendo presente che se H_0 è vera si dispone in effetti di un unico campione da $N(\mu, \nu)$, si ottiene

$$l(\mu, \mu, \nu) \leq l(\bar{x}, \bar{x}, s^2) = \max_{(\mu_1, \mu_2, \nu) \in \Theta_0} l(\mu_1, \mu_2, \nu),$$

dove

$$\bar{x} = \frac{1}{n} \sum_{j=1}^2 n_j \bar{x}_j$$

e

$$s^2 = \frac{1}{n} \sum_{j=1}^2 \sum_{i=1}^{n_j} (x_{ji} - \bar{x})^2.$$

Inoltre risulta

$$s^2 = \frac{1}{n} \sum_{j=1}^2 \sum_{i=1}^{n_j} ((x_{ji} - \bar{x}_j) + (\bar{x}_j - \bar{x}))^2 = \frac{1}{n} \sum_{j=1}^2 n_j s_j^2 + \frac{1}{n} \sum_{j=1}^2 n_j (\bar{x}_j - \bar{x})^2 = s_w^2 + s_b^2,$$

dove si è tenuto presente che $\sum_{i=1}^{n_j} (x_{ji} - \bar{x}_j) = 0$ e dove si è posto

$$s_b^2 = \frac{1}{n} \sum_{j=1}^2 n_j (\bar{x}_j - \bar{x})^2.$$

La precedente relazione evidenzia la classica scomposizione della varianza totale s^2 nella varianza all'interno dei gruppi s_w^2 e nella varianza fra i gruppi s_b^2 . È facile inoltre verificare che

$$s_b^2 = \frac{n_1 n_2}{n^2} (\bar{x}_2 - \bar{x}_1)^2.$$

Dunque, risulta

$$l(\bar{x}, \bar{x}, s^2) = \log c - \frac{n}{2} \log s^2 - \frac{ns_w^2}{2s^2} - \frac{ns_b^2}{2s^2} = \log c - \frac{n}{2} \log s^2 - \frac{n}{2}.$$

Di conseguenza, la determinazione campionaria di R è data da

$$\begin{aligned} r &= e^{l(\bar{x}, \bar{s}^2) - l(\bar{x}_1, \bar{x}_2, s_w^2)} = e^{-\frac{n}{2} \log s^2 + \frac{n}{2} \log s_w^2} \\ &= \left(1 + \frac{s_b^2}{s_w^2}\right)^{-n/2} = \left(1 + \frac{t^2}{n-2}\right)^{-n/2}, \end{aligned}$$

dove

$$t = \sqrt{n-2} \frac{s_b}{s_w}.$$

Le quantità s_b^2 e s_w^2 sono le determinazioni campionarie di due statistiche S_b^2 e S_w^2 indipendenti. Questo risultato è una generalizzazione dell'Esempio 1.2.5, in quanto le statistiche S_b^2 e S_w^2 sono trasformate rispettivamente delle medie e delle varianze campionarie di campioni indipendenti. Inoltre, dal momento che

$$\frac{nS_w^2}{v} = \sum_{j=1}^2 \frac{n_j S_j^2}{v},$$

dove le statistiche $n_j S_j^2/v \sim \chi_{n_j-1}^2$ sono indipendenti in quanto i campioni sono indipendenti, per la proprietà i) della §A.3.2, si ha $nS_w^2/v \sim \chi_{n-2}^2$. Dal momento che risulta

$$\frac{nS^2}{v} = \frac{nS_w^2}{v} + \frac{nS_b^2}{v}$$

e che $nS^2/v \sim \chi_{n-1}^2$, ancora per la proprietà i) della §A.3.2 si deve necessariamente concludere che $nS_b^2/v \sim \chi_1^2$. Dunque, se H_0 è vera, t è la determinazione campionaria della statistica

$$T = \sqrt{n-2} \frac{S_b}{S_w} = \sqrt{\frac{nS_b^2/v}{(nS_w^2/v)/(n-2)}} = \sqrt{\frac{n_1 n_2 (n-2)}{n^2}} \frac{\bar{X}_2 - \bar{X}_1}{S_w},$$

distribuita come una t di Student con $(n-2)$ gradi di libertà. Inoltre r è una funzione monotona decrescente di $|t|$ e quindi i test costruiti su R e $|T|$ sono equivalenti. Tenendo presente la relazione fra le statistiche T e $|T|$, dal momento che rifiutare H_0 per piccole determinazioni di R è equivalente a rifiutare per quelle elevate di $|T|$, la regione critica del test è data da

$$\mathcal{T}_1 = \{t : t \leq -t_{n-2, 1-\alpha/2}, t \geq t_{n-2, 1-\alpha/2}\}.$$

Dato che la distribuzione di T è simmetrica, se si è osservato il valore campionario t , il livello di significatività osservato risulta infine

$$\alpha_{oss} = 2 \Pr(t_{n-2} \geq |t|).$$

Anche se in questo caso la distribuzione della statistica test è nota per campioni finiti, si noti tuttavia che per grandi campioni la quantità

$$-2 \log r = n \log \left(1 + \frac{t^2}{n-2}\right)$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_1^2$, in quanto vi sono tre parametri, di cui due da stimare sotto ipotesi di base.

• **Esempio 4.3.3.** In un esperimento è stata misurata la secrezione di tromboglobulina urinaria in 12 pazienti sani e in 12 pazienti diabetici. I dati relativi all'esperimento clinico sono riportati nella Tavola 4.3.3.

Tavola 4.3.3. Secrezione di tromboglobulina.

| paziente sano | x_{1i} | paziente malato | x_{2i} |
|---------------|----------|-----------------|----------|
| 1 | 4.1 | 1 | 11.5 |
| 2 | 6.3 | 2 | 12.1 |
| 3 | 7.8 | 3 | 16.1 |
| 4 | 8.5 | 4 | 17.8 |
| 5 | 8.9 | 5 | 24.0 |
| 6 | 10.4 | 6 | 28.8 |
| 7 | 11.5 | 7 | 33.9 |
| 8 | 12.0 | 8 | 40.7 |
| 9 | 13.8 | 9 | 51.3 |
| 10 | 17.6 | 10 | 56.2 |
| 11 | 24.3 | 11 | 61.7 |
| 12 | 37.2 | 12 | 69.2 |

Fonte: van Oost, Veldhayzen, Timmermans e Sixma (1983)

Si sospetta che i pazienti diabetici abbiano una secrezione di tromboglobulina differente dai pazienti sani, ovvero si vuole verificare il sistema di ipotesi $H_0 : \mu_1 = \mu_2 = \mu$ contro $H_1 : \mu_1 \neq \mu_2$. Dal momento che $\bar{x}_1 = 13.5333$, $\bar{x}_2 = 35.2750$, $s_1^2 = 77.4939$ e $s_2^2 = 376.6335$, allora $s_w = 227.0637$, da cui $t = 3.3838$. Di conseguenza, il livello di significatività osservato risulta

$$\alpha_{oss} = 2 \Pr(t_{22} \geq 3.3838) = 0.0027 .$$

Dato che la significatività osservata è molto bassa, vi è una forte indicazione a concludere che i pazienti diabetici abbiano una secrezione di tromboglobulina differente dai pazienti sani. Infatti, si può respingere H_0 ad ogni livello di significatività $\alpha > 0.0027$. Risulta $-2 \log r = 10.0561$ e dunque si ha

$$\alpha_{n,oss} = \Pr(\chi_1^2 \geq 10.0561) = 0.0015 ,$$

un risultato che conferma le conclusioni precedenti. □

4.3.5. Il test F di Snedecor per l'omogeneità delle varianze. Si supponga di considerare due campioni casuali indipendenti $(x_{11}, \dots, x_{1n_1})$ e $(x_{21}, \dots, x_{2n_2})$ rispettivamente da $X_1 \sim N(\mu_1, v_1)$ e $X_2 \sim N(\mu_2, v_2)$, ovvero da due variabili casuali Normali indipendenti con differenti medie e varianze. Se inoltre $n = n_1 + n_2$ è la numerosità campionaria globale, il modello statistico è dato da

$$\mathcal{F}_{\mu_1, \mu_2, v_1, v_2} = \left\{ f_n : f_n(x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{2n_2}) = \prod_{j=1}^2 \prod_{i=1}^{n_j} \frac{1}{\sqrt{v_j}} \phi \left(\frac{x_{ji} - \mu_j}{\sqrt{v_j}} \right) \right\} ,$$

dove $(\mu_1, \mu_2, v_1, v_2) \in \mathbb{R}^2 \times \mathbb{R}^+ \times \mathbb{R}^+$ e dove per semplicità di notazione si è eliminato la dipendenza di f_n dai parametri. Si vuole verificare il sistema di ipotesi $H_0 : v_1 = v_2 = v$ contro $H_1 : v_1 \neq v_2$. Dunque, si ha

$$\Theta = \{(\mu_1, \mu_2, v_1, v_2) : (\mu_1, \mu_2, v_1, v_2) \in \mathbb{R}^2 \times \mathbb{R}^+ \times \mathbb{R}^+\}$$

e

$$\Theta_0 = \{(\mu_1, \mu_2, v_1, v_2) : v_1 = v_2, \mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}\} .$$

Se \bar{x}_j e s_j^2 rappresentano rispettivamente le realizzazioni campionarie della media campionaria \bar{X}_j e della varianza campionaria S_j^2 per il j -esimo campione, la funzione di log-verosimiglianza è data da

$$l(\mu_1, \mu_2, v_1, v_2) = \log c + \sum_{j=1}^2 \left(-\frac{n_j}{2} \log v_j - \frac{n_j}{2v_j} (s_j^2 + (\bar{x}_j - \mu_j)^2) \right) .$$

In maniera analoga all'Esempio 3.1.2, adoperando la disuguaglianza del medesimo esempio rispettivamente con $d = s_1^2$ e $d = s_2^2$, si ha

$$\begin{aligned} l(\mu_1, \mu_2, v_1, v_2) &\leq \log c + \sum_{j=1}^2 \left(-\frac{n_j}{2} \log v_j - \frac{n_j s_j^2}{2v_j} \right) = l(\bar{x}_1, \bar{x}_2, v_1, v_2) \\ &\leq \log c + \sum_{j=1}^2 \left(-\frac{n_j}{2} \log s_j^2 - \frac{n_j}{2} \right) \\ &= l(\bar{x}_1, \bar{x}_2, s_1^2, s_2^2) = \max_{(\mu_1, \mu_2, v_1, v_2) \in \Theta} l(\mu_1, \mu_2, v_1, v_2). \end{aligned}$$

Inoltre, tenendo presente la definizione di s_w^2 data nella §4.3.4, si ha

$$\begin{aligned} l(\mu_1, \mu_2, v, v) &\leq \log c - \frac{n}{2} \log v - \frac{ns_w^2}{2v} = l(\bar{x}_1, \bar{x}_2, v, v) \\ &\leq \log c - \frac{n}{2} \log s_w^2 - \frac{n}{2} = l(\bar{x}_1, \bar{x}_2, s_w^2, s_w^2) = \max_{(\mu_1, \mu_2, v_1, v_2) \in \Theta_0} l(\mu_1, \mu_2, v_1, v_2). \end{aligned}$$

Di conseguenza, la determinazione campionaria di R è data da

$$\begin{aligned} r &= e^{l(\bar{x}_1, \bar{x}_2, s_w^2, s_w^2) - l(\bar{x}_1, \bar{x}_2, s_1^2, s_2^2)} = e^{-\frac{n}{2} \log s_w^2 + \sum_{j=1}^2 \frac{n_j}{2} \log s_j^2} = e^{-\sum_{j=1}^2 \frac{n_j}{2} (\log s_w^2 - \log s_j^2)} \\ &= \prod_{j=1}^2 \left(\frac{s_w^2}{s_j^2} \right)^{-\frac{1}{2}n_j} = \left(\frac{n_1}{n} \right)^{-\frac{1}{2}n_1} \left(\frac{n_2}{n} \right)^{-\frac{1}{2}n_2} \left(1 + \frac{n_2 - 1}{(n_1 - 1)f} \right)^{-\frac{1}{2}n_1} \left(\frac{(n_1 - 1)f}{n_2 - 1} + 1 \right)^{-\frac{1}{2}n_2}, \end{aligned}$$

dove

$$f = \frac{s_{c1}^2}{s_{c2}^2}.$$

Dal momento che le statistiche $(n_j - 1)S_{c_j}^2/v \sim \chi_{n_j-1}^2$ sono indipendenti in quanto i campioni sono indipendenti, se H_0 è vera, f è una determinazione della statistica

$$F = \frac{S_{c1}^2}{S_{c2}^2} = \frac{((n_1 - 1)S_{c1}^2/v)/(n_1 - 1)}{((n_2 - 1)S_{c2}^2/v)/(n_2 - 1)},$$

distribuita come una F di Snedecor con $(n_1 - 1)$ e $(n_2 - 1)$ gradi di libertà. In questo caso, r non è una funzione biunivoca di f e quindi i test costruiti su R e F non sono equivalenti. Tuttavia, notando che $r = r(f)$ è una funzione prima crescente e successivamente decrescente, allora la regione critica del test è del tipo $\mathcal{T}_1 = \{f : f \leq c_1, f \geq c_2\}$, dove c_1 e c_2 vanno scelti in modo tale che $r(c_1) = r(c_2)$ e $\Pr(c_1 < F_{n_1-1, n_2-1} < c_2) = 1 - \alpha$. Questa procedura risulta piuttosto complessa e quindi usualmente si adotta come regione critica

$$\mathcal{T}_1 = \{f : f \leq F_{n_1-1, n_2-1, \alpha/2}, f \geq F_{n_1-1, n_2-1, 1-\alpha/2}\}.$$

Dato che la distribuzione di F non è simmetrica, il livello di significatività osservato non è definito. Tuttavia, dato che per n_1 e n_2 elevati la distribuzione di F è approssimativamente simmetrica, se si è osservato il valore campionario f , usualmente si pone

$$\alpha_{oss} = 2 \min(\Pr(F_{n_1-1, n_2-1} \leq f), \Pr(F_{n_1-1, n_2-1} \geq f)).$$

Per grandi campioni la quantità

$$-2 \log r = \sum_{j=1}^2 n_j \log \left(\frac{s_w^2}{s_j^2} \right)$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_1^2$, in quanto vi sono quattro parametri, di cui tre da stimare sotto ipotesi di base.

Quando non si è sicuri che l'ipotesi di omoschedasticità nella §4.3.4 risulti valida, si dovrebbe eseguire in maniera preliminare il test analizzato in questa sezione e successivamente, se l'ipotesi di base è accettata, si dovrebbe procedere con la verifica di ipotesi sulle medie. Tuttavia, se i due test sono effettuati rispettivamente ai livelli di significatività α_1 e α_2 , allora l'ipotesi di uguaglianza delle medie verrà accettata al livello $(1 - \alpha_1)(1 - \alpha_2)$, dal momento che si può dimostrare che le due statistiche test sono indipendenti. Dunque, in effetti il livello di significatività complessivo sarà dato da $\alpha = \alpha_1 + \alpha_2 - \alpha_1\alpha_2$. Questa serie di test successivi è detta test in cascata. In generale le statistiche test di un test in cascata non sono indipendenti e quindi diventa estremamente complesso calcolare il livello di significatività globale.

• **Esempio 4.3.4.** Una fabbrica produce un tipo specifico di sfere di acciaio del diametro di 1 micron. Alla fine di una giornata di produzione sono stati estratti due campioni casuali di 10 sfere fra quelle prodotte da due macchinari ed è stato misurato il loro diametro, ottenendo i dati della Tavola 4.3.4.

Tavola 4.3.4. Diametro delle sfere (in micron).

| sfera mac. 1 | x_{1i} | sfera mac. 2 | x_{2i} |
|--------------|----------|--------------|----------|
| 1 | 1.18 | 1 | 1.72 |
| 2 | 1.42 | 2 | 1.62 |
| 3 | 0.69 | 3 | 1.69 |
| 4 | 0.88 | 4 | 0.79 |
| 5 | 1.62 | 5 | 1.79 |
| 6 | 1.09 | 6 | 0.77 |
| 7 | 1.53 | 7 | 1.44 |
| 8 | 1.02 | 8 | 1.29 |
| 9 | 1.19 | 9 | 1.96 |
| 10 | 1.32 | 10 | 0.99 |

Fonte: Romano (1977)

Si vuole controllare se la produzione è sotto controllo, ovvero se i due macchinari producono sfere con la medesima precisione, e dunque si vuole verificare il sistema di ipotesi $H_0 : v_1 = v_2 = v$ contro $H_1 : v_1 \neq v_2$. Dal momento che per questi dati è immediato verificare che $\bar{x}_1 = 1.1940$, $\bar{x}_2 = 1.4060$, $s_{c1}^2 = 0.0755$, $s_{c2}^2 = 0.1651$, allora si ha $f = 0.4574$. Di conseguenza, il livello di significatività osservato risulta

$$\alpha_{oss} = 2 \Pr(F_{9,9} \geq 0.4574) = 0.2592 .$$

Dato che la significatività osservata è abbastanza alta, si può concludere che i due macchinari producono sfere con la medesima precisione. Infatti, si può accettare H_0 ad ogni livello di significatività $\alpha < 0.2592$. Dal momento che $-2 \log r = 1.4918$, si ha

$$\alpha_{n,oss} = \Pr(\chi_1^2 \geq 1.4918) = 0.2219 ,$$

un risultato che conferma le conclusioni precedenti. Supponiamo di voler verificare inoltre il sistema di ipotesi $H_0 : \mu_1 = \mu_2 = \mu$ contro $H_1 : \mu_1 \neq \mu_2$. Dal momento che si ha $s_w = 0.1203$, allora $t = 1.2965$. Il livello di significatività osservato risulta

$$\alpha_{oss} = 2 \Pr(t_{18} \geq 1.2965) = 0.2112 .$$

Dato che la significatività osservata è abbastanza alta, si può concludere che i due macchinari producono sfere con il medesimo diametro. I due test sono in cascata e si può scegliere i livelli di significatività pari a

$$\alpha_1 = \alpha_2 = 1 - \sqrt{1 - 0.05} \simeq 0.0253$$

per accettare l'ipotesi di base ad un livello di significatività globale pari a $\alpha = 0.05$. \square

4.3.6. Il test t di Student per dati appaiati. Si supponga di considerare due campioni casuali dipendenti (x_{11}, \dots, x_{1n}) e (x_{21}, \dots, x_{2n}) , relativi ad un gruppo di n soggetti su cui è stata osservata una certa variabile casuale prima e dopo un trattamento. In questo caso l'obiettivo è quello di valutare l'efficacia del trattamento. I dati di questo tipo sono detti dati appaiati. Al fine di analizzare queste osservazioni si costruisce il nuovo campione casuale (d_1, \dots, d_n) , dove $d_i = x_{2i} - x_{1i}$. Supponendo che questo ultimo campione casuale provenga da $D \sim N(\mu, v)$, la verifica dell'efficacia del trattamento si riduce a considerare il sistema di ipotesi $H_0 : \mu = 0$ contro $H_1 : \mu \neq 0$ (o eventualmente l'opportuno sistema di ipotesi direzionale). In questo caso, è sufficiente applicare le procedure di verifica di ipotesi basate sul test t di Student discusse nelle §4.3.1 e §4.3.2.

• **Esempio 4.3.5.** Su 8 pazienti con anemia cronica grave è stato misurato l'indice di infarto prima e dopo un trattamento medico e i relativi dati sono stati riportati nella Tavola 4.3.5.

Tavola 4.3.5. Indice di infarto dei pazienti con anemia (in ml/battito/m²).

| paziente | prima | dopo | d_i |
|----------|-------|------|-------|
| 1 | 109 | 56 | - 53 |
| 2 | 57 | 44 | - 13 |
| 3 | 53 | 55 | 2 |
| 4 | 57 | 40 | - 17 |
| 5 | 68 | 62 | - 6 |
| 6 | 72 | 46 | - 26 |
| 7 | 51 | 49 | - 2 |
| 8 | 65 | 41 | - 24 |

Fonte: Bhatia, Manchanda e Roy (1969)

Per vedere se il trattamento è stato effettivo e quindi per determinare se l'indice di infarto è diminuito, si vuole verificare dunque il sistema di ipotesi $H_0 : \mu = 0$ contro $H_1 : \mu < 0$. Dal momento che si ha $\bar{d} = -17.3750$ e $s_c = 16.3845$, allora $t = -2.8055$. Dunque, il livello di significatività osservato risulta

$$\alpha_{oss} = \Pr(t_7 \leq -2.8055) = 0.0132.$$

Dato che la significatività osservata è piuttosto bassa, si deve concludere che il trattamento è stato effettivo, dal momento che si può respingere H_0 ad ogni livello di significatività $\alpha > 0.0132$. \square

4.3.7. L'analisi della varianza ad un criterio. Si supponga di considerare m campioni casuali indipendenti $(x_{j1}, \dots, x_{jn_j})$ rispettivamente dalle variabili casuali $X_j \sim N(\mu_j, v)$, ovvero da m variabili casuali Normali indipendenti con differenti medie e uguale varianza. Sia inoltre $n = \sum_{j=1}^m n_j$ la numerosità campionaria globale. In questo caso, il modello statistico è dato da

$$\mathcal{F}_{\mu_1, \dots, \mu_m, v} = \left\{ f_n : f_n(x_{11}, \dots, x_{1n_1}, \dots, x_{m1}, \dots, x_{mn_m}) = \prod_{j=1}^m \prod_{i=1}^{n_j} \frac{1}{\sqrt{v}} \phi \left(\frac{x_{ji} - \mu_j}{\sqrt{v}} \right) \right\},$$

dove $(\mu_1, \dots, \mu_m, v) \in \mathbb{R}^m \times \mathbb{R}^+$ e dove per semplicità di notazione si è eliminato la dipendenza di f_n dai parametri. Si vuole verificare il sistema di ipotesi $H_0 : \mu_1 = \dots = \mu_m = \mu$ contro $H_1 : \mu_j \neq \mu_l, \exists j \neq l = 1, \dots, m$. Dunque, si ha

$$\Theta = \{(\mu_1, \dots, \mu_m, v) : (\mu_1, \dots, \mu_m, v) \in \mathbb{R}^m \times \mathbb{R}^+\},$$

mentre

$$\Theta_0 = \{(\mu_1, \dots, \mu_m, v) : \mu_1 = \dots = \mu_m, v \in \mathbb{R}^+\}.$$

Se \bar{x}_j e s_j^2 rappresentano rispettivamente le realizzazioni campionarie della media campionaria \bar{X}_j e della varianza campionaria S_j^2 per il j -esimo campione, allora estendendo i risultati della §4.3.4 la funzione di log-verosimiglianza è data da

$$\begin{aligned} l(\mu_1, \dots, \mu_m, \nu) &= \log c + \sum_{j=1}^m \left(-\frac{n_j}{2} \log \nu - \frac{n_j}{2\nu} (s_j^2 + (\bar{x}_j - \mu_j)^2) \right) \\ &= \log c - \frac{n}{2} \log \nu - \frac{ns_w^2}{2\nu} - \frac{1}{2\nu} \sum_{j=1}^m n_j (\bar{x}_j - \mu_j)^2, \end{aligned}$$

dove

$$s_w^2 = \frac{1}{n} \sum_{j=1}^m n_j s_j^2.$$

In maniera analoga alla §4.3.4, si ha

$$\begin{aligned} l(\mu_1, \dots, \mu_m, \nu) &\leq \log c - \frac{n}{2} \log \nu - \frac{ns_w^2}{2\nu} = l(\bar{x}_1, \dots, \bar{x}_m, \nu) \\ &\leq \log c - \frac{n}{2} \log s_w^2 - \frac{n}{2} = l(\bar{x}_1, \dots, \bar{x}_m, s_w^2) \\ &= \max_{(\mu_1, \dots, \mu_m, \nu) \in \Theta} l(\mu_1, \dots, \mu_m, \nu). \end{aligned}$$

Inoltre, tenendo presente che se H_0 è vera si dispone in effetti di un unico campione da una $N(\mu, \nu)$, allora

$$l(\mu, \dots, \mu, \nu) \leq l(\bar{x}, \dots, \bar{x}, s^2) = \max_{(\mu, \dots, \mu, \nu) \in \Theta_0} l(\mu, \dots, \mu, \nu),$$

dove

$$\bar{x} = \frac{1}{n} \sum_{j=1}^m n_j \bar{x}_j$$

e

$$s^2 = \frac{1}{n} \sum_{j=1}^m \sum_{i=1}^{n_j} (x_{ji} - \bar{x})^2.$$

In completa analogia con la §4.3.4, si ha dunque $s^2 = s_w^2 + s_b^2$, dove

$$s_b^2 = \frac{1}{n} \sum_{j=1}^m n_j (\bar{x}_j - \bar{x})^2.$$

Di conseguenza, risulta

$$l(\bar{x}, \dots, \bar{x}, s^2) = \log c - \frac{n}{2} \log s^2 - \frac{n}{2}.$$

In questo caso, la determinazione campionaria di R è data da

$$\begin{aligned} r &= e^{l(\bar{x}, \dots, \bar{x}, s^2) - l(\bar{x}_1, \dots, \bar{x}_m, s_w^2)} = e^{-\frac{n}{2} \log s^2 + \frac{n}{2} \log s_w^2} \\ &= \left(1 + \frac{s_b^2}{s_w^2} \right)^{-n/2} = \left(1 + \frac{(m-1)f}{n-m} \right)^{-n/2}, \end{aligned}$$

dove

$$f = \frac{(n - m)s_b^2}{(m - 1)s_w^2}.$$

Tenendo presente i risultati della §4.3.4, s_b^2 e s_w^2 sono le determinazioni campionarie di due statistiche S_b^2 e S_w^2 indipendenti. Inoltre, dal momento che $nS_w^2/v = \sum_{j=1}^m n_j S_j^2/v$, dove le statistiche $n_j S_j^2/v \sim \chi_{n_j-1}^2$ sono indipendenti, per la proprietà i) della §A.3.2, si ha $nS_w^2/v \sim \chi_{n-m}^2$. Risulta

$$\frac{nS^2}{v} = \frac{nS_w^2}{v} + \frac{nS_b^2}{v}$$

mentre si ha $nS^2/v \sim \chi_{n-1}^2$ di nuovo per la proprietà i) della §A.3.2, e quindi si deve necessariamente concludere che $nS_b^2/v \sim \chi_{m-1}^2$. Dunque, se H_0 è vera, f è la determinazione campionaria della statistica

$$F = \frac{(n - m)S_b^2}{(m - 1)S_w^2} = \frac{(nS_b^2/v)/(m - 1)}{(nS_w^2/v)/(n - m)},$$

distribuita come una F di Snedecor con $(m - 1)$ e $(n - m)$ gradi di libertà. In questo caso, r è una funzione monotona decrescente di f e quindi i test costruiti su R e F sono equivalenti. Dal momento che rifiutare H_0 per piccole determinazioni di R è equivalente a rifiutare per determinazioni elevate di F , la regione critica del test è data da

$$\mathcal{T}_1 = \{f : f \geq F_{m-1, n-m, 1-\alpha}\}.$$

Se si è osservato il valore campionario f , il livello di significatività osservato risulta infine

$$\alpha_{oss} = \Pr(F_{m-1, n-m} \geq f).$$

Anche se la distribuzione della statistica test è nota per campioni finiti, per grandi campioni

$$-2 \log r = n \log \left(1 + \frac{(m - 1)f}{n - m} \right),$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_{m-1}^2$, in quanto vi sono $(m + 1)$ parametri, di cui due da stimare sotto ipotesi di base.

• **Esempio 4.3.6.** In tre zone nei pressi di Bimini Lagoon alle Bahamas sono stati effettuati alcuni prelievi di acqua e sono stati misurati i relativi coefficienti di salinità (numero di parti per mille) ottenendo in questa maniera i dati della Tavola 4.3.6.

Tavola 4.3.6. Indici di salinità (numero di parti per mille).

| zona 1 | x_{1i} | zona 2 | x_{2i} | zona 3 | x_{3i} |
|--------|----------|--------|----------|--------|----------|
| 1 | 37.54 | 1 | 40.17 | 1 | 39.04 |
| 2 | 37.01 | 2 | 40.80 | 2 | 39.21 |
| 3 | 36.71 | 3 | 39.76 | 3 | 39.05 |
| 4 | 37.03 | 4 | 39.70 | 4 | 38.24 |
| 5 | 37.32 | 5 | 40.79 | 5 | 38.53 |
| 6 | 37.01 | 6 | 40.44 | 6 | 38.71 |
| 7 | 37.03 | 7 | 39.79 | 7 | 38.89 |
| 8 | 37.70 | 8 | 39.38 | 8 | 38.66 |
| 9 | 37.36 | | | 9 | 38.51 |
| 10 | 36.75 | | | 10 | 40.08 |
| 11 | 37.45 | | | | |
| 12 | 38.85 | | | | |

Fonte: Till (1974)

Si è interessati a determinare se la salinità è identica nelle tre zone considerate, ovvero il sistema di ipotesi $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu$ contro $H_1 : \mu_j \neq \mu_l, \exists j \neq l = 1, 2, 3$. Dal momento che $\bar{x}_1 = 37.3133$, $\bar{x}_2 = 40.1038$, $\bar{x}_3 = 38.8920$, $s_1^2 = 0.3008$, $s_2^2 = 0.2471$ e $s_3^2 = 0.2348$, allora $s_w^2 = 0.2645$ e $s_b^2 = 1.2934$, da cui $f = 66.0210$. Di conseguenza, il livello di significatività osservato risulta

$$\alpha_{oss} = \Pr(F_{2,27} \geq 66.0210) \simeq 0.$$

Dato che la significatività osservata è estremamente bassa, vi è una forte indicazione a concludere che le salinità sono differenti nelle tre zone considerate. Inoltre $-2 \log r = 53.2000$ e dunque si ha

$$\alpha_{n,oss} = \Pr(\chi_2^2 \geq 53.2000) \simeq 0,$$

un risultato che conferma le conclusioni precedenti. \square

4.3.8. La regressione lineare. Si consideri il modello di regressione lineare dell'Esempio 1.1.8 e si supponga di voler verificare il sistema di ipotesi $H_0 : b = 0$ contro $H_1 : b \neq 0$. Dunque,

$$\Theta = \{(a, b, v) : (a, b, v) \in \mathbb{R}^2 \times \mathbb{R}^+\}$$

e

$$\Theta_0 = \{(a, b, v) : b = 0, a \in \mathbb{R}, v \in \mathbb{R}^+\}.$$

In base ai risultati dell'Esempio 3.1.5, si ha

$$\max_{(a,b,v) \in \Theta} l(a, b, v) = l(\hat{a}, \hat{b}, s_r^2),$$

dove $\hat{a} = \bar{y}$, $\hat{b} = s_{zy}/s_z^2$ e $s_r^2 = s_y^2 - s_{zy}^2/s_z^2$. Inoltre, tenendo presente la disuguaglianza dell'Esempio 3.1.2 con $d = s_y^2$, risulta

$$\begin{aligned} l(a, 0, v) &= \log c - \frac{n}{2} \log v - \frac{n}{2v} [s_y^2 + (\bar{y} - a)^2] \leq \log c - \frac{n}{2} \log v - \frac{ns_y^2}{2v} = l(\bar{y}, 0, v) \\ &\leq \log c - \frac{n}{2} \log s_y^2 - \frac{n}{2} = l(\bar{y}, 0, s_y^2) = \max_{(a,b,v) \in \Theta_0} l(a, b, v). \end{aligned}$$

Inoltre, si ha

$$s_y^2 = s_y^2 - \frac{s_{zy}^2}{s_z^2} + \frac{s_{zy}^2}{s_z^2} = s_y^2(1 - \eta^2) + s_y^2\eta^2 = s_r^2 + s_s^2,$$

dove $s_r^2 = s_y^2(1 - \eta^2)$ e $s_s^2 = s_y^2\eta^2$, mentre

$$\eta^2 = \frac{s_{zy}^2}{s_z^2 s_y^2}$$

è il coefficiente di determinazione lineare. È noto che l'indice η^2 rappresenta la percentuale di variabilità spiegata dal modello lineare (vedi Fattorini, 1997). Dunque, la varianza totale s_y^2 può essere scomposta in una parte spiegata dal modello lineare, ovvero s_s^2 , ed in una parte residua, ovvero s_r^2 . In questo caso, la determinazione campionaria di R è data da

$$\begin{aligned} r &= e^{l(\bar{y}, 0, s_y^2) - l(\hat{a}, \hat{b}, s_r^2)} = e^{-\frac{n}{2} \log(s_r^2 + s_s^2) + \frac{n}{2} \log s_r^2} \\ &= \left(1 + \frac{s_s^2}{s_r^2}\right)^{-n/2} = \left(1 + \frac{f}{n-2}\right)^{-n/2}, \end{aligned}$$

dove

$$f = (n - 2) \frac{s_s^2}{s_r^2}.$$

Tenendo presente i risultati della §3.1.5, essendo $s_s^2 = s_z^2 \hat{b}^2$, allora s_r^2 e s_s^2 sono le determinazioni campionarie di due statistiche S_r^2 e S_s^2 indipendenti. Inoltre, dal momento che $nS_r^2/v \sim \chi_{n-2}^2$ e $nS_s^2/v = ns_z^2 \hat{B}^2/v \sim \chi_1^2$, se H_0 è vera, f è la determinazione campionaria di una statistica

$$F = (n - 2) \frac{S_s^2}{S_r^2} = \frac{nS_s^2/v}{(nS_r^2/v)/(n - 2)},$$

distribuita come una F di Snedecor con 1 e $(n - 2)$ gradi di libertà. In questo caso, r è una funzione monotona decrescente di f e quindi i test costruiti su R e F sono equivalenti. Dal momento che rifiutare H_0 per piccole determinazioni di R è equivalente a rifiutare per determinazioni elevate di F , la regione critica del test è data da

$$\mathcal{T}_1 = \{f : f \geq F_{1,n-2,1-\alpha}\}.$$

Se si è osservato il valore campionario f , il livello di significatività osservato risulta infine

$$\alpha_{oss} = \Pr(F_{1,n-2} \geq f).$$

• **Esempio 4.3.7.** Su 36 piante di eucalipto sono state rilevate rispettivamente la densità e la durezza di Janka, ottenendo in questa maniera i dati della Tavola 4.3.7.

Tavola 4.3.7. Densità e durezza degli eucalipti.

| eucalipto | densità | durezza | eucalipto | densità | durezza |
|-----------|---------|---------|-----------|---------|---------|
| 1 | 24.7 | 484 | 19 | 42.9 | 1270 |
| 2 | 24.8 | 427 | 20 | 45.8 | 1180 |
| 3 | 27.3 | 413 | 21 | 46.9 | 1400 |
| 4 | 28.4 | 517 | 22 | 48.2 | 1760 |
| 5 | 28.4 | 549 | 23 | 51.5 | 1710 |
| 6 | 29.0 | 648 | 24 | 51.5 | 2010 |
| 7 | 30.3 | 587 | 25 | 53.4 | 1880 |
| 8 | 32.7 | 704 | 26 | 56.0 | 1980 |
| 9 | 35.6 | 979 | 27 | 56.5 | 1820 |
| 10 | 38.5 | 914 | 28 | 57.3 | 2020 |
| 11 | 38.8 | 1070 | 29 | 57.6 | 1980 |
| 12 | 39.3 | 1020 | 30 | 59.2 | 2310 |
| 13 | 39.4 | 1210 | 31 | 59.8 | 1940 |
| 14 | 39.9 | 989 | 32 | 66.0 | 3260 |
| 15 | 40.3 | 1160 | 33 | 67.4 | 2700 |
| 16 | 40.6 | 1010 | 34 | 68.8 | 2890 |
| 17 | 40.7 | 1100 | 35 | 69.1 | 2740 |
| 18 | 40.7 | 1130 | 36 | 69.1 | 3140 |

Fonte: Williams (1959)

Si noti che la durezza di Janka è una importante caratteristica del legname che risulta tuttavia difficilmente rilevabile, mentre la densità può essere misurata in modo semplice. Sarebbe dunque opportuna l'esistenza di un legame lineare fra le due quantità, in modo tale che la durezza di Janka possa essere prevista sulla base della densità. Di conseguenza, si è interessati a verificare il sistema di ipotesi $H_0 : b = 0$ contro $H_1 : b \neq 0$. Una volta considerati gli scarti delle misurazioni relative alla densità, è immediato verificare che $\hat{a} = 1469.4722$ e $\hat{b} = 57.5067$. Inoltre, dal momento che $\eta^2 = 0.9493$, si ottiene $f = 636.9794$. Dunque, il livello di significatività osservato risulta

$$\alpha_{oss} = \Pr(F_{1,34} \geq 636.9794) \simeq 0.$$

Dato che la significatività osservata è estremamente bassa, vi è una netta indicazione a concludere che esiste un forte legame lineare fra la durezza di Janka e la densità. In effetti l'indice η^2 evidenzia che circa il 95% della variabilità è spiegata dal modello lineare.

I dati della Tavola 4.3.7 sono rappresentati graficamente nella Figura 4.3.1 insieme con la retta di regressione. \square

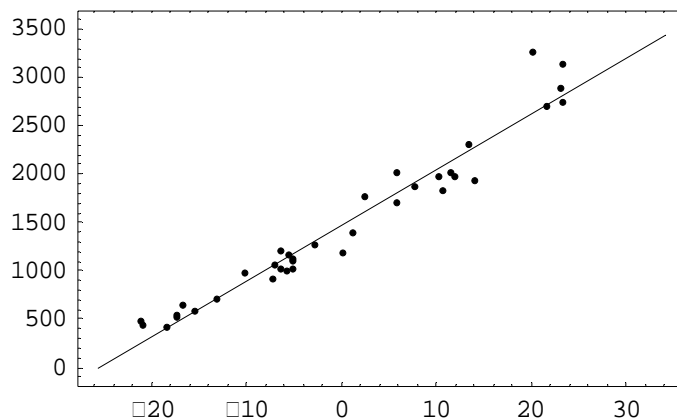


Figura 4.3.1. Diagramma per punti e retta di regressione per i dati della Tavola 4.3.7.

4.3.9. I test per dati categoriali. Si supponga di disporre di un campione casuale di n osservazioni da una variabile qualitativa X che assume valori su k categorie esaustive e mutuamente esclusive. Sia inoltre p_j la probabilità di osservare un valore nella j -esima categoria e sia n_j la frequenza assoluta di osservazioni nella medesima categoria. Ovviamente, si ha $\sum_{j=1}^k n_j = n$. L'informazione campionaria è contenuta nel vettore di frequenze (n_1, \dots, n_k) , dette appunto frequenze osservate. Dal momento che il campione è casuale e che le categorie sono esaustive e mutuamente esclusive, allora (n_1, \dots, n_k) è la realizzazione di un vettore casuale (N_1, \dots, N_k) distribuito come $Mu(n; p_1, \dots, p_k)$. Tenendo presente che $p_k = 1 - \sum_{j=1}^{k-1} p_j$ e che quindi solo i primi $(k-1)$ parametri sono distinti, il modello statistico in questo caso è dato da

$$\mathcal{F}_{p_1, \dots, p_{k-1}} = \left\{ f_n : f_n(n_1, \dots, n_k; p_1, \dots, p_{k-1}) = \binom{n}{n_1 \dots n_k} \prod_{j=1}^k p_j^{n_j} \mathbf{1}_A(n_1, \dots, n_k) \right\},$$

dove l'insieme A è definito nella §A.2.2, mentre $(p_1, \dots, p_{k-1}) \in S_{k-1}$ e

$$S_{k-1} = \left\{ (p_1, \dots, p_{k-1}) : p_j \in (0, 1), \sum_{j=1}^{k-1} p_j \leq 1, j = 1, \dots, k-1 \right\}.$$

Si vuole verificare il sistema di ipotesi $H_0 : p_j = p_{0j}, j = 1, \dots, k-1$, contro $H_1 : p_j \neq p_{0j}, \exists j = 1, \dots, k-1$, dove $(p_{01}, \dots, p_{0(k-1)}) \in S_{k-1}$ e $p_{0k} = 1 - \sum_{j=1}^{k-1} p_{0j}$. Si ha

$$\Theta = \{(p_1, \dots, p_{k-1}) : (p_1, \dots, p_{k-1}) \in S_{k-1}\},$$

mentre risulta

$$\Theta_0 = \{(p_1, \dots, p_{k-1}) : p_j = p_{0j}, j = 1, \dots, k-1\}.$$

Per evidenti motivi le componenti del vettore $(np_{01}, \dots, np_{0k})$ sono dette frequenze attese. Quindi, se si pone $\hat{p}_j = n_j/n$, la funzione di log-verosimiglianza è data da

$$l(p_1, \dots, p_{k-1}) = \log c + n \sum_{j=1}^k \hat{p}_j \log p_j, (p_1, \dots, p_{k-1}) \in S_{k-1}.$$

Tenendo presente la disuguaglianza della teoria dell'informazione

$$\sum_{j=1}^k a_j \log b_j \leq \sum_{j=1}^k a_j \log a_j,$$

dove $a_j > 0, b_j > 0$ e $\sum_{j=1}^k a_j = \sum_{j=1}^k b_j = 1$ (vedi Rao, 1973), si ottiene

$$\begin{aligned} l(p_1, \dots, p_{k-1}) &= \log c + n \sum_{j=1}^k \hat{p}_j \log p_j \leq \log c + n \sum_{j=1}^k \hat{p}_j \log \hat{p}_j \\ &= l(\hat{p}_1, \dots, \hat{p}_{k-1}) = \max_{(p_1, \dots, p_{k-1}) \in \Theta} l(p_1, \dots, p_{k-1}). \end{aligned}$$

Inoltre, è immediato verificare che

$$l(p_{01}, \dots, p_{0(k-1)}) = \max_{(p_1, \dots, p_{k-1}) \in \Theta_0} l(p_1, \dots, p_{k-1}).$$

La determinazione campionaria di R è data da

$$\begin{aligned} r &= e^{l(p_{01}, \dots, p_{0(k-1)}) - l(\hat{p}_1, \dots, \hat{p}_{k-1})} = e^{n \sum_{j=1}^k \hat{p}_j \log p_{0j} - n \sum_{j=1}^k \hat{p}_j \log \hat{p}_j} \\ &= \prod_{j=1}^k \left(\frac{p_{0j}}{\hat{p}_j} \right)^{n \hat{p}_j}. \end{aligned}$$

La distribuzione per campioni finiti di R non può essere ottenuta in modo semplice e quindi è conveniente adoperare i risultati per grandi campioni. Dunque, la quantità

$$-2 \log r = -2n \sum_{j=1}^k \hat{p}_j \log \left(\frac{p_{0j}}{\hat{p}_j} \right)$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_{k-1}^2$, in quanto vi sono $(k-1)$ parametri.

Tenendo presente le proprietà della Multinomiale (vedi §A.2.2), si ha

$$\begin{aligned} \mathbf{I}_n(p_{01}, \dots, p_{0(k-1)}) &= -\mathbf{E} \left[\begin{pmatrix} -\frac{N_1}{p_{01}^2} - \frac{N_k}{p_{0k}^2} & -\frac{N_k}{p_{0k}^2} & \dots & -\frac{N_k}{p_{0k}^2} \\ -\frac{N_k}{p_{0k}^2} & -\frac{N_2}{p_{02}^2} - \frac{N_k}{p_{0k}^2} & \dots & -\frac{N_k}{p_{0k}^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{N_k}{p_{0k}^2} & -\frac{N_k}{p_{0k}^2} & \dots & -\frac{N_{k-1}}{p_{0(k-1)}^2} - \frac{N_k}{p_{0k}^2} \end{pmatrix} \right] \\ &= \begin{pmatrix} \frac{n}{p_{01}} + \frac{n}{p_{0k}} & \frac{n}{p_{0k}} & \dots & \frac{n}{p_{0k}} \\ \frac{n}{p_{0k}} & \frac{n}{p_{02}} + \frac{n}{p_{0k}} & \dots & \frac{n}{p_{0k}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{n}{p_{0k}} & \frac{n}{p_{0k}} & \dots & \frac{n}{p_{0(k-1)}} + \frac{n}{p_{0k}} \end{pmatrix}. \end{aligned}$$

Dunque, sulla base della discussione fatta nella §4.2.3, una statistica che fornisce un test equivalente a $-2 \log R_n$ per grandi campioni è data da

$$\begin{aligned}
Q_n &= (\hat{p}_1 - p_{01}, \dots, \hat{p}_{k-1} - p_{0(k-1)})^T \mathbf{I}_n(p_{01}, \dots, p_{0(k-1)}) (\hat{p}_1 - p_{01}, \dots, \hat{p}_{k-1} - p_{0(k-1)}) \\
&= \sum_{j=1}^{k-1} \frac{n(\hat{p}_j - p_{0j})^2}{p_{0j}} + \sum_{j=1}^{k-1} \sum_{l=1}^{k-1} \frac{n(\hat{p}_j - p_{0j})(\hat{p}_l - p_{0l})}{p_{0k}} \\
&= \sum_{j=1}^{k-1} \frac{(n\hat{p}_j - np_{0j})^2}{np_{0j}} + \frac{n}{p_{0k}} \left(\sum_{j=1}^{k-1} (\hat{p}_j - p_{0j}) \right)^2 \\
&= \sum_{j=1}^{k-1} \frac{(n\hat{p}_j - np_{0j})^2}{np_{0j}} + \frac{n(\hat{p}_k - p_{0k})^2}{p_{0k}} = \sum_{j=1}^k \frac{(n_j - np_{0j})^2}{np_{0j}}.
\end{aligned}$$

La statistica test Q_n è la cosiddetta statistica di Pearson. Dato che Q_n ha le medesime proprietà per grandi campioni di $-2 \log R_n$ allora risulta $Q_n \xrightarrow{d} \chi_{k-1}^2$. Tuttavia, si può dimostrare che Q_n converge più rapidamente di $-2 \log R_n$ (vedi Ferguson, 1996). Infatti, una numerosità campionaria di $n \geq 25$ è solitamente adeguata per l'approssimazione per grandi campioni, qualora nessuna delle frequenze attese sia minore di 1. In caso contrario, si può raggruppare le categorie in modo da ottenere frequenze attese maggiori di 1 per ogni categoria. Infine, la regione critica per grandi campioni di Q_n è data da

$$\mathcal{T}_{n,1} = \{q : q \geq \chi_{k-1,1-\alpha}^2\},$$

mentre per un dato valore campionario q il livello di significatività osservato risulta

$$\alpha_{n,oss} = \Pr(\chi_{k-1}^2 \geq q).$$

• **Esempio 4.3.8.** In un esperimento di genetica sono stati considerati ibridi di pomodoro con un rapporto atteso di quattro fenotipi pari a 9 : 3 : 3 : 1. I dati della Tavola 4.3.8 sono relativi al numero di piante generate per ogni fenotipo.

Tavola 4.3.8. Numero di piante di pomodoro per fenotipo.

| Fenotipo | Frequenze osservate |
|-----------------------|---------------------|
| Alto, foglia corta | 926 |
| Alto, foglia a patata | 288 |
| Nano, foglia corta | 293 |
| Nano, foglia a patata | 104 |

Fonte: McArthur (1931)

Si vuole verificare sperimentalmente i risultati della teoria genetica, ovvero il sistema di ipotesi $H_0 : p_1 = 9/16, p_2 = 3/16, p_3 = 3/16$ contro $H_1 : p_j \neq p_{0j}, \exists j \neq l = 1, 2, 3$. Ovviamente risulta $p_4 = 1/16$. Dal momento che $n = 1611$, allora è immediato ricavare che $\hat{p}_1 = 0.5748, \hat{p}_2 = 0.1788, \hat{p}_3 = 0.1819$ e $\hat{p}_4 = 0.0645$. Dunque, risulta $-2 \log r = 1.4776$, con un livello di significatività osservato per grandi campioni

$$\alpha_{n,oss} = \Pr(\chi_3^2 \geq 1.4776) = 0.6875.$$

Inoltre, si ha $q = 1.4687$, con un livello di significatività osservato per grandi campioni

$$\alpha_{n,oss} = \Pr(\chi_3^2 \geq 1.4687) = 0.6895.$$

Data l'elevata numerosità campionaria i due test forniscono ovviamente risultati quasi equivalenti, e quindi si deve concludere che i risultati teorici sono in questo caso supportati dall'evidenza empirica. \square

4.3.10. I test per l'indipendenza nelle tavole di contingenza. Si supponga di disporre di un campione casuale di n osservazioni da una coppia di variabili qualitative (X, Y) che assumono valori rispettivamente su k e h categorie esaustive e mutuamente esclusive. Sia inoltre p_{jl} la probabilità di

osservare un valore contemporaneamente nella j -esima categoria di X e nella l -esima categoria di Y e sia n_{jl} la frequenza assoluta di osservazioni nella medesima categoria. Ovviamente, risulta che $\sum_{j=1}^k \sum_{l=1}^h n_{jl} = n$. Inoltre, si pone $\sum_{l=1}^h n_{jl} = n_{j+}$ e $\sum_{j=1}^k n_{jl} = n_{+l}$. Questo tipo di osservazioni campionarie viene usualmente organizzato nella cosiddetta tabella di contingenza (vedi Tavola 4.3.9).

Tavola 4.3.9. Tabella di contingenza.

| | 1 | 2 | ... | l | ... | h | Totale |
|--------|----------|----------|-----|----------|-----|----------|----------|
| 1 | n_{11} | n_{12} | ... | n_{1l} | ... | n_{1h} | n_{1+} |
| 2 | n_{21} | n_{22} | ... | n_{2l} | ... | n_{2h} | n_{2+} |
| ... | ... | ... | ... | ... | ... | ... | ... |
| j | n_{j1} | n_{j2} | ... | n_{jl} | ... | n_{jh} | n_{j+} |
| ... | ... | ... | ... | ... | ... | ... | ... |
| k | n_{k1} | n_{k2} | ... | n_{kl} | ... | n_{kh} | n_{k+} |
| Totale | n_{+1} | n_{+2} | ... | n_{+l} | ... | n_{+h} | n |

Si noti che l'informazione campionaria è contenuta nel vettore di frequenze $(n_{11}, \dots, n_{1h}, \dots, n_{k1}, \dots, n_{kh})$. Dal momento che il campione è casuale e che le categorie sono esaustive e mutuamente esclusive, allora $(n_{11}, \dots, n_{1h}, \dots, n_{k1}, \dots, n_{kh})$ è la realizzazione di un vettore casuale $Mu(n; p_{11}, \dots, p_{1h}, \dots, p_{k1}, \dots, p_{kh})$. Tenendo presente che solo i primi $(kh - 1)$ parametri sono distinti, il modello statistico in questo caso è dato da

$$\mathcal{F}_{p_{11}, \dots, p_{1h}, \dots, p_{k1}, \dots, p_{k(h-1)}} = \left\{ f_n : f_n(n_{11}, \dots, n_{1h}, \dots, n_{k1}, \dots, n_{kh}) = \right. \\ \left. = \binom{n}{n_{11} \dots n_{1h} \dots n_{k1} \dots n_{kh}} \prod_{j=1}^k \prod_{l=1}^h p_{jl}^{n_{jl}} \mathbf{1}_A(n_{11}, \dots, n_{1h}, \dots, n_{k1}, \dots, n_{kh}) \right\},$$

dove per semplicità di notazione si è eliminato la dipendenza di f_n dai parametri e dove l'insieme A è definito nella §A.2.2, mentre $(p_{11}, \dots, p_{1h}, \dots, p_{k1}, \dots, p_{k(h-1)}) \in S_{k(h-1)}$ e

$$S_{k(h-1)} = \left\{ (p_{11}, \dots, p_{1h}, \dots, p_{k1}, \dots, p_{k(h-1)}) : p_{jl} \in (0, 1), \sum_{j=1}^k \sum_{l=1}^{h-1} p_{jl} \leq 1 \right\}.$$

Si denoti ora con p_{j+} la probabilità di avere una osservazione nella j -esima categoria di X , ovvero $p_{j+} = \sum_{l=1}^h p_{jl}$, e con p_{+l} la probabilità di avere una osservazione nella l -esima categoria di Y , ovvero $p_{+l} = \sum_{j=1}^k p_{jl}$. Risulta dunque interessante verificare l'indipendenza delle due variabili categoriali, ovvero il sistema di ipotesi $H_0 : p_{jl} = p_{j+}p_{+l}, j = 1, \dots, k, l = 1, \dots, h$, contro $H_1 : p_{jl} \neq p_{j+}p_{+l}, \exists j = 1, \dots, k, l = 1, \dots, h$. Dunque, si ha

$$\Theta = \{(p_{11}, \dots, p_{1h}, \dots, p_{k1}, \dots, p_{k(h-1)}) : (p_{11}, \dots, p_{1h}, \dots, p_{k1}, \dots, p_{k(h-1)}) \in S_{k(h-1)}\}$$

e

$$\Theta_0 = \{(p_{11}, \dots, p_{1h}, \dots, p_{k1}, \dots, p_{k(h-1)}) : p_{jl} = p_{j+}p_{+l}, j = 1, \dots, k, l = 1, \dots, h\}.$$

Quindi, se si pone $\hat{p}_{jl} = n_{jl}/n$, la funzione di log-verosimiglianza è data da

$$l(p_{11}, \dots, p_{1h}, \dots, p_{k1}, \dots, p_{k(h-1)}) = \log c + n \sum_{j=1}^k \sum_{l=1}^h \hat{p}_{jl} \log p_{jl},$$

con $(p_{11}, \dots, p_{1h}, \dots, p_{k1}, \dots, p_{k(h-1)}) \in S_{k(h-1)}$. Tenendo presente i risultati della §4.3.9, si ottiene inoltre

$$\begin{aligned}
l(p_{11}, \dots, p_{1h}, \dots, p_{k1}, \dots, p_{k(h-1)}) &= \log c + n \sum_{j=1}^k \sum_{l=1}^h \hat{p}_{jl} \log p_{jl} \\
&\leq \log c + n \sum_{j=1}^k \sum_{l=1}^h \hat{p}_{jl} \log \hat{p}_{jl} = l(\hat{p}_{11}, \dots, \hat{p}_{1h}, \dots, \hat{p}_{k1}, \dots, \hat{p}_{k(h-1)}) \\
&= \max_{(p_{11}, \dots, p_{1h}, \dots, p_{k1}, \dots, p_{k(h-1)}) \in \Theta} l(p_{11}, \dots, p_{1h}, \dots, p_{k1}, \dots, p_{k(h-1)}).
\end{aligned}$$

In modo analogo alla §4.3.9, e tenendo presente che sotto ipotesi di base la verosimiglianza dipende solo da $(p_{1+}, \dots, p_{(k-1)+}, p_{+1}, \dots, p_{+(h-1)})$, posto $\hat{p}_{j+} = n_{j+}/n$ e $\hat{p}_{+l} = n_{+l}/n$ si ha

$$\begin{aligned}
l(p_{1+}, \dots, p_{(k-1)+}, p_{+1}, \dots, p_{+(h-1)}) &= \log c + n \sum_{j=1}^k \hat{p}_{j+} \log p_{j+} + n \sum_{l=1}^h \hat{p}_{+l} \log p_{+l} \\
&\leq \log c + n \sum_{j=1}^k \hat{p}_{j+} \log \hat{p}_{j+} + n \sum_{l=1}^h \hat{p}_{+l} \log \hat{p}_{+l} = l(\hat{p}_{1+}, \dots, \hat{p}_{(k-1)+}, \hat{p}_{+1}, \dots, \hat{p}_{+(h-1)}) \\
&= \max_{(p_{1+}, \dots, p_{(k-1)+}, p_{+1}, \dots, p_{+(h-1)}) \in \Theta_0} l(p_{1+}, \dots, p_{(k-1)+}, p_{+1}, \dots, p_{+(h-1)}).
\end{aligned}$$

La determinazione campionaria di R è data da

$$\begin{aligned}
r &= e^{l(\hat{p}_{1+}, \dots, \hat{p}_{(k-1)+}, \hat{p}_{+1}, \dots, \hat{p}_{+(h-1)}) - l(\hat{p}_{11}, \dots, \hat{p}_{1h}, \dots, \hat{p}_{k1}, \dots, \hat{p}_{k(h-1)})} \\
&= e^{n \sum_{j=1}^k \hat{p}_{j+} \log \hat{p}_{j+} + n \sum_{l=1}^h \hat{p}_{+l} \log \hat{p}_{+l} - n \sum_{j=1}^k \sum_{l=1}^h \hat{p}_{jl} \log \hat{p}_{jl}} = \frac{\prod_{j=1}^k \hat{p}_{j+}^{n \hat{p}_{j+}} \prod_{l=1}^h \hat{p}_{+l}^{n \hat{p}_{+l}}}{\prod_{j=1}^k \prod_{l=1}^h \hat{p}_{jl}^{n \hat{p}_{jl}}}
\end{aligned}$$

È evidente che la distribuzione per campioni finiti di R non può essere ottenuta in modo semplice e quindi è conveniente adoperare i risultati per grandi campioni. Dunque la quantità

$$-2 \log r = -2n \left(\sum_{j=1}^k \hat{p}_{j+} \log \hat{p}_{j+} + \sum_{l=1}^h \hat{p}_{+l} \log \hat{p}_{+l} - \sum_{j=1}^k \sum_{l=1}^h \hat{p}_{jl} \log \hat{p}_{jl} \right)$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_{(k-1)(h-1)}^2$, in quanto vi sono $(kh - 1)$ parametri, di cui $(k + h - 2)$ da stimare sotto ipotesi di base.

Con una dimostrazione analoga a quella della §4.3.9, una statistica che fornisce un test equivalente a $-2 \log R_n$ per grandi campioni è data da

$$Q_n = \sum_{j=1}^k \sum_{l=1}^h \frac{(n_{jl} - n \hat{p}_{j+} \hat{p}_{+l})^2}{n \hat{p}_{j+} \hat{p}_{+l}},$$

detta statistica di Pearson per l'indipendenza. Dato che Q_n ha le medesime proprietà per grandi campioni di $-2 \log R_n$ allora si ha $Q_n \xrightarrow{d} \chi_{(k-1)(h-1)}^2$. Anche in questo caso una numerosità campionaria di $n \geq 25$ è solitamente adeguata per l'approssimazione per grandi campioni, qualora nessuna delle frequenze attese sia minore di 1. Infine, la regione critica per grandi campioni di Q_n è data da

$$\mathcal{T}_{n,1} = \{q : q \geq \chi_{(k-1)(h-1), 1-\alpha}^2\},$$

mentre per un dato valore campionario q il livello di significatività osservato risulta

$$\alpha_{n,oss} = \Pr(\chi_{(k-1)(h-1)}^2 \geq q).$$

• **Esempio 4.3.9.** I dati della tabella di contingenza della Tavola 4.3.10 sono stati raccolti durante uno studio della malattia di Hodgkin. Ognuno dei 538 malati è stato classificato per tipologie istologiche (indicate con le sigle LP, NS, MC e LD) e per la risposta dopo tre mesi di trattamento.

Tavola 4.3.10.

| | positiva | parziale | nessuna | Totale |
|--------|----------|----------|---------|--------|
| LP | 74 | 18 | 12 | 104 |
| NS | 68 | 16 | 12 | 96 |
| MC | 154 | 54 | 58 | 266 |
| LD | 18 | 10 | 44 | 72 |
| Totale | 314 | 98 | 126 | 538 |

Fonte: Dunsmore e Daly (1987)

Si vuole verificare che la risposta al trattamento è indipendente dalla tipologia istologica del malato, ovvero il sistema di ipotesi $H_0 : p_{jl} = p_{j+}p_{+l}, j = 1, 2, 3, 4, l = 1, 2, 3$, contro $H_1 : p_{jl} \neq p_{j+}p_{+l}, \exists j = 1, 2, 3, 4, l = 1, 2, 3$. È immediato ricavare che $-2 \log r = 68.2955$, con un livello di significatività osservato per grandi campioni

$$\alpha_{n,oss} = \Pr(\chi_6^2 \geq 68.2955) \simeq 0.$$

Inoltre, si ha $q = 75.8901$, con un livello di significatività osservato per grandi campioni

$$\alpha_{n,oss} = \Pr(\chi_6^2 \geq 75.8901) \simeq 0.$$

Data l'elevata numerosità campionaria due test forniscono ovviamente risultati quasi equivalenti, ovvero si deve concludere che la risposta al trattamento dipende dalla tipologia istologica del malato. \square

Capitolo 5

La stima per intervalli

5.1. Gli intervalli di confidenza

5.1.1. La definizione di intervallo di confidenza. Si consideri un modello statistico \mathcal{F}_θ , dove θ è un parametro univariato. Piuttosto che selezionare sulla base del campione (x_1, \dots, x_n) un unico valore come nella stima per punti, può essere utile dal punto di vista operativo ottenere un insieme di valori plausibili di θ .

Supponiamo che esista una funzione

$$P : \mathcal{C}_n \times \Theta \rightarrow \mathbb{R},$$

tale che $P = P(X_1, \dots, X_n; \theta)$ abbia una distribuzione che non dipende da θ . La variabile casuale P , detta quantità pivot, non è una statistica dal momento che dipende dal parametro θ . Se c_1 e c_2 sono due valori tali che

$$\Pr_\theta(c_1 < P(X_1, \dots, X_n; \theta) < c_2) = 1 - \alpha,$$

e se $L = L(X_1, \dots, X_n)$ e $U = U(X_1, \dots, X_n)$ sono statistiche tali che

$$\begin{aligned} \{(x_1, \dots, x_n) : c_1 < P(x_1, \dots, x_n; \theta) < c_2\} &\Leftrightarrow \\ \Leftrightarrow \{(x_1, \dots, x_n) : L(x_1, \dots, x_n) < \theta < U(x_1, \dots, x_n)\}, \end{aligned}$$

allora $(L(x_1, \dots, x_n), U(x_1, \dots, x_n))$ è detto intervallo di confidenza per θ al livello di confidenza $(1 - \alpha)$. In questo contesto, il termine “confidenza” è da intendersi nel senso di sicurezza.

La nozione di intervallo di confidenza deve essere adoperata con una certa cautela, ovvero non si deve affermare che il vero valore del parametro θ_0 è contenuto in un intervallo con probabilità pari a $(1 - \alpha)$. In termini rigorosi di probabilità, una volta che l'intervallo di confidenza è stato determinato sul campione, questo contiene il vero valore θ_0 con probabilità 0 o 1. Si può affermare invece che l'intervallo di confidenza è la determinazione di un procedimento casuale che sceglie intervalli in modo tale che la probabilità di ottenere un intervallo contenente θ_0 è pari a $(1 - \alpha)$.

È possibile ottenere un intervallo di confidenza solo quando $P(x_1, \dots, x_n; \theta)$ è funzione monotona di θ . Tuttavia, il concetto di intervallo di confidenza può essere generalizzato. Data una quantità pivot P , se $B \subset \mathbb{R}$ è un insieme tale che

$$\Pr_\theta(P(X_1, \dots, X_n; \theta) \in B) = 1 - \alpha,$$

e se $E(X_1, \dots, X_n)$ è un insieme casuale tale che

$$\{(x_1, \dots, x_n) : P(x_1, \dots, x_n; \theta) \in B\} \Leftrightarrow \{(x_1, \dots, x_n) : \theta \in E(x_1, \dots, x_n)\},$$

allora $E(x_1, \dots, x_n)$ è detto insieme di confidenza per θ al livello di confidenza $(1 - \alpha)$. Ovviamente, la nozione di intervallo di confidenza è molto più utile dal punto di vista operativo di quella di insieme di confidenza.

Infine, quando la quantità pivot è una variabile casuale discreta, allora esiste un numero finito o contabile di livelli di confidenza ottenibili, che vengono detti livelli di confidenza naturali.

• **Esempio 5.1.1.** Se si considera un campione casuale da $X \sim N(\mu, 1)$, una possibile quantità pivot è data da $\sqrt{n}(\bar{X} - \mu)$. Questa variabile casuale è in effetti una quantità pivot, dal momento che la sua

distribuzione non dipende da μ essendo $\sqrt{n}(\bar{X} - \mu) \sim N(0, 1)$. Scelto un livello di confidenza α , risulta

$$\Pr_{\mu}(-z_{1-\alpha/2} < \sqrt{n}(\bar{X} - \mu) < z_{1-\alpha/2}) = 1 - \alpha,$$

da cui

$$\begin{aligned} & \{(x_1, \dots, x_n) : -z_{1-\alpha/2} < \sqrt{n}(\bar{x} - \mu) < z_{1-\alpha/2}\} \Leftrightarrow \\ & \Leftrightarrow \left\{ (x_1, \dots, x_n) : \bar{x} - z_{1-\alpha/2} \frac{1}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2} \frac{1}{\sqrt{n}} \right\}. \end{aligned}$$

Si deve dunque concludere che

$$\left(\bar{x} - z_{1-\alpha/2} \frac{1}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \frac{1}{\sqrt{n}} \right)$$

è un intervallo di confidenza per μ al livello di confidenza $(1 - \alpha)$. \square

5.1.2. La relazione fra gli intervalli di confidenza e il test statistico. Viene ora evidenziata la stretta connessione che esiste tra il problema della stima per intervalli e quello della verifica di ipotesi. Questa è anche la ragione per cui la stima per intervalli è stata analizzata successivamente alla teoria relativa alla verifica di ipotesi.

Dato un modello statistico \mathcal{F}_{θ} e il sistema di ipotesi $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$, si consideri un test corretto al livello di significatività α . Se \mathcal{C}_0 è la regione di accettazione del test, è immediato verificare che \mathcal{C}_0 dipende dal valore prefissato θ_0 di θ . Dunque risulta $\mathcal{C}_0 = \mathcal{C}_0(\theta)$. Inversamente, per un data realizzazione campionaria (x_1, \dots, x_n) deve esistere un insieme $E(x_1, \dots, x_n)$ di valori di θ per i quali si accetta l'ipotesi di base. I due insiemi sono equivalenti, ovvero

$$\{(x_1, \dots, x_n) : (x_1, \dots, x_n) \in \mathcal{C}_0(\theta)\} \Leftrightarrow \{(x_1, \dots, x_n) : \theta \in E(x_1, \dots, x_n)\}.$$

Dal momento che

$$\{(x_1, \dots, x_n) : (x_1, \dots, x_n) \in \mathcal{C}_0(\theta)\} = \{(x_1, \dots, x_n) : \mathbf{1}_{\mathcal{C}_0(\theta)}(x_1, \dots, x_n) = 1\}$$

e

$$\Pr_{\theta}((X_1, \dots, X_n) \in \mathcal{C}_0(\theta)) = \Pr_{\theta}(\mathbf{1}_{\mathcal{C}_0(\theta)}(X_1, \dots, X_n) = 1) = 1 - \alpha,$$

se come quantità pivot si considera la variabile casuale $\mathbf{1}_{\mathcal{C}_0(\theta)}(X_1, \dots, X_n)$, allora $E(x_1, \dots, x_n)$ è un insieme di confidenza al livello di confidenza $(1 - \alpha)$. Quindi, l'insieme di tutti i valori θ per cui si accetta l'ipotesi di base costituisce un insieme di confidenza per θ . Nel caso che $E(x_1, \dots, x_n)$ sia un intervallo, con la precedente procedura si ottiene ovviamente un intervallo di confidenza e questa situazione è frequente con i modelli usualmente adoperati in pratica.

Partendo dunque da un test opportuno è possibile costruire un intervallo di confidenza al livello di confidenza desiderato. Inoltre, più il test prescelto ha funzione potenza elevata più l'intervallo di confidenza risultante è desiderabile, dal momento che *a priori* dal campionamento la probabilità che esso contenga un qualunque valore θ diverso da θ_0 è bassa. È dunque buona regola costruire intervalli di confidenza a partire da un test che gode di buone proprietà.

Nel caso che si costruisca insiemi di confidenza a partire dal test del rapporto delle verosimiglianze, la regione di accettazione per il sistema di ipotesi $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$ risulta

$$\mathcal{C}_0 = \{(x_1, \dots, x_n) : L(\theta; x_1, \dots, x_n) > c_{\alpha}\},$$

dove

$$c_\alpha = r_\alpha \max_{\theta \in \Theta} L(\theta; x_1, \dots, x_n)$$

e $\Pr_\theta(R \leq r_\alpha) = \alpha$. Dunque, il relativo insieme di confidenza $E(x_1, \dots, x_n)$ al livello di confidenza $(1 - \alpha)$ è dato da tutti i valori di $\theta \in \Theta$ per cui la verosimiglianza $L(\theta; x_1, \dots, x_n)$ è maggiore di c_α . Quando la verosimiglianza è una funzione prima crescente e successivamente decrescente, si ottiene un intervallo di confidenza.

Se il modello statistico risulta \mathcal{F}_θ dove $\theta = (\theta_A, \theta_B^T)^T$, dove θ_A è unidimensionale, supponiamo di essere interessati a costruire un insieme di confidenza per θ_A a partire dal test del rapporto delle verosimiglianze. In questo caso la regione di accettazione per il sistema di ipotesi $H_0 : \theta_A = \theta_{0A}$ contro $H_1 : \theta_A \neq \theta_{0A}$ è data da

$$C_0 = \{(x_1, \dots, x_n) : L_p(\theta_A; x_1, \dots, x_n) > c_\alpha\}.$$

Di conseguenza, il relativo insieme di confidenza $E(x_1, \dots, x_n)$ al livello di confidenza $(1 - \alpha)$ è dato da tutti i valori di $\theta_A \in \Theta_A$ per cui la verosimiglianza profilo $L_p(\theta_A; x_1, \dots, x_n)$ è maggiore di

$$c_\alpha = r_\alpha \max_{\theta \in \Theta} L(\theta; x_1, x_2, \dots, x_n)$$

dove $\Pr_\theta(R \leq r_\alpha) = \alpha$. Anche in questo caso, quando la verosimiglianza profilo è una funzione prima crescente e successivamente decrescente, si ottiene un intervallo di confidenza.

• **Esempio 5.1.2.** Dato un campione casuale da $X \sim N(\mu, v)$, si vuole costruire un intervallo di confidenza per μ . Se si considera il sistema di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$, dalla §4.3.1 è noto che la regione di accettazione per il test del rapporto delle verosimiglianze risulta

$$C_0 = \left\{ (x_1, \dots, x_n) : -t_{n-1, 1-\alpha/2} < \frac{\sqrt{n}(\bar{x} - \mu)}{s_c} < t_{n-1, 1-\alpha/2} \right\},$$

che è equivalente all'insieme

$$\left\{ (x_1, \dots, x_n) : \bar{x} - t_{n-1, 1-\alpha/2} \frac{s_c}{\sqrt{n}} < \mu < \bar{x} + t_{n-1, 1-\alpha/2} \frac{s_c}{\sqrt{n}} \right\}.$$

Dunque,

$$\left(\bar{x} - t_{n-1, 1-\alpha/2} \frac{s_c}{\sqrt{n}}, \bar{x} + t_{n-1, 1-\alpha/2} \frac{s_c}{\sqrt{n}} \right)$$

costituisce un intervallo di confidenza per μ al livello di confidenza $(1 - \alpha)$.

Se si considera dunque di nuovo i dati dell'Esempio 4.3.1, si vuole costruire un intervallo di confidenza al livello di confidenza del 95% per il rapporto dei lati dei rettangoli usati dai nativi americani Shoshoni. Dunque, si sceglie $\alpha = 0.05$, e tenendo presente i risultati dell'Esempio 4.3.1 si ha

$$\left(\bar{x} - t_{19, 0.975} \frac{s_c}{\sqrt{20}}, \bar{x} + t_{19, 0.975} \frac{s_c}{\sqrt{20}} \right) = (0.6172, 0.7038). \quad \square$$

• **Esempio 5.1.3.** Dato un campione casuale da $X \sim N(\mu, v)$, si vuole costruire un intervallo di confidenza per v . Se si considera il sistema di ipotesi $H_0 : v = v_0$ contro $H_1 : v \neq v_0$, dalla §4.3.3 è noto che la regione di accettazione per il test del rapporto delle verosimiglianze può essere scelta come

$$C_0 = \left\{ (x_1, \dots, x_n) : \chi_{n-1, \alpha/2}^2 < \frac{ns^2}{v} < \chi_{n-1, 1-\alpha/2}^2 \right\}$$

che è equivalente all'insieme

$$\left\{ (x_1, \dots, x_n) : \frac{ns^2}{\chi_{n-1, 1-\alpha/2}^2} < v < \frac{ns^2}{\chi_{n-1, \alpha/2}^2} \right\}.$$

Di conseguenza,

$$\left(\frac{ns^2}{\chi_{n-1, 1-\alpha/2}^2}, \frac{ns^2}{\chi_{n-1, \alpha/2}^2} \right)$$

costituisce un intervallo di confidenza per v al livello di confidenza $(1 - \alpha)$. Sulla base della discussione dell'Esempio 4.3.3, questo intervallo di confidenza non corrisponde esattamente a quello si sarebbe ottenuto a partire dal test del rapporto delle verosimiglianze a causa della scelta semplificata fatta per i quantili.

Se si considera dunque di nuovo i dati dell'Esempio 4.3.1, un intervallo di confidenza per v al livello di confidenza del 95% è dato da

$$\left(\frac{20s^2}{\chi_{19, 0.975}^2}, \frac{20s^2}{\chi_{19, 0.025}^2} \right) = (0.0046, 0.0173). \quad \square$$

• **Esempio 5.1.4.** Dato il modello di regressione lineare dell'Esempio 1.1.8, si vuole costruire un intervallo di confidenza per b . Se si considera il sistema di ipotesi $H_0 : b = b_0$ contro $H_1 : b \neq b_0$, in maniera analoga alla §4.3.8 la regione di accettazione per il test del rapporto delle verosimiglianze è data da

$$C_0 = \left\{ (x_1, \dots, x_n) : (n-2)(\hat{b} - b)^2 \frac{s_z^2}{s_r^2} < F_{1, n-2, 1-\alpha} \right\}.$$

Essendo $t_{n, 1-\alpha/2} = \sqrt{F_{1, n, 1-\alpha}}$ per la relazione fra la t_n e la $F_{1, n}$ (vedi §A.3.4), allora la precedente regione di accettazione può essere anche espressa come

$$C_0 = \left\{ (x_1, \dots, x_n) : -t_{n-2, 1-\alpha/2} < \sqrt{n-2}(\hat{b} - b) \frac{s_z}{s_r} < t_{n-2, 1-\alpha/2} \right\},$$

che è equivalente all'insieme

$$\left\{ (x_1, \dots, x_n) : \hat{b} - t_{n-2, 1-\alpha/2} \frac{s_r}{s_z \sqrt{n-2}} < b < \hat{b} + t_{n-2, 1-\alpha/2} \frac{s_r}{s_z \sqrt{n-2}} \right\}.$$

Di conseguenza,

$$\left(\hat{b} - t_{n-2, 1-\alpha/2} \frac{s_r}{s_z \sqrt{n-2}}, \hat{b} + t_{n-2, 1-\alpha/2} \frac{s_r}{s_z \sqrt{n-2}} \right)$$

è un intervallo di confidenza per b al livello di confidenza $(1 - \alpha)$.

Se si considera di nuovo i dati dell'Esempio 4.3.7, un intervallo di confidenza per b al livello di confidenza del 95% è dato da

$$\left(\hat{b} - t_{34,0.975} \frac{s_r}{s_z \sqrt{34}}, \hat{b} + t_{34,0.975} \frac{s_r}{s_z \sqrt{34}} \right) = (52.8800, 62.1333). \quad \square$$

• **Esempio 5.1.5.** Dato un campione casuale da $X \sim Bi(1, p)$, si vuole costruire un intervallo di confidenza per p . Se si considera il sistema di ipotesi $H_0 : p = p_0$ contro $H_1 : p \neq p_0$, dall'Esempio 4.2.5 si ha che $r = r(t)$, dove $t = n\bar{x}$, è una funzione prima crescente e successivamente decrescente. Dunque, la regione di accettazione per il test del rapporto delle verosimiglianze è del tipo

$$\mathcal{C}_0 = \{(x_1, \dots, x_n) : t_1(p) < t < t_2(p)\},$$

dove

$$\Pr_p(t_1(p) < T < t_2(p)) = 1 - \alpha$$

e $r(t_1(p)) = r(t_2(p))$. Tuttavia, in maniera analoga alla §4.3.3, è conveniente scegliere $t_1(p)$ e $t_2(p)$ in modo che

$$\Pr_p(T \leq t_1(p)) = \Pr_p(T \geq t_2(p)) = \frac{\alpha}{2}.$$

Dal momento che $T \sim Bi(n, p)$, allora risulta

$$\Pr_p(T \geq t) = \sum_{i=t}^n \binom{n}{i} p^i (1-p)^{n-i} = \Pr(Z \leq p),$$

dove $Z \sim Be(0, 1; t, n-t+1)$ (vedi §A.1.5). In modo analogo si ha $\Pr_p(T \leq t) = \Pr(Z \geq p)$, dove in questo caso $Z \sim Be(0, 1; t+1, n-t)$. Inoltre, essendo

$$Be(0, 1; t, n-t+1) = \frac{t}{t + (n-t+1)F_{2(n-t+1), 2t}}$$

e

$$Be(0, 1; t+1, n-t) = \frac{t+1}{t+1 + (n-t)F_{2(n-t), 2(t+1)}}$$

(vedi §A.3.4), la regione di accettazione è equivalente a

$$\left\{ (x_1, \dots, x_n) : \frac{t}{t + (n-t+1)F_{2(n-t+1), 2t, 1-\alpha/2}} < p < \frac{t+1}{t+1 + (n-t)F_{2(n-t), 2(t+1), \alpha/2}} \right\}$$

Si deve dunque concludere che

$$\left(\frac{t}{t + (n-t+1)F_{2(n-t+1), 2t, 1-\alpha/2}}, \frac{t+1}{t+1 + (n-t)F_{2(n-t), 2(t+1), \alpha/2}} \right)$$

è un intervallo di confidenza per p al livello di confidenza $(1 - \alpha)$.

Se $n = 10$ e $t = 6$ un intervallo di confidenza per p al livello di confidenza del 95% è dato da

$$\left(\frac{6}{6 + 5F_{10, 12, 0.975}}, \frac{7}{7 + 4F_{8, 14, 0.025}} \right) = (0.2626, 0.8782). \quad \square$$

5.2. La stima per intervalli con grandi campioni

5.2.1. Gli intervalli di confidenza con grandi campioni. Molte volte non è nota la distribuzione esatta del rapporto delle verosimiglianze e non è possibile costruire l'intervallo di confidenza a partire dal relativo test. In questo caso, si può tuttavia fare riferimento alle proprietà per grandi campioni del rapporto delle verosimiglianze.

Dato il modello statistico \mathcal{F}_θ e il sistema di ipotesi $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$, supponendo verificate le condizioni della §3.3.2, sulla base dei risultati contenuti nella §4.2.3, la statistica test $-2 \log R_n$ ha la stessa distribuzione per grandi campioni della statistica test $\hat{I}_n(\hat{\Theta}_n)(\hat{\Theta}_n - \theta_0)^2$. I test basati sulle due statistiche sono equivalenti e godono quindi delle stesse proprietà ottimali per grandi campioni, anche se la seconda statistica è più opportuna in questo contesto teorico. Dal momento che risulta

$$\lim_{n \rightarrow \infty} \Pr_{\theta_0}(\hat{I}_n(\hat{\Theta}_n)(\hat{\Theta}_n - \theta_0)^2 < \chi_{1,1-\alpha}^2) = 1 - \alpha,$$

la regione di accettazione per grandi campioni del test basato sulla statistica precedente è data dall'insieme

$$\mathcal{C}_{n,0}(\theta) = \{(x_1, \dots, x_n) : \hat{I}_n(\hat{\theta}_n)(\hat{\theta}_n - \theta)^2 < \chi_{1,1-\alpha}^2\}.$$

Inoltre, tenendo presente la relazione fra la $N(0, 1)$ e la χ_1^2 (vedi Esempio 4.2.2) si ha l'equivalenza

$$\begin{aligned} & \{(x_1, \dots, x_n) : \hat{I}_n(\hat{\theta}_n)(\hat{\theta}_n - \theta)^2 < \chi_{1,1-\alpha}^2\} \Leftrightarrow \\ \Leftrightarrow & \left\{ (x_1, \dots, x_n) : \hat{\theta}_n - z_{1-\alpha/2} \frac{1}{\sqrt{\hat{I}_n(\hat{\theta}_n)}} < \theta < \hat{\theta}_n + z_{1-\alpha/2} \frac{1}{\sqrt{\hat{I}_n(\hat{\theta}_n)}} \right\}. \end{aligned}$$

Dunque, si deve concludere che

$$\left(\hat{\theta}_n - z_{1-\alpha/2} \frac{1}{\sqrt{\hat{I}_n(\hat{\theta}_n)}}, \hat{\theta}_n + z_{1-\alpha/2} \frac{1}{\sqrt{\hat{I}_n(\hat{\theta}_n)}} \right)$$

è un intervallo di confidenza per θ per grandi campioni al livello di confidenza $(1 - \alpha)$. Mentre questa costruzione di intervalli di confidenza è sempre possibile, difficilmente si ottiene l'intervallo di confidenza partendo dalla statistica test $-2 \log R_n$. Infine, la precedente procedura fornisce intervalli ottimi, almeno per grandi campioni, dal momento che il test basato sulla successione di statistiche $(\hat{I}_n(\hat{\Theta}_n)(\hat{\Theta}_n - \theta_0)^2)_{n \geq 1}$ è uniformemente più potente in senso asintotico.

Se il modello statistico risulta \mathcal{F}_θ dove $\theta = (\theta_A, \theta_B^T)^T$ dove θ_A è unidimensionale, sulla base dei risultati della §4.2.3, un intervallo di confidenza per θ_A per grandi campioni al livello di confidenza $(1 - \alpha)$ è dato da

$$\left(\hat{\theta}_{A,n} - z_{1-\alpha/2} \frac{1}{\sqrt{\hat{I}_{A,n}(\hat{\theta}_n)}}, \hat{\theta}_{A,n} + z_{1-\alpha/2} \frac{1}{\sqrt{\hat{I}_{A,n}(\hat{\theta}_n)}} \right),$$

dove con $\hat{I}_{A,n}(\hat{\theta}_n)$ si denota l'informazione osservata di Fisher relativa a θ_A .

• **Esempio 5.2.1.** Se si considera un campione casuale da $X \sim N(\mu, \nu)$, tenendo presenti i risultati dell'Esempio 3.3.10, si ha che

$$\left(\bar{x} - z_{1-\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \frac{s}{\sqrt{n}} \right)$$

e

$$\left(s^2 - z_{1-\alpha/2} \sqrt{\frac{2s^4}{n}}, s^2 + z_{1-\alpha/2} \sqrt{\frac{2s^4}{n}} \right)$$

sono intervalli di confidenza per μ e v per grandi campioni al livello di confidenza $(1 - \alpha)$. Questo risultato serve unicamente ad esemplificare, dal momento che si dispone degli intervalli di confidenza per campioni finiti. \square

• **Esempio 5.2.2.** Se si considera un campione casuale da $X \sim Bi(1, p)$ tenendo presenti i risultati dell'Esempio 3.1.8, allora risulta che

$$\left(\bar{x} - z_{1-\alpha/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \bar{x} + z_{1-\alpha/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} \right)$$

è un intervallo di confidenza per p per grandi campioni al livello di confidenza $(1 - \alpha)$.

Se $n = 10$ e $n\bar{x} = 6$, un intervallo di confidenza per p per grandi campioni al livello di confidenza del 95% è dato da

$$\left(\bar{x} - z_{0.975} \sqrt{\frac{\bar{x}(1-\bar{x})}{10}}, \bar{x} + z_{0.975} \sqrt{\frac{\bar{x}(1-\bar{x})}{10}} \right) = (0.2963, 0.9036).$$

Tenendo presente i risultati dell'Esempio 5.1.5, è evidente che questo intervallo fornisce una buona approssimazione dell'intervallo per campioni finiti.

In una indagine è stata considerata la patologia del russare su un campione casuale di 2484 individui. È risultato che 1104 dei soggetti esaminati soffrono di questa patologia in modo più o meno grave (fonte: Norton e Dunn, 1985). Dal momento che si ha $\bar{x} = 1104/2484 = 0.4444$, un intervallo di confidenza per grandi campioni al livello di confidenza del 95% per la proporzione di individui che russano è dato da

$$\left(\bar{x} - z_{0.975} \sqrt{\frac{\bar{x}(1-\bar{x})}{2484}}, \bar{x} + z_{0.975} \sqrt{\frac{\bar{x}(1-\bar{x})}{2484}} \right) = (0.4249, 0.4640). \quad \square$$

• **Esempio 5.2.3.** Se si considera un campione casuale da $X \sim W(0, 1; p)$, dal momento che non si può esprimere in forma esplicita lo stimatore di massima verosimiglianza, allora non è possibile ottenere una forma generale per l'intervallo di confidenza. Tuttavia, per il campione specifico dell'Esempio 3.1.10, essendo in questo caso $\hat{p} = 2.83525$ e $\hat{I}_n(\hat{p}) = 3.29202$, scegliendo $\alpha = 0.05$, allora $(1.7550, 3.9155)$ è un intervallo di confidenza per p per grandi campioni al livello di confidenza del 95%. \square

5.2.2. Le trasformazioni stabilizzatrici della varianza e gli intervalli di confidenza. Dal momento che l'impiego della distribuzione per grandi campioni comporta un'approssimazione, può essere conveniente considerare una riparametrizzazione per migliorare il grado di approssimazione. Supponendo verificate le condizioni della §3.3.2, se

$$g(\theta) = \int I(\theta)^{1/2} d\theta$$

è la trasformazione stabilizzatrice della varianza, allora è immediato dimostrare che

$$n(g(\hat{\Theta}_n) - g(\theta_0))^2 \xrightarrow{d} \chi_1^2.$$

Tenendo presente la discussione della sezione precedente, questa statistica è equivalente al rapporto delle verosimiglianze e quindi gode di proprietà ottime per costruire intervalli di confidenza per $g(\theta)$ per grandi campioni. Dal momento che risulta

$$\lim_{n \rightarrow \infty} \Pr_{\theta_0}(n(g(\hat{\Theta}_n) - g(\theta_0))^2 < \chi_{1,1-\alpha}^2) = 1 - \alpha,$$

la regione di accettazione per grandi campioni del test basato su questa statistica è data dall'insieme

$$\mathcal{C}_{n,0}(\theta) = \{(x_1, \dots, x_n) : n(g(\hat{\theta}_n) - g(\theta))^2 < \chi_{1,1-\alpha}^2\}.$$

Tenendo presente che g è una funzione biunivoca, si ha l'equivalenza

$$\begin{aligned} & \{(x_1, \dots, x_n) : n(g(\hat{\theta}_n) - g(\theta))^2 < \chi_{1,1-\alpha}^2\} \Leftrightarrow \\ & \Leftrightarrow \left\{ (x_1, \dots, x_n) : g(\hat{\theta}_n) - z_{1-\alpha/2} \frac{1}{\sqrt{n}} < g(\theta) < g(\hat{\theta}_n) + z_{1-\alpha/2} \frac{1}{\sqrt{n}} \right\} \Leftrightarrow \\ & \Leftrightarrow \left\{ (x_1, \dots, x_n) : g^{-1} \left(g(\hat{\theta}_n) - z_{1-\alpha/2} \frac{1}{\sqrt{n}} \right) < \theta < g^{-1} \left(g(\hat{\theta}_n) + z_{1-\alpha/2} \frac{1}{\sqrt{n}} \right) \right\}, \end{aligned}$$

ovvero

$$\left(g^{-1} \left(g(\hat{\theta}_n) - z_{1-\alpha/2} \frac{1}{\sqrt{n}} \right), g^{-1} \left(g(\hat{\theta}_n) + z_{1-\alpha/2} \frac{1}{\sqrt{n}} \right) \right)$$

è intervallo di confidenza per θ per grandi campioni. Questa procedura consente di eliminare il problema della stima della varianza e quindi rende più stabili gli intervalli di confidenza ottenuti.

• **Esempio 5.2.4.** Se si considera un campione casuale da $X \sim Bi(1, p)$, la trasformazione stabilizzatrice della varianza è data da

$$g(p) = \int \frac{1}{\sqrt{p(1-p)}} dp = 2 \sin^{-1} \sqrt{p}.$$

Di conseguenza, si ha che

$$\left(\sin^2 \left(\sin^{-1} \sqrt{\bar{x}} - z_{1-\alpha/2} \frac{1}{2\sqrt{n}} \right), \sin^2 \left(\sin^{-1} \sqrt{\bar{x}} + z_{1-\alpha/2} \frac{1}{2\sqrt{n}} \right) \right)$$

è un intervallo di confidenza per θ per grandi campioni al livello di confidenza $(1 - \alpha)$. Questo intervallo di confidenza seleziona insiemi di valori solo in $(0, 1)$, ovvero all'interno dello spazio parametrico. L'intervallo ottenuto nell'Esempio 5.2.2 invece può selezionare valori al di fuori dello spazio parametrico e questo motivo lo rende poco opportuno.

Se $n = 10$ e $n\bar{x} = 6$, un intervallo di confidenza per p per grandi campioni al livello di confidenza del 95% è dato da

$$\left(\sin^2 \left(\sin^{-1} \sqrt{\bar{x}} - z_{0.975} \frac{1}{2\sqrt{10}} \right), \sin^2 \left(\sin^{-1} \sqrt{\bar{x}} + z_{0.975} \frac{1}{2\sqrt{10}} \right) \right) = (0.2968, 0.8659),$$

Tenendo presente i risultati dell'Esempio 5.1.5, questo intervallo fornisce una approssimazione migliore di quella dell'Esempio 5.2.2.

Se si considera di nuovo i dati dell'Esempio 5.2.2, un intervallo di confidenza per grandi campioni al livello di confidenza del 95% per la proporzione di individui che russano risulta (0.4247, 0.4641). \square

5.3. Le regioni di confidenza

5.3.1. La definizione di regione di confidenza. In questa sezione viene generalizzato il concetto di intervallo di confidenza nel caso che si disponga di un vettore di parametri θ . Sia \mathcal{F}_θ un modello statistico, e P una quantità pivot. Se $B \subset \mathbb{R}$ è un insieme tale che

$$\Pr_\theta(P(X_1, \dots, X_n; \theta) \in B) = 1 - \alpha,$$

e se $E(X_1, \dots, X_n) \subset \mathbb{R}^k$ è un insieme casuale tale che per ogni θ

$$\{(x_1, \dots, x_n) : P(x_1, \dots, x_n; \theta) \in B\} \Leftrightarrow \{(x_1, \dots, x_n) : \theta \in E(x_1, \dots, x_n)\},$$

allora $E(x_1, \dots, x_n)$ è detta regione di confidenza per θ al livello di confidenza $(1 - \alpha)$. Le quantità pivot più opportune sono quelle che forniscono regioni di confidenza convesse o almeno connesse, dal momento che sono le più facili da usare in modo operativo.

Si può dimostrare che esiste un legame fra la regione di confidenza e la regione di accettazione del test per il sistema di ipotesi $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$. Partendo da un test opportuno è possibile costruire una regione di confidenza al livello di confidenza desiderato.

Nel caso che si costruisca regioni di confidenza a partire dal test del rapporto delle verosimiglianze, in maniera analoga al caso univariato, si può verificare che la regione di confidenza $E(x_1, \dots, x_n)$ al livello di confidenza $(1 - \alpha)$ è data da tutti i valori di $\theta \in \Theta$ per cui la verosimiglianza $L(\theta; x_1, \dots, x_n)$ è maggiore di

$$c_\alpha = r_\alpha \max_{\theta \in \Theta} L(\theta; x_1, \dots, x_n),$$

dove $\Pr_\theta(R \leq r_\alpha) = \alpha$. Inoltre, se si ha $\theta = (\theta_A^T, \theta_B^T)^T$, dove θ_A è un parametro a $(k - h)$ componenti, la regione di confidenza $E(x_1, \dots, x_n)$ al livello di confidenza $(1 - \alpha)$ è data da tutti i valori di $\theta_A \in \Theta_A$ per cui la verosimiglianza profilo $L_p(\theta_A; x_1, \dots, x_n)$ è maggiore di c_α .

• **Esempio 5.3.1.** Dato un campione casuale da $X \sim N(\mu, \nu)$, si vuole costruire una regione di confidenza per (μ, ν) . Supponiamo di scegliere come quantità pivot $nS^2/\nu + n(\bar{X} - \mu)^2/\nu$. Tenendo presente i risultati dell'Esempio 1.2.5 è noto che $nS^2/\nu + n(\bar{X} - \mu)^2/\nu \sim \chi_n^2$ e scelto quindi un livello di confidenza α si ha

$$\Pr_{\mu, \nu} \left(\frac{nS^2}{\nu} + n \frac{(\bar{X} - \mu)^2}{\nu} < \chi_{n, 1-\alpha}^2 \right) = 1 - \alpha,$$

da cui

$$\left\{ (x_1, \dots, x_n) : \frac{ns^2}{\nu} + n \frac{(\bar{x} - \mu)^2}{\nu} < \chi_{n, 1-\alpha}^2 \right\} \Leftrightarrow \left\{ (x_1, \dots, x_n) : s^2 + (\bar{x} - \mu)^2 < \chi_{n, 1-\alpha}^2 \frac{\nu}{n} \right\}$$

In questo caso la regione di confidenza $E(x_1, \dots, x_n)$ per (μ, ν) al livello di confidenza $(1 - \alpha)$ è data dall'epigrafico di una parabola nel piano (μ, ν) con vertice $(\bar{x}, ns^2/\chi_{n, 1-\alpha}^2)$. Questa regione di confidenza è convessa, ma tuttavia è illimitata. Per questo modello è molto laborioso ottenere la regione di confidenza a partire dal test del rapporto delle verosimiglianze. \square

5.3.2. Le regioni di confidenza per grandi campioni. È evidente che è piuttosto difficile determinare una regione di confidenza partendo dal test del rapporto delle verosimiglianze, come è stato evidenziato nella precedente sezione. Tuttavia, nel caso di grandi campioni la teoria si semplifica in modo elegante.

Dato il modello statistico \mathcal{F}_θ e il sistema di ipotesi $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$, supponendo verificate le condizioni della §3.3.2, nella 4.2.3 si è evidenziato che la statistica test $(\hat{\theta}_n - \theta_0)^T \hat{\mathbf{I}}_n(\hat{\theta}_n)(\hat{\theta}_n - \theta_0)$ è equivalente alla statistica test $-2 \log R_n$. Dal momento che risulta

$$\lim_{n \rightarrow \infty} \Pr_{\theta_0}((\hat{\theta}_n - \theta_0)^T \hat{\mathbf{I}}_n(\hat{\theta}_n)(\hat{\theta}_n - \theta_0) < \chi_{k,1-\alpha}^2) = 1 - \alpha,$$

la regione di accettazione per grandi campioni del test basato sulla statistica precedente è data dall'insieme

$$C_{n,0}(\theta) = \{(x_1, \dots, x_n) : (\hat{\theta}_n - \theta)^T \hat{\mathbf{I}}_n(\hat{\theta}_n)(\hat{\theta}_n - \theta) < \chi_{k,1-\alpha}^2\}.$$

La regione di confidenza per θ per grandi campioni al livello di confidenza $(1 - \alpha)$ è data da

$$\{\theta : (\theta - \hat{\theta}_n)^T \hat{\mathbf{I}}_n(\hat{\theta}_n)(\theta - \hat{\theta}_n) < \chi_{k,1-\alpha}^2\},$$

ovvero dall'insieme dei punti contenuti in un ellissoide di centro $\hat{\theta}_n$ con direzioni degli assi date dagli autovettori di $\hat{\mathbf{I}}_n(\hat{\theta}_n)$ e con lunghezza degli assi pari a $2\sqrt{\chi_{k,1-\alpha}^2/\alpha_j}$, dove α_j è il j -esimo autovalore di $\hat{\mathbf{I}}_n(\hat{\theta}_n)$. Questa regione di confidenza è anche detto ellissoide di confidenza. Per gli stessi motivi esposti nella §5.2.1, questa procedura fornisce regioni di confidenza ottime per grandi campioni.

Infine, se risulta $\theta = (\theta_A^T, \theta_B^T)^T$, dove θ_A è un parametro a $(k - h)$ componenti, la regione di confidenza per θ_A per grandi campioni al livello di confidenza $(1 - \alpha)$ è data da

$$\{\theta_A : (\theta_A - \hat{\theta}_{An})^T \hat{\mathbf{I}}_{An}(\hat{\theta}_n)(\theta_A - \hat{\theta}_{An}) < \chi_{k-h,1-\alpha}^2\},$$

ovvero dall'insieme dei punti contenuti in un ellissoide di centro $\hat{\theta}_{An}$ con direzioni degli assi date dagli autovettori di $\hat{\mathbf{I}}_{An}(\hat{\theta}_n)$ e con lunghezza degli assi pari a $2\sqrt{\chi_{k-h,1-\alpha}^2/\alpha_j}$, dove α_j è il j -esimo autovalore di $\hat{\mathbf{I}}_{An}(\hat{\theta}_n)$ e $\hat{\mathbf{I}}_{An}(\hat{\theta}_n)$ è l'informazione osservata di Fisher relativa a θ_A .

• **Esempio 5.3.2.** Dato il terzo modello statistico dell'Esempio 1.1.7, tenendo presente i risultati dell'Esempio 3.3.10, si ha che

$$\left\{ (\mu, v) : n \frac{(\mu - \bar{x})^2}{s^2} + n \frac{(v - s^2)^2}{2s^4} < \chi_{2,1-\alpha}^2 \right\}$$

costituisce una ellisse di confidenza per (μ, v) per grandi campioni al livello di confidenza $(1 - \alpha)$. Questa ellisse è centrata in (\bar{x}, s^2) ed ha gli assi paralleli agli assi cartesiani di lunghezze rispettivamente pari a $2\sqrt{s^2 \chi_{2,1-\alpha}^2/n}$ e $2\sqrt{2s^4 \chi_{2,1-\alpha}^2/n}$. Se si considera di nuovo i dati dell'Esempio 4.3.1, una ellisse di confidenza per (μ, v) per grandi campioni al livello di confidenza del 95% è centrata in $(0.6605, 0.0081)$ ed ha gli assi paralleli agli assi cartesiani di lunghezze rispettivamente pari a 0.0987 e 0.0125. \square

• **Esempio 5.3.3.** Dato il modello di regressione lineare dell'Esempio 1.1.8, si vuole costruire una regione di confidenza per (a, b) per grandi campioni. È evidente che

$$\left\{ (a, b) : n \frac{(a - \hat{a})^2}{s_r^2} + n \frac{(b - \hat{b})^2}{s_r^2/s_z^2} < \chi_{2,1-\alpha}^2 \right\}$$

costituisce una ellisse di confidenza per (a, b) per grandi campioni al livello di confidenza $(1 - \alpha)$. Questa ellisse è centrata in (\hat{a}, \hat{b}) ed ha gli assi paralleli agli assi cartesiani di lunghezze rispettivamente pari a $2\sqrt{s_r^2 \chi_{2,1-\alpha}^2/n}$ e $2\sqrt{s_r^2 \chi_{2,1-\alpha}^2/(ns_z^2)}$. Se si considera di nuovo i dati dell'Esempio 4.3.8, allora una ellisse di confidenza per (a, b) per grandi campioni al livello di

confidenza del 95% è centrata in (1469.4722, 57.5067) ed ha gli assi paralleli agli assi cartesiani di lunghezze pari a 145.1871 e 10.8428. \square

5.4. Alcuni esempi di intervalli di confidenza

5.4.1. Gli intervalli di confidenza e la distribuzione Esponenziale. Dato un campione casuale da $X \sim E(0, \sigma)$, si vuole costruire un intervallo di confidenza per σ . Se si considera il sistema di ipotesi $H_0 : \sigma = \sigma_0$ contro $H_1 : \sigma \neq \sigma_0$, dall'Esempio 4.2.4 è noto che il rapporto delle verosimiglianze è funzione di $u = 2n\bar{x}/\sigma_0$. Quando H_0 è vera, u è la determinazione campionaria di una statistica U distribuita come una Chi-quadrato con $2n$ gradi di libertà, essendo $n\bar{X}/\sigma_0$ distribuita come una $G(0, 1; n)$ (vedi Esempio 1.2.2). Tuttavia, r non è una funzione biunivoca di u e quindi i test costruiti su R e U non sono equivalenti. Dal momento che $r = r(u)$ è una funzione prima crescente e successivamente decrescente, risulta che la regione di accettazione del test è del tipo $\{u : c_1 \leq u \leq c_2\}$, dove c_1 e c_2 vanno scelti in modo tale che $r(c_1) = r(c_2)$ e $\Pr(c_1 \leq \chi_{2n}^2 \leq c_2) = 1 - \alpha$. Tuttavia, tenendo presente la discussione fatta nell'Esempio 4.3.3, questa procedura risulta complessa ed è conveniente invece adottare come regione di accettazione approssimata per il test del rapporto delle verosimiglianze

$$C_0 = \left\{ (x_1, \dots, x_n) : \chi_{2n, \alpha/2}^2 < \frac{2n\bar{x}}{\sigma} < \chi_{2n, 1-\alpha/2}^2 \right\},$$

che è equivalente all'insieme

$$\left\{ (x_1, \dots, x_n) : \frac{2n\bar{x}}{\chi_{2n, 1-\alpha/2}^2} < \sigma < \frac{2n\bar{x}}{\chi_{2n, \alpha/2}^2} \right\}.$$

Di conseguenza,

$$\left(\frac{2n\bar{x}}{\chi_{2n, 1-\alpha/2}^2}, \frac{2n\bar{x}}{\chi_{2n, \alpha/2}^2} \right).$$

costituisce un intervallo di confidenza per σ al livello di confidenza $(1 - \alpha)$.

Tenendo presenti i risultati dell'Esempio 3.3.9, si ha che

$$\left(\bar{x} - z_{1-\alpha/2} \frac{\bar{x}}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \frac{\bar{x}}{\sqrt{n}} \right)$$

costituisce un intervallo di confidenza per σ per grandi campioni al livello di confidenza $(1 - \alpha)$. La trasformazione stabilizzatrice della varianza è data da

$$g(\sigma) = \int \frac{1}{\sigma} d\sigma = \log \sigma.$$

Di conseguenza, risulta che

$$\left(\bar{x} e^{-z_{1-\alpha/2}/\sqrt{n}}, \bar{x} e^{z_{1-\alpha/2}/\sqrt{n}} \right)$$

è un intervallo di confidenza per σ per grandi campioni al livello di confidenza $(1 - \alpha)$. Questo intervallo di confidenza seleziona insiemi di valori solo in $(0, \infty)$, ovvero lo spazio parametrico. L'intervallo per grandi campioni ottenuto in precedenza invece può selezionare valori negativi (e quindi al di fuori dello spazio parametrico) e questo motivo lo rende poco opportuno.

Al fine di illustrare i precedenti risultati teorici sono stati considerati i dati della Tavola 5.4.1. In questa tavola sono stati riportati i tempi di rottura in ore (operative) di un campione casuale di 20 componenti elettronici.

Tavola 5.4.1. Tempi di rottura (in ore).

| componente | x_i | componente | x_i |
|------------|-------|------------|-------|
| 1 | 6278 | 11 | 3212 |
| 2 | 3113 | 12 | 9003 |
| 3 | 5236 | 13 | 3523 |
| 4 | 11584 | 14 | 12888 |
| 5 | 12628 | 15 | 9460 |
| 6 | 7725 | 16 | 13341 |
| 7 | 8604 | 17 | 17809 |
| 8 | 14266 | 18 | 2812 |
| 9 | 6125 | 19 | 11825 |
| 10 | 9350 | 20 | 2398 |

Fonte: Angus (1982).

Dal momento che si ha $\bar{x} = 8563.5$, un intervallo di confidenza al livello di confidenza del 95% per σ è dato da

$$\left(\frac{40\bar{x}}{\chi_{40,0.975}^2}, \frac{40\bar{x}}{\chi_{40,0.025}^2} \right) = (5772.3029, 14019.5637).$$

Dalla precedente discussione si ha inoltre che due intervalli di confidenza per σ approssimativamente al livello del 95% sono dati da

$$\left(\bar{x} - z_{0.975} \frac{\bar{x}}{\sqrt{20}}, \bar{x} + z_{0.975} \frac{\bar{x}}{\sqrt{20}} \right) = (4809.6147, 12317.3853)$$

e

$$\left(\bar{x} e^{-z_{0.975}/\sqrt{20}}, e^{z_{0.975}/\sqrt{20}} \bar{x} \right) = (5524.2651, 13274.8033).$$

L'intervallo di confidenza ottenuto con la trasformazione stabilizzatrice della varianza fornisce una migliore approssimazione. \square

5.4.2. Gli intervalli di confidenza e la distribuzione Uniforme. Dato un campione casuale da $X \sim U(0, \delta)$, si vuole costruire un intervallo di confidenza per δ . Se si considera il sistema di ipotesi $H_0 : \delta = \delta_0$ contro $H_1 : \delta \neq \delta_0$, tenendo presente l'Esempio 3.1.4, la determinazione campionaria del rapporto delle verosimiglianze è data da

$$r = \frac{c \delta_0^{-n} \mathbf{1}_{(0,1)}(x_{(n)}/\delta_0)}{c x_{(n)}^{-n}} = \left(\frac{x_{(n)}}{\delta_0} \right)^n \mathbf{1}_{(0,1)} \left(\frac{x_{(n)}}{\delta_0} \right),$$

Si noti che r è funzione di $x_{(n)}/\delta_0$. Dunque, quando H_0 è vera, $x_{(n)}/\delta_0$ è la determinazione campionaria di una statistica $X_{(n)}/\delta_0$ distribuita come una $Be(0, 1; n, 1)$ (vedi Esempio 1.2.6). È facile verificare che r è una funzione monotona crescente di $x_{(n)}/\delta_0$ e quindi i test costruiti su R e $X_{(n)}/\delta_0$ sono equivalenti. Dal momento che il quantile di ordine α di una $Be(0, 1; n, 1)$ è dato da $\alpha^{1/n}$, la regione di accettazione per il test del rapporto delle verosimiglianze risulta

$$\mathcal{C}_0 = \left\{ (x_1, \dots, x_n) : \alpha^{1/n} < \frac{x_{(n)}}{\delta} < 1 \right\},$$

che è equivalente all'insieme

$$\{(x_1, \dots, x_n) : x_{(n)} < \delta < x_{(n)}\alpha^{-1/n}\}.$$

Di conseguenza, $(x_{(n)}, x_{(n)}\alpha^{-1/n})$ è un intervallo di confidenza per δ al livello di confidenza $(1 - \alpha)$. Non si può ottenere un intervallo per grandi campioni sulla base dei risultati della §5.2.1, dal momento che non risultano verificate le condizioni della §3.3.2.

Capitolo 6

L'inferenza multivariata

6.1. Il campionamento con osservazioni multivariate

6.1.1. La matrice campionaria e il modello statistico multivariato. In questo capitolo saranno considerati per semplicità solo i risultati relativi ad un campionamento casuale, anche se i concetti esposti possono essere facilmente estesi a contesti più ampi. Dunque, se invece di n realizzazioni indipendenti di una variabile casuale X , si dispone di n realizzazioni indipendenti di un vettore casuale $\mathcal{X} = (X_1, \dots, X_d)^T$ con funzione di ripartizione congiunta F , allora l'esperimento produce la cosiddetta matrice campionaria di ordine $(d \times n)$ data da

$$\begin{pmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{d1} & \dots & x_{dn} \end{pmatrix} = (\mathbf{x}_1, \dots, \mathbf{x}_n),$$

dove

$$\mathbf{x}_i = \begin{pmatrix} x_{1i} \\ \vdots \\ x_{di} \end{pmatrix},$$

rappresenta la i -esima osservazione campionaria. La j -esima riga della matrice campionaria, ovvero (x_{j1}, \dots, x_{jn}) , costituisce un campione casuale univariato di n osservazioni relativo alla variabile casuale X_j .

Con considerazioni simili a quelle fatte nella §1.1.1, risulta evidente che prima del rilevamento ogni osservazione campionaria \mathcal{X}_i è un vettore casuale la cui distribuzione è identica a quella del vettore casuale \mathcal{X} . Di conseguenza, la funzione di ripartizione congiunta di $(\mathcal{X}_1, \dots, \mathcal{X}_n)$ risulta

$$F_n(\mathbf{x}_1, \dots, \mathbf{x}_n) = \prod_{i=1}^n F(\mathbf{x}_i).$$

La definizione di modello statistico si estende in modo ovvio alla struttura multivariata. Supponiamo che la funzione di ripartizione $F(\cdot; \boldsymbol{\theta})$ relativa al vettore casuale \mathcal{X} sia specificata a meno di un insieme di parametri $\boldsymbol{\theta} \in \Theta$ e che in corrispondenza di F esista una funzione di densità o di probabilità congiunta $f(\cdot; \boldsymbol{\theta})$. Dunque, se $f_n(\mathbf{x}_1, \dots, \mathbf{x}_n; \boldsymbol{\theta})$ è la funzione di densità o di probabilità congiunta relativa a $F_n(\mathbf{x}_1, \dots, \mathbf{x}_n; \boldsymbol{\theta})$ e se il campionamento è casuale, allora il modello statistico multivariato è dato da

$$\mathcal{F}_\theta = \left\{ f_n : f_n(\mathbf{x}_1, \dots, \mathbf{x}_n; \boldsymbol{\theta}) = \prod_{i=1}^n f(\mathbf{x}_i; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \right\}.$$

Infine, lo spazio campionario risulta

$$\mathcal{C}_n = \bigcup_{\boldsymbol{\theta} \in \Theta} \mathcal{S}_{n, \boldsymbol{\theta}},$$

dove $\mathcal{S}_{n, \boldsymbol{\theta}} = \mathcal{S}_\theta \times \dots \times \mathcal{S}_\theta$ e \mathcal{S}_θ è il supporto di $f(\mathbf{x}; \boldsymbol{\theta})$.

• **Esempio 6.1.1.** Nella statistica inferenziale univariata esistono numerosi modelli statistici alternativi a quello relativo ad un campionamento casuale da una variabile casuale normale. Come è stato visto nei precedenti capitoli, questi modelli vengono ampiamente applicati e per essi sono disponibili numerose procedure inferenziali ben collaudate e di facile utilizzo. Al contrario, nella statistica multivariata il modello relativo ad un campionamento casuale da un vettore normale multivariato, ovvero da $\mathcal{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, svolge un ruolo veramente fondamentale e spesso costituisce l'unico modello disponibile. Questo modello può essere espresso come

$$\mathcal{F}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}} = \left\{ f_n : f_n(\mathbf{x}_1, \dots, \mathbf{x}_n; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^n \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})} \right\},$$

dove $\boldsymbol{\mu} \in \mathbb{R}^d$ e $\boldsymbol{\Sigma} > \mathbf{0}$. La notazione $\boldsymbol{\Sigma} > \mathbf{0}$ serve ad indicare che la matrice $\boldsymbol{\Sigma}$ è definita positiva. Dal momento che la matrice di varianza-covarianza contiene solo $(d(d+1)/2)$ parametri distinti, in questo modello sono presenti $(d + d(d+1)/2)$ parametri in totale. Inoltre, lo spazio parametrico è dato dal cartesiano di \mathbb{R}^d con lo spazio costituito dall'insieme delle matrici simmetriche definite positive di ordine d . \square

6.1.2. Il vettore medio campionario e la matrice di varianza-covarianza campionaria. La definizione di statistica data nella §1.2.1 si applica in modo immediato al presente contesto multivariato. Dato un modello statistico \mathcal{F}_θ , si definisce come statistica un vettore di trasformazioni misurabili a k componenti

$$\mathbf{T} : \mathcal{C}_n \rightarrow \mathbb{R}^k$$

che non dipende da θ , mentre un valore $\mathbf{t} = \mathbf{T}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ è detto realizzazione campionaria di \mathbf{T} relativa alla matrice campionaria $(\mathbf{x}_1, \dots, \mathbf{x}_n)$.

È opportuno ora considerare in maggiore dettaglio le generalizzazioni multivariate delle statistiche media e varianza campionaria. Si consideri dunque un modello statistico \mathcal{F}_θ relativo ad un campionamento casuale dal vettore casuale \mathcal{X} per cui esistono il vettore medio $E[\mathcal{X}] = \boldsymbol{\mu}$ e la matrice di varianza-covarianza $\text{Var}[\mathcal{X}] = \boldsymbol{\Sigma}$. La realizzazione campionaria del vettore medio campionario è data

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_d \end{pmatrix},$$

dove

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ji}$$

rappresenta la media campionaria relativa alle n determinazioni della variabile casuale X_j . Il vettore medio campionario può anche essere scritto come combinazione lineare delle n osservazioni campionarie, ovvero

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i.$$

Da un punto di vista probabilistico, $\bar{\mathbf{x}}$ è dunque una realizzazione del vettore casuale

$$\bar{\mathcal{X}} = \frac{1}{n} \sum_{i=1}^n \mathcal{X}_i$$

ottenuto dalla combinazione lineare dei vettori casuali $(\mathcal{X}_1, \dots, \mathcal{X}_n)$. Di conseguenza risulta

$$E[\bar{\mathbf{x}}] = \frac{1}{n} \sum_{i=1}^n E[\mathbf{x}_i] = \boldsymbol{\mu},$$

e tenendo presente l'indipendenza dei vettori casuali $(\mathbf{x}_1, \dots, \mathbf{x}_n)$, si ha

$$\text{Var}[\bar{\mathbf{x}}] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[\mathbf{x}_i] = \frac{1}{n} \boldsymbol{\Sigma}.$$

• **Esempio 6.1.2.** Dato un campione casuale da $\mathbf{x} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, dal momento che $\bar{\mathbf{x}}$ è una combinazione lineare di vettori casuali indipendenti ed ugualmente distribuiti $\mathbf{x}_i \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, in base alla proprietà *i*) della §A.4.1, si ha $\bar{\mathbf{x}} \sim N_d(\boldsymbol{\mu}, n^{-1}\boldsymbol{\Sigma})$, ovvero i risultati ottenuti in questa sezione risultano verificati. \square

La realizzazione campionaria della matrice di varianza-covarianza campionaria è data

$$\mathbf{S} = \begin{pmatrix} s_{11} & \dots & s_{1d} \\ \vdots & \ddots & \vdots \\ s_{d1} & \dots & s_{dd} \end{pmatrix},$$

ovvero una matrice quadrata di ordine d , simmetrica e semidefinita positiva, il cui generico elemento di posto (j, l) risulta

$$s_{jl} = \frac{1}{n} \sum_{i=1}^n (x_{ji} - \bar{x}_j)(x_{li} - \bar{x}_l).$$

In altri termini, gli elementi della diagonale di \mathbf{S} sono le varianze delle determinazioni di ogni singola variabile casuale, mentre gli altri elementi sono le covarianze tra le determinazioni di ogni coppia di variabili casuali. La matrice \mathbf{S} può essere anche espressa come

$$\mathbf{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

e di conseguenza la matrice di varianza-covarianza campionaria è data della matrice casuale

$$\mathcal{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T.$$

Al fine di derivare le proprietà della matrice di varianza-covarianza campionaria, risulta conveniente esprimere \mathbf{S} nel modo seguente

$$\begin{aligned} \mathbf{S} &= \frac{1}{n} \sum_{i=1}^n ((\mathbf{x}_i - \boldsymbol{\mu}) - (\bar{\mathbf{x}} - \boldsymbol{\mu}))((\mathbf{x}_i - \boldsymbol{\mu}) - (\bar{\mathbf{x}} - \boldsymbol{\mu}))^T \\ &= \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T - (\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})^T. \end{aligned}$$

Dunque, la matrice di varianza-covarianza campionaria è data da

$$\mathcal{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T - (\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})^T.$$

Di conseguenza, si ha

$$\begin{aligned} E[\mathcal{S}] &= \frac{1}{n} \sum_{i=1}^n E[(\mathcal{X}_i - \boldsymbol{\mu})(\mathcal{X}_i - \boldsymbol{\mu})^\top] - E[(\bar{\mathcal{X}} - \boldsymbol{\mu})(\bar{\mathcal{X}} - \boldsymbol{\mu})^\top] \\ &= \frac{1}{n} \sum_{i=1}^n \text{Var}[\mathcal{X}_i] - \text{Var}[\bar{\mathcal{X}}] = \frac{n-1}{n} \boldsymbol{\Sigma}. \end{aligned}$$

• **Esempio 6.1.3.** Dato un campione casuale da $\mathcal{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, si consideri i vettori scarto

$$\mathcal{D}_i = \mathcal{X}_i - \bar{\mathcal{X}} = \sum_{j=1}^n c_{ij} \mathcal{X}_j,$$

dove le costanti c_{ij} sono definite nell'Esempio 1.2.5. Dal momento che

$$\bar{\mathcal{X}} = \sum_{j=1}^n d_j \mathcal{X}_j,$$

dove di nuovo le costanti d_j sono definite nell'Esempio 1.2.5, in analogia con il medesimo esempio sulla base della proprietà *ii*) della §A.4.1, i vettori casuali \mathcal{D}_i e $\bar{\mathcal{X}}$ sono indipendenti. Di conseguenza, essendo

$$\mathcal{S} = \frac{1}{n} \sum_{i=1}^n \mathcal{D}_i \mathcal{D}_i^\top,$$

una trasformata dei vettori scarto, allora anche \mathcal{S} e $\bar{\mathcal{X}}$ sono indipendenti. Inoltre, dall'espressione della matrice di varianza-covarianza si ottiene

$$\sum_{i=1}^n (\mathcal{X}_i - \boldsymbol{\mu})(\mathcal{X}_i - \boldsymbol{\mu})^\top = n\mathcal{S} + n(\bar{\mathcal{X}} - \boldsymbol{\mu})(\bar{\mathcal{X}} - \boldsymbol{\mu})^\top.$$

Tenendo presente che $\mathcal{X}_i \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ sono indipendenti, dalla §A.4.2 si ha

$$\sum_{i=1}^n (\mathcal{X}_i - \boldsymbol{\mu})(\mathcal{X}_i - \boldsymbol{\mu})^\top \sim W_d(\boldsymbol{\Sigma}, n),$$

mentre, essendo $\bar{\mathcal{X}} \sim N_d(\boldsymbol{\mu}, n^{-1}\boldsymbol{\Sigma})$, si ha

$$n(\bar{\mathcal{X}} - \boldsymbol{\mu})(\bar{\mathcal{X}} - \boldsymbol{\mu})^\top \sim W_d(\boldsymbol{\Sigma}, 1).$$

Dal momento che \mathcal{S} e $\bar{\mathcal{X}}$ sono indipendenti, per la proprietà *i*) della §A.4.2, si deve concludere che

$$n\mathcal{S} \sim W_d(\boldsymbol{\Sigma}, n-1).$$

Dunque i risultati di questa sezione sono confermati, dal momento che in base alla §A.4.2 si ha

$$E[\mathcal{S}] = \frac{1}{n} E[n\mathcal{S}] = \frac{n-1}{n} \boldsymbol{\Sigma}. \quad \square$$

6.1.3. La funzione di verosimiglianza con matrici campionarie. Dato un modello statistico \mathcal{F}_θ relativo ad un campionamento casuale, una volta che la matrice campionaria $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ è stata osservata, è immediato estendere il concetto di funzione di verosimiglianza al caso multivariato. Di nuovo la funzione di verosimiglianza è la funzione $L : \Theta \rightarrow \mathbb{R}^+ \cup \{0\}$ data da

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = c f_n(\mathbf{x}_1, \dots, \mathbf{x}_n; \boldsymbol{\theta}) = c \prod_{i=1}^n f(\mathbf{x}_i; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta,$$

mentre la funzione di log-verosimiglianza è data da

$$l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \log c + \sum_{i=1}^n \log f(\mathbf{x}_i; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta.$$

• **Esempio 6.1.4.** Dato un campione casuale da $\mathcal{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, la verosimiglianza risulta

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= c \prod_{i=1}^n \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})} \\ &= c \det(\boldsymbol{\Sigma})^{-\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})}, \boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\Sigma} > \mathbf{0}. \end{aligned}$$

Si ha

$$\begin{aligned} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}) &= \sum_{i=1}^n \text{tr}(\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T) \\ &= \text{tr}\left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T\right) \\ &= \text{tr}(\boldsymbol{\Sigma}^{-1}(n\mathbf{S} + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})^T)) \\ &= n \text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) + n \text{tr}(\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})^T) \\ &= n \text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) + n(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}), \end{aligned}$$

dove si è tenuta presente la relazione ottenuta nella §6.1.2. Dunque, risulta infine

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = c \det(\boldsymbol{\Sigma})^{-\frac{n}{2}} e^{-\frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - \frac{n}{2}(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})}.$$

Inoltre, la log-verosimiglianza è data da

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - \frac{n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}). \quad \square$$

6.2. La stima per punti con osservazioni multivariate

6.2.1. Gli stimatori con matrici campionarie e le relative proprietà. I concetti considerati nel Capitolo 2 si estendono in modo ovvio al contesto multivariato. Dunque, dato un modello statistico \mathcal{F}_θ relativo ad un campionamento casuale, il procedimento di stima per punti fa corrispondere ad ogni matrice campionaria $(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathcal{C}_n$ un valore $\boldsymbol{\theta} \in \Theta$ per mezzo dello stimatore, che risulta di nuovo definito come

$$\tilde{\boldsymbol{\theta}} : \mathcal{C}_n \rightarrow \Theta.$$

Di conseguenza, la realizzazione campionaria dello stimatore, ovvero la stima, è data da $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}(\mathbf{x}_1, \dots, \mathbf{x}_n)$.

Per quanto riguarda le condizioni di regolarità date nella §2.1.2, se si considera un modello statistico \mathcal{F}_θ relativo ad un campionamento casuale, la condizione a) si riduce a richiedere che le distribuzioni $f(\mathbf{x}; \boldsymbol{\theta})$ specificate dal modello statistico siano distinte per valori distinti di $\boldsymbol{\theta} \in \Theta$, mentre la condizione b) rimane inalterata. Inoltre, la condizione c) si riduce all'esistenza delle derivate

$$\frac{\partial}{\partial \boldsymbol{\theta}} f(\mathbf{x}; \boldsymbol{\theta})$$

e

$$\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} f(\mathbf{x}; \boldsymbol{\theta})$$

per ogni $\boldsymbol{\theta} \in \Theta$ e per ogni $\mathbf{x} \in \mathcal{S}_\theta$ (a meno di insiemi di probabilità nulla), mentre la condizione *d*) risulta

$$\int_{\mathcal{S}_\theta} \frac{\partial}{\partial \boldsymbol{\theta}} f(\mathbf{x}; \boldsymbol{\theta}) d\nu = \frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathcal{S}_\theta} f(\mathbf{x}; \boldsymbol{\theta}) d\nu$$

e

$$\int_{\mathcal{S}_\theta} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} f(\mathbf{x}; \boldsymbol{\theta}) d\nu = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \int_{\mathcal{S}_\theta} f(\mathbf{x}; \boldsymbol{\theta}) d\nu .$$

La definizione di correttezza e di coerenza di una stimatore rimangono inalterate nel contesto multivariato, ovvero si dice che lo stimatore $\tilde{\boldsymbol{\Theta}}$ è corretto per $\boldsymbol{\theta}$ se

$$E[\tilde{\boldsymbol{\Theta}}] = \boldsymbol{\theta} ,$$

per ogni $\boldsymbol{\theta} \in \Theta$, mentre se $\mathcal{F}_{\theta,n}$ è un modello statistico per ogni n , allora si dice che lo stimatore $\tilde{\boldsymbol{\Theta}}_n$ è coerente per $\boldsymbol{\theta}$ se

$$\tilde{\boldsymbol{\Theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0 ,$$

dove $\boldsymbol{\theta}_0$ è il vero valore del parametro nella popolazione.

In modo ovvio si può estendere anche il concetto di vettore di funzioni punteggio, ovvero nel caso di campionamento casuale si ha

$$\mathbf{s}_n(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \mathbf{s}_n(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{s}(\boldsymbol{\theta}; \mathbf{x}_i) ,$$

dove

$$\mathbf{s}(\boldsymbol{\theta}; \mathbf{x}_i) = \mathbf{s}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{x}_i; \boldsymbol{\theta})$$

è il vettore delle funzioni punteggio relative all'*i*-esima osservazione campionaria. In questo caso, la matrice di informazione di Fisher può essere espressa come

$$\mathbf{I}_n(\boldsymbol{\theta}) = n \text{Var}[\mathbf{s}(\boldsymbol{\theta}; \boldsymbol{\mathcal{X}})] = -n E \left[\frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{s}(\boldsymbol{\theta}; \boldsymbol{\mathcal{X}}) \right] = n \mathbf{I}(\boldsymbol{\theta}) ,$$

dove $\mathbf{I}(\boldsymbol{\theta}) = \text{Var}[\mathbf{s}(\boldsymbol{\theta}; \boldsymbol{\mathcal{X}})]$ è la matrice di informazione di Fisher relativa ad una sola osservazione campionaria. Il limite inferiore di Rao-Cramér risulta dunque

$$\text{Var}[\tilde{\boldsymbol{\Theta}}] \geq \frac{1}{n} \mathbf{I}(\boldsymbol{\theta})^{-1} ,$$

per cui l'efficienza di uno stimatore $\tilde{\boldsymbol{\Theta}}$ corretto per $\boldsymbol{\theta}$ è data da

$$\text{eff}(\tilde{\boldsymbol{\Theta}}) = \det(\text{Var}(\tilde{\boldsymbol{\Theta}}) \mathbf{I}_n(\boldsymbol{\theta}))^{-1} .$$

Per quanto riguarda la sufficienza nel contesto multivariato, uno stimatore $\tilde{\boldsymbol{\Theta}}$ è detto sufficiente per $\boldsymbol{\theta}$ se per ogni $(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathcal{C}_n$ e $(\mathbf{y}_1, \dots, \mathbf{y}_n) \in \mathcal{C}_n$ si ha

$$\tilde{\Theta}(\mathbf{x}_1, \dots, \mathbf{x}_n) = \tilde{\Theta}(\mathbf{y}_1, \dots, \mathbf{y}_n) \Rightarrow L(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) \propto L(\boldsymbol{\theta}; \mathbf{y}_1, \dots, \mathbf{y}_n)$$

per ogni $\boldsymbol{\theta} \in \Theta$. In questo caso, il criterio di fattorizzazione di Neyman richiede che

$$f_n(\mathbf{x}_1, \dots, \mathbf{x}_n; \boldsymbol{\theta}) = h_n(\mathbf{x}_1, \dots, \mathbf{x}_n)g(\tilde{\boldsymbol{\theta}}; \boldsymbol{\theta}).$$

• **Esempio 6.2.1.** Dato un campione casuale da $\mathcal{X} \sim N_d(\boldsymbol{\mu}, \mathbf{I}_d)$, ovvero data una versione ridotta del modello statistico dell'Esempio 6.1.1, allora risulta

$$\mathcal{F}_\mu = \left\{ f_n : f_n(\mathbf{x}_1, \dots, \mathbf{x}_n; \boldsymbol{\mu}) = \prod_{i=1}^n (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^\top (\mathbf{x}_i - \boldsymbol{\mu})} \right\},$$

dove $\boldsymbol{\mu} \in \mathbb{R}^d$. In questo caso la verosimiglianza è data da

$$L(\boldsymbol{\mu}) = c e^{-\frac{n}{2}(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top (\bar{\mathbf{x}} - \boldsymbol{\mu})},$$

per cui il vettore delle funzioni punteggio risulta

$$\mathbf{s}_n(\boldsymbol{\mu}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{\partial}{\partial \boldsymbol{\mu}} \left(\log c - \frac{n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})^\top (\bar{\mathbf{x}} - \boldsymbol{\mu}) \right) = n(\bar{\mathbf{x}} - \boldsymbol{\mu}),$$

mentre la matrice d'informazione di Fisher è data da

$$\mathbf{I}_n(\boldsymbol{\mu}) = -\mathbb{E} \left[\frac{\partial}{\partial \boldsymbol{\mu}^\top} (n(\bar{\mathcal{X}} - \boldsymbol{\mu})) \right] = n\mathbf{I}_d.$$

Di conseguenza, il limite inferiore di Rao-Cramér risulta

$$\text{Var}[\tilde{\boldsymbol{\mu}}] \geq \frac{1}{n} \mathbf{I}_d,$$

per cui, essendo $\mathbb{E}[\bar{\mathcal{X}}] = \boldsymbol{\mu}$ e $\text{Var}[\bar{\mathcal{X}}] = n^{-1}\mathbf{I}_d$ in base ai risultati della §6.1.2, $\bar{\mathcal{X}}$ è uno stimatore efficiente per $\boldsymbol{\mu}$. Inoltre è immediato verificare mediante il criterio di fattorizzazione che $\bar{\mathcal{X}}$ è anche sufficiente. \square

6.2.2. Gli stimatori di massima verosimiglianza con matrici campionarie. Se si considera un modello statistico \mathcal{F}_θ relativo ad un campionamento casuale, data la matrice campionaria $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ si dice stima di massima verosimiglianza di $\boldsymbol{\theta}$ quel valore $\hat{\boldsymbol{\theta}} \in \Theta$ tale che

$$L(\hat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}),$$

che costituisce ovviamente una condizione equivalente a

$$l(\hat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta} \in \Theta} l(\boldsymbol{\theta}).$$

La stima di massima verosimiglianza è la realizzazione campionaria $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\Theta}}(\mathbf{x}_1, \dots, \mathbf{x}_n)$ dello stimatore di massima verosimiglianza $\hat{\boldsymbol{\Theta}}$. Inoltre, se $\boldsymbol{\theta}$ può essere ripartito in due componenti $\boldsymbol{\theta} = (\boldsymbol{\theta}_A^\top, \boldsymbol{\theta}_B^\top)^\top$, tali che $\boldsymbol{\theta}_A \in \Theta_A$ e $\boldsymbol{\theta}_B \in \Theta_B$, la verosimiglianza profilo relativa a $\boldsymbol{\theta}_A$ risulta

$$L_p(\boldsymbol{\theta}_A) = L_p(\boldsymbol{\theta}_A; \mathbf{x}_1, \dots, \mathbf{x}_n) = \max_{\boldsymbol{\theta}_B \in \Theta_B} L(\boldsymbol{\theta}_A, \boldsymbol{\theta}_B).$$

Supponiamo ora che siano soddisfatte le condizioni di regolarità sul modello statistico e che per semplicità lo stima di massima verosimiglianza coincida con l'unica soluzione ottenuta dall'equazione di verosimiglianza

$$\mathbf{s}_n(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \mathbf{0}.$$

In questo caso, sulla base dei risultati si può facilmente dimostrare che lo stimatore di massima verosimiglianza è coerente per $\boldsymbol{\theta}$, ovvero si ha

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0,$$

dove $\boldsymbol{\theta}_0$ è il vero valore del parametro nella popolazione. Inoltre, se l'informazione di Fisher $\mathbf{I}(\boldsymbol{\theta})$ è definita positiva, allora si ha anche

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N_k(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}_0)^{-1}).$$

• **Esempio 6.2.2.** Dato un campione casuale da $\mathcal{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, si vuole determinare la stima di massima verosimiglianza di $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$. Dall'Esempio 6.1.4 si ha

$$\begin{aligned} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}) - \frac{n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\ &\leq \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}) = l(\bar{\mathbf{x}}, \boldsymbol{\Sigma}) = \max_{\boldsymbol{\mu} \in \mathbb{R}^d} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \end{aligned}$$

Nella precedente espressione si è tenuto presente che $\boldsymbol{\Sigma} > \mathbf{0}$, per cui anche $\boldsymbol{\Sigma}^{-1} > \mathbf{0}$, in modo tale che la forma quadratica $(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$ è definita positiva e quindi raggiunge il minimo quando $\boldsymbol{\mu} = \bar{\mathbf{x}}$.

Per $\boldsymbol{\Sigma} > \mathbf{0}$ si ha

$$-\log \det(\boldsymbol{\Sigma}) - \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{D}) \leq -\log \det(\mathbf{D}) - d,$$

dove $\mathbf{D} > \mathbf{0}$ è una matrice quadrata di costanti di ordine d . Al fine di dimostrare la precedente disuguaglianza, si noti che dalle assunzioni fatte si ha anche $\boldsymbol{\Sigma}^{-1} \mathbf{D} > \mathbf{0}$, per cui la matrice $\boldsymbol{\Sigma}^{-1} \mathbf{D}$ ha autovalori $(\alpha_1, \dots, \alpha_d)$ tali che $\alpha_j > 0$. Dunque, si ha

$$\begin{aligned} -\log \det(\boldsymbol{\Sigma}) - \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{D}) &= \log \det(\mathbf{D}) - \log \det(\mathbf{D}) - \log \det(\boldsymbol{\Sigma}) - \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{D}) \\ &= -\log \det(\mathbf{D}) + \log \det(\boldsymbol{\Sigma}^{-1} \mathbf{D}) - \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{D}) \\ &= -\log \det(\mathbf{D}) + \log \prod_{j=1}^d \alpha_j - \sum_{j=1}^d \alpha_j \\ &= -\log \det(\mathbf{D}) + \sum_{j=1}^d (\log \alpha_j - \alpha_j) \leq -\log \det(\mathbf{D}) - d, \end{aligned}$$

dove si è tenuto presente che

$$\log \alpha_j - \alpha_j \leq -1.$$

Adoperando la disuguaglianza con $\mathbf{D} = \mathbf{S}$, si ha

$$\begin{aligned} l(\bar{\mathbf{x}}, \boldsymbol{\Sigma}) &= \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}) \\ &\leq \log c - \frac{n}{2} \log \det(\mathbf{S}) - \frac{nd}{2} = l(\bar{\mathbf{x}}, \mathbf{S}) = \max_{\boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\Sigma} > \mathbf{0}} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \end{aligned}$$

Dunque $(\bar{\mathbf{x}}, \mathbf{S})$ costituisce la stima di massima verosimiglianza di $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, per cui lo stimatore di massima verosimiglianza è dato da $(\bar{\mathcal{X}}, \mathcal{S})$. Tenendo presente i risultati dell'Esempio 6.1.3 si ha inoltre che $\bar{\mathcal{X}} \sim N_d(\boldsymbol{\mu}, n^{-1} \boldsymbol{\Sigma})$ e $n\mathcal{S} \sim W_d(\boldsymbol{\Sigma}, n-1)$ e che $\bar{\mathcal{X}}$ e \mathcal{S} sono indipendenti. Si può verificare che $(\bar{\mathcal{X}}, \mathcal{S})$ è asintoticamente corretto, coerente e sufficiente. \square

• **Esempio 6.2.3.** Dato un campione casuale da $\mathcal{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, si vuole determinare la stima di massima verosimiglianza degli autovalori e degli autovettori di $\boldsymbol{\Sigma}$. Dunque, se $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T$ dove $\boldsymbol{\Gamma}$ è una matrice ortogonale di ordine d e $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_d)$, supponendo senza perdita di generalità che $\lambda_1 \geq \dots \geq \lambda_d$, allora l'obiettivo è quello di ottenere la stima di massima verosimiglianza di $\boldsymbol{\Lambda}$ e $\boldsymbol{\Gamma}$.

Nel caso in cui tutti gli autovalori sono distinti, ovvero se $\lambda_1 \neq \dots \neq \lambda_d$, allora la rappresentazione di $\boldsymbol{\Sigma}$ è unica e gli autovalori e gli autovettori sono funzioni biunivoche di $\boldsymbol{\Sigma}$. Di conseguenza, a meno di matrici campionarie di probabilità nulla, la matrice di varianza-covarianza campionaria ammette una rappresentazione unica $\mathbf{S} = \mathbf{G}\mathbf{L}\mathbf{G}^T$, dove \mathbf{G} è una matrice ortogonale di ordine d , mentre $\mathbf{L} = \text{diag}(l_1, \dots, l_d)$ e (l_1, \dots, l_d) sono gli autovalori campionari di \mathbf{S} , che senza perdita di generalità possono essere scelti come $l_1 \geq \dots \geq l_d$. In base alla proprietà di equivarianza della stima di massima verosimiglianza, allora le stime di massima verosimiglianza di $\boldsymbol{\Lambda}$ e $\boldsymbol{\Gamma}$ risultano rispettivamente \mathbf{L} e \mathbf{G} .

Quando invece solo i primi $e < d$ autovalori sono distinti, ovvero se $\lambda_1 \neq \dots \neq \lambda_e$ e $\lambda_{e+1} = \dots = \lambda_d = \lambda$, allora la rappresentazione di $\boldsymbol{\Sigma}$ non è unica e la proprietà di equivarianza non può essere applicata. In questo caso, tenendo presente l'Esempio 6.1.4, la log-verosimiglianza può essere espressa come

$$l(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) = \log c - \frac{n}{2} \log \det(\boldsymbol{\Lambda}) - \frac{n}{2} \text{tr}(\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}^T\mathbf{G}\mathbf{L}\mathbf{G}^T) - \frac{n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}^T (\bar{\mathbf{x}} - \boldsymbol{\mu})$$

e massimizzando in maniera analoga all'Esempio 6.2.1 si ha

$$l(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) \leq \log c - \frac{n}{2} \log \det(\boldsymbol{\Lambda}) - \frac{n}{2} \text{tr}(\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}^T\mathbf{G}\mathbf{L}\mathbf{G}^T) = l(\bar{\mathbf{x}}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) = \max_{\boldsymbol{\mu} \in \mathbb{R}^d} l(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}).$$

Per le assunzioni si ha $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\Lambda}_1, \lambda\mathbf{I}_{d-e})$ con $\boldsymbol{\Lambda}_1 = \text{diag}(\lambda_1, \dots, \lambda_e)$. Inoltre, sia $\boldsymbol{\Gamma}_1$ la matrice di ordine $(d \times e)$ costituita dai primi e autovettori di $\boldsymbol{\Gamma}$ e sia $\boldsymbol{\Gamma}_2$ la matrice di ordine $(d \times (d - e))$ costituita dai restanti $(d - e)$ autovettori di $\boldsymbol{\Gamma}$. Analogamente, sia $\mathbf{L}_1 = \text{diag}(l_1, \dots, l_e)$, sia \mathbf{G}_1 la matrice di ordine $(d \times e)$ costituita dai primi e autovettori di \mathbf{G} e sia \mathbf{G}_2 la matrice di ordine $(d \times (d - e))$ costituita dai restanti $(d - e)$ autovettori di \mathbf{G} . Di conseguenza, dalla precedente espressione si ha

$$\begin{aligned} l(\bar{\mathbf{x}}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) &= \log c - \frac{n}{2} \log \det(\text{diag}(\boldsymbol{\Lambda}_1, \lambda\mathbf{I}_{d-e})) - \frac{n}{2} \text{tr}(\boldsymbol{\Gamma}\text{diag}(\boldsymbol{\Lambda}_1^{-1}, \lambda^{-1}\mathbf{I}_{d-e})\boldsymbol{\Gamma}^T\mathbf{G}\mathbf{L}\mathbf{G}^T) \\ &= \log c - \frac{n}{2} \log \det(\boldsymbol{\Lambda}_1) - \frac{n(d-e)}{2} \log \lambda \\ &\quad - \frac{n}{2} \text{tr}(\boldsymbol{\Gamma}\text{diag}(\boldsymbol{\Lambda}_1^{-1}, \mathbf{0})\boldsymbol{\Gamma}^T\mathbf{G}\mathbf{L}\mathbf{G}^T) - \frac{n}{2\lambda} \text{tr}(\boldsymbol{\Gamma}\text{diag}(\mathbf{0}, \mathbf{I}_{d-e})\boldsymbol{\Gamma}^T\mathbf{G}\mathbf{L}\mathbf{G}^T) \\ &= \log c - \frac{n}{2} \log \det(\boldsymbol{\Lambda}_1) - \frac{n(d-e)}{2} \log \lambda \\ &\quad - \frac{n}{2} \text{tr}(\boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_1^{-1}\boldsymbol{\Gamma}_1^T\mathbf{G}\mathbf{L}\mathbf{G}^T) - \frac{n}{2\lambda} \text{tr}(\text{diag}(\mathbf{0}, \mathbf{I}_{d-e})\mathbf{G}\mathbf{L}\mathbf{G}^T). \end{aligned}$$

Dal momento che risulta

$$\text{tr}(\text{diag}(\mathbf{0}, \mathbf{I}_{d-e})\mathbf{G}\mathbf{L}\mathbf{G}^T) = \text{tr}(\mathbf{G}^T\text{diag}(\mathbf{0}, \mathbf{I}_{d-e})\mathbf{G}\mathbf{L}) = \text{tr}(\text{diag}(\mathbf{0}, \mathbf{I}_{d-e})\mathbf{L}) = (d - e)\bar{l},$$

dove $\bar{l} = (d - e)^{-1} \sum_{j=e+1}^d l_j$, allora si ha

$$l(\bar{\mathbf{x}}, \boldsymbol{\Lambda}, \boldsymbol{\Gamma}) = \log c - \frac{n}{2} \log \det(\boldsymbol{\Lambda}_1) - \frac{n}{2} \text{tr}(\boldsymbol{\Lambda}_1^{-1}\boldsymbol{\Gamma}_1^T\mathbf{G}\mathbf{L}^T\mathbf{G}\boldsymbol{\Gamma}_1) - \frac{n(d-e)}{2} \log \lambda - \frac{n(d-e)\bar{l}}{2\lambda}.$$

La precedente espressione non dipende da $\boldsymbol{\Gamma}_2$ e quindi qualsiasi matrice $\hat{\boldsymbol{\Gamma}}_2$ di ordine $(d \times (d - e))$, le cui colonne sono ortogonali, costituisce la stima di massima verosimiglianza di $\boldsymbol{\Gamma}_2$ anche se solitamente per convenzione si adotta $\hat{\boldsymbol{\Gamma}}_2 = \mathbf{G}_2$. Inoltre, tenendo presente che gli elementi sulla diagonale di $\boldsymbol{\Lambda}_1^{-1}$ sono in ordine decrescente, quelli sulla diagonale di \mathbf{L} sono in ordine crescente e che $\mathbf{G}^T\boldsymbol{\Gamma}_1$ è una matrice le cui colonne sono ortogonali, allora

$$\text{tr}(\mathbf{\Lambda}_1^{-1} \mathbf{\Gamma}_1^T \mathbf{G} \mathbf{L} \mathbf{G}^T \mathbf{\Gamma}_1) \geq \text{tr}(\mathbf{\Lambda}_1^{-1} \mathbf{L}_1),$$

e l'identità viene ottenuta quando $\mathbf{\Gamma}_1 = \mathbf{G}_1$ (vedi Muirhead, 1982). Dunque,

$$\begin{aligned} l(\bar{\mathbf{x}}, \mathbf{\Lambda}, \mathbf{\Gamma}) &\leq \log c - \frac{n}{2} \log \det(\mathbf{\Lambda}_1) - \frac{n}{2} \text{tr}(\mathbf{\Lambda}_1^{-1} \mathbf{L}_1) - \frac{n(d-e)}{2} \log \lambda - \frac{n(d-e)\bar{l}}{2\lambda} \\ &= l(\bar{\mathbf{x}}, \mathbf{\Lambda}, \widehat{\mathbf{\Gamma}}) = \max_{\mu \in \mathbb{R}^d, \Gamma \in \mathbb{O}} l(\mu, \mathbf{\Lambda}, \mathbf{\Gamma}), \end{aligned}$$

dove $\widehat{\mathbf{\Gamma}} = (\mathbf{G}_1, \widehat{\mathbf{\Gamma}}_2)$, mentre \mathbb{O} rappresenta il gruppo delle matrici ortogonali di ordine d . Applicando opportunamente alla precedente espressione la disuguaglianza ottenuta nell'Esempio 6.2.2 e quella ottenuta nell'Esempio 3.1.2 rispettivamente con $\mathbf{D} = \mathbf{L}_1$ e $d = \bar{l}$, si ha

$$\begin{aligned} l(\bar{\mathbf{x}}, \mathbf{\Lambda}, \widehat{\mathbf{\Gamma}}) &\leq \log c - \frac{n}{2} \log \det(\mathbf{L}_1) - \frac{ne}{2} - \frac{n(d-e)}{2} \log \bar{l} - \frac{n(d-e)}{2} \\ &= \log c - \frac{n}{2} \log \det(\mathbf{L}_1) - \frac{n(d-e)}{2} \log \bar{l} - \frac{nd}{2} \\ &= \log c - \frac{n}{2} \log \det(\text{diag}(\mathbf{L}_1, \bar{l} \mathbf{I}_{d-e})) - \frac{nd}{2} \\ &= l(\bar{\mathbf{x}}, \widehat{\mathbf{\Lambda}}, \widehat{\mathbf{\Gamma}}) = \max_{\mu \in \mathbb{R}^d, \mathbf{\Lambda} \in \mathbb{D}_1, \Gamma \in \mathbb{O}} l(\mu, \mathbf{\Lambda}, \mathbf{\Gamma}), \end{aligned}$$

dove $\widehat{\mathbf{\Lambda}} = \text{diag}(\mathbf{L}_1, \bar{l} \mathbf{I}_{d-e})$, mentre \mathbb{D}_1 rappresenta il gruppo delle matrici diagonali di ordine d i cui ultimi $(d-e)$ elementi sulla diagonale sono uguali. Dunque, si deve concludere che le stime di massima verosimiglianza dei primi e autovalori di $\mathbf{\Sigma}$ sono date dai primi e autovalori di \mathbf{S} , mentre la stima dei restanti $(d-e)$ autovalori coincidenti di $\mathbf{\Sigma}$ è data dalla media aritmetica \bar{l} dei corrispondenti autovalori di \mathbf{S} . Inoltre, le stime di massima verosimiglianza dei primi e autovettori di $\mathbf{\Sigma}$ sono date dai primi e autovettori di \mathbf{S} , mentre le stime di massima verosimiglianza dei restanti $(d-e)$ autovettori non sono uniche, anche se solitamente per convenzione si adottano gli ultimi $(d-e)$ autovettori di \mathbf{S} .

Infine, si noti che risulta molto complicato determinare la distribuzione degli stimatori di massima verosimiglianza degli autovalori e degli autovettori (vedi Muirhead, 1982). \square

6.3. La verifica di ipotesi con osservazioni multivariate

6.3.1. Il test del rapporto delle verosimiglianze con matrici campionarie. Si consideri un modello statistico \mathcal{F}_θ relativo ad un campionamento casuale e il sistema di ipotesi $H_0 : \theta \in \Theta_0$ contro $H_1 : \theta \in \Theta_1$. Il test del rapporto delle verosimiglianze è basato sulla statistica test R la cui realizzazione campionaria è data da

$$r = \frac{\max_{\theta \in \Theta_0} L(\theta; \mathbf{x}_1, \dots, \mathbf{x}_n)}{\max_{\theta \in \Theta} L(\theta; \mathbf{x}_1, \dots, \mathbf{x}_n)}.$$

La regione critica del test è data da $\mathcal{T}_1 = \{r : r \leq r_\alpha\}$, dove

$$\sup_{\theta \in \Theta_0} \Pr_\theta(R \leq r_\alpha) = \alpha,$$

mentre, avendo osservato il valore campionario r , il livello di significatività osservato del test è dato da

$$\alpha_{oss} = \sup_{\theta \in \Theta_0} \Pr_\theta(R \leq r).$$

Se inoltre si considera il sistema di ipotesi $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ contro $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, supponendo verificate le condizioni della §4.2.3 per ottenere la normalità per grandi campioni degli stimatori di massima verosimiglianza, allora se H_0 è vera risulta

$$-2 \log R_n \xrightarrow{d} \chi_k^2.$$

Di conseguenza, la regione critica per grandi campioni del test basato sul rapporto delle verosimiglianze è data da

$$\mathcal{T}_{n,1} = \{r : -2 \log r \geq \chi_{k,1-\alpha}^2\},$$

mentre, per un dato valore campionario r , il livello di significatività osservato per grandi campioni risulta

$$\alpha_{n,oss} = \Pr(\chi_k^2 \geq -2 \log r).$$

Se si suppone infine che $\boldsymbol{\theta} = (\boldsymbol{\theta}_A^T, \boldsymbol{\theta}_B^T)^T$, dove $\boldsymbol{\theta}_A$ è un vettore a $(k - h)$ componenti e $\boldsymbol{\theta}_B$ è un vettore a h componenti e si considera il sistema di ipotesi $H_0 : \boldsymbol{\theta}_A = \boldsymbol{\theta}_{0A}$ contro $H_1 : \boldsymbol{\theta}_A \neq \boldsymbol{\theta}_{0A}$, allora se H_0 è vera risulta

$$-2 \log R_n \xrightarrow{d} \chi_{k-h}^2.$$

Dunque, la regione critica per grandi campioni del test basato sul rapporto delle verosimiglianze è data da

$$\mathcal{T}_{n,1} = \{r : -2 \log r \geq \chi_{k-h,1-\alpha}^2\},$$

mentre per un dato valore campionario r , il livello di significatività osservato per grandi campioni risulta

$$\alpha_{n,oss} = \Pr(\chi_{k-h}^2 \geq -2 \log r).$$

Nelle prossime sezioni vengono considerate alcune applicazioni del test del rapporto delle verosimiglianze quando si dispone di matrici campionarie provenienti da popolazioni Normali multivariate.

6.3.2. Il test di Hotelling. Dato un campione casuale da $\mathcal{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, si supponga di voler verificare il sistema di ipotesi $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ contro $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. Si ha

$$\Theta = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\Sigma} > \mathbf{0}\}$$

e

$$\Theta_0 = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} = \boldsymbol{\mu}_0, \boldsymbol{\Sigma} > \mathbf{0}\}.$$

Dai risultati dell'Esempio 6.2.2, è immediato ricavare che

$$\max_{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = l(\bar{\mathbf{x}}, \mathbf{S}).$$

Inoltre, si ha

$$\begin{aligned} l(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}) &= \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}) - \frac{n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \\ &= \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T) \\ &= \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} (\mathbf{S} + (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T)), \end{aligned}$$

per cui adoperando nella precedente espressione la disuguaglianza ottenuta nell'Esempio 6.2.2 con $\mathbf{D} = \mathbf{S} + (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T$, si ottiene

$$\begin{aligned} l(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}) &\leq \log c - \frac{n}{2} \log \det(\mathbf{S} + (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T) - \frac{nd}{2} \\ &= l(\boldsymbol{\mu}_0, \mathbf{S} + (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T) = \max_{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta_0} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \end{aligned}$$

Tenendo presente che per le proprietà dei determinanti si ha

$$\det(\mathbf{S} + (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T) = \det(\mathbf{S})(\mathbf{1} + (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0))$$

allora

$$\begin{aligned} l(\boldsymbol{\mu}_0, \mathbf{S} + (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T) &= \log c - \frac{n}{2} \log(\det(\mathbf{S})(\mathbf{1} + (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0))) - \frac{nd}{2} \\ &= \log c - \frac{n}{2} \log \det(\mathbf{S}) - \frac{n}{2} \log \left(1 + \frac{h}{n-1} \right) - \frac{nd}{2}, \end{aligned}$$

dove $h = (n-1)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$. Di conseguenza, la determinazione campionaria di R risulta

$$r = e^{l(\boldsymbol{\mu}_0, \mathbf{S} + (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T) - l(\bar{\mathbf{x}}, \mathbf{S})} = \left(1 + \frac{h}{n-1} \right)^{-\frac{1}{2}n}.$$

Se H_0 è vera, h è la determinazione campionaria di una statistica H distribuita come una $T_{d, n-1}$ di Hotelling dal momento che

$$H = (n-1)(\bar{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mu}_0) = (n-1) \sqrt{n} (\bar{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mu}_0)^T (n\mathbf{S})^{-1} \sqrt{n} (\bar{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mu}_0)$$

e che inoltre $\sqrt{n}(\bar{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mu}_0) \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$ e $n\mathbf{S} \sim W_d(\boldsymbol{\Sigma}, n-1)$ sono indipendenti. Tenendo presente le proprietà della T di Hotelling date nella §A.4.2 si ha inoltre

$$H \sim \frac{(n-1)d}{n-d} F_{d, n-d}.$$

Si noti che r è una funzione monotona decrescente di h e quindi i test costruiti su R e H sono equivalenti. Dal momento che rifiutare H_0 per piccole determinazioni di R è equivalente a rifiutare H_0 per determinazioni elevate di H , la regione critica del test risulta di conseguenza

$$\mathcal{T}_1 = \left\{ h : h \geq \frac{(n-1)d}{n-d} F_{d, n-d, 1-\alpha} \right\}.$$

Se si è osservato il valore campionario h , il livello di significatività osservato risulta infine

$$\alpha_{oss} = \Pr \left(F_{d, n-d} \geq \frac{n-d}{(n-1)d} h \right).$$

Anche se la distribuzione della statistica test è nota per campioni finiti, per grandi campioni quando H_0 è vera la quantità

$$-2 \log r = n \log \left(1 + \frac{h}{n-1} \right)$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_d^2$, in quanto vi sono $(d + d(d+1)/2)$ parametri in totale, di cui $(d(d+1)/2)$ parametri da stimare sotto ipotesi di base.

Questa struttura di verifica di ipotesi può essere impiegata quando si dispone di dati appaiati, in maniera analoga a quanto visto nella §4.3.6. Si supponga di considerare due matrici campionarie $(\mathbf{x}_{11}, \dots, \mathbf{x}_{1n})$ e $(\mathbf{x}_{21}, \dots, \mathbf{x}_{2n})$, relative ad un gruppo di n soggetti su cui sono state osservate d

variabili casuali prima e dopo un trattamento. Analogamente alla §4.3.6, l'obiettivo è quello di valutare l'efficacia del trattamento. Dunque, si costruisce la nuova matrice campionaria $(\mathbf{d}_1, \dots, \mathbf{d}_n)$, dove $\mathbf{d}_i = \mathbf{x}_{2i} - \mathbf{x}_{1i}$, e si suppone che provenga da $\mathcal{D} \sim N(\boldsymbol{\mu}, \Sigma)$. La verifica dell'efficacia del trattamento si riduce a considerare il sistema di ipotesi $H_0 : \boldsymbol{\mu} = \mathbf{0}$ contro $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$ e di conseguenza possono essere applicate le procedure di verifica di ipotesi discusse in questa sezione.

• **Esempio 6.3.1.** Per confrontare due tipi di vernice anticorrosione, sono stati verniciati 15 pezzi di tubo di uguale lunghezza con la prima vernice e altri 15 sempre di uguale lunghezza con la seconda vernice. Successivamente, i tubi sono stati sepolti per un certo tempo a coppie di due in 15 differenti locazioni. Su ogni tubo sono state rilevate successivamente due variabili legate alla corrosione, ovvero la massima profondità delle macchie di corrosione e il numero di tali macchie, e le relative osservazioni sono state riportate nella Tavola 6.3.1.

Tavola 6.3.1. Massima profondità (in millesimi di pollice) e numero di macchie di corrosione.

| locazione | prima vernice | | seconda vernice | | differenza | |
|-----------|---------------|--------|-----------------|--------|------------|--------|
| | profondità | numero | profondità | numero | profondità | numero |
| 1 | 73 | 31 | 51 | 35 | 22 | -4 |
| 2 | 43 | 19 | 41 | 14 | 2 | 5 |
| 3 | 47 | 22 | 43 | 19 | 4 | 3 |
| 4 | 53 | 26 | 41 | 29 | 12 | -3 |
| 5 | 58 | 36 | 47 | 34 | 11 | 2 |
| 6 | 47 | 30 | 32 | 26 | 15 | 4 |
| 7 | 52 | 29 | 24 | 19 | 28 | 10 |
| 8 | 38 | 36 | 43 | 37 | -5 | -1 |
| 9 | 61 | 34 | 53 | 24 | 8 | 10 |
| 10 | 56 | 33 | 52 | 27 | 4 | 6 |
| 11 | 56 | 19 | 57 | 14 | -1 | 5 |
| 12 | 34 | 19 | 44 | 19 | -10 | 0 |
| 13 | 55 | 26 | 57 | 30 | -2 | -4 |
| 14 | 65 | 15 | 40 | 7 | 25 | 8 |
| 15 | 75 | 18 | 68 | 13 | 7 | 5 |

Fonte: Kramer e Jensen (1969)

Si vuole determinare se le due vernici sono ugualmente efficaci nel prevenire la corrosione, ovvero si vuole verificare il sistema di ipotesi $H_0 : \boldsymbol{\mu} = \mathbf{0}$ contro $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$. Dal momento che per questi dati si può verificare che

$$\bar{\mathbf{d}} = \begin{pmatrix} 8.0000 \\ 3.0667 \end{pmatrix}$$

e

$$\mathbf{S} = \begin{pmatrix} 113.4667 & 15.9333 \\ 15.9333 & 20.3289 \end{pmatrix},$$

allora si ha $h = 10.8189$. Dunque il livello di significatività osservato risulta

$$\alpha_{oss} = \Pr(F_{2,13} \geq 5.0231) = 0.0242.$$

Dato che la significatività osservata è piuttosto bassa, si deve concludere che le due vernici hanno diverse prestazioni, dal momento che si può respingere H_0 ad ogni livello di significatività $\alpha > 0.0242$. Si ha inoltre $-2 \log r = 8.5882$ e dunque

$$\alpha_{n,oss} = \Pr(\chi_2^2 \geq 8.5882) = 0.0136,$$

un risultato che conferma le conclusioni precedenti. □

6.3.3. Il test per la verifica di una matrice di varianza-covarianza diagonale a blocchi. Dato un campione casuale da $\mathcal{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, si supponga di voler verificare il sistema di ipotesi $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 = \text{diag}(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})$ contro $H_1 : \boldsymbol{\Sigma} > \mathbf{0}$, dove $\boldsymbol{\Sigma}_{11}$ è la matrice di varianza-covarianza relativa alle prime $e < d$ componenti (X_1, \dots, X_e) del vettore casuale \mathcal{X} , mentre $\boldsymbol{\Sigma}_{22}$ è la matrice di varianza-covarianza relativa alle restanti $(d - e)$ componenti $(X_e, X_{e+1}, \dots, X_d)$ del vettore casuale \mathcal{X} . Questa ipotesi di base implica l'indipendenza dei due gruppi di componenti. Si ha

$$\Theta = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\Sigma} > \mathbf{0}\}$$

e

$$\Theta_0 = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 = \text{diag}(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})\}.$$

Dai risultati dell'Esempio 6.2.2, è immediato ricavare che

$$\max_{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = l(\bar{\mathbf{x}}, \mathbf{S}).$$

Inoltre, si ha

$$\begin{aligned} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0) &= \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}_0) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}_0^{-1} \mathbf{S}) - \frac{n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_0^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\ &\leq \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}_0) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}_0^{-1} \mathbf{S}) = l(\bar{\mathbf{x}}, \boldsymbol{\Sigma}_0) = \max_{\boldsymbol{\mu} \in \mathbb{R}^d} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0), \end{aligned}$$

dal momento che $(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_0^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \geq 0$ (vedi Esempio 6.2.2). Tenendo presente che

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix},$$

dove \mathbf{S}_{11} è la matrice di varianza-covarianza campionaria relativa alle prime e componenti del vettore casuale \mathcal{X} , \mathbf{S}_{22} è la matrice di varianza-covarianza campionaria relativa alle restanti $(d - e)$ componenti, mentre $\mathbf{S}_{12} = \mathbf{S}_{21}^T$ è la matrice delle covarianze fra le prime e componenti e le restanti $(d - e)$ componenti, allora

$$\begin{aligned} l(\bar{\mathbf{x}}, \boldsymbol{\Sigma}_0) &= \log c - \frac{n}{2} \log \det(\text{diag}(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})) - \frac{n}{2} \text{tr}(\text{diag}(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})^{-1} \mathbf{S}) \\ &= \log c - \frac{n}{2} \log(\det(\boldsymbol{\Sigma}_{11}) \det(\boldsymbol{\Sigma}_{22})) - \frac{n}{2} \text{tr}(\text{diag}(\boldsymbol{\Sigma}_{11}^{-1} \mathbf{S}_{11}, \boldsymbol{\Sigma}_{22}^{-1} \mathbf{S}_{22})) \\ &= \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}_{11}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}_{11}^{-1} \mathbf{S}_{11}) - \frac{n}{2} \log \det(\boldsymbol{\Sigma}_{22}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}_{22}^{-1} \mathbf{S}_{22}). \end{aligned}$$

Adoperando nella precedente espressione due volte la disuguaglianza ottenuta nell'Esempio 6.2.2, rispettivamente con $\mathbf{D} = \mathbf{S}_{11}$ e $\mathbf{D} = \mathbf{S}_{22}$, si ottiene

$$\begin{aligned} l(\bar{\mathbf{x}}, \boldsymbol{\Sigma}_0) &\leq \log c - \frac{n}{2} \log \det(\mathbf{S}_{11}) - \frac{ne}{2} - \frac{n}{2} \log \det(\mathbf{S}_{22}) - \frac{n(d - e)}{2} \\ &= \log c - \frac{n}{2} \log \det(\text{diag}(\mathbf{S}_{11}, \mathbf{S}_{22})) - \frac{nd}{2} = l(\bar{\mathbf{x}}, \text{diag}(\mathbf{S}_{11}, \mathbf{S}_{22})) = \max_{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta_0} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \end{aligned}$$

Di conseguenza, la determinazione campionaria di R risulta

$$r = e^{l(\bar{\mathbf{x}}, \text{diag}(\mathbf{S}_{11}, \mathbf{S}_{22})) - l(\bar{\mathbf{x}}, \mathbf{S})} = \left(\frac{\det(\mathbf{S})}{\det(\mathbf{S}_{11}) \det(\mathbf{S}_{22})} \right)^{\frac{1}{2}n} = \lambda^{\frac{1}{2}n},$$

dove

$$\lambda = \frac{\det(\mathbf{S})}{\det(\mathbf{S}_{11}) \det(\mathbf{S}_{22})}.$$

Per le proprietà dei determinanti si ha

$$\det(\mathbf{S}) = \det(\mathbf{S}_{11})\det(\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}),$$

da cui

$$\lambda = \frac{\det(\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})}{\det(\mathbf{S}_{22})} = \frac{\det(\mathbf{E})}{\det(\mathbf{E} + \mathbf{H})},$$

dove $\mathbf{E} = n(\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})$ e $\mathbf{H} = n\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$. Dal momento che $n\mathbf{S} \sim W_d(\boldsymbol{\Sigma}, n-1)$, allora per la proprietà *ii*) della §6.3.6, \mathbf{E} e \mathbf{H} sono le determinazioni campionarie di due matrici casuali $\mathcal{E} \sim W_{d-e}(\boldsymbol{\Sigma}_{22}, n-1-e)$ e $\mathcal{H} \sim W_{d-e}(\boldsymbol{\Sigma}_{22}, e)$ indipendenti. Dunque, se H_0 è vera, λ è la determinazione campionaria della statistica

$$\Lambda = \frac{\det(\mathcal{E})}{\det(\mathcal{E} + \mathcal{H})},$$

distribuita come una $\Lambda_{d-e, n-1-e, e}$ di Wilks. Inoltre, r è una funzione monotona crescente di λ e quindi i test costruiti su R e Λ sono equivalenti. Dal momento che rifiutare H_0 per piccole determinazioni di R è equivalente a rifiutare per piccole determinazioni di Λ , la regione critica del test è data da

$$\mathcal{T}_1 = \{\lambda : \lambda \leq \Lambda_{d-e, n-1-e, e, \alpha}\}.$$

Se si è osservato il valore campionario λ , il livello di significatività osservato risulta infine

$$\alpha_{oss} = \Pr(\Lambda_{d-e, n-1-e, e} \leq \lambda).$$

Anche se la distribuzione della statistica test è nota per campioni finiti, per grandi campioni, quando H_0 è vera, la quantità

$$-2 \log r = n \log \left(\frac{\det(\mathbf{S}_{11})\det(\mathbf{S}_{22})}{\det(\mathbf{S})} \right),$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_{e(d-e)}^2$, in quanto vi sono $(d + d(d+1)/2)$ parametri in totale, di cui $(d + e(e+1)/2 + (d-e)(d-e+1)/2)$ parametri da stimare sotto ipotesi di base. Di conseguenza, la regione critica per grandi campioni è data da

$$\mathcal{T}_{n,1} = \{r : -2 \log r \geq \chi_{e(d-e), 1-\alpha}^2\},$$

mentre la significatività osservata per grandi campioni risulta

$$\alpha_{n,oss} = \Pr(\chi_{e(d-e)}^2 \geq -2 \log r).$$

• **Esempio 6.3.2.** Sono state considerate 25 famiglie in cui erano presenti coppie di fratelli. Su ogni fratello di ogni coppia sono state rilevate la lunghezza e il diametro della testa, ottenendo le osservazioni della Tavola 6.3.1.

Si vuole verificare se le misurazioni relative al primo fratello sono indipendenti dalle quelle relative al secondo fratello, ovvero si vuole verificare il sistema di ipotesi $H_0 : \boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})$ contro $H_1 : \boldsymbol{\Sigma} > \mathbf{0}$, dove $\boldsymbol{\Sigma}_{11}$ è la matrice di varianza-covarianza relativa alle componenti (X_1, X_2) , mentre $\boldsymbol{\Sigma}_{22}$ è la matrice di varianza-covarianza relativa alle restanti componenti (X_3, X_4) . Risulta

$$\bar{\mathbf{x}} = \begin{pmatrix} 185.72 \\ 151.12 \\ 183.84 \\ 149.24 \end{pmatrix}.$$

e

$$\mathbf{S} = \begin{pmatrix} 91.4816 & 50.7536 & 66.8752 & 44.2672 \\ 50.7536 & 52.1856 & 49.2592 & 33.6512 \\ 66.8752 & 49.2592 & 96.7744 & 54.2784 \\ 44.2672 & 33.6512 & 54.2784 & 43.2244 \end{pmatrix}.$$

Tavola 6.3.2. Lunghezza e diametro della testa di fratelli (in cm).

| famiglia | primo fratello | | secondo fratello | |
|----------|----------------|----------|------------------|----------|
| | lunghezza | diametro | lunghezza | diametro |
| 1 | 191 | 155 | 179 | 145 |
| 2 | 195 | 149 | 201 | 152 |
| 3 | 181 | 148 | 185 | 149 |
| 4 | 183 | 153 | 188 | 149 |
| 5 | 176 | 144 | 171 | 142 |
| 6 | 208 | 157 | 192 | 152 |
| 7 | 189 | 150 | 190 | 149 |
| 8 | 197 | 159 | 189 | 152 |
| 9 | 188 | 152 | 197 | 159 |
| 10 | 192 | 150 | 187 | 151 |
| 11 | 179 | 158 | 186 | 148 |
| 12 | 183 | 147 | 174 | 147 |
| 13 | 174 | 150 | 185 | 152 |
| 14 | 190 | 159 | 195 | 157 |
| 15 | 188 | 151 | 187 | 158 |
| 16 | 163 | 137 | 161 | 130 |
| 17 | 195 | 155 | 183 | 158 |
| 18 | 186 | 153 | 173 | 148 |
| 19 | 181 | 145 | 182 | 146 |
| 20 | 175 | 140 | 165 | 137 |
| 21 | 192 | 154 | 185 | 152 |
| 22 | 174 | 143 | 178 | 147 |
| 23 | 176 | 139 | 176 | 143 |
| 24 | 197 | 167 | 200 | 158 |
| 25 | 190 | 163 | 187 | 150 |

Fonte: Frets (1921)

Inoltre si ha

$$\mathbf{S}_{11} = \begin{pmatrix} 91.4816 & 50.7336 \\ 50.7536 & 52.1858 \end{pmatrix}$$

e

$$\mathbf{S}_{22} = \begin{pmatrix} 96.7744 & 54.2784 \\ 54.2784 & 43.2244 \end{pmatrix}.$$

Dal momento che $\det(\mathbf{S}) = 102525$, $\det(\mathbf{S}_{11}) = 1236.68$ e $\det(\mathbf{S}_{22}) = 2198.09$, allora risulta $\lambda = 0.3772$. Tenendo presenti i risultati della §A.4.2, il livello di significatività osservato è dato da

$$\alpha_{oss} = \Pr(\Lambda_{2,22,2} \leq 0.3772) = \Pr(F_{4,42} \geq 6.5964) \simeq 0.$$

Dal momento che la significatività osservata estremamente bassa, si deve concludere che i due gruppi di variabili non sono indipendenti. Inoltre si ha $-2 \log r = 6.5964$ e dunque

$$\alpha_{n,oss} = \Pr(\chi_4^2 \geq 24.377) \simeq 0,$$

un risultato che conferma le conclusioni precedenti. □

6.3.4. Il test per la verifica di una matrice di varianza-covarianza diagonale. Dato un campione casuale da $\mathcal{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, si supponga di voler verificare il sistema di ipotesi $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 = \text{diag}(\sigma_{11}, \dots, \sigma_{dd})$ contro $H_1 : \boldsymbol{\Sigma} > \mathbf{0}$. Questa ipotesi di base implica l'indipendenza delle componenti del vettore \mathcal{X} . Si ha

$$\Theta = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\Sigma} > \mathbf{0}\}$$

e

$$\Theta_0 = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 = \text{diag}(\sigma_{11}, \dots, \sigma_{dd})\}.$$

Dai risultati dell'Esempio 6.2.2, si ricava che

$$\max_{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = l(\bar{\mathbf{x}}, \mathbf{S}).$$

In completa analogia con la §6.3.3 si ha

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0) \leq l(\bar{\mathbf{x}}, \boldsymbol{\Sigma}_0) = \max_{\boldsymbol{\mu} \in \mathbb{R}^d} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0).$$

Inoltre, tenendo presente la disuguaglianza ottenuta nell'Esempio 3.1.2 si ha

$$\begin{aligned} l(\bar{\mathbf{x}}, \boldsymbol{\Sigma}_0) &= \log c - \frac{n}{2} \log \det(\text{diag}(\sigma_{11}, \dots, \sigma_{dd})) - \frac{n}{2} \text{tr}(\text{diag}(\sigma_{11}, \dots, \sigma_{dd})^{-1} \mathbf{S}) \\ &= \log c - \frac{n}{2} \log \prod_{j=1}^d \sigma_{jj} - \frac{n}{2} \sum_{j=1}^d \frac{s_{jj}}{\sigma_{jj}} = \log c + \frac{n}{2} \sum_{j=1}^d \left(-\log \sigma_{jj} - \frac{s_{jj}}{\sigma_{jj}} \right) \\ &\leq \log c + \frac{n}{2} \sum_{j=1}^d (-\log s_{jj} - 1) = \log c - \frac{n}{2} \log \det(\text{diag}(\mathbf{S})) - \frac{nd}{2} \\ &= l(\bar{\mathbf{x}}, \text{diag}(\mathbf{S})) = \max_{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta_0} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \end{aligned}$$

per cui la determinazione campionaria di R risulta

$$r = e^{l(\bar{\mathbf{x}}, \text{diag}(\mathbf{S})) - l(\bar{\mathbf{x}}, \mathbf{S})} = \left(\frac{\det(\text{diag}(\mathbf{S}))}{\det(\mathbf{S})} \right)^{-\frac{1}{2}n}.$$

In questo caso è complesso ottenere la distribuzione esatta di R o di una sua trasformata. Tuttavia, per grandi campioni, quando H_0 è vera, la quantità

$$-2 \log r = n \log \left(\frac{\det(\text{diag}(\mathbf{S}))}{\det(\mathbf{S})} \right)$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_{d(d-1)/2}^2$, in quanto vi sono in totale $(d + d(d + 1)/2)$ parametri, di cui $(2d)$ parametri da stimare sotto ipotesi di base. Di conseguenza, la regione critica per grandi campioni è data da

$$\mathcal{T}_{n,1} = \{r : -2 \log r \geq \chi_{d(d-1)/2, 1-\alpha}^2\},$$

mentre la significatività osservata per grandi campioni risulta

$$\alpha_{n,oss} = \Pr(\chi_{d(d-1)/2}^2 \geq -2 \log r).$$

6.3.5. Il test per la verifica della sfericità. Dato un campione casuale da $\mathcal{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, si supponga di voler verificare il sistema di ipotesi $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 = \sigma \mathbf{I}_d$ contro $H_1 : \boldsymbol{\Sigma} > \mathbf{0}$. Questa ipotesi di base

implica l'indipendenza e l'omoschedasticità delle componenti del vettore \mathcal{X} e per questo motivo è detta ipotesi di sfericità. Si ha

$$\Theta = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\Sigma} > \mathbf{0}\}$$

e

$$\Theta_0 = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \boldsymbol{\mu} \in \mathbb{R}^d, \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 = \sigma \mathbf{I}_d\}.$$

Dai risultati dell'Esempio 6.2.2, si ricava che

$$\max_{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = l(\bar{\mathbf{x}}, \mathbf{S}).$$

In completa analogia con la §6.3.3 risulta

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0) \leq l(\bar{\mathbf{x}}, \boldsymbol{\Sigma}_0) = \max_{\boldsymbol{\mu} \in \mathbb{R}^d} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0).$$

Inoltre, tenendo presente la disuguaglianza ottenuta nell'Esempio 3.1.2 si ha

$$\begin{aligned} l(\bar{\mathbf{x}}, \boldsymbol{\Sigma}_0) &= \log c - \frac{nd}{2} \log \sigma - \frac{n}{2\sigma} \text{tr}(\mathbf{S}) \leq \log c - \frac{nd}{2} \log \hat{\sigma} - \frac{nd}{2} \\ &= \log c - \frac{n}{2} \log \det(\hat{\sigma} \mathbf{I}_d) - \frac{nd}{2} = l(\bar{\mathbf{x}}, \hat{\sigma} \mathbf{I}_d) = \max_{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta_0} l(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \end{aligned}$$

dove $\hat{\sigma} = d^{-1} \text{tr}(\mathbf{S})$. La determinazione campionaria di R risulta dunque

$$r = e^{l(\bar{\mathbf{x}}, \hat{\sigma} \mathbf{I}_d) - l(\bar{\mathbf{x}}, \mathbf{S})} = \left(\frac{\hat{\sigma}^d}{\det(\mathbf{S})} \right)^{-\frac{1}{2}n}.$$

In questo caso è complesso ottenere la distribuzione esatta di R o di una sua trasformata. Tuttavia, per grandi campioni, quando H_0 è vera, la quantità

$$-2 \log r = n \log \left(\frac{\hat{\sigma}^d}{\det(\mathbf{S})} \right),$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_{(d+2)(d-1)/2}^2$, in quanto vi sono in totale $(d + d(d+1)/2)$ parametri, di cui $(d+1)$ parametri da stimare sotto ipotesi di base. Di conseguenza, la regione critica per grandi campioni è data da

$$\mathcal{T}_{n,1} = \{r : -2 \log r \geq \chi_{(d+2)(d-1)/2, 1-\alpha}^2\},$$

mentre la significatività osservata per grandi campioni risulta

$$\alpha_{n,oss} = \Pr(\chi_{(d+2)(d-1)/2}^2 \geq -2 \log r).$$

6.3.6. Il test per la verifica dell'omogeneità degli ultimi $(d-e)$ autovalori. Dato un campione casuale da $\mathcal{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, supponiamo che $\mathcal{X} = \mathcal{Y} + \mathcal{E}$, dove $\mathcal{Y} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Phi})$ e $\mathcal{E} \sim N_d(\mathbf{0}, \sigma^2 \mathbf{I}_d)$ sono vettori casuali indipendenti. Inoltre, supponiamo che $r(\boldsymbol{\Phi}) = e < d$, ovvero il vettore casuale \mathcal{Y} è degenere. Questa situazione campionaria si può verificare in pratica quando, pur essendo il vettore casuale \mathcal{Y} degenere, a causa di errori stocastici di misura \mathcal{E} , il vettore casuale risultante \mathcal{X} non appare degenere. In questo caso, pur essendo solo e le variabili che descrivono il fenomeno, si può essere indotti a considerarne d , rendendo inutilmente complessa l'analisi statistica.

Se Λ e Γ rappresentano rispettivamente la matrice diagonale degli autovalori e la matrice degli autovettori di Σ , allora risulta $\Sigma = \Gamma \Lambda \Gamma^T$. Tenendo presente l'indipendenza di \mathcal{Y} e \mathcal{E} , si ha la seguente ulteriore rappresentazione di Σ

$$\Sigma = \Phi + \sigma^2 \mathbf{I}_d = \Gamma \Delta \Gamma^T + \sigma^2 \Gamma \Gamma^T = \Gamma (\Delta + \sigma^2 \mathbf{I}_d) \Gamma^T,$$

dove Δ rappresenta la matrice diagonale degli autovalori di Φ . Quindi, risulta che $\Lambda = \Delta + \sigma^2 \mathbf{I}_d$. Tuttavia, solo i primi e autovalori di Φ sono non nulli, per cui

$$\Delta = \text{diag}(\Delta_1, \mathbf{0}),$$

dove $\Delta_1 = \text{diag}(\delta_1, \dots, \delta_e)$, ovvero

$$\Lambda = \text{diag}(\Lambda_1, \lambda \mathbf{I}_{d-e}),$$

dove $\Lambda_1 = \text{diag}(\lambda_1, \dots, \lambda_e) = \text{diag}(\delta_1 + \sigma^2, \dots, \delta_e + \sigma^2)$ mentre $\lambda = \sigma^2$.

Al fine di decidere se esistono componenti ridondanti nel vettore casuale \mathcal{X} , si deve dunque verificare il sistema di ipotesi $H_0 : \Sigma = \Gamma \text{diag}(\Lambda_1, \lambda \mathbf{I}_{d-e}) \Gamma^T$ contro $H_1 : \Sigma = \Gamma \Lambda \Gamma^T$. Tenendo presente la notazione dell'Esempio 6.2.3, si ha

$$\Theta = \{(\boldsymbol{\mu}, \Lambda, \Gamma) : \boldsymbol{\mu} \in \mathbb{R}^d, \Lambda \in \mathbb{D}, \Gamma \in \mathbb{O}\}$$

e

$$\Theta_0 = \{(\boldsymbol{\mu}, \Lambda, \Gamma) : \boldsymbol{\mu} \in \mathbb{R}^d, \Lambda \in \mathbb{D}_1, \Gamma \in \mathbb{O}\}.$$

Sulla base dei risultati dell'Esempio 6.2.3 si ha dunque

$$\max_{\boldsymbol{\mu} \in \mathbb{R}^d, \Lambda \in \mathbb{D}, \Gamma \in \mathbb{O}} l(\boldsymbol{\mu}, \Lambda, \Gamma) = l(\bar{\mathbf{x}}, \mathbf{L}, \mathbf{G}) = \log c - \frac{n}{2} \log \det(\mathbf{L}) - \frac{nd}{2},$$

mentre

$$\max_{\boldsymbol{\mu} \in \mathbb{R}^d, \Lambda \in \mathbb{D}_1, \Gamma \in \mathbb{O}} l(\boldsymbol{\mu}, \Lambda, \Gamma) = l(\bar{\mathbf{x}}, \hat{\Lambda}, \hat{\Gamma}).$$

La determinazione campionaria di R risulta

$$r = e^{l(\bar{\mathbf{x}}, \hat{\Lambda}, \hat{\Gamma}) - l(\bar{\mathbf{x}}, \mathbf{L}, \mathbf{G})} = \left(\frac{\det(\text{diag}(\mathbf{L}_1, \bar{l} \mathbf{I}_{d-e}))}{\det(\mathbf{L})} \right)^{-\frac{1}{2}n} = \left(\frac{\bar{l}}{\tilde{l}} \right)^{-\frac{n}{2}(d-e)}$$

dove $\tilde{l} = (\prod_{j=1}^{d-e} l_j)^{1/(d-e)}$ rappresenta la media geometrica degli ultimi $(d-e)$ autovalori campionari. In questo caso è complesso ottenere la distribuzione esatta di R o di una sua trasformata. Tuttavia, per grandi campioni, quando H_0 è vera, la quantità

$$-2 \log r = n(d-e) \log \left(\frac{\bar{l}}{\tilde{l}} \right),$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_{(d-e+2)(d-e-1)/2}^2$. Infatti, sotto ipotesi di base Σ è determinata da e autovalori distinti, un autovalore coincidente ed e autovettori a d componenti corrispondenti agli autovalori distinti, i quali sono inoltre vincolati da $(e(e+1)/2)$ condizioni di ortonormalità. Dunque, considerando anche il vettore delle medie, sotto ipotesi di base vi sono $(d+e+1+de-e(e+1)/2)$ parametri in totale. Inoltre, sotto ipotesi alternativa Σ è determinata da d autovalori e d autovettori, i quali sono inoltre vincolati da $(d(d+1)/2)$ condizioni di ortonormalità. Di conseguenza, vi sono $(d+d+d^2-d(d+1)/2) = (d+d(d+1)/2)$ parametri in totale. La regione critica per grandi campioni è dunque data da

$$\mathcal{T}_{n,1} = \{r : -2 \log r \geq \chi_{(d-e+2)(d-e-1)/2, 1-\alpha}^2\},$$

mentre la significatività osservata per grandi campioni risulta

$$\alpha_{n,oss} = \Pr(\chi_{(d-e+2)(d-e-1)/2}^2 \geq -2 \log r).$$

6.3.6. Il test di Hotelling a due campioni. Si supponga di considerare due campioni casuali indipendenti $(\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1})$ e $(\mathbf{x}_{21}, \dots, \mathbf{x}_{2n_2})$ rispettivamente dai vettori casuali $\mathcal{X}_1 \sim N_d(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ e $\mathcal{X}_2 \sim N_d(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, ovvero si suppone l'omoschedasticità dei due vettori casuali. Sia inoltre $n = \sum_{l=1}^2 n_l$ la numerosità campionaria globale. In questo caso, il modello statistico è dato da

$$\begin{aligned} \mathcal{F}_{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}} &= \left\{ f_n : f_n(\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1}, \mathbf{x}_{21}, \dots, \mathbf{x}_{2n_2}; \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) = \right. \\ &= \left. \prod_{l=1}^2 \prod_{i=1}^{n_l} \det(2\pi \boldsymbol{\Sigma})^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}_{li} - \boldsymbol{\mu}_l)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_{li} - \boldsymbol{\mu}_l)} \right\}, \end{aligned}$$

dove $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^d$ e $\boldsymbol{\Sigma} > \mathbf{0}$. Si vuole verificare il sistema di ipotesi $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}$ contro $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$. Si ha

$$\Theta = \{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) : \boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^d, \boldsymbol{\Sigma} > \mathbf{0}\}$$

e

$$\Theta_0 = \{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2, \boldsymbol{\Sigma} > \mathbf{0}\}.$$

Se $\bar{\mathbf{x}}_l$ e \mathbf{S}_l rappresentano rispettivamente le realizzazioni campionarie del vettore medio campionario $\bar{\mathcal{X}}_l$ e della matrice di varianza-covarianza campionaria \mathcal{S}_l per l' l -esimo campione, allora la funzione di log-verosimiglianza è data da

$$\begin{aligned} l(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) &= \log c + \sum_{l=1}^2 \left(-\frac{n_l}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n_l}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}_l) - \frac{n_l}{2} (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l)^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l) \right) \\ &= \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{W}) - \frac{1}{2} \sum_{l=1}^2 n_l (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l)^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l), \end{aligned}$$

dove

$$\mathbf{W} = \frac{1}{n} \sum_{l=1}^2 n_l \mathbf{S}_l.$$

In maniera analoga all'Esempio 6.2.2, adoperando la disuguaglianza del medesimo esempio con $\mathbf{D} = \mathbf{W}$, si ha

$$\begin{aligned} l(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) &\leq \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{W}) = l(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \boldsymbol{\Sigma}) = \max_{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^d} l(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) \\ &\leq \log c - \frac{n}{2} \log \det(\mathbf{W}) - \frac{nd}{2} = l(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{W}) = \max_{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) \in \Theta} l(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}). \end{aligned}$$

Inoltre, tenendo presente che se H_0 è vera si dispone in effetti di un unico campione da $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, si ottiene

$$l(\boldsymbol{\mu}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \leq l(\bar{\mathbf{x}}, \bar{\mathbf{x}}, \mathbf{S}) = \max_{(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) \in \Theta_0} l(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}),$$

dove

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{l=1}^2 n_l \bar{\mathbf{x}}_l$$

e

$$\mathbf{S} = \frac{1}{n} \sum_{l=1}^2 \sum_{i=1}^{n_l} (\mathbf{x}_{li} - \bar{\mathbf{x}})(\mathbf{x}_{li} - \bar{\mathbf{x}})^T.$$

Si ha

$$\begin{aligned} \mathbf{S} &= \frac{1}{n} \sum_{l=1}^2 \sum_{i=1}^{n_l} ((\mathbf{x}_{li} - \bar{\mathbf{x}}_l) + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}))((\mathbf{x}_{li} - \bar{\mathbf{x}}_l) + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}))^T \\ &= \frac{1}{n} \sum_{l=1}^2 \sum_{i=1}^{n_l} (\mathbf{x}_{li} - \bar{\mathbf{x}}_l)(\mathbf{x}_{li} - \bar{\mathbf{x}}_l)^T + \frac{1}{n} \sum_{l=1}^2 \sum_{i=1}^{n_l} (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^T \\ &= \frac{1}{n} \sum_{l=1}^2 n_l \mathbf{S}_l + \frac{1}{n} \sum_{l=1}^2 n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^T = \mathbf{W} + \mathbf{B}, \end{aligned}$$

dove si è tenuto presente che $\sum_{i=1}^{n_l} (\mathbf{x}_{li} - \bar{\mathbf{x}}_l) = \mathbf{0}$ e dove si è posto

$$\mathbf{B} = \frac{1}{n} \sum_{l=1}^2 n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^T.$$

La precedente relazione è evidentemente una generalizzazione della relazione ottenuta nella §4.3.4, ovvero costituisce la scomposizione della matrice di varianza-covarianza totale \mathbf{S} nella matrice di varianza-covarianza all'interno dei gruppi \mathbf{W} e nella matrice di varianza-covarianza fra i gruppi \mathbf{B} . È facile verificare che

$$\mathbf{B} = \frac{n_1 n_2}{n^2} (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^T.$$

Dunque,

$$\begin{aligned} l(\bar{\mathbf{x}}, \bar{\mathbf{x}}, \mathbf{S}) &= \log c - \frac{n}{2} \log \det(\mathbf{S}) - \frac{n}{2} \text{tr}(\mathbf{S}^{-1} \mathbf{W}) - \frac{1}{2} \sum_{l=1}^2 n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) \\ &= \log c - \frac{n}{2} \log \det(\mathbf{S}) - \frac{n}{2} \text{tr}(\mathbf{S}^{-1} \mathbf{W}) - \frac{1}{2} \sum_{l=1}^2 n_l \text{tr}(\mathbf{S}^{-1} (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^T) \\ &= \log c - \frac{n}{2} \log \det(\mathbf{S}) - \frac{n}{2} \text{tr}(\mathbf{S}^{-1} \mathbf{W}) - \frac{1}{2} \text{tr} \left(\mathbf{S}^{-1} \sum_{l=1}^2 n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^T \right) \\ &= \log c - \frac{n}{2} \log \det(\mathbf{S}) - \frac{n}{2} \text{tr}(\mathbf{S}^{-1} \mathbf{W}) - \frac{n}{2} \text{tr}(\mathbf{S}^{-1} \mathbf{B}) \\ &= \log c - \frac{n}{2} \log \det(\mathbf{S}) - \frac{n}{2} \text{tr}(\mathbf{S}^{-1} (\mathbf{W} + \mathbf{B})) = \log c - \frac{n}{2} \log \det(\mathbf{S}) - \frac{nd}{2}. \end{aligned}$$

Di conseguenza, la determinazione campionaria di R è data da

$$\begin{aligned} r &= e^{l(\bar{\mathbf{x}}, \bar{\mathbf{x}}, \mathbf{S}) - l(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{W})} = e^{-\frac{n}{2} \log \det(\mathbf{S}) + \frac{n}{2} \log \det(\mathbf{W})} \\ &= \left(\frac{\det(\mathbf{W})}{\det(\mathbf{W} + \mathbf{B})} \right)^{\frac{1}{2}n} = \lambda^{\frac{1}{2}n}, \end{aligned}$$

dove

$$\lambda = \frac{\det(\mathbf{W})}{\det(\mathbf{W} + \mathbf{B})}.$$

Le matrici \mathbf{B} e \mathbf{W} sono le determinazioni campionarie di due matrici casuali \mathbf{B} e \mathbf{W} indipendenti. Questo risultato è una generalizzazione dell'Esempio 6.1.3, in quanto le statistiche \mathbf{B} e \mathbf{W} sono trasformate rispettivamente dei vettori medi campionari e delle matrici di varianza-covarianza campionarie relative a matrici campionarie indipendenti. Inoltre, dal momento che

$$n\mathbf{W} = \sum_{l=1}^2 n_l \mathbf{S}_l,$$

dove le matrici casuali $n_l \mathbf{S}_l \sim W_d(\boldsymbol{\Sigma}, n_l - 1)$ sono indipendenti in quanto i campioni sono indipendenti, per la proprietà *i*) della §A.4.2, si ha $n\mathbf{W} \sim W_d(\boldsymbol{\Sigma}, n - 2)$. Essendo

$$n\mathbf{S} = n\mathbf{W} + n\mathbf{B}$$

e $n\mathbf{S} \sim W_d(\boldsymbol{\Sigma}, n - 1)$, ancora per la proprietà *i*) della §A.4.2 si deve necessariamente concludere che $n\mathbf{B} \sim W_d(\boldsymbol{\Sigma}, 1)$. Dunque, se H_0 è vera, λ è la determinazione campionaria della statistica

$$\Lambda = \frac{\det(\mathbf{W})}{\det(\mathbf{W} + \mathbf{B})},$$

distribuita come una Λ di Wilks con $(n - 2)$ e 1 gradi di libertà. Inoltre r è una funzione monotona crescente di λ e quindi i test costruiti su R e Λ sono equivalenti. Dunque, dal momento che rifiutare H_0 per piccole determinazioni di R è equivalente a rifiutare per piccole determinazioni di Λ , la regione critica del test è data da

$$T_1 = \{\lambda : \lambda \leq \Lambda_{d, n-2, 1, \alpha}\}.$$

Se si è osservato il valore campionario λ , il livello di significatività osservato risulta infine

$$\alpha_{oss} = \Pr(\Lambda_{d, n-2, 1} \leq \lambda).$$

Partendo dall'espressione di r , si può ottenere una formulazione alternativa della statistica test. Infatti, tenendo presente la rappresentazione di \mathbf{B} , si ha

$$\begin{aligned} r &= \left(\frac{\det(\mathbf{W} + \mathbf{B})}{\det(\mathbf{W})} \right)^{-\frac{1}{2}n} = (\det(\mathbf{W}^{-1}) \det(\mathbf{W} + \mathbf{B}))^{-\frac{1}{2}n} = \det(\mathbf{I}_d + \mathbf{W}^{-1}\mathbf{B})^{-\frac{1}{2}n} \\ &= \det\left(\mathbf{I}_d + \frac{n_1 n_2}{n^2} \mathbf{W}^{-1} (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1) (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^T\right)^{-\frac{1}{2}n} \\ &= \left(1 + \frac{n_1 n_2}{n^2} (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^T \mathbf{W}^{-1} (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)\right)^{-\frac{1}{2}n} = \left(1 + \frac{h}{n-2}\right)^{-\frac{1}{2}n}, \end{aligned}$$

dove

$$h = \frac{n_1 n_2 (n-2)}{n^2} (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^T \mathbf{W}^{-1} (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1).$$

La quantità $\sqrt{n_1 n_2 / n} (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)$ è la determinazione campionaria di un vettore casuale

$$\sqrt{\frac{n_1 n_2}{n}} (\bar{\mathbf{X}}_2 - \bar{\mathbf{X}}_1) \sim N_d(\mathbf{0}, \boldsymbol{\Sigma}),$$

per la proprietà *i*) della §A.4.1, dal momento che $\bar{\mathbf{X}}_1$ e $\bar{\mathbf{X}}_2$ sono indipendenti in quanto i campioni sono indipendenti e $\sqrt{n}(\bar{\mathbf{X}}_l - \boldsymbol{\mu}) \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$. Inoltre, il precedente vettore casuale è indipendente da \mathbf{W} in quanto le due statistiche sono trasformate rispettivamente dei vettori medi campionari e delle

matrici di varianza-covarianza campionarie relative a matrici campionarie indipendenti. Dunque, se H_0 è vera, h è la determinazione campionaria di una statistica H distribuita come una $T_{d,n-2}$ di Hotelling dal momento che

$$H = \frac{n_1 n_2 (n-2)}{n^2} (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^T \boldsymbol{\mathcal{W}}^{-1} (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1) \\ = (n-2) \sqrt{\frac{n_1 n_2}{n}} (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^T (n \boldsymbol{\mathcal{W}})^{-1} \sqrt{\frac{n_1 n_2}{n}} (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1).$$

Tenendo presente le proprietà della statistica di Hotelling si ha inoltre

$$H \sim \frac{(n-2)d}{n-d-1} F_{d,n-d-1}.$$

La quantità r è una funzione monotona decrescente di h e quindi i test costruiti su R e H sono equivalenti. Dal momento che rifiutare H_0 per piccole determinazioni di R è equivalente a rifiutare H_0 per determinazioni elevate di H , la regione critica del test risulta di conseguenza

$$\mathcal{T}_1 = \left\{ h : h \geq \frac{(n-2)d}{n-d-1} F_{d,n-d-1,1-\alpha} \right\}.$$

Se si è osservato il valore campionario h , il livello di significatività osservato risulta infine

$$\alpha_{oss} = \Pr \left(F_{d,n-d-1} \geq \frac{n-d-1}{(n-2)d} h \right).$$

Anche se la distribuzione della statistica test è nota per campioni finiti, per grandi campioni, quando H_0 è vera, la quantità

$$-2 \log r = n \log \left(1 + \frac{h}{n-2} \right),$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_d^2$, in quanto vi sono $(2d + d(d+1)/2)$ parametri in totale, di cui $(d + d(d+1)/2)$ parametri da stimare sotto ipotesi di base.

• **Esempio 6.3.3.** Sono stati considerati 20 atleti di valore internazionale di cui 12 fondisti e 8 maratoneti. Su ogni atleta è stato misurato la massa grassa, la massa magra, il VO_2 durante lo sforzo massimale, la concentrazione di acido lattico durante uno sforzo sub-massimale e il VO_2 durante lo sforzo sub-massimale e le osservazioni sono state riportate nella Tavola 6.3.3. Per inciso, si noti che il VO_2 costituisce una misura dell'efficienza aerobica di un atleta.

Tavola 6.3.3. Caratteristiche fisiologiche dei fondisti e dei maratoneti.

| fondisti | massa grassa | massa magra | VO_2 max | acido lattico | VO_2 submax | maratoneti | massa grassa | massa magra | VO_2 max | acido lattico | VO_2 submax |
|----------|--------------|-------------|------------|---------------|---------------|------------|--------------|-------------|------------|---------------|---------------|
| 1 | 8.3 | 65.8 | 5.77 | 30 | 4.71 | 1 | 5.5 | 58.7 | 5.07 | 17 | 4.32 |
| 2 | 3.3 | 57.6 | 4.51 | 33 | 4.15 | 2 | 4.8 | 60.2 | 4.74 | 34 | 4.00 |
| 3 | 2.9 | 59.9 | 4.82 | 46 | 4.20 | 3 | 2.1 | 53.7 | 4.04 | 38 | 3.40 |
| 4 | 1.4 | 56.6 | 4.38 | 37 | 3.99 | 4 | 2.8 | 61.5 | 4.77 | 40 | 4.20 |
| 5 | 7.1 | 59.2 | 5.30 | 26 | 4.29 | 5 | 1.4 | 55.6 | 4.39 | 44 | 3.99 |
| 6 | 3.2 | 58.9 | 4.96 | 24 | 4.07 | 6 | 2.0 | 59.3 | 4.37 | 19 | 3.71 |
| 7 | 4.1 | 65.8 | 5.41 | 19 | 4.57 | 7 | 5.8 | 55.5 | 4.47 | 27 | 3.85 |
| 8 | 1.8 | 67.6 | 5.29 | 32 | 4.44 | 8 | 1.9 | 64.2 | 4.84 | 29 | 4.10 |
| 9 | 4.0 | 49.6 | 4.33 | 39 | 3.53 | | | | | | |
| 10 | 3.8 | 62.5 | 5.60 | 16 | 4.79 | | | | | | |
| 11 | 4.7 | 54.5 | 4.67 | 24 | 3.96 | | | | | | |
| 12 | 0.9 | 60.1 | 5.04 | 31 | 4.12 | | | | | | |

Fonte: Pollock, Jackson e Pate (1980)

Si vuole verificare se i due gruppi di atleti hanno simili caratteristiche fisiologiche, ovvero si vuole verificare il sistema di ipotesi $H_0 : \mu_1 = \mu_2 = \mu$ contro $H_1 : \mu_1 \neq \mu_2$. Si ha

$$\bar{x}_1 = \begin{pmatrix} 3.7917 \\ 59.8417 \\ 5.0067 \\ 29.7500 \\ 4.2350 \end{pmatrix}$$

e

$$\bar{x}_2 = \begin{pmatrix} 3.2875 \\ 58.5875 \\ 4.5863 \\ 31.0000 \\ 3.9463 \end{pmatrix}.$$

Inoltre, risulta

$$S_1 = \begin{pmatrix} 4.3058 & 1.0645 & 0.4476 & -4.6771 & 0.2263 \\ 1.0645 & 24.0491 & 1.8711 & -13.9479 & 1.4712 \\ 0.4476 & 1.8711 & 0.2094 & -2.1642 & 0.1398 \\ -4.6771 & -13.9479 & -2.1642 & 67.0208 & -1.4679 \\ 0.2263 & 1.4712 & 0.1398 & -1.4679 & 0.1144 \end{pmatrix}$$

e

$$S_2 = \begin{pmatrix} 2.7861 & -0.4689 & 0.2333 & -6.7593 & 0.1500 \\ -0.4689 & 10.6811 & 0.7153 & -6.3750 & 0.5498 \\ 0.2333 & 0.7153 & 0.0941 & -1.0600 & 0.0783 \\ -6.9375 & -6.3750 & -1.0600 & 83.5000 & -0.3913 \\ 0.1500 & 0.5498 & 0.0783 & -0.3913 & 0.0745 \end{pmatrix}.$$

Dalle precedenti espressioni risulta

$$W = \begin{pmatrix} 3.6979 & 0.4512 & 0.3619 & -5.5812 & 0.1958 \\ 0.4512 & 18.7019 & 1.4088 & -10.9188 & 1.1027 \\ 0.3619 & 1.4088 & 0.1633 & -1.7225 & 0.1152 \\ -5.5812 & -10.9188 & -1.7225 & 73.6125 & -1.0372 \\ 0.1958 & 1.1027 & 0.1152 & -1.0373 & 0.0984 \end{pmatrix}$$

e

$$B = \begin{pmatrix} 0.0610 & 0.1518 & 0.0509 & -0.1513 & 0.0349 \\ 0.1518 & 0.3775 & 0.1265 & -0.3763 & 0.0869 \\ 0.0509 & 0.1265 & 0.0424 & -0.1261 & 0.0291 \\ -0.1513 & -0.3763 & -0.1261 & 0.3750 & -0.0866 \\ 0.0349 & 0.0869 & 0.0291 & -0.0866 & 0.0200 \end{pmatrix},$$

per cui $\lambda = 0.5499$. Di conseguenza il livello di significatività osservato è dato da

$$\alpha_{oss} = \Pr(\Lambda_{5,18,1} \leq 0.5499) = \Pr(F_{10,28} \geq 2.2921) = 0.0410.$$

Visto il livello di significatività osservata, vi è una certa indicazione a rifiutare l'ipotesi di base, ovvero i due gruppi di atleti sembrano differire sulla base delle caratteristiche fisiologiche. Si ha inoltre $-2 \log r = 11.9515$ e di conseguenza la significatività osservata per grandi campioni risulta

$$\alpha_{n,oss} = \Pr(\chi_5^2 \geq 11.9615) = 0.0353,$$

un risultato che conferma le conclusioni precedenti. \square

6.3.7. L'analisi multivariata della varianza. Si supponga di considerare m campioni casuali indipendenti $(\mathbf{x}_{l1}, \dots, \mathbf{x}_{ln_l})$ rispettivamente dai vettori casuali $\mathcal{X}_l \sim N_d(\boldsymbol{\mu}_l, \boldsymbol{\Sigma})$, ovvero da m vettori casuali normali omoschedastici. Sia inoltre $n = \sum_{l=1}^m n_l$ la numerosità campionaria globale. Il modello statistico è dato da

$$\begin{aligned} \mathcal{F}_{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}} &= \left\{ f_n : f_n(\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1}, \dots, \mathbf{x}_{m1}, \dots, \mathbf{x}_{mn_m}; \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}) = \right. \\ &= \left. \prod_{l=1}^m \prod_{i=1}^{n_l} \det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}_{li} - \boldsymbol{\mu}_l)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{li} - \boldsymbol{\mu}_l)} \right\}, \end{aligned}$$

dove $\boldsymbol{\mu}_l \in \mathbb{R}^d$ e $\boldsymbol{\Sigma} > \mathbf{0}$. Si vuole verificare il sistema di ipotesi $H_0 : \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_m = \boldsymbol{\mu}$ contro $H_1 : \boldsymbol{\mu}_l \neq \boldsymbol{\mu}_j, \exists l \neq j = 1, \dots, m$. Dunque, si ha

$$\Theta = \{(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}) : \boldsymbol{\mu}_l \in \mathbb{R}^d, \boldsymbol{\Sigma} > \mathbf{0}, l = 1, \dots, m\}$$

e

$$\Theta_0 = \{(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}) : \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_m, \boldsymbol{\Sigma} > \mathbf{0}\}.$$

Se $\bar{\mathbf{x}}_l$ e \mathbf{S}_l rappresentano rispettivamente le realizzazioni campionarie del vettore medio campionario $\bar{\mathcal{X}}_l$ e della matrice di varianza-covarianza campionaria \mathcal{S}_l per l' l -esimo campione, estendendo in maniera ovvia i risultati della §6.3.6, la funzione di log-verosimiglianza è data da

$$l(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}) = \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{W}) - \frac{1}{2} \sum_{l=1}^m n_l (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l)^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l),$$

dove

$$\mathbf{W} = \frac{1}{n} \sum_{l=1}^m n_l \mathbf{S}_l.$$

In maniera analoga alla §6.3.6, si ha

$$\begin{aligned} l(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}) &\leq \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{W}) \\ &= l(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_m, \boldsymbol{\Sigma}) = \max_{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m \in \mathbb{R}^d} l(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}) \\ &\leq \log c - \frac{n}{2} \log \det(\mathbf{W}) - \frac{nd}{2} = l(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_m, \mathbf{W}) \\ &= \max_{(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}) \in \Theta} l(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}). \end{aligned}$$

Inoltre, tenendo presente che se H_0 è vera si dispone in effetti di un unico campione da una $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, allora si ottiene

$$l(\boldsymbol{\mu}, \dots, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \leq l(\bar{\mathbf{x}}, \dots, \bar{\mathbf{x}}, \mathbf{S}) = \max_{(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}) \in \Theta_0} l(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}),$$

dove

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{l=1}^m n_l \bar{\mathbf{x}}_l$$

e

$$\mathbf{S} = \frac{1}{n} \sum_{l=1}^m \sum_{i=1}^{n_l} (\mathbf{x}_{li} - \bar{\mathbf{x}})(\mathbf{x}_{li} - \bar{\mathbf{x}})^T.$$

In completa analogia con la §6.3.6 si ha $\mathbf{S} = \mathbf{W} + \mathbf{B}$, dove

$$\mathbf{B} = \frac{1}{n} \sum_{l=1}^m n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^T.$$

Dunque,

$$l(\bar{\mathbf{x}}, \dots, \bar{\mathbf{x}}, \mathbf{S}) = \log c - \frac{n}{2} \log \det(\mathbf{S}) - \frac{nd}{2}.$$

Di conseguenza, la determinazione campionaria di R è data da

$$\begin{aligned} r &= e^{l(\bar{\mathbf{x}}, \dots, \bar{\mathbf{x}}, \mathbf{S}) - l(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_m, \mathbf{W})} = e^{-\frac{n}{2} \log \det(\mathbf{S}) + \frac{n}{2} \log \det(\mathbf{W})} \\ &= \left(\frac{\det(\mathbf{W})}{\det(\mathbf{W} + \mathbf{B})} \right)^{\frac{1}{2}n} = \lambda^{\frac{1}{2}n}, \end{aligned}$$

dove

$$\lambda = \frac{\det(\mathbf{W})}{\det(\mathbf{W} + \mathbf{B})}.$$

In maniera analoga alla §6.3.6, \mathbf{B} e \mathbf{W} sono le determinazioni campionarie di due matrici casuali \mathbf{B} e \mathbf{W} indipendenti. Inoltre, dal momento che

$$n\mathbf{W} = \sum_{l=1}^m n_l \mathbf{S}_l,$$

dove le matrici casuali $n_l \mathbf{S}_l \sim W_d(\boldsymbol{\Sigma}, n_l - 1)$ sono indipendenti in quanto i campioni sono indipendenti, per la proprietà *i*) della §A.4.1, si ha $n\mathbf{W} \sim W_d(\boldsymbol{\Sigma}, n - m)$. Dal momento che risulta

$$n\mathbf{S} = n\mathbf{W} + n\mathbf{B}$$

e che $n\mathbf{S} \sim W_d(\boldsymbol{\Sigma}, n - 1)$, ancora per la proprietà *i*) della §A.4.1 si deve necessariamente concludere che $n\mathbf{B} \sim W_d(\boldsymbol{\Sigma}, m - 1)$. Dunque, se H_0 è vera, λ è la determinazione campionaria della statistica

$$\Lambda = \frac{\det(\mathbf{W})}{\det(\mathbf{W} + \mathbf{B})},$$

distribuita come una Λ di Wilks con $(n - m)$ e $(m - 1)$ gradi di libertà. Inoltre r è una funzione monotona crescente di λ e quindi i test costruiti su R e Λ sono equivalenti. Dal momento che rifiutare H_0 per piccole determinazioni di R è equivalente a rifiutare per piccole determinazioni di Λ , la regione critica del test è data da

$$\mathcal{T}_1 = \{\lambda : \lambda \leq \Lambda_{d, n-m, m-1, \alpha}\}.$$

Se si è osservato il valore campionario λ , il livello di significatività osservato risulta infine

$$\alpha_{oss} = \Pr(\Lambda_{d, n-m, m-1} \leq \lambda).$$

Anche se in questo caso la distribuzione della statistica test è nota per campioni finiti, per grandi campioni, quando H_0 è vera, la quantità

$$-2 \log r = -n \log \lambda,$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi^2_{(m-1)d}$, in quanto è facile verificare che vi sono in totale $(md + d(d + 1)/2)$ parametri, di cui $(d + d(d + 1)/2)$ parametri da stimare sotto ipotesi di base.

• **Esempio 6.3.4.** Tre gruppi di pesci, ognuno dei quali era composto da 12 pesci, sono stati cucinati con tre diversi metodi di cottura. Successivamente, alcuni giudici hanno assaggiato ciascun pesce e hanno assegnato un punteggio sulla base di 4 variabili, ovvero l'aroma (A), il sapore (S), la consistenza (C) e il grado di umidità (U). Nella Tavola 6.3.4 sono riportati i punteggi medi dei giudici per ogni pesce e ogni metodo di cottura.

Tavola 6.3.4. Punteggi dei giudici per i tre metodi di cottura dei pesci.

| metodo 1 | A | B | C | D | metodo 2 | A | B | C | D | metodo 3 | A | B | C | D |
|----------|-----|-----|-----|-----|----------|-----|-----|-----|-----|----------|-----|-----|-----|-----|
| 1 | 5.4 | 6.0 | 6.3 | 6.7 | 1 | 5.0 | 5.3 | 5.3 | 6.5 | 1 | 4.8 | 5.0 | 6.5 | 7.0 |
| 2 | 5.2 | 6.2 | 6.0 | 5.8 | 2 | 4.8 | 4.9 | 4.2 | 5.6 | 2 | 5.4 | 5.0 | 6.0 | 6.4 |
| 3 | 6.1 | 5.9 | 6.0 | 7.0 | 3 | 3.9 | 4.0 | 4.4 | 5.0 | 3 | 4.9 | 5.1 | 5.9 | 6.5 |
| 4 | 4.8 | 5.0 | 4.9 | 5.0 | 4 | 4.0 | 5.1 | 4.8 | 5.8 | 4 | 5.7 | 5.2 | 6.4 | 6.4 |
| 5 | 5.0 | 5.7 | 5.0 | 6.5 | 5 | 5.6 | 5.4 | 5.1 | 6.2 | 5 | 4.2 | 4.6 | 5.3 | 6.3 |
| 6 | 5.7 | 6.1 | 6.0 | 6.6 | 6 | 6.0 | 5.5 | 5.7 | 6.0 | 6 | 6.0 | 5.3 | 5.8 | 6.4 |
| 7 | 6.0 | 6.0 | 5.8 | 6.0 | 7 | 5.2 | 4.8 | 5.4 | 6.0 | 7 | 5.1 | 5.2 | 6.2 | 6.5 |
| 8 | 4.0 | 5.0 | 4.0 | 5.0 | 8 | 5.3 | 5.1 | 5.8 | 6.4 | 8 | 4.8 | 4.6 | 5.7 | 5.7 |
| 9 | 5.7 | 5.4 | 4.9 | 5.0 | 9 | 5.9 | 6.1 | 5.7 | 6.0 | 9 | 5.3 | 5.4 | 6.8 | 6.6 |
| 10 | 5.6 | 5.2 | 5.4 | 5.8 | 10 | 6.1 | 6.0 | 6.1 | 6.2 | 10 | 4.6 | 4.4 | 5.7 | 5.6 |
| 11 | 5.8 | 6.1 | 5.2 | 6.4 | 11 | 6.2 | 5.7 | 5.9 | 6.0 | 11 | 4.5 | 4.0 | 5.0 | 5.9 |
| 12 | 5.3 | 5.9 | 5.8 | 6.0 | 12 | 5.1 | 4.9 | 5.3 | 4.8 | 12 | 4.4 | 4.2 | 5.6 | 5.5 |

Fonte: Baten, Tack e Baeder (1958)

Si vuole verificare se i metodi di cottura non differiscono sulla base dei punteggi assegnati dai giudici, ovvero si vuole verificare il sistema di ipotesi $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu$ contro $H_1 : \mu_i \neq \mu_j, \exists i \neq j = 1, 2, 3$. Si ha

$$\bar{\mathbf{x}}_1 = \begin{pmatrix} 5.3833 \\ 5.7083 \\ 5.4417 \\ 5.9833 \end{pmatrix}$$

e

$$\bar{\mathbf{x}}_2 = \begin{pmatrix} 5.2583 \\ 5.2333 \\ 5.3083 \\ 5.8750 \end{pmatrix},$$

mentre

$$\bar{\mathbf{x}}_3 = \begin{pmatrix} 4.9750 \\ 4.8333 \\ 5.9083 \\ 6.2333 \end{pmatrix}.$$

Inoltre

$$\mathbf{S}_1 = \begin{pmatrix} 0.3131 & 0.1410 & 0.2399 & 0.2114 \\ 0.1410 & 0.1791 & 0.2105 & 0.2093 \\ 0.2399 & 0.2105 & 0.4041 & 0.2999 \\ 0.2114 & 0.2093 & 0.2299 & 0.4447 \end{pmatrix}$$

e

$$\mathbf{S}_2 = \begin{pmatrix} 0.5341 & 0.3289 & 0.3412 & 0.1756 \\ 0.3289 & 0.3022 & 0.2289 & 0.1633 \\ 0.3412 & 0.2289 & 0.3241 & 0.1477 \\ 0.1756 & 0.1633 & 0.1477 & 0.2452 \end{pmatrix},$$

mentre

$$\mathbf{S}_3 = \begin{pmatrix} 0.2702 & 0.1783 & 0.1419 & 0.1017 \\ 0.1783 & 0.1939 & 0.1739 & 0.1497 \\ 0.1419 & 0.1739 & 0.2391 & 0.1389 \\ 0.1017 & 0.1497 & 0.1389 & 0.1906 \end{pmatrix}.$$

Dalle precedenti espressioni risulta

$$\mathbf{W} = \begin{pmatrix} 0.3725 & 0.2161 & 0.2410 & 0.1629 \\ 0.2161 & 0.2251 & 0.2044 & 0.1741 \\ 0.2410 & 0.2044 & 0.3224 & 0.1955 \\ 0.1629 & 0.1741 & 0.1955 & 0.2935 \end{pmatrix}$$

e

$$\mathbf{B} = \begin{pmatrix} 0.0292 & 0.0589 & -0.0382 & -0.0211 \\ 0.0589 & 0.1280 & -0.0650 & -0.0345 \\ -0.0382 & -0.0650 & 0.0662 & 0.0385 \\ -0.0211 & -0.0345 & 0.0385 & 0.0225 \end{pmatrix},$$

per cui $\lambda = 0.2200$. Di conseguenza il livello di significatività osservato risulta

$$\alpha_{oss} = \Pr(A_{4,33,2} \leq 0.2200) = \Pr(F_{8,60} \geq 8.4901) \simeq 0.$$

Dato che la significatività osservata estremamente bassa, si deve concludere che i tre metodi di cottura differiscono sostanzialmente. Inoltre $-2 \log r = 54.5008$ e dunque si ha

$$\alpha_{n,oss} = \Pr(\chi_8^2 \geq 54.5008) \simeq 0,$$

un risultato che conferma le conclusioni precedenti. \square

6.3.8. Il test per la verifica della omoschedasticità. Si supponga di considerare m campioni casuali indipendenti $(\mathbf{x}_{l1}, \dots, \mathbf{x}_{ln_l})$ rispettivamente dai vettori casuali $\mathcal{X}_l \sim N_d(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$. Sia inoltre $n = \sum_{l=1}^m n_l$ la numerosità campionaria globale. In questo caso, il modello statistico è dato da

$$\mathcal{F}_{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m} = \left\{ f_n : f_n(\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1}, \dots, \mathbf{x}_{m1}, \dots, \mathbf{x}_{mn_m}; \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m) \right. \\ \left. = \prod_{l=1}^m \prod_{i=1}^{n_l} \det(2\pi \boldsymbol{\Sigma}_l)^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}_{li} - \boldsymbol{\mu}_l)^T \boldsymbol{\Sigma}_l^{-1} (\mathbf{x}_{li} - \boldsymbol{\mu}_l)} \right\},$$

dove $\boldsymbol{\mu}_l \in \mathbb{R}^d$ e $\boldsymbol{\Sigma}_l > \mathbf{0}$. Si vuole verificare il sistema di ipotesi $H_0 : \boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_m = \boldsymbol{\Sigma}$ contro $H_1 : \boldsymbol{\Sigma}_l \neq \boldsymbol{\Sigma}_j, \exists l \neq j = 1, \dots, m$. Si ha

$$\Theta = \{(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m) : \boldsymbol{\mu}_l \in \mathbb{R}^d, \boldsymbol{\Sigma}_l > \mathbf{0}, l = 1, \dots, m\}$$

e

$$\Theta_0 = \{(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m) : \boldsymbol{\mu}_l \in \mathbb{R}^d, \boldsymbol{\Sigma}_1 = \dots = \boldsymbol{\Sigma}_m = \boldsymbol{\Sigma} > \mathbf{0}\}.$$

Se $\bar{\mathbf{x}}_l$ e \mathbf{S}_l rappresentano rispettivamente le realizzazioni campionarie del vettore medio campionario $\bar{\mathbf{X}}_l$ e della matrice di varianza-covarianza campionaria \mathbf{S}_l per l' l -esimo campione, allora la funzione di log-verosimiglianza è data da

$$\begin{aligned} l(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m) &= \\ &= \log c + \sum_{l=1}^m \left(-\frac{n_l}{2} \log \det(\boldsymbol{\Sigma}_l) - \frac{n_l}{2} \text{tr}(\boldsymbol{\Sigma}_l^{-1} \mathbf{S}_l) - \frac{n_l}{2} (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l)^\top \boldsymbol{\Sigma}_l^{-1} (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l) \right). \end{aligned}$$

Tenendo presente che ogni $(\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l)^\top \boldsymbol{\Sigma}_l^{-1} (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l)$ raggiunge il minimo per $\boldsymbol{\mu}_l = \bar{\mathbf{x}}_l$ e adoperando nella precedente espressione la disuguaglianza ottenuta nell'Esempio 6.2.2 in maniera ripetuta con $\mathbf{D} = \mathbf{S}_l$, si ottiene

$$\begin{aligned} l(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m) &\leq \log c + \sum_{l=1}^m \left(-\frac{n_l}{2} \log \det(\boldsymbol{\Sigma}_l) - \frac{n_l}{2} \text{tr}(\boldsymbol{\Sigma}_l^{-1} \mathbf{S}_l) \right) \\ &= l(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_m, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m) \\ &\leq \log c + \sum_{l=1}^m \left(-\frac{n_l}{2} \log \det(\mathbf{S}_l) - \frac{n_l d}{2} \right) \\ &= \log c - \frac{1}{2} \sum_{l=1}^m n_l \log \det(\mathbf{S}_l) - \frac{nd}{2} \\ &= l(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_m, \mathbf{S}_1, \dots, \mathbf{S}_m) \\ &= \max_{(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m) \in \Theta} l(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m). \end{aligned}$$

Inoltre, se H_0 è vera, allora si ha

$$\begin{aligned} l(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}, \dots, \boldsymbol{\Sigma}) &= \\ &= \log c + \sum_{l=1}^m \left(-\frac{n_l}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n_l}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}_l) - \frac{n_l}{2} (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l)^\top \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l) \right) \\ &= \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{W}) - \frac{1}{2} \sum_{l=1}^m n_l (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l)^\top \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l). \end{aligned}$$

Di conseguenza, adoperando la disuguaglianza ottenuta nell'Esempio 6.2.2 rispettivamente con $\mathbf{D} = \mathbf{W}$, si ottiene

$$\begin{aligned} l(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}, \dots, \boldsymbol{\Sigma}) &\leq \log c - \frac{n}{2} \log \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{W}) = l(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_m, \boldsymbol{\Sigma}, \dots, \boldsymbol{\Sigma}) \\ &\leq \log c - \frac{n}{2} \log \det(\mathbf{W}) - \frac{nd}{2} = l(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_m, \mathbf{W}, \dots, \mathbf{W}) \\ &= \max_{(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m) \in \Theta_0} l(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_m). \end{aligned}$$

Dunque, la determinazione campionaria di R è data da

$$r = e^{l(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_m, \mathbf{W}, \dots, \mathbf{W}) - l(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_m, \mathbf{S}_1, \dots, \mathbf{S}_m)} = \prod_{l=1}^m \left(\frac{\det(\mathbf{W})}{\det(\mathbf{S}_l)} \right)^{-n_l/2}$$

In questo caso è complesso ottenere la distribuzione esatta di R o di una sua trasformata. Tuttavia, per grandi campioni, quando H_0 è vera, la quantità

$$-2 \log r = \sum_{l=1}^m n_l \log \left(\frac{\det(\mathbf{W})}{\det(\mathbf{S}_l)} \right),$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_{(m-1)d(d+1)/2}^2$, in quanto vi sono in totale $(md + md(d+1)/2)$ parametri, di cui $(md + d(d+1)/2)$ parametri da stimare sotto ipotesi di base. Di conseguenza, la regione critica per grandi campioni è data da

$$\mathcal{T}_{n,1} = \{r : -2 \log r \geq \chi_{(m-1)d(d+1)/2, 1-\alpha}^2\},$$

mentre la significatività osservata risulta

$$\alpha_{n,oss} = \Pr(\chi_{(m-1)d(d+1)/2}^2 \geq -2 \log r).$$

• **Esempio 6.3.5.** Con gli stessi dati dell'Esempio 6.3.4 si vuole verificare il sistema di ipotesi $H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3$ contro $H_1 : \Sigma_l \neq \Sigma_j, \exists l \neq j = 1, 2, 3$. In questo caso è facile verificare che $-2 \log r = 17.7733$ e di conseguenza il livello di significatività osservato per grandi campioni risulta

$$\alpha_{n,oss} = \Pr(\chi_{20}^2 \geq 17.7733) = 0.6023.$$

Questo valore è un valore elevato di $\alpha_{n,oss}$ che porta dunque ad accettare dunque l'ipotesi di omoschedasticità. \square

6.3.9. La regressione lineare multivariata. Supponiamo che la matrice campionaria $(\mathbf{y}_1, \dots, \mathbf{y}_n)$ sia la determinazione di un insieme di vettori casuali indipendenti $(\mathcal{Y}_1, \dots, \mathcal{Y}_n)$, tali che $\mathcal{Y}_i \sim N_d(\mathbf{a} + \mathbf{B}\mathbf{z}_i, \Sigma)$, dove \mathbf{a} è un vettore di parametri di ordine d , \mathbf{B} è una matrice di parametri di ordine $(d \times e)$, mentre $(\mathbf{z}_1, \dots, \mathbf{z}_n)$ costituisce una matrice di quantità note e fisse di ordine $(e \times n)$. Ovviamente, in questo caso il campionamento non è casuale ed è immediato verificare che questa struttura campionaria costituisce una estensione multivariata del modello di regressione lineare. Analogamente all'Esempio 1.1.8, si suppone che la parametrizzazione sia fatta in modo tale che

$$\bar{\mathbf{z}} = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i = \mathbf{0}.$$

In questo caso il modello statistico è dato da

$$\mathcal{F}_{\mathbf{a}, \mathbf{B}, \Sigma} = \left\{ f_n : f_n(\mathbf{y}_1, \dots, \mathbf{y}_n; \mathbf{a}, \mathbf{B}, \Sigma) = \prod_{i=1}^n \det(2\pi\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y}_i - \mathbf{a} - \mathbf{B}\mathbf{z}_i)^\top \Sigma^{-1}(\mathbf{y}_i - \mathbf{a} - \mathbf{B}\mathbf{z}_i)} \right\},$$

dove $\mathbf{a} \in \mathbb{R}^d$, $\mathbf{B} \in \mathbb{R}^{de}$, $\Sigma > \mathbf{0}$. Si vuole verificare il sistema di ipotesi $H_0 : \mathbf{B} = \mathbf{0}$ contro $H_1 : \mathbf{B} \neq \mathbf{0}$. Si ha

$$\Theta = \{(\mathbf{a}, \mathbf{B}, \Sigma) : \mathbf{a} \in \mathbb{R}^d, \mathbf{B} \in \mathbb{R}^{de}, \Sigma > \mathbf{0}\}$$

e

$$\Theta_0 = \{(\mathbf{a}, \mathbf{B}, \Sigma) : \mathbf{B} = \mathbf{0}, \mathbf{a} \in \mathbb{R}^d, \Sigma > \mathbf{0}\}.$$

La log-verosimiglianza può essere espressa come

$$l(\mathbf{a}, \mathbf{B}, \Sigma) = \log c - \frac{n}{2} \det(\Sigma) - \frac{1}{2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{a} - \mathbf{B}\mathbf{z}_i)^\top \Sigma^{-1} (\mathbf{y}_i - \mathbf{a} - \mathbf{B}\mathbf{z}_i).$$

Si indichi rispettivamente con $\bar{\mathbf{y}}$ e \mathbf{S}_y il vettore medio campionario e la matrice di varianza-covarianza campionaria relative a $(\mathbf{y}_1, \dots, \mathbf{y}_n)$ e con \mathbf{S}_z la matrice di varianza-covarianza relativa a $(\mathbf{z}_1, \dots, \mathbf{z}_n)$. Sia inoltre

$$\mathbf{S}_{zy} = \frac{1}{n} \sum_{i=1}^n (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{y}_i - \bar{\mathbf{y}})^T = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{y}_i^T$$

la matrice di covarianza relativa a $(\mathbf{z}_1, \dots, \mathbf{z}_n)$ e $(\mathbf{y}_1, \dots, \mathbf{y}_n)$, mentre sia $\mathbf{S}_{yz} = \mathbf{S}_{zy}^T$. Si ha

$$\begin{aligned} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{a} - \mathbf{B}\mathbf{z}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \mathbf{a} - \mathbf{B}\mathbf{z}_i) &= \\ &= \sum_{i=1}^n ((\mathbf{y}_i - \bar{\mathbf{y}}) - \mathbf{B}\mathbf{z}_i + (\bar{\mathbf{y}} - \mathbf{a}))^T \boldsymbol{\Sigma}^{-1} ((\mathbf{y}_i - \bar{\mathbf{y}}) - \mathbf{B}\mathbf{z}_i + (\bar{\mathbf{y}} - \mathbf{a})) \\ &= \sum_{i=1}^n \text{tr}(\boldsymbol{\Sigma}^{-1} ((\mathbf{y}_i - \bar{\mathbf{y}}) - \mathbf{B}\mathbf{z}_i + (\bar{\mathbf{y}} - \mathbf{a})) ((\mathbf{y}_i - \bar{\mathbf{y}}) - \mathbf{B}\mathbf{z}_i + (\bar{\mathbf{y}} - \mathbf{a}))^T) \\ &= n \text{tr}(\boldsymbol{\Sigma}^{-1} (\mathbf{S}_y + \mathbf{B}\mathbf{S}_z\mathbf{B}^T + (\bar{\mathbf{y}} - \mathbf{a})(\bar{\mathbf{y}} - \mathbf{a})^T - \mathbf{S}_{yz}\mathbf{B}^T - \mathbf{B}\mathbf{S}_{zy})), \end{aligned}$$

dove è stato tenuto presente che $\sum_{i=1}^n \mathbf{z}_i = \mathbf{0}$ e $\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}}) = \mathbf{0}$. Inoltre, sommando e sottraendo la matrice $\mathbf{S}_{yz}\mathbf{S}_z^{-1}\mathbf{S}_{zy}$ nella precedente espressione e raccogliendo opportunamente si ha

$$\begin{aligned} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{a} - \mathbf{B}\mathbf{z}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \mathbf{a} - \mathbf{B}\mathbf{z}_i) &= \\ &= n \text{tr}(\boldsymbol{\Sigma}^{-1} (\mathbf{S}_y - \mathbf{S}_{yz}\mathbf{S}_z^{-1}\mathbf{S}_{zy} + (\mathbf{B} - \mathbf{S}_{yz}\mathbf{S}_z^{-1})\mathbf{S}_z(\mathbf{B} - \mathbf{S}_{yz}\mathbf{S}_z^{-1})^T + (\bar{\mathbf{y}} - \mathbf{a})(\bar{\mathbf{y}} - \mathbf{a})^T)) \\ &= n \text{tr}(\boldsymbol{\Sigma}^{-1} (\mathbf{S}_y - \mathbf{S}_{yz}\mathbf{S}_z^{-1}\mathbf{S}_{zy})) + n \text{tr}(\boldsymbol{\Sigma}^{-1} (\mathbf{B} - \mathbf{S}_{yz}\mathbf{S}_z^{-1})\mathbf{S}_z(\mathbf{B} - \mathbf{S}_{yz}\mathbf{S}_z^{-1})^T) \\ &\quad + n(\bar{\mathbf{y}} - \mathbf{a})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{y}} - \mathbf{a}), \end{aligned}$$

Dunque, si ottiene infine

$$\begin{aligned} l(\mathbf{a}, \mathbf{B}, \boldsymbol{\Sigma}) &= \log c - \frac{n}{2} \det(\boldsymbol{\Sigma}) - \frac{n}{2} (\text{tr}(\boldsymbol{\Sigma}^{-1} (\mathbf{S}_y - \mathbf{S}_{yz}\mathbf{S}_z^{-1}\mathbf{S}_{zy})) \\ &\quad + \text{tr}(\boldsymbol{\Sigma}^{-1} (\mathbf{B} - \mathbf{S}_{yz}\mathbf{S}_z^{-1})\mathbf{S}_z(\mathbf{B} - \mathbf{S}_{yz}\mathbf{S}_z^{-1})^T) + (\bar{\mathbf{y}} - \mathbf{a})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{y}} - \mathbf{a})). \end{aligned}$$

Sulla base di questi risultati, si ha dunque

$$l(\mathbf{a}, \mathbf{B}, \boldsymbol{\Sigma}) \leq \log c - \frac{n}{2} \det(\boldsymbol{\Sigma}) - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} (\mathbf{S}_y - \mathbf{S}_{yz}\mathbf{S}_z^{-1}\mathbf{S}_{zy})) = l(\hat{\mathbf{a}}, \hat{\mathbf{B}}, \boldsymbol{\Sigma}),$$

dove $\hat{\mathbf{a}} = \bar{\mathbf{y}}$ e $\hat{\mathbf{B}} = \mathbf{S}_{yz}\mathbf{S}_z^{-1}$ che ovviamente costituiscono le stime di massima verosimiglianza di \mathbf{a} e \mathbf{B} . Nella precedente espressione si è tenuto presente che $\boldsymbol{\Sigma} > \mathbf{0}$ e dunque anche $\boldsymbol{\Sigma}^{-1} > \mathbf{0}$, in modo tale che la forma quadratica $(\bar{\mathbf{y}} - \mathbf{a})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{y}} - \mathbf{a})$ è definita positiva e quindi raggiunge il minimo quando $\mathbf{a} = \bar{\mathbf{y}}$. Nella stessa espressione si è notato che, essendo \mathbf{S}_z definita positiva in quanto matrice di varianza-covarianza, allora la matrice $(\mathbf{B} - \mathbf{S}_{yz}\mathbf{S}_z^{-1})\mathbf{S}_z(\mathbf{B} - \mathbf{S}_{yz}\mathbf{S}_z^{-1})^T$ è definita positiva e quindi anche la matrice $\boldsymbol{\Sigma}^{-1} (\mathbf{B} - \mathbf{S}_{yz}\mathbf{S}_z^{-1})\mathbf{S}_z(\mathbf{B} - \mathbf{S}_{yz}\mathbf{S}_z^{-1})^T$ è definita positiva. Di conseguenza, la traccia di questa matrice è maggiore o uguale a 0 e raggiunge quindi il minimo in corrispondenza della matrice nulla, ovvero quando $\mathbf{B} - \mathbf{S}_{yz}\mathbf{S}_z^{-1} = \mathbf{0}$. Tenendo presente la disuguaglianza ottenuta nell'Esempio 6.2.2 con $\mathbf{D} = \mathbf{S}_y - \mathbf{S}_{yz}\mathbf{S}_z^{-1}\mathbf{S}_{zy}$, si ottiene dunque

$$l(\hat{\mathbf{a}}, \hat{\mathbf{B}}, \boldsymbol{\Sigma}) \leq \log c - \frac{n}{2} \det(\mathbf{S}_y - \mathbf{S}_{yz}\mathbf{S}_z^{-1}\mathbf{S}_{zy}) - \frac{nd}{2} = l(\hat{\mathbf{a}}, \hat{\mathbf{B}}, \mathbf{E}) = \max_{(\mathbf{a}, \mathbf{B}, \boldsymbol{\Sigma}) \in \Theta} l(\mathbf{a}, \mathbf{B}, \boldsymbol{\Sigma}),$$

dove $\mathbf{E} = \mathbf{S}_y - \mathbf{S}_{yz}\mathbf{S}_z^{-1}\mathbf{S}_{zy}$ è la stima di massima verosimiglianza di $\boldsymbol{\Sigma}$. È facile dimostrare che

$$\mathbf{E} = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{z}_i)(\mathbf{y}_i - \hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{z}_i)^T$$

e dunque in effetti \mathbf{E} costituisce la matrice delle medie degli scarti al quadrato dei valori osservati \mathbf{y}_i dai valori stimati $\hat{\mathbf{a}} - \hat{\mathbf{B}}\mathbf{z}_i$. Di conseguenza, in maniera analoga alla §4.3.8, la matrice \mathbf{E} rappresenta la matrice di varianza-covarianza residua.

Tenendo presente le note fatte in precedenza, si ha inoltre che

$$\begin{aligned} l(\mathbf{a}, \mathbf{0}, \boldsymbol{\Sigma}) &= \log c - \frac{n}{2} \det(\boldsymbol{\Sigma}) - \frac{n}{2} (\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{S}_y) + (\bar{\mathbf{y}} - \mathbf{a})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{y}} - \mathbf{a})) \\ &\leq \log c - \frac{n}{2} \det(\mathbf{S}_y) - \frac{nd}{2} = l(\bar{\mathbf{y}}, \mathbf{0}, \mathbf{S}_y) = \max_{(\mathbf{a}, \mathbf{B}, \boldsymbol{\Sigma}) \in \Theta_0} l(\mathbf{a}, \mathbf{B}, \boldsymbol{\Sigma}). \end{aligned}$$

In maniera analoga alla §4.3.8, si noti che si ha la seguente scomposizione

$$\mathbf{S}_y = \mathbf{E} + \mathbf{H},$$

dove $\mathbf{H} = \mathbf{S}_{yz} \mathbf{S}_z^{-1} \mathbf{S}_{zy}$ rappresenta la matrice di varianza-covarianza spiegata dal modello lineare. Di conseguenza, si ha anche $\mathbf{E} = \mathbf{S}_y - \mathbf{S}_{yz} \mathbf{S}_z^{-1} \mathbf{S}_{zy}$. La determinazione campionaria di R è dunque data da

$$r = e^{l(\bar{\mathbf{y}}, \mathbf{0}, \mathbf{S}_y) - l(\hat{\mathbf{a}}, \hat{\mathbf{B}}, \mathbf{E})} = \left(\frac{\det(\mathbf{S}_y - \mathbf{S}_{yz} \mathbf{S}_z^{-1} \mathbf{S}_{zy})}{\det(\mathbf{S}_y)} \right)^{\frac{1}{2}n} = \lambda^{\frac{1}{2}n}$$

dove

$$\lambda = \frac{\det(\mathbf{S}_y - \mathbf{S}_{yz} \mathbf{S}_z^{-1} \mathbf{S}_{zy})}{\det(\mathbf{S}_y)} = \frac{\det(\mathbf{E})}{\det(\mathbf{E} + \mathbf{H})}.$$

Dal momento che $n\mathbf{S}_y \sim W_d(\boldsymbol{\Sigma}, n-1)$, per la proprietà *iii*) della §A.4.2, \mathbf{E} e \mathbf{H} sono le determinazioni campionarie di due matrici casuali $n\boldsymbol{\mathcal{E}} \sim W_d(\boldsymbol{\Sigma}, n-1-e)$ e $n\boldsymbol{\mathcal{H}} \sim W_d(\boldsymbol{\Sigma}, e)$ indipendenti. Dunque, se H_0 è vera, λ è la determinazione campionaria della statistica

$$\Lambda = \frac{\det(\boldsymbol{\mathcal{E}})}{\det(\boldsymbol{\mathcal{E}} + \boldsymbol{\mathcal{H}})},$$

distribuita come una $\Lambda_{d, n-1-e, e}$ di Wilks. Inoltre, r è una funzione monotona crescente di λ e quindi i test costruiti su R e Λ sono equivalenti. Dunque, dal momento che rifiutare H_0 per piccole determinazioni di R è equivalente a rifiutare per piccole determinazioni di Λ , la regione critica del test è data da

$$\mathcal{T}_1 = \{\lambda : \lambda \leq \Lambda_{d, n-1-e, e, \alpha}\}.$$

Avendo osservato il valore campionario λ , il livello di significatività osservato risulta infine

$$\alpha_{oss} = \Pr(\Lambda_{d, n-1-e, e} \leq \lambda).$$

Anche se in questo caso la distribuzione della statistica test è nota per campioni finiti, per grandi campioni, quando H_0 è vera, la quantità

$$-2 \log r = n \log \left(\frac{\det(\mathbf{S}_y - \mathbf{S}_{yz} \mathbf{S}_z^{-1} \mathbf{S}_{zy})}{\det(\mathbf{S}_y)} \right),$$

è la determinazione di una statistica $-2 \log R_n \xrightarrow{d} \chi_{de}^2$, in quanto vi sono $(d + de + d(d+1)/2)$ parametri in totale, di cui $(d + d(d+1)/2)$ parametri da stimare sotto ipotesi di base. Di conseguenza, la regione critica per grandi campioni è data da

$$\mathcal{T}_{n,1} = \{r : -2 \log r \geq \chi_{de, 1-\alpha}^2\},$$

mentre la significatività osservata risulta

$$\alpha_{n,oss} = \Pr(\chi_{de}^2 \geq -2 \log r) .$$

• **Esempio 6.3.6.** Sono stati condotti 19 esperimenti che riguardavano una certa reazione chimica. Le variabili tenute sotto controllo sono state la temperatura, la concentrazione e il tempo. Si vuole verificare che esiste una relazione lineare di queste variabili con le variabili di risposta della reazione chimica, ovvero la quantità di materiale iniziale non trasformato, la quantità di materiale iniziale trasformato e la quantità di sottoprodotto non desiderato. Le osservazioni relative agli esperimenti sono riportati nella Tavola 6.3.5.

Tavola 6.3.5. Risultati degli esperimenti relativi alle reazioni chimiche.

| Esperimento | variabili di risposta | | | variabili sotto controllo | | |
|-------------|-----------------------|-------------|---------------|---------------------------|----------------|-------|
| | non trasformato | trasformato | sottoprodotto | temperatura | concentrazione | tempo |
| 1 | 41.5 | 45.9 | 11.2 | 162 | 23.0 | 3.0 |
| 2 | 33.8 | 53.3 | 11.2 | 162 | 23.0 | 8.0 |
| 3 | 27.7 | 57.5 | 12.7 | 162 | 30.0 | 5.0 |
| 4 | 21.7 | 58.8 | 16.0 | 162 | 30.0 | 8.0 |
| 5 | 19.9 | 60.6 | 16.2 | 172 | 25.0 | 5.0 |
| 6 | 15.0 | 58.0 | 22.6 | 172 | 25.0 | 8.0 |
| 7 | 12.2 | 58.6 | 24.5 | 172 | 30.0 | 5.0 |
| 8 | 4.3 | 52.4 | 38.0 | 172 | 30.0 | 8.0 |
| 9 | 19.3 | 56.9 | 21.3 | 167 | 27.5 | 6.5 |
| 10 | 6.4 | 55.4 | 30.8 | 177 | 27.5 | 6.5 |
| 11 | 37.6 | 46.9 | 14.7 | 157 | 27.5 | 6.5 |
| 12 | 18.0 | 57.3 | 22.2 | 167 | 32.5 | 6.5 |
| 13 | 26.3 | 55.0 | 18.3 | 167 | 22.5 | 6.5 |
| 14 | 9.9 | 58.9 | 28.0 | 167 | 27.5 | 9.5 |
| 15 | 25.0 | 50.3 | 22.1 | 167 | 27.5 | 3.5 |
| 16 | 14.1 | 61.1 | 23.0 | 177 | 20.0 | 6.5 |
| 17 | 15.2 | 62.9 | 20.7 | 177 | 20.0 | 6.5 |
| 18 | 15.9 | 60.0 | 22.1 | 160 | 34.0 | 7.5 |
| 19 | 19.6 | 60.6 | 19.3 | 160 | 34.0 | 7.5 |

Fonte: Box e Youle (1955)

Si vuole verificare se esiste o meno una relazione lineare fra le variabili di risposta e le variabili sotto controllo, ovvero si vuole verificare il sistema di ipotesi $H_0 : \mathbf{B} = \mathbf{0}$ contro $H_1 : \mathbf{B} \neq \mathbf{0}$. Si ha

$$\mathbf{S}_z = \begin{pmatrix} 36.6371 & -11.5319 & -0.2105 \\ -11.5319 & 17.0055 & 1.1579 \\ -0.2105 & 1.1579 & 2.5263 \end{pmatrix}$$

e

$$\mathbf{S}_y = \begin{pmatrix} 94.0701 & -27.0650 & -55.9214 \\ -27.0650 & 21.0771 & 5.4795 \\ -55.9214 & 5.4795 & 42.8888 \end{pmatrix},$$

mentre

$$\mathbf{S}_{yz} = \begin{pmatrix} -39.7407 & -8.9882 & -6.9763 \\ 11.2252 & 1.5143 & 2.8658 \\ 22.8681 & 6.0924 & 3.7632 \end{pmatrix}.$$

Risulta quindi

$$\hat{\mathbf{a}} = \begin{pmatrix} 20.1789 \\ 56.3368 \\ 20.7842 \end{pmatrix}$$

e

$$\hat{\mathbf{B}} = \begin{pmatrix} -1.5460 & -1.4256 & -2.2374 \\ 0.4046 & 0.2930 & 1.0338 \\ 0.9139 & 0.8995 & 1.1535 \end{pmatrix}.$$

Inoltre si ha

$$\mathbf{E} = \begin{pmatrix} 4.2197 & -1.1423 & -3.4696 \\ -1.1423 & 13.1296 & -9.4472 \\ -3.4696 & -9.4472 & 12.1683 \end{pmatrix},$$

mentre

$$\mathbf{H} = \begin{pmatrix} 89.8504 & -25.9227 & -52.4518 \\ -25.9227 & 7.9475 & 14.9267 \\ -52.4518 & 14.9267 & 30.7204 \end{pmatrix},$$

per cui $\lambda = 0.0332$. Di conseguenza il livello di significatività osservato risulta

$$\alpha_{oss} = \Pr(A_{3,15,3} \leq 0.0332) \simeq 0.$$

Dato che la significatività osservata è estremamente bassa, si deve concludere che esiste un legame lineare fra le variabili di risposta e quelle di controllo per la reazione chimica considerata. Inoltre, $-2 \log r = 64.7232$ e dunque si ha

$$\alpha_{n,oss} = \Pr(\chi_9^2 \geq 64.7232) \simeq 0,$$

un risultato che conferma le conclusioni precedenti. □

6.4. Le regioni di confidenza con osservazioni multivariate

6.4.1. Le regioni di confidenza con matrici campionarie. Se \mathcal{F}_θ è un modello statistico relativo ad un campionamento casuale, allora si definisce quantità pivot P la trasformata

$$P : \mathcal{C}_n \times \Theta \rightarrow \mathbb{R},$$

tale che $P = P(\mathcal{X}_1, \dots, \mathcal{X}_n; \theta)$ ha una distribuzione che non dipende da θ . Se $B \subset \mathbb{R}$ è un insieme tale che

$$\Pr_\theta(P(\mathcal{X}_1, \dots, \mathcal{X}_n; \theta) \in B) = 1 - \alpha,$$

e se $E(\mathcal{X}_1, \dots, \mathcal{X}_n) \subset \mathbb{R}^k$ è un insieme casuale tale che per ogni θ

$$\{(\mathbf{x}_1, \dots, \mathbf{x}_n) : P(\mathbf{x}_1, \dots, \mathbf{x}_n; \theta) \in B\} \Leftrightarrow \{(\mathbf{x}_1, \dots, \mathbf{x}_n) : \theta \in E(\mathbf{x}_1, \dots, \mathbf{x}_n)\},$$

allora $E(\mathbf{x}_1, \dots, \mathbf{x}_n)$ è detta regione di confidenza per θ al livello di confidenza $(1 - \alpha)$.

Anche nel caso multivariato si può dimostrare che esiste un legame fra la regione di confidenza e la regione di accettazione del test per il sistema di ipotesi $H_0 : \theta = \theta_0$ contro $H_1 : \theta \neq \theta_0$. Nel caso che si costruisca regioni di confidenza a partire dal test del rapporto delle verosimiglianze si può verificare che la regione di confidenza $E(\mathbf{x}_1, \dots, \mathbf{x}_n)$ al livello di confidenza $(1 - \alpha)$ è data da tutti i valori di $\theta \in \Theta$ per cui la verosimiglianza $L(\theta; \mathbf{x}_1, \dots, \mathbf{x}_n)$ è maggiore di

$$c_\alpha = r_\alpha \max_{\theta \in \Theta} L(\theta; \mathbf{x}_1, \dots, \mathbf{x}_n),$$

dove $\Pr_{\theta}(R \leq r_\alpha) = \alpha$. Inoltre, se si ha $\theta = (\theta_A^T, \theta_B^T)^T$, dove θ_A è un parametro a $(k - h)$ componenti, si verifica che la regione di confidenza $E(\mathbf{x}_1, \dots, \mathbf{x}_n)$ al livello di confidenza $(1 - \alpha)$ è data da tutti i valori di $\theta_A \in \Theta_A$ per cui la verosimiglianza profilo $L_p(\theta_A; \mathbf{x}_1, \dots, \mathbf{x}_n)$ è maggiore di c_α .

6.4.2. Gli ellissoidi di confidenza. Dato un campione casuale da $\mathcal{X} \sim N_d(\mu, \Sigma)$, si vuole costruire una regione di confidenza per μ . Se si considera il sistema di ipotesi $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$ dalla §6.3.2 è noto che la regione di accettazione per il test del rapporto delle verosimiglianze è data da

$$\mathcal{X}_0 = \left\{ (\mathbf{x}_1, \dots, \mathbf{x}_n) : (\bar{\mathbf{x}} - \mu)^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \mu) < \frac{d}{n - d} F_{d, n-d, 1-\alpha} \right\}.$$

Di conseguenza, la regione di confidenza per μ al livello di confidenza $(1 - \alpha)$ è data da

$$\left\{ \mu : (\mu - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mu - \bar{\mathbf{x}}) < \frac{d}{n - d} F_{d, n-d, 1-\alpha} \right\},$$

ovvero dall'insieme dei punti contenuti in un ellissoide di centro $\bar{\mathbf{x}}$ con direzioni degli assi date dagli autovettori di \mathbf{S} e con lunghezza degli assi pari a $2\sqrt{\alpha_j d F_{d, n-d, 1-\alpha} / (n - d)}$, dove α_j è il j -esimo autovalore di \mathbf{S} . Questa regione di confidenza è detto ellissoide di confidenza del vettore medio.

• **Esempio 6.4.1.** Su 31 ciliegi del parco nazionale di Allegheny in Pennsylvania sono state misurate rispettivamente l'altezza e il volume, ottenendo in questa maniera le osservazioni della Tavola 6.4.1.

Tavola 6.4.1. Altezza e volume dei ciliegi (in piedi e piedi cubi).

| ciliegio | altezza | volume | ciliegio | altezza | volume |
|----------|---------|--------|----------|---------|--------|
| 1 | 70 | 10.3 | 17 | 65 | 10.3 |
| 2 | 63 | 10.2 | 18 | 72 | 16.4 |
| 3 | 81 | 18.8 | 19 | 83 | 19.7 |
| 4 | 66 | 15.6 | 20 | 75 | 18.2 |
| 5 | 80 | 22.6 | 21 | 75 | 19.9 |
| 6 | 79 | 24.2 | 22 | 76 | 21.0 |
| 7 | 76 | 21.4 | 23 | 69 | 21.3 |
| 8 | 75 | 19.1 | 24 | 74 | 22.2 |
| 9 | 85 | 33.8 | 25 | 86 | 27.4 |
| 10 | 71 | 25.7 | 26 | 64 | 24.9 |
| 11 | 78 | 34.5 | 27 | 80 | 31.7 |
| 12 | 74 | 36.3 | 28 | 72 | 38.3 |
| 13 | 77 | 42.6 | 29 | 81 | 55.4 |
| 14 | 82 | 55.7 | 30 | 80 | 58.3 |
| 15 | 80 | 51.5 | 31 | 80 | 51.0 |
| 16 | 87 | 77.0 | | | |

Fonte: Ryan, Joiner e Ryan (1985)

Risulta

$$\bar{\mathbf{x}} = \begin{pmatrix} 76.0000 \\ 30.1710 \end{pmatrix}$$

e

$$\mathbf{S} = \begin{pmatrix} 39.2903 & 60.6387 \\ 60.6387 & 261.4866 \end{pmatrix}.$$

Dunque, i due autovalori di \mathbf{S} sono dati da 23.8189 e 276.9580, mentre gli autovettori di \mathbf{S} sono $(0.2472, 0.9690)$ e $(0.9690, -0.2472)$. Di conseguenza, una ellisse di confidenza per (μ_1, μ_2) al livello di confidenza del 95% è centrata in $(76.00, 30.17)$ ed ha gli assi in direzione degli autovettori di lunghezze pari a 4.6760 e 15.9449. Il diagramma di dispersione e il grafico dell'ellissi di confidenza sono riportati in Figura 6.4.1. \square

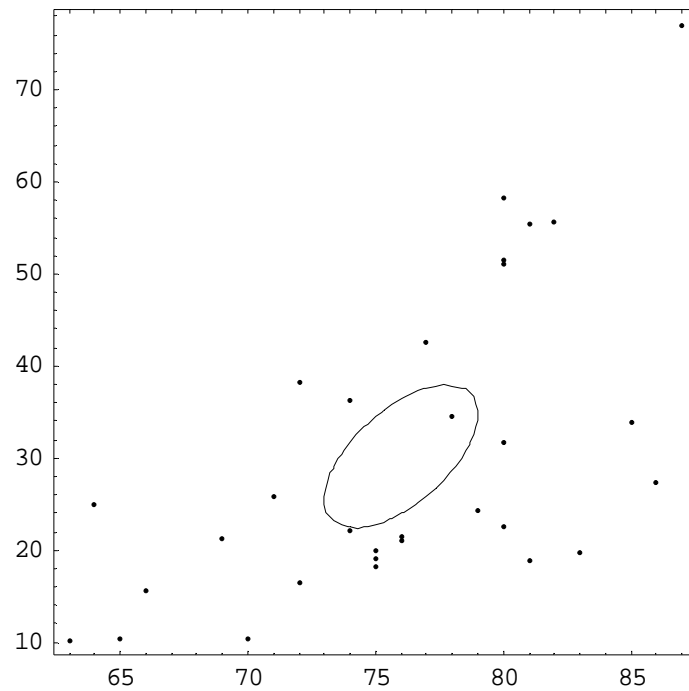


Figura 6.4.1. Ellissi di confidenza per (μ_1, μ_2) al livello di confidenza del 95%.

Appendice

A.1. Alcune distribuzioni continue

A.1.1. Le famiglie di posizione e scala. Nelle prossime sezioni vengono presentate a grandi linee le caratteristiche di alcune distribuzioni continue di uso frequente (per approfondimenti sulle distribuzioni continue univariate si può consultare i due manuali di Johnson e Kotz, 1972, vol. I e II). Nel seguito una variabile casuale Z viene analizzata nella sua forma standard. La funzione di ripartizione e la funzione di densità di Z vengono indicate rispettivamente con F_Z e f_Z .

A partire da una variabile casuale standard Z , attraverso la trasformazione lineare

$$X = \lambda + \delta Z,$$

dove $\lambda \in \mathbb{R}$ e $\delta \in \mathbb{R}^+$, si può generare la cosiddetta famiglia (di distribuzioni) di posizione e di scala. Questa famiglia è caratterizzata appunto dai parametri di posizione λ e di scala δ .

La variabile casuale non standard X possiede funzione di ripartizione e funzione di densità date da

$$F_X(x) = F_Z\left(\frac{x - \lambda}{\delta}\right)$$

e

$$f_X(x) = \frac{1}{\delta} f_Z\left(\frac{x - \lambda}{\delta}\right).$$

Inoltre, supponendo $E[Z^2] < \infty$, risulta

$$\mu = E[X] = \lambda + \delta E[Z]$$

e

$$\sigma^2 = \text{Var}[X] = \delta^2 \text{Var}[Z].$$

In particolare, se $E[Z] = 0$ e $\text{Var}[Z] = 1$, i parametri di posizione e di scala coincidono rispettivamente con la media e lo scarto quadratico medio. Per evidenziare questa proprietà, nel seguito si adottano i simboli μ e σ al posto di λ e δ . Infine, i parametri rimanenti di una distribuzione sono detti parametri di forma.

A.1.2. La distribuzione Normale. La variabile casuale Z è detta Normale standard se ammette funzione di densità $f_Z = \phi$, dove

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}.$$

Il grafico della funzione di densità ϕ è riportato in Figura A.1.1. La funzione di ripartizione di Z viene indicata con Φ e non può essere ottenuta in forma analitica.

I momenti della Normale standard risultano

$$E[Z^r] = \begin{cases} 0 & r \text{ dispari} \\ 1 \times 3 \times \dots \times (r-1) & r \text{ pari} \end{cases}, r = 1, 2, \dots$$

Di conseguenza, si ha $E[Z] = 0$ e $\text{Var}[Z] = 1$. Quindi, per la Normale non standard X i parametri di posizione e di scala coincidono con la media μ e con lo scarto quadratico medio σ .

Per indicare che Z ha distribuzione Normale standard si adotta la notazione $Z \sim N(0, 1)$, mentre se X ha distribuzione Normale non standard si scrive $X \sim N(\mu, \sigma^2)$. Il quantile di ordine α della Normale standard viene indicato con z_α .

Per la Normale valgono le seguenti tre proprietà (vedi Barabesi, 2020):

i) Se $X \sim N(\mu, \sigma^2)$, allora

$$Y = a + bX \sim N(a + b\mu, b^2\sigma^2),$$

dove a e b sono costanti.

ii) Se (X_1, \dots, X_k) è un vettore di variabili casuali indipendenti tale che $X_i \sim N(\mu_i, \sigma_i^2)$, allora

$$Y = \sum_{i=1}^k (a_i + b_i X_i) \sim N\left(\sum_{i=1}^k (a_i + b_i \mu_i), \sum_{i=1}^k b_i^2 \sigma_i^2\right),$$

dove (a_1, \dots, a_k) e (b_1, \dots, b_k) sono vettori di costanti.

iii) Se (X_1, \dots, X_k) è un vettore di variabili casuali indipendenti tale che $X_i \sim N(\mu, \sigma^2)$, allora le variabili casuali

$$Y_1 = \sum_{i=1}^k c_i X_i, Y_2 = \sum_{i=1}^k d_i X_i,$$

dove (c_1, \dots, c_k) e (d_1, \dots, d_k) sono vettori di costanti, sono indipendenti se e solo se

$$\sum_{i=1}^k c_i d_i = 0.$$

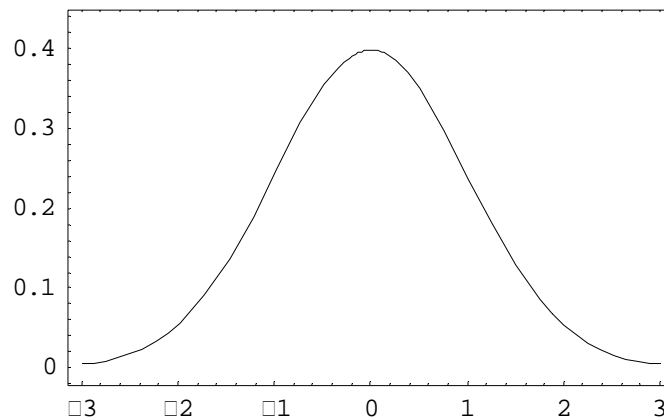


Figura A.1.1. Funzione di densità della Normale standard.

A.1.3. La distribuzione Uniforme. La variabile casuale Z è detta Uniforme standard se ammette funzione di densità

$$f_Z(z) = \mathbf{1}_{[0,1]}(z).$$

Il grafico della funzione di densità di Z è riportato in Figura A.1.2.

La funzione di ripartizione di Z risulta

$$F_Z(z) = z\mathbf{1}_{[0,1]}(z) + \mathbf{1}_{[1,\infty)}(z).$$

I momenti della Uniforme standard sono dati da

$$E[Z^r] = \frac{1}{r+1}, r = 1, 2, \dots$$

Di conseguenza, risulta $E[Z] = 1/2$ e $\text{Var}[Z] = 1/12$.

Per indicare che Z ha distribuzione Uniforme standard si adotta la notazione $Z \sim U(0, 1)$, mentre se X ha distribuzione Uniforme non standard si scrive $X \sim U(\lambda, \lambda + \delta)$. La parametrizzazione in termini di λ e $\lambda + \delta$ si usa per evidenziare che il supporto della variabile casuale non standard X risulta $(\lambda, \lambda + \delta)$.

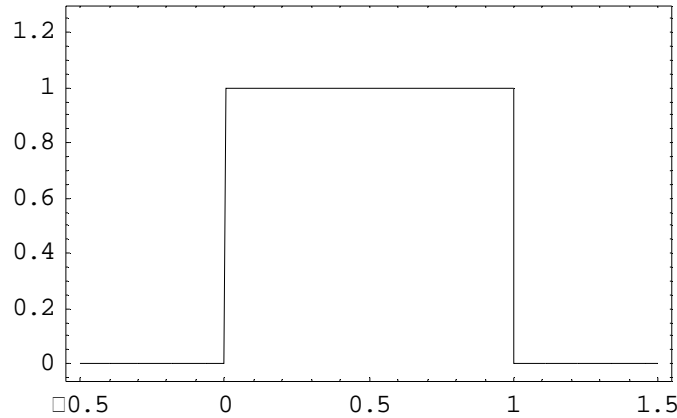


Figura A.1.2. Funzione di densità della Uniforme standard.

A.1.4. La distribuzione Gamma. La variabile casuale Z è detta Gamma standard se ammette funzione di densità

$$f_Z(z) = \frac{1}{\Gamma(p)} z^{p-1} e^{-z} \mathbf{1}_{[0, \infty)}(z),$$

dove $p \in \mathbb{R}^+$ è un parametro di forma. Per $p = 1$ la variabile casuale Z è detta Esponenziale standard. I grafici delle funzioni di densità di Z per alcuni p sono riportati in Figura A.1.3.

La funzione di ripartizione di Z è ottenibile in forma chiusa solo se p è intero, nel qual caso

$$F_Z(z) = \left(1 - e^{-z} \sum_{i=0}^{p-1} \frac{z^i}{i!} \right) \mathbf{1}_{[0, \infty)}(z), p \in \mathbb{N}^+.$$

I momenti della Gamma standard risultano

$$E[Z^r] = \frac{\Gamma(p+r)}{\Gamma(p)}, r = 1, 2, \dots$$

Di conseguenza, si ha $E[Z] = p$ e $\text{Var}[Z] = p$.

Per indicare che Z ha distribuzione Gamma standard con parametro di forma p si adopera la notazione $Z \sim G(0, 1; p)$, mentre se X ha distribuzione Gamma non standard si scrive $X \sim G(\lambda, \delta; p)$. Per indicare che Z ha distribuzione Esponenziale standard si adopera invece la notazione $Z \sim E(0, 1)$, mentre se X ha distribuzione Esponenziale non standard si scrive $X \sim E(\lambda, \sigma)$, dal momento che in questo caso il parametro di scala coincide con lo scarto quadratico medio.

Per la Gamma valgono le seguenti due proprietà (vedi Barabesi, 2020):

i) Se $X \sim G(\lambda, \delta; p)$, allora

$$Y = a + bX \sim G(a + \lambda b, b\delta; p),$$

dove a e b sono costanti.

ii) Se (X_1, \dots, X_k) è un vettore di variabili casuali indipendenti con $X_i \sim G(\lambda, \delta; p_i)$, allora

$$Y = \sum_{i=1}^k X_i \sim G\left(\lambda, \delta; \sum_{i=1}^k p_i\right).$$

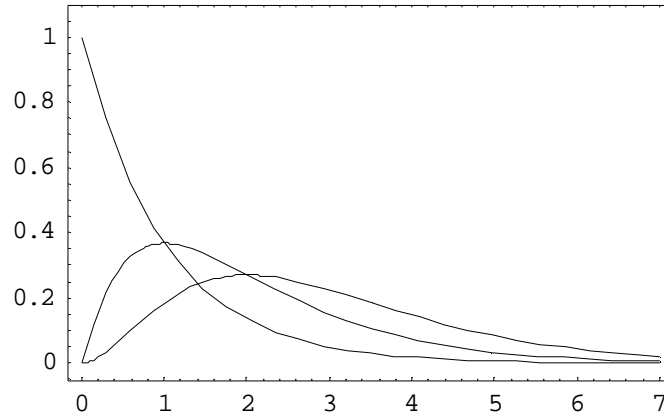


Figura A.1.3. Funzioni di densità della Gamma standard per $p = 1, 2, 3$.

A.1.5. La distribuzione Beta. La variabile casuale Z è detta Beta standard se ammette funzione di densità

$$f_Z(z) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} z^{p-1}(1-z)^{q-1} \mathbf{1}_{[0,1]}(z),$$

dove $p \in \mathbb{R}^+$ e $q \in \mathbb{R}^+$ sono parametri di forma. Se $p = q = 1$, la Beta standard si riduce alla Uniforme standard. I grafici della funzione di densità di Z per alcuni valori di p e q sono riportati nelle Figure A.1.4 e A.1.5.

La funzione di ripartizione di Z può essere ottenuta in forma chiusa solo se p e q sono interi, nel qual caso

$$F_Z(z) = \sum_{i=p}^{p+q-1} \binom{p+q-1}{i} z^i (1-z)^{p+q-1-i} \mathbf{1}_{[0,1]}(z) + \mathbf{1}_{[1,\infty)}(z), \quad p \in \mathbb{N}^+, \quad q \in \mathbb{N}^+.$$

Per la Beta standard si ha

$$E[Z^r] = \frac{\Gamma(p+q)\Gamma(p+r)}{\Gamma(p)\Gamma(p+q+r)}, \quad r = 1, 2, \dots,$$

da cui

$$E[Z] = \frac{p}{p+q}$$

e

$$\text{Var}[Z] = \frac{pq}{(p+q)^2(p+q+1)}.$$

Per indicare che Z ha distribuzione Beta standard con parametri di forma p e q si adotta la notazione $Z \sim Be(0, 1; p, q)$, mentre se X ha distribuzione Beta non standard si scrive $X \sim Be(\lambda, \lambda + \delta; p, q)$. La parametrizzazione in termini di λ e $\lambda + \delta$ viene impiegata per evidenziare che il supporto della variabile casuale non standard X risulta $(\lambda, \lambda + \delta)$.

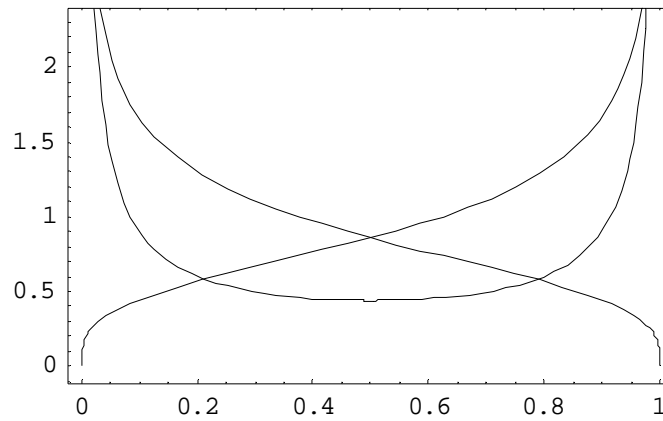


Figura A.1.4. Funzioni di densità della Beta standard per $(p, q) = (0.3, 0.3), (1.3, 0.7), (0.7, 1.3)$.

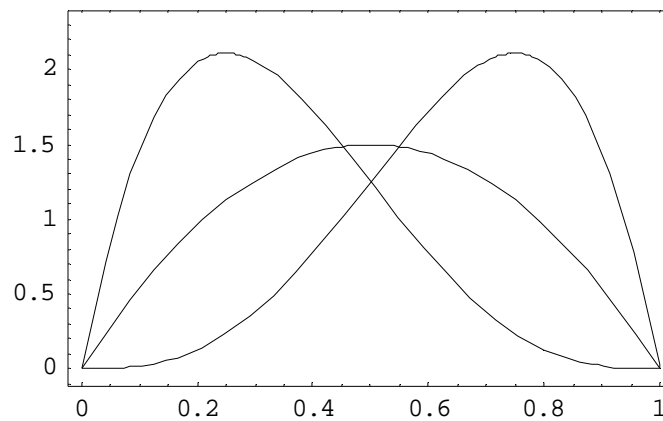


Figura A.1.5. Funzioni di densità della Beta standard per $(p, q) = (2, 2), (2, 4), (4, 2)$.

A.1.6. La distribuzione di Cauchy. La variabile casuale Z è detta di Cauchy standard se ammette funzione di densità

$$f_Z(z) = \frac{1}{\pi(1 + z^2)} .$$

Il grafico della funzione di densità di Z è riportato in Figura A.1.6.

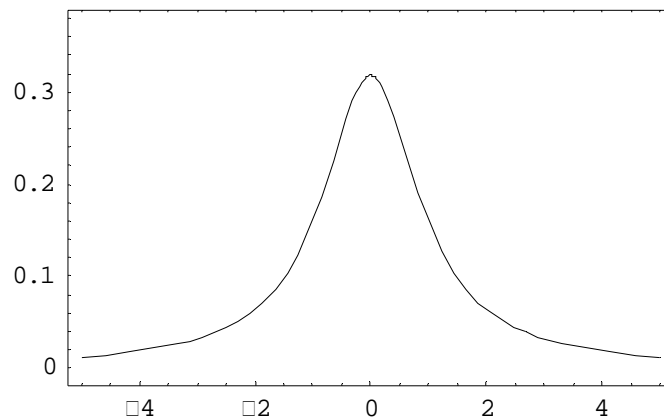


Figura A.1.6. Funzione di densità della Cauchy standard.

La funzione di ripartizione di Z risulta

$$F_Z(z) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(z) .$$

La Cauchy standard non possiede momenti di alcun ordine. Per indicare che Z ha distribuzione Cauchy standard si adotta la notazione $Z \sim C(0, 1)$, mentre se X ha distribuzione Cauchy non standard si scrive $X \sim C(\lambda, \delta)$.

A.1.7. La distribuzione di Laplace. La variabile casuale Z è detta di Laplace standard se ammette funzione di densità

$$f_Z(z) = \frac{1}{2} e^{-|z|}.$$

Il grafico della funzione di densità di Z è riportato in Figura A.1.7.

La funzione di ripartizione di Z risulta

$$F_Z(z) = \frac{1}{2} e^{-|z|} \mathbf{1}_{(-\infty, 0]}(z) + \frac{1}{2} (2 - e^{-|z|}) \mathbf{1}_{(0, \infty)}(z).$$

Per la Laplace standard si ha

$$E[Z^r] = \begin{cases} 0 & r \text{ dispari} \\ \Gamma(r+1) & r \text{ pari} \end{cases}, r = 1, 2, \dots$$

Di conseguenza, risulta $E[Z] = 0$ e $\text{Var}[Z] = 2$. Quindi, per la Laplace non standard il parametro di posizione coincide con la media μ . Per indicare che Z ha distribuzione Laplace standard si adotta la notazione $Z \sim L(0, 1)$, mentre se X ha distribuzione Laplace non standard si scrive $X \sim L(\mu, \delta)$.

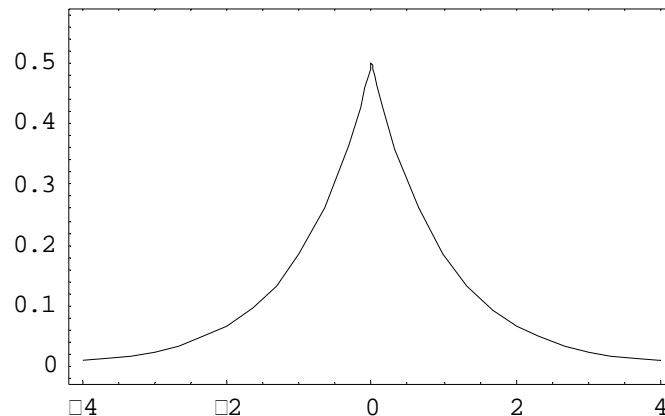


Figura A.1.7. Funzione di densità della Laplace standard.

A.1.8. La distribuzione di Weibull. La variabile casuale Z è detta di Weibull standard se ammette funzione di densità

$$f_Z(z) = pz^{p-1} e^{-z^p} \mathbf{1}_{[0, \infty)}(z),$$

dove $p \in \mathbb{R}^+$ è un parametro di forma. I grafici della funzione di densità di Z per alcuni valori di p sono riportati in Figura A.1.8.

La funzione di ripartizione di Z risulta

$$F_Z(z) = (1 - e^{-z^p}) \mathbf{1}_{[0, \infty)}(z).$$

I momenti della Weibull standard sono dati da

$$E[Z^r] = \Gamma\left(\frac{r}{p} + 1\right), r = 1, 2, \dots$$

Di conseguenza, risulta

$$E[Z] = \Gamma\left(\frac{1}{p} + 1\right)$$

e

$$\text{Var}[Z] = \Gamma\left(\frac{2}{p} + 1\right) - \Gamma\left(\frac{1}{p} + 1\right)^2 .$$

Per indicare che Z ha distribuzione Weibull standard si adotta la notazione $Z \sim W(0, 1; p)$, mentre se X ha distribuzione Weibull non standard si scrive $X \sim W(\lambda, \delta; p)$.

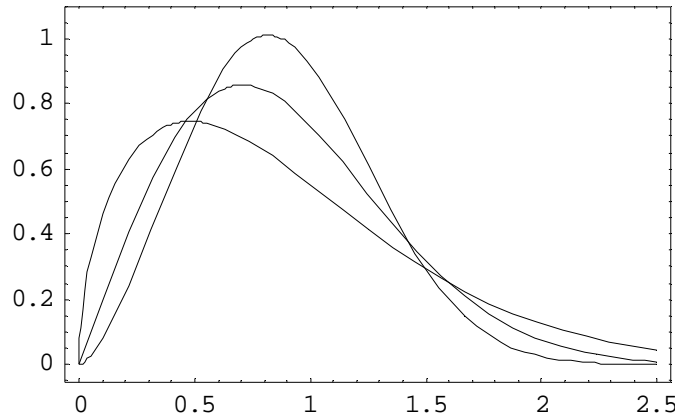


Figura A.1.8. Funzioni di densità della Weibull standard per $p = 1.5, 2.0, 2.5$.

A.2. Alcune distribuzioni discrete

A.2.1. Premessa. Nelle prossime sezione vengono analizzate brevemente le caratteristiche di alcune distribuzioni discrete di importanza fondamentale. Per ulteriori approfondimenti delle distribuzioni discrete è opportuno consultare il manuale di Johnson e Kotz (1972, vol. IV).

A.2.2. La distribuzione Binomiale. La variabile casuale Z è detta Binomiale se ammette funzione di probabilità

$$p_Z(z) = \binom{n}{z} p^z (1 - p)^{n-z} \mathbf{1}_{\{0,1,\dots,n\}}(z) ,$$

dove $p \in (0, 1)$ e $n \in \mathbb{N}^+$. I grafici della funzione di probabilità di Z per alcuni valori di p e n sono riportati in Figura A.2.1.

Per la Binomiale si ha $E[Z] = np$ e $\text{Var}[Z] = np(1 - p)$. Per indicare che Z ha distribuzione Binomiale si adotta la notazione $Z \sim Bi(n, p)$.

Per la Binomiale vale la seguente proprietà (vedi Barabesi, 2020):

i) Se (Z_1, \dots, Z_k) è un vettore di variabili casuali indipendenti con $Z_i \sim Bi(n_i, p)$, allora

$$Y = \sum_{i=1}^k Z_i \sim Bi\left(\sum_{i=1}^k n_i, p\right) .$$

La variabile casuale Binomiale può essere generalizzata al caso multivariato ottenendo la distribuzione Multinomiale. Il vettore di variabili casuali discrete (Z_1, \dots, Z_k) è detto Multinomiale se ammette funzione di probabilità congiunta

$$p_{(Z_1, \dots, Z_k)}(z_1, \dots, z_k) = \binom{n}{z_1 \dots z_k} \prod_{j=1}^k p_j^{z_j} \mathbf{1}_A(z_1, \dots, z_k),$$

dove

$$A = \left\{ (z_1, \dots, z_k) : z_j \in \mathbb{N}_+, \sum_{j=1}^k z_j = n \right\}$$

e (p_1, \dots, p_k) è un vettore di parametri tale che $p_j \in (0, 1)$ e che $\sum_{j=1}^k p_j = 1$.

Evidentemente, (Z_1, \dots, Z_k) è un vettore degenere di variabili casuali e inoltre sono distinti solo $(k-1)$ parametri nel vettore (p_1, \dots, p_k) . Si può dimostrare che $Z_j \sim Bi(n, p_j)$, mentre $E[Z_j] = np_j$, $\text{Var}[Z_j] = np_j(1-p_j)$ e $\text{Cov}[Z_j, Z_l] = -np_j p_l$ (vedi Barabesi, 2020). Infine, per indicare che (Z_1, \dots, Z_k) ha distribuzione Multinomiale si adotta la notazione $(Z_1, \dots, Z_k) \sim Mu(n; p_1, \dots, p_k)$.

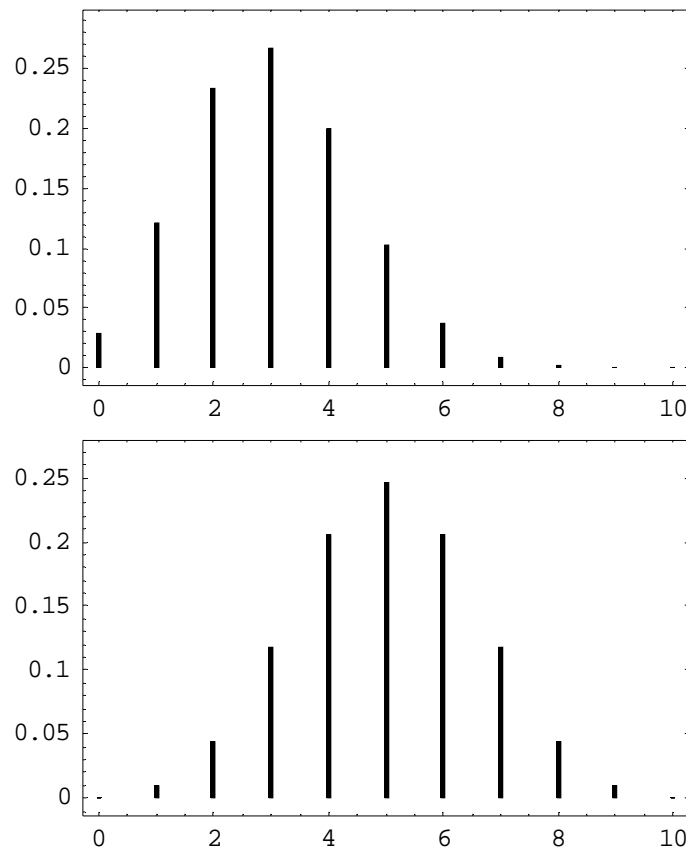


Figura A.2.1. Funzioni di probabilità della Binomiale per $p = 0.3, 0.5$ e $n = 10$.

A.2.3. La distribuzione di Poisson. La variabile casuale Z è detta di Poisson se ammette funzione di probabilità

$$p_Z(z) = e^{-\mu} \frac{\mu^z}{z!} \mathbf{1}_{\{0,1,\dots\}}(z),$$

dove $\mu \in \mathbb{R}^+$. I grafici della funzione di probabilità di Z per alcuni valori di μ sono riportati in Figura A.2.2.

Per la Poisson si ha $E[Z] = \mu$ e $\text{Var}[Z] = \mu$. Per indicare che Z ha distribuzione di Poisson si adotta la notazione $Z \sim Po(\mu)$.

Per la Poisson vale la seguente proprietà (vedi Barabesi, 2020):

i) Se (Z_1, \dots, Z_k) è un vettore di variabili casuali indipendenti tali che $Z_i \sim Po(\mu_i)$, allora

$$Y = \sum_{i=1}^k Z_i \sim Po\left(\sum_{i=1}^k \mu_i\right).$$

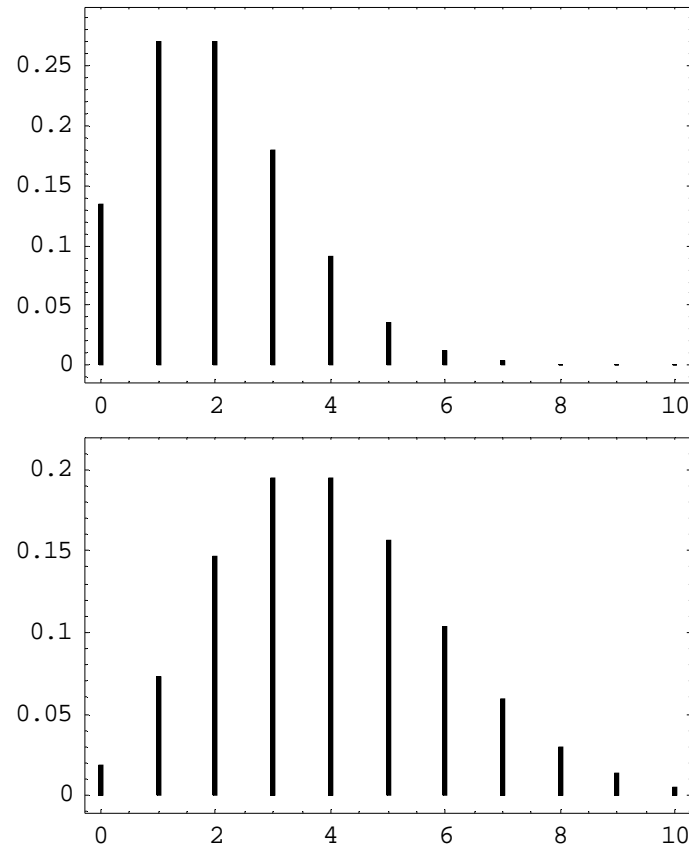


Figura A.2.2. Funzioni di probabilità della Poisson per $\mu = 2, 4$.

A.2.4. La distribuzione Ipergeometrica. La variabile casuale Z è detta Ipergeometrica se ammette funzione di probabilità

$$p_Z(z) = \frac{\binom{M}{z} \binom{N-M}{n-z}}{\binom{N}{n}} \mathbf{1}_{\{\max(0, n-N+M), \dots, \min(n, M)\}}(z),$$

dove $n \in \mathbb{N}^+$, $M \in \mathbb{N}^+$, $N \in \mathbb{N}^+$ sono tali che $n \leq N$ e $M \leq N$. I grafici della funzione di probabilità di Z per alcuni valori di n , M e N sono riportati in Figura A.2.3.

Si ha

$$E[Z] = \frac{nM}{N}$$

e

$$\text{Var}[Z] = \frac{nM(N-M)(N-n)}{N^2(N-1)}.$$

Per indicare che Z ha distribuzione Ipergeometrica si adotta la notazione $Z \sim I(n, M, N)$.

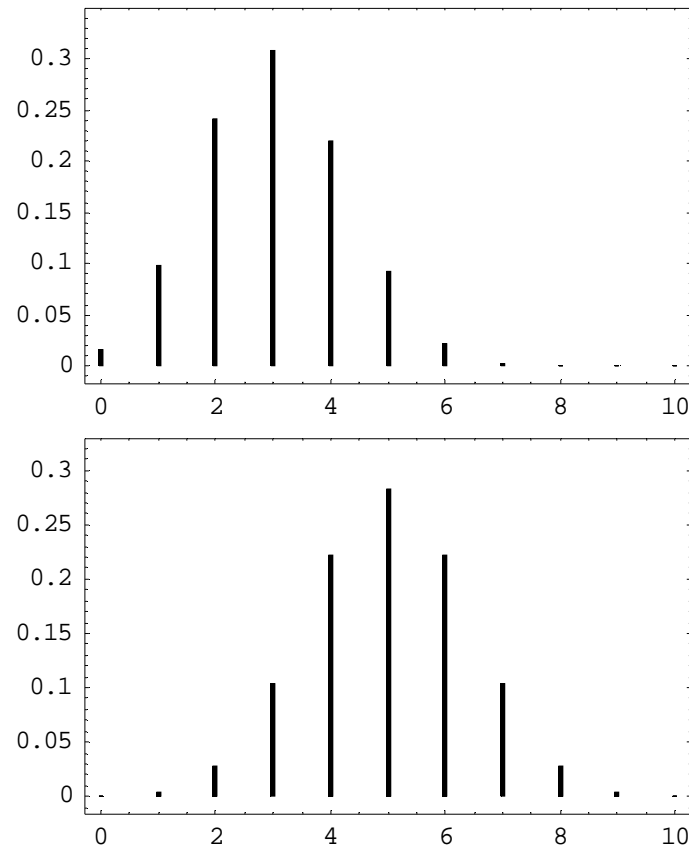


Figura A.2.3. Funzioni di probabilità della Ipergeometrica per $M = 12, 20$, $N = 40$ e $n = 10$.

A.3. Le distribuzioni Chi-quadrato, t e F

A.3.1. Premessa. In questa sezione si considera alcune distribuzioni fondamentali nella statistica inferenziale connesse alla Normale. Per i dettagli teorici della genesi di queste distribuzioni si veda Barabesi (2020).

A.3.2. La distribuzione Chi-quadrato. Se (Z_1, \dots, Z_n) è un vettore di variabili casuali indipendenti tali che $Z_i \sim N(0, 1)$, allora la trasformata

$$U = \sum_{i=1}^n Z_i^2$$

è detta variabile casuale Chi-quadrato con n gradi di libertà. Per indicare che U ha distribuzione Chi-quadrato con n gradi di libertà si adotta la notazione $U \sim \chi_n^2$. Dal momento che $\chi_n^2 \sim G(0, 2; n/2)$ (vedi Barabesi, 2020), allora la funzione di densità di U risulta

$$f_U(u) = \frac{1}{2\Gamma(\frac{n}{2})} \left(\frac{u}{2}\right)^{\frac{1}{2}n-1} e^{-\frac{1}{2}u} \mathbf{1}_{[0, \infty)}(u).$$

I grafici della funzione di densità di U per alcuni valori di n sono riportati in Figura A.3.1.

Per la Chi-quadrato risulta $E[U] = n$ e $\text{Var}[U] = 2n$. Si noti che il quantile di ordine α della Chi-quadrato con n gradi di libertà viene indicato con $\chi_{n, \alpha}^2$.

Per la Chi-quadrato vale la seguente proprietà (vedi Barabesi, 2020):

i) Se (U_1, \dots, U_k) è un vettore di variabili casuali indipendenti con $U_i \sim \chi_{n_i}^2$, allora

$$Y = \sum_{i=1}^k U_i \sim \chi_n^2,$$

dove $n = \sum_{i=1}^k n_i$.

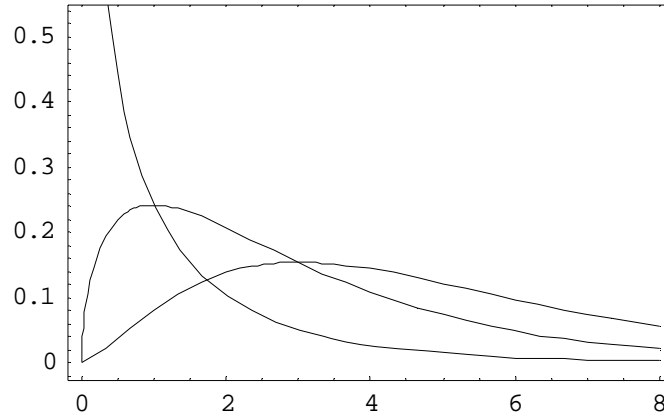


Figura A.3.1. Funzioni di densità della Chi-quadrato per $n = 1, 3, 5$.

A.3.3. La distribuzione t di Student. Se $Z \sim N(0, 1)$ e $U \sim \chi_n^2$ sono indipendenti, allora la trasformata

$$T = \frac{Z}{\sqrt{U/n}}$$

è detta variabile casuale t di Student con n gradi di libertà. Per indicare che T è t di Student con n gradi di libertà si adotta la notazione $T \sim t_n$. Si può dimostrare la funzione di densità di T risulta (vedi Barabesi, 2020)

$$f_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)}.$$

I grafici della funzione di densità di T per alcuni valori di n sono riportati in Figura A.3.2.

Per $n = 1$ la t di Student si riduce alla Cauchy standard. Per la t di Student si ha inoltre $E[T] = 0$ se $n > 1$, e $\text{Var}[T] = n/(n - 2)$ se $n > 2$. Il quantile di ordine α della t di Student con n gradi di libertà viene indicato con $t_{n,\alpha}$. Infine, per $n \rightarrow \infty$ risulta $t_n \xrightarrow{d} N(0, 1)$ (vedi Barabesi, 2020).

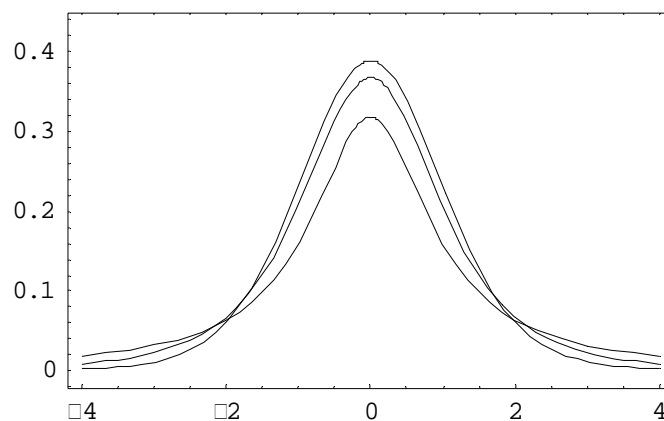


Figura A.3.2. Funzioni di densità della t di Student per $n = 1, 3, 10$.

A.3.4. La distribuzione F di Snedecor. Se $U \sim \chi_m^2$ e $V \sim \chi_n^2$ sono indipendenti, allora la trasformata

$$F = \frac{U/m}{V/n}$$

è detta variabile casuale F di Snedecor con m e n gradi di libertà. Per indicare che F ha distribuzione F di Snedecor con m e n gradi di libertà si adotta la notazione $F \sim F_{m,n}$. Si può dimostrare la funzione di densità di F risulta (vedi Barabesi, 2020)

$$f_F(f) = \frac{\Gamma(\frac{m+n}{2})(\frac{m}{n})^{\frac{1}{2}m}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} f^{\frac{1}{2}m-1} \left(1 + \frac{mf}{n}\right)^{-\frac{1}{2}(m+n)} \mathbf{1}_{[0,\infty)}(f).$$

I grafici della funzione di densità di F per alcuni valori di m e n sono riportati in Figura A.3.3. Per la F di Snedecor risulta rispettivamente

$$E[F] = \frac{n}{n-2}$$

se $n > 2$, e

$$\text{Var}[F] = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$$

se $n > 4$. Il quantile di ordine α della F di Snedecor con m e n gradi di libertà viene indicato con $F_{m,n,\alpha}$. Infine, si ha $t_n = \sqrt{F_{1,n}}$ e $Be(0, 1; n/2, m/2) = n/(n + mF_{m,n})$.

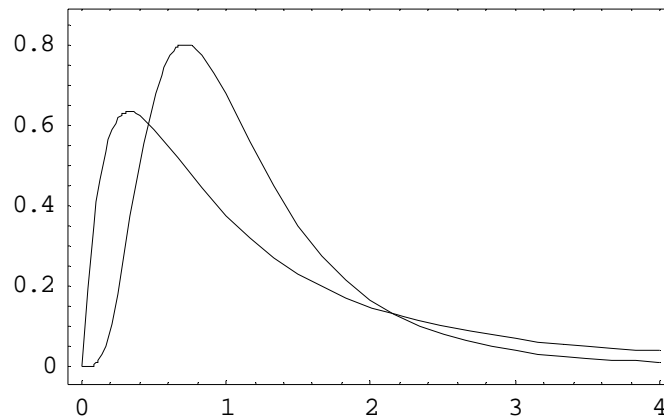


Figura A.3.3. Funzioni di densità della F di Snedecor per $(m, n) = (4, 4), (12, 12)$.

A.4. Alcune distribuzioni multivariate

A.4.1. La distribuzione Normale multivariata. La variabile casuale Normale può essere generalizzata al caso multivariato. Si dice che il vettore di variabili casuali $\mathcal{X} = (X_1, \dots, X_d)^T$ è Normale multivariato se ammette funzione di densità congiunta

$$f_{\mathcal{X}}(\mathbf{x}) = \det(2\pi\Sigma)^{-1/2} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})},$$

dove $\boldsymbol{\mu} \in \mathbb{R}^d$ e $\Sigma > \mathbf{0}$ è una matrice simmetrica definita positiva di ordine d .

La Normale multivariata possiede vettore medio $E[\mathcal{X}] = \boldsymbol{\mu}$ e matrice di varianza-covarianza $\text{Var}[\mathcal{X}] = \Sigma$. Per indicare che \mathcal{X} ha distribuzione Normale multivariata si adotta la notazione $\mathcal{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$. I grafici della funzione di densità di \mathcal{X} per $d = 2$ e alcuni valori di $\boldsymbol{\mu}$ e Σ sono riportati nelle Figure A.4.1, A.4.2 e A.4.3.

Per la Normale multivariata valgono le seguenti due proprietà (vedi Muirhead, 1982):

i) Se $(\mathcal{X}_1, \dots, \mathcal{X}_k)$ sono vettori di variabili casuali indipendenti tali che $\mathcal{X}_i \sim N_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, allora

$$\mathcal{Y} = \sum_{i=1}^k (a_i + b_i \mathcal{X}_i) \sim N_d \left(\sum_{i=1}^k (a_i + b_i \boldsymbol{\mu}_i), \sum_{i=1}^k b_i^2 \boldsymbol{\Sigma}_i \right),$$

dove (a_1, \dots, a_k) e (b_1, \dots, b_k) sono vettori di costanti.

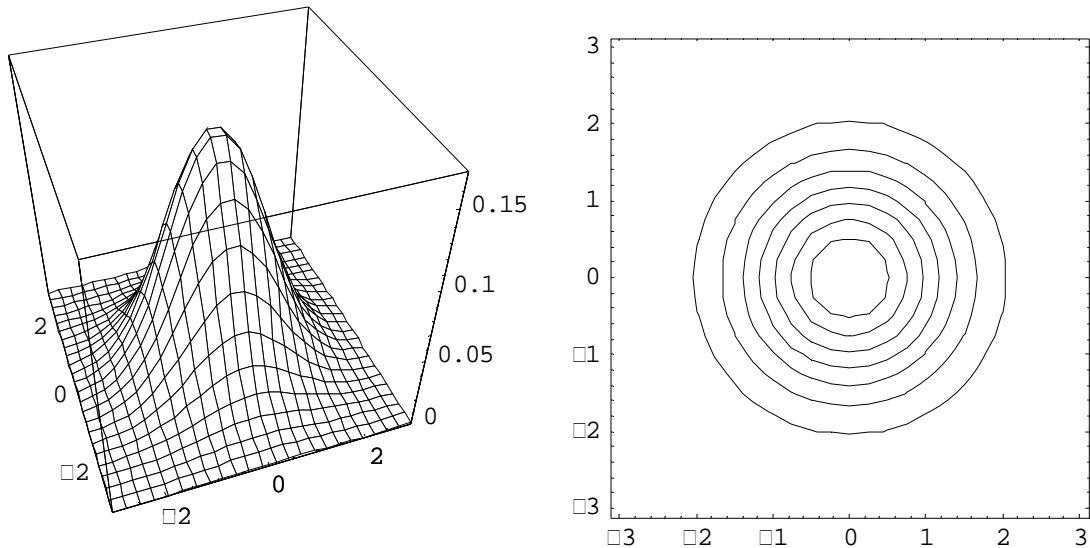


Figura A.4.1. Funzione di densità (con grafico di contorno) di una Normale multivariata $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ e $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

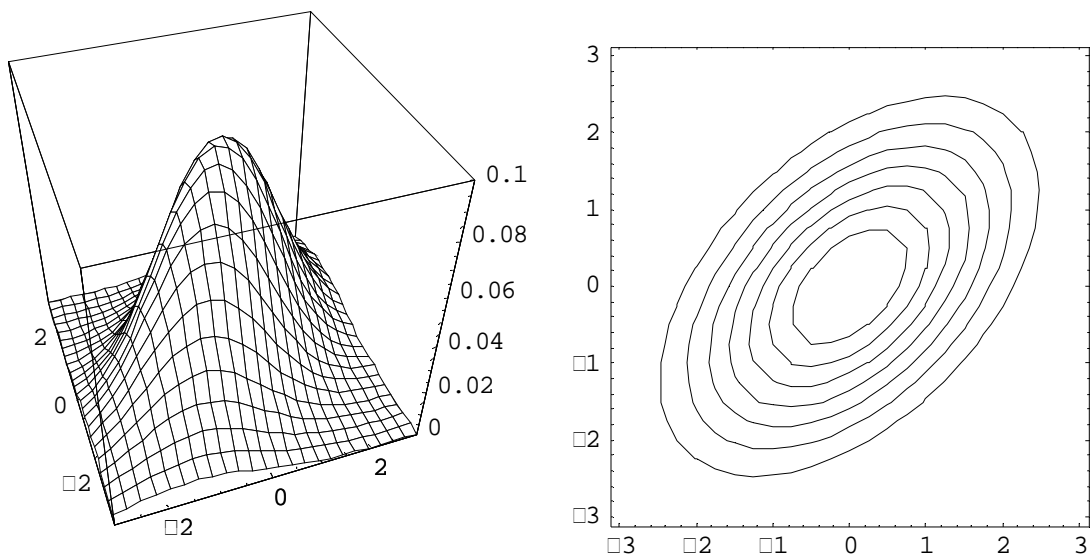


Figura A.4.2. Funzione di densità (con grafico di contorno) di una Normale multivariata $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ e $\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

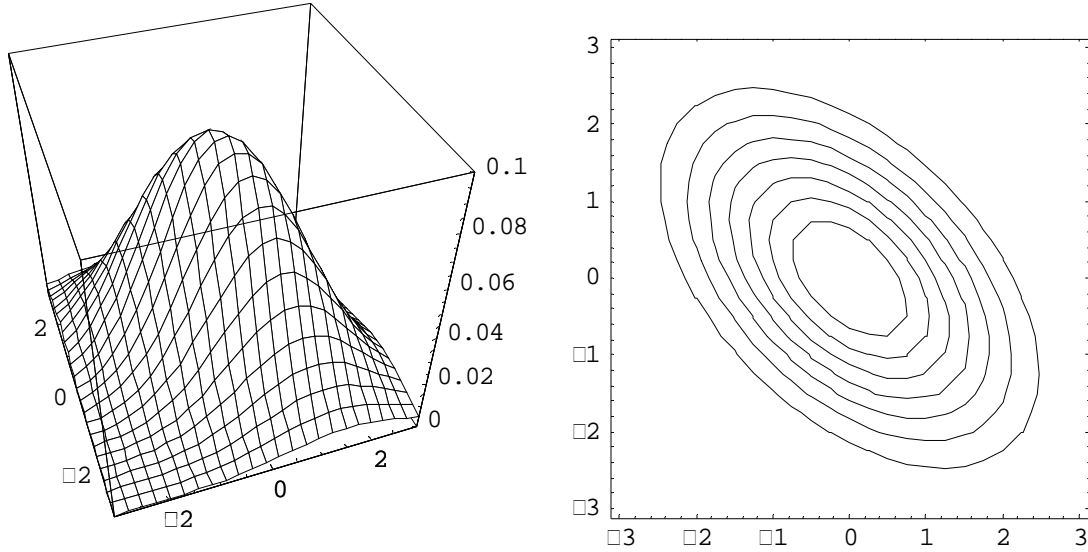


Figura A.4.3. Funzione di densità (con grafico di contorno) di una Normale multivariata $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ e $\boldsymbol{\Sigma} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$.

ii) Se $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ sono vettori di variabili casuali indipendenti tali che $\boldsymbol{x}_i \sim N_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, allora i vettori di variabili casuali

$$\boldsymbol{y}_1 = \sum_{i=1}^k c_i \boldsymbol{x}_i, \quad \boldsymbol{y}_2 = \sum_{i=1}^k d_i \boldsymbol{x}_i,$$

dove (c_1, \dots, c_k) e (d_1, \dots, d_k) sono vettori di costanti, sono indipendenti se e solo se

$$\sum_{i=1}^k c_i d_i = 0.$$

A.4.2. La distribuzione di Wishart. Se $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ sono vettori di variabili casuali indipendenti tali che $\boldsymbol{x}_i \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$, allora la trasformata

$$\boldsymbol{M} = \sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^T$$

è detta matrice casuale di Wishart con n gradi di libertà e parametro $\boldsymbol{\Sigma}$. Per indicare che \boldsymbol{M} ha distribuzione di Wishart si adotta la notazione $\boldsymbol{M} \sim W_d(n, \boldsymbol{\Sigma})$. Se $n > d$, allora \boldsymbol{M} risulta quasi certamente positiva definita con funzione di densità congiunta

$$f_{\boldsymbol{M}}(\boldsymbol{M}) = \frac{\det(2\boldsymbol{\Sigma})^{-n/2}}{\Gamma_d(\frac{n}{2})} \det(\boldsymbol{M})^{\frac{1}{2}(n-d-1)} e^{-\frac{1}{2}\text{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{M})}, \quad \boldsymbol{M} > \mathbf{0},$$

dove $\Gamma_d(\frac{n}{2}) = \pi^{\frac{1}{4}d(d-1)} \prod_{j=1}^d \Gamma(\frac{n-j-1}{2})$. Se $\boldsymbol{\Sigma} = \mathbf{I}_d$ e $d = 1$, allora la $W_d(n, \boldsymbol{\Sigma})$ si riduce alla χ_n^2 . Inoltre, risulta $E[\boldsymbol{M}] = \boldsymbol{\Sigma}$.

Per la distribuzione di Wishart valgono le seguenti proprietà (vedi Muirhead, 1982):

i) Se $(\boldsymbol{M}_1, \dots, \boldsymbol{M}_k)$ sono matrici casuali indipendenti con $\boldsymbol{M}_i \sim W_d(n_i, \boldsymbol{\Sigma})$, allora

$$\mathcal{Y} = \sum_{i=1}^k \mathcal{M}_i \sim W_d(n, \Sigma),$$

dove $n = \sum_{i=1}^k n_i$.

ii) Se $\mathcal{M} \sim W_d(n, \Sigma)$ con $\Sigma = \text{diag}(\Sigma_{11}, \Sigma_{22})$ dove Σ_{11} è di ordine e e Σ_{22} è di ordine $(d - e)$ e se

$$\mathcal{M} = \begin{pmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} \\ \mathcal{M}_{21} & \mathcal{M}_{22} \end{pmatrix},$$

allora

$$\mathcal{E} = \mathcal{M}_{22} - \mathcal{M}_{21} \mathcal{M}_{11}^{-1} \mathcal{M}_{12} \sim W_{d-e}(\Sigma_{22}, n - e)$$

e

$$\mathcal{H} = \mathcal{M}_{21} \mathcal{M}_{11}^{-1} \mathcal{M}_{12} \sim W_{d-e}(\Sigma_{22}, e),$$

sono indipendenti.

iii) Se $\mathcal{M}_x = \sum_{i=1}^n \mathcal{X}_i \mathcal{X}_i^T \sim W_d(n, \Sigma)$ e se $(\mathbf{z}_1, \dots, \mathbf{z}_n)$ costituisce una matrice di coefficienti di ordine $(e \times n)$ tale che $\sum_{i=1}^n \mathbf{z}_i = \mathbf{0}$, allora

$$\mathcal{E} = \mathcal{M}_x - \mathcal{M}_{xz} \mathbf{M}_z^{-1} \mathcal{M}_{zx} \sim W_d(\Sigma, n - e)$$

e

$$\mathcal{H} = \mathcal{M}_{xz} \mathbf{M}_z^{-1} \mathcal{M}_{zx} \sim W_d(\Sigma, e),$$

sono indipendenti, dove $\mathcal{M}_{xz} = \mathcal{M}_{zx}^T = \sum_{i=1}^n \mathcal{X}_i \mathbf{z}_i^T$ e $\mathbf{M}_z = \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T$.

Una trasformata della matrice casuale di Wishart è la variabile casuale T di Hotelling. Se il vettore casuale $\mathcal{X} \sim N_d(\mathbf{0}, \Sigma)$ e la matrice casuale $\mathcal{M} \sim W_d(n, \Sigma)$ sono indipendenti, allora la trasformata

$$H = n \mathcal{X}^T \mathcal{M}^{-1} \mathcal{X}$$

è detta variabile casuale T di Hotelling con n gradi di libertà. Per indicare che H ha distribuzione T di Hotelling si adotta la notazione $H \sim T_{d,n}$. Si può dimostrare (vedi Muirhead, 1982) che fra la T di Hotelling e la F di Snedecor esiste la seguente relazione

$$T_{d,n} = \frac{nd}{n - d + 1} F_{d,n-d+1}.$$

Evidentemente, per $d = 1$ si ha $t_n = \sqrt{T_{1,n}}$.

Una ulteriore trasformata della matrice casuale di Wishart è la variabile casuale Λ di Wilks. Se le matrici casuali $\mathcal{M}_1 \sim W_d(m, \Sigma)$ e $\mathcal{M}_2 \sim W_d(n, \Sigma)$ sono indipendenti, allora la trasformata

$$\Lambda = \frac{\det(\mathcal{M}_1)}{\det(\mathcal{M}_1 + \mathcal{M}_2)}$$

è detta Λ di Wilks con m e n gradi di libertà. Per indicare la distribuzione di Λ si adotta la notazione $\Lambda \sim \Lambda_{d,m,n}$. Si può dimostrare (vedi Muirhead, 1982) che la $\Lambda_{d,m,n}$ di Wilks vale la seguente relazione

$$\Lambda_{d,m,n} = \prod_{i=1}^n Be\left(0, 1; \frac{m - d + i}{2}, \frac{d}{2}\right),$$

dove le variabili casuali Beta nel prodotto sono indipendenti. Come casi particolari della precedente relazione si ha inoltre

$$\Lambda_{d,m,1} = Be\left(0, 1; \frac{m-d+1}{2}, \frac{d}{2}\right) = \frac{m-d+1}{m-d+1+dF_{d,m-d+1}}$$

e

$$\Lambda_{d,m,2} = Be(0, 1; m-d+1, d)^2 = \left(\frac{m-d+1}{m-d+1+dF_{2d,2m-2d+2}}\right)^2,$$

mentre

$$\Lambda_{1,m,n} = Be\left(0, 1; \frac{m}{2}, \frac{n}{2}\right) = \frac{m}{m+nF_{n,m}}$$

e

$$\Lambda_{2,m,n} = Be(0, 1; m-1, n)^2 = \left(\frac{m-1}{m-1+nF_{2n,2m-2}}\right)^2.$$

Infine, il quantile di ordine α della Λ di Wilks con m e n gradi di libertà viene indicato con $\Lambda_{d,m,n,\alpha}$.

Bibliografia

Bibliografia di base

- Agresti, A. (1995) *An Introduction to Categorical Data Analysis*, Wiley, New York.
- Barabesi, L. (2020) *Fondamenti di Probabilità*, Università di Siena
- Brown, L.D. (1986) *Fundamentals of Statistical Exponential Families*, Institute of Mathematical statistics, Hayward, CA.
- Ferguson, T.S. (1996) *A Course in Large Sample Theory*, Chapman and Hall, London.
- Johnson, N.L. e Kotz, S. (1972) *Distributions in Statistics*, vol. I-IV, Wiley, New York.
- Lehmann, E.L. (1983) *The Theory of Point Estimation*, Wiley, New York.
- Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*, Wiley, New York.
- Rao, C.R. (1973) *Linear Statistical Inference and its Applications*, Wiley, New York.
- Seber, G.A.F. (1977) *Linear Regression Analysis*, Wiley, New York.
- Wilks, S.S. (1962) *Mathematical Statistics*, Wiley, New York.

Riferimenti bibliografici

- Angus, J.E. (1982) Goodness-of-fit tests for exponentiality based on a loss-of-memory type functional equation, *Journal of Statistical Planning and Inference* **6**, 241-251.
- Baten, W.D., Tack, P.I. e Baeder H.A. (1958) Testing for differences between methods of preparing fish by use of a discriminant function, *Industrial Quality Control* **14**, 6-10.
- Bissel, A.F. (1972) A negative binomial model with varying element sizes, *Biometrika* **59**, 435-441.
- Bhatia, M.L., Manchanda, S.C. e Roy, S.B. (1969) Coronary haemodynamic studies in chronic severe anaemia, *British Heart Journal* **31**, 365-374.
- Box, G.E.P. e Youle, P.V. (1955) The exploration of response surfaces: an example of the link between the fitted surface and the basic mechanism of the system, *Biometrics* **11**, 287-323.
- Dalal, S.R., Fowlkes, E.B. e Hoadley, B. (1989) Risk analysis of the space shuttle: pre-challenger prediction of failure, *Journal of the American Statistical Association* **84**, 945-957.
- Dubois, C. (1970) *Lowie's selected papers in anthropology*, University of California Press, California.
- Dunsmore, I.R. e Daly, F. (1987) *M345 Statistical methods, Unit 9: categorical data*, The Open University, Milton Keynes.
- Frets, G.P. (1921) Heredity of head form in man, *Genetica* **3**, 193-384.
- Kendall, D.G. (1951) Some problems in the theory of queues, *Journal of the Royal Statistical Society, Serie B* **13**, 151-185.
- Kramer, F. e Jensen, K. (1969) Fundamentals of multivariate analysis, Part II. Inference about two treatments, *Journal of Quality Technology* **2**, 32-40.
- McArthur, J.W. (1931) Linkage studies with the tomato III. Fifteen factors in six groups, *Transaction of the Royal Canadian Institute* **18**, 1-19.
- Norton, P.G e Dunn, E.V. (1985) Snoring as a risk factor for disease: an epidemiological survey, *British Medical Journal* **291**, 630-632.
- Odum, E.P. e Pontin, A.J. (1961) Population density of the underground ant *Lasius Flavus*, as determined by tagging with P^{32} , *Ecology* **42**, 186-188.
- Pollock, M.L., Jackson, A.S. e Pate, R.R. (1980) Discriminant analysis of physiological differences between good and elite runners, *Research Quarterly* **51**, 521-532.
- Romano, A. (1977) *Applied statistics for science and industry*, Allyn and Bacon, Boston.

- Ryan, T.A. Jr, Joiner, B.L. e Ryan, B.F. (1985) *The Minitab student handbook*, Duxbury Press, Boston.
- Seber, G.A.F. (1973) *The estimation of animal abundance*, Griffin, London.
- Selvin, S. (1991) *Statistical analysis of epidemiological data*, Oxford University Press, New York.
- Till, R. (1974) *Statistical methods for the earth scientist*, McMillan, London.
- Van Oost, B.A., Veldhayzen, B., Timmermans, A.P.M. e Sixma, J.J. (1983) Increased urinary β -thromboglobulin excretion in diabetics assayed with a modified RIA kit-technique, *Thrombosis and Haemostasis* **9**, 18-20.
- Williams, E.J. (1959) *Regression analysis*, Wiley, New York.